Quantum Mechanics

Mathematical Structure and Physical Structure Part I

John R. Boccio
Emeritus Professor of Physics
Swarthmore College

Copyright © 2020 by Professor John Boccio
Do not reproduce or distribute copies of this book without permission.

July 29, 2023

Contents

1 Motivation, Motivation, Motivation 1

2 The Mathematics of Quantum Physics: Dirac Language 5

2.1 Mappings 6
2.2 Linear Vector Spaces 6
2.3 Linear Spaces and Linear Functionals 7
2.4 Inner Products 14
2.5 Linear Functionals 16
2.6 The Dirac language of Kets and Bras 18
2.7 Gram-Schmidt Orthogonalization Process 19
2.8 Linear Operators 21
2.9 An Aside : The Connection to Matrices 22
2.10 More about Vectors, Linear Functionals, Operators 23
2.11 Self-Adjoint Operators 28
2.12 Eigenvalues and Eigenvectors 29
2.13 Completeness 31
2.14 Expand Our Knowledge - Selected Topics 32
2.14.1 Hilbert Space 32
2.14.2 Bounded Operators 34
2.14.3 Inverses 34
2.14.4 Unitary Operators 35
2.14.5 More on Matrix Representations 36
2.14.6 Projection Operators 37
2.14.7 Unbounded Operators 39
2.15 Eigenvalues and Eigenvectors of Unitary Operators 40
2.16 Eigenvalue Decomposition 40
2.17 Extension to Infinite-Dimensional Spaces 41
2.18 Spectral Decomposition - More Details 51
2.19 Functions of Operators (general case); Stone's Theorem 63
2.20 Commuting Operators 69
2.21 Another Continuous Spectrum Operator 74
2.22 Problems 80
2.22.1 Simple Basis Vectors 80

2.22.2 Eigenvalues and Eigenvectors 80
2.22.3 Orthogonal Basis Vectors 80
2.22.4 Operator Matrix Representation 80
2.22.5 Matrix Representation and Expectation Value 81
2.22.6 Projection Operator Representation 81
2.22.7 Operator Algebra 81
2.22.8 Functions of Operators 81
2.22.9 A Symmetric Matrix 82
2.22.10 Determinants and Traces 82
2.22.11 Function of a Matrix 82
2.22.12 More Gram-Schmidt 82
2.22.13 Infinite Dimensions 82
2.22.14 Spectral Decomposition 83
2.22.15 Measurement Results 83
2.22.16 Expectation Values 83
2.22.17 Eigenket Properties 84
2.22.18 The World of Hard/Soft Particles 84
2.22.19 Things in Hilbert Space 84
2.22.20 A 2-Dimensional Hilbert Space 85
2.22.21 Find the Eigenvalues 86
2.22.22 Operator Properties 86
2.22.23 Ehrenfest's Relations 86
2.22.24 Solution of Coupled Linear ODEs 87
2.22.25 Spectral Decomposition Practice 87
2.22.26 More on Projection Operators 87

3 Probability **89**

3.1 Probability Concepts 89
3.1.1 Standard Thinking 89
3.1.2 Bayesian Thinking 93
3.2 Probability Interpretation 97
3.3 First hints of "subversive" or "Bayesian" thinking..... 100
3.3.1 The Problem of Prior Probabilities 106
3.4 Testable Information:
The Principle of Maximum Entropy 110
3.5 Discussion 117
3.6 Problems 119
3.6.1 Simple Probability Concepts 119
3.6.2 Playing Cards 120
3.6.3 Birthdays 120
3.6.4 Is there life? 121
3.6.5 Law of large Numbers 121
3.6.6 Bayes 121
3.6.7 Psychological Tests 121
3.6.8 Bayes Rules, Gaussians and Learning 122
3.6.9 Berger's Burgers-Maximum Entropy Ideas 122

3.6.10 Extended Menu at Berger's Burgers 124
3.6.11 The Poisson Probability Distribution 125
3.6.12 Modeling Dice: Observables and Expectation Values . . . 126
3.6.13 Conditional Probabilities for Dice 127
3.6.14 Matrix Observables for Classical Probability 127

4 The Formulation of Quantum Mechanics 129
4.1 Introduction 129
4.2 Two Postulates:
Presentation, Rationalization and Meaning 130
4.3 States and Probabilities 135
4.3.1 States 135
4.3.2 Probabilities 137
4.4 Quantum Pictures 144
4.5 Transformations of States and Observables
The way it must be..... 145
4.5.1 Antiunitary/Antilinear Operators 145
4.5.2 Wigner's Theorem 146
4.5.3 The Transformation Operator and its Generator 147
4.6 The Schrodinger Picture 148
4.7 The Heisenberg Picture 153
4.8 Interaction Picture 154
4.9 Symmetries of Space-Time 156
4.10 Generators of the Group Transformations 158
4.11 Commutators and Identities 161
4.12 Identification of Operators with Observables 166
4.13 Multiparticle Systems 179
4.14 Equations of Motion Revisited and Finished 185
4.15 Symmetries, Conservation Laws and Stationary States 188
4.16 The Collapse or Reduction Postulate? 191
4.17 Putting Some of These Ideas Together 198
4.17.1 Composite Quantum Systems; Tensor Product 198
4.17.2 Quantum Entanglement and the EPR *Paradox* 204
4.17.3 Entanglement and Communication 208
4.17.4 Nonlocality and Tests of Quantum Entanglement 210
4.18 Expanding on the Density Operator and the Statistical Description of Quantum Systems 214
4.18.1 Statistical Description of Quantum Systems and the Nonexistence of an Averaged Quantum State 215
4.18.2 Open Quantum Systems and the Reduced Density Operator 218
4.19 Problems 220
4.19.1 Can It Be Written? 220
4.19.2 Pure and Nonpure States 220
4.19.3 Probabilities 221
4.19.4 Acceptable Density Operators 221

4.19.5 Is it a Density Matrix? . . . 221
4.19.6 Unitary Operators . . . 222
4.19.7 More Density Matrices . . . 222
4.19.8 Scale Transformation . . . 222
4.19.9 Operator Properties . . . 223
4.19.10 An Instantaneous Boost . . . 223
4.19.11 A Very Useful Identity . . . 223
4.19.12 A Very Useful Identity with some help.... . . . 224
4.19.13 Another Very Useful Identity . . . 224
4.19.14 Pure to Nonpure? . . . 224
4.19.15 Schur's Lemma . . . 225
4.19.16 More About the Density Operator . . . 225
4.19.17 Entanglement and the Purity of a Reduced Density Operator . . . 226
4.19.18 The Controlled-Not Operator . . . 226
4.19.19 Creating Entanglement via Unitary Evolution . . . 226
4.19.20 Tensor-Product Bases . . . 227
4.19.21 Matrix Representations . . . 227
4.19.22 Practice with Dirac Language for Joint Systems . . . 228
4.19.23 More Mixed States . . . 228
4.19.24 Complete Sets of Commuting Observables . . . 229
4.19.25 Conserved Quantum Numbers . . . 229

5 How Does It really Work: Photons, K-Mesons and Stern-Gerlach 231

5.1 Introduction . . . 231
5.1.1 Photon Interference . . . 238
5.2 Photon Polarization . . . 243
5.2.1 How Many Basis Sets ? . . . 246
5.2.2 Projection Operators . . . 253
5.2.3 Amplitudes and Probabilities . . . 254
5.2.4 Pure States, Unpure States and Density Operators . . . 258
5.2.5 Unpolarized Light . . . 260
5.2.6 How Does the Polarization State Vector Change? . . . 261
5.2.7 Calculating the Transition Probability . . . 264
5.2.8 Some More Bayesian Thoughts . . . 265
5.3 The Strange World of Neutral K-Mesons . . . 269
5.4 Stern-Gerlach Experiments and Measurement . . . 278
5.5 Problems . . . 289
5.5.1 Change the Basis . . . 289
5.5.2 Polaroids . . . 289
5.5.3 Calcite Crystal . . . 290
5.5.4 Turpentine . . . 290
5.5.5 What QM is all about - Two Views . . . 290
5.5.6 Photons and Polarizers . . . 291

5.5.7 Time Evolution . . . 291
5.5.8 K-Meson oscillations . . . 292
5.5.9 What comes out? . . . 292
5.5.10 Orientations . . . 293
5.5.11 Find the phase angle . . . 293
5.5.12 Quarter-wave plate . . . 294
5.5.13 What is happening? . . . 294
5.5.14 Interference . . . 294
5.5.15 More Interference . . . 295
5.5.16 The Mach-Zender Interferometer and Quantum Interference 296
5.5.17 More Mach-Zender . . . 297

6 Schrodinger Wave equation 1-Dimensional Quantum Systems 299

6.1 The Coordinate Representation . . . 299
6.2 The Free Particle and Wave Packets . . . 306
6.3 Derivation of the Uncertainty Relations in General . . . 310
6.3.1 The Meaning of the Indeterminacy Relations . . . 312
6.3.2 Time-Energy Uncertainty Relations . . . 313
6.4 The Wave Function and Its Meaning . . . 315
6.5 One-Dimensional Systems . . . 319
6.5.1 One-Dimensional Barrier . . . 321
6.5.2 Tunneling . . . 329
6.5.3 Bound States . . . 332
6.5.4 Delta-Function Potentials . . . 346
6.6 Harmonic Oscillators . . . 354
6.6.1 Differential Equation Method . . . 354
6.6.2 Algebraic Method . . . 359
6.7 Green's Functions . . . 369
6.8 Charged Particles in Electromagnetic Fields . . . 375
6.9 Time Evolution of Probabilities . . . 381
6.10 Numerical Techniques . . . 385
6.11 Translation Invariant Potential - Bands . . . 390
6.12 Closed Commutator Algebras and Glauber's Theorem . . . 393
6.13 Quantum Interference with Slits . . . 396
6.13.1 Introduction . . . 396
6.13.2 Probability functions and quantum interference . . . 397
6.13.3 Scattering from a narrow slit . . . 398
6.13.4 Scattering from a double slit(narrow) . . . 399
6.13.5 Scattering from a slit of finite width . . . 400
6.13.6 Scattering from a double finite-width slit . . . 402
6.13.7 Conclusions . . . 404
6.14 Algebraic Methods - Supersymmetric Quantum Mechanics . . . 404
6.14.1 Generalized Ladder Operators . . . 404
6.14.2 Examples . . . 410

6.14.3 Generalizations . . . 413
6.14.4 Examples . . . 416
6.14.5 Supersymmetric Quantum Mechanics . . . 424
6.14.6 Additional Thoughts . . . 425
6.15 Problems . . . 426
6.15.1 Delta function in a well . . . 426
6.15.2 Properties of the wave function . . . 426
6.15.3 Repulsive Potential . . . 427
6.15.4 Step and Delta Functions . . . 427
6.15.5 Atomic Model . . . 427
6.15.6 A confined particle . . . 428
6.15.7 $1/x$ potential . . . 429
6.15.8 Using the commutator . . . 429
6.15.9 Matrix Elements for Harmonic Oscillator . . . 429
6.15.10 A matrix element . . . 429
6.15.11 Correlation function . . . 430
6.15.12 Instantaneous Force . . . 430
6.15.13 Coherent States . . . 430
6.15.14 Oscillator with Delta Function . . . 430
6.15.15 Measurement on a Particle in a Box . . . 432
6.15.16 Aharonov-Bohm experiment . . . 432
6.15.17 A Josephson Junction . . . 433
6.15.18 Eigenstates using Coherent States . . . 434
6.15.19 Bogoliubov Transformation . . . 435
6.15.20 Harmonic oscillator . . . 435
6.15.21 Another oscillator . . . 435
6.15.22 The coherent state . . . 436
6.15.23 Neutrino Oscillations . . . 437
6.15.24 Generating Function . . . 438
6.15.25 Given the wave function 438
6.15.26 What is the oscillator doing? . . . 438
6.15.27 Coupled oscillators . . . 439
6.15.28 Interesting operators 439
6.15.29 What is the state? . . . 439
6.15.30 Things about a particle in a box . . . 439
6.15.31 Handling arbitrary barriers..... . . . 440
6.15.32 Deuteron model . . . 442
6.15.33 Use Matrix Methods . . . 442
6.15.34 Finite Square Well Encore . . . 443
6.15.35 Half-Infinite Half-Finite Square Well Encore . . . 443
6.15.36 Nuclear α Decay . . . 444
6.15.37 One Particle, Two Boxes . . . 445
6.15.38 A half-infinite/half-leaky box . . . 447
6.15.39 Neutrino Oscillations Redux . . . 448
6.15.40 Is it in the ground state? . . . 449
6.15.41 Some Thoughts on T-Violation . . . 449

6.15.42 Kronig-Penney Model 450
6.15.43 Operator Moments and Uncertainty 451
6.15.44 Uncertainty and Dynamics 451

7 Angular Momentum; 2- and 3-Dimensions 453
7.1 Angular Momentum Eigenvalues and Eigenvectors 453
7.1.1 Derivation of Eigenvalues 455
7.2 Transformations and Generators; Spherical Harmonics 458
7.2.1 Eigenfunctions; Eigenvalues; Position Representation . . . 463
7.3 Spin 466
7.3.1 Spin 1/2 467
7.3.2 Superselection Rules 476
7.3.3 A Box with 2 Sides 479
7.4 Magnetic Resonance 483
7.4.1 Spin Resonance 489
7.5 Addition of Angular Momentum 492
7.5.1 Addition of Two Spin = 1/2 Angular Momenta 492
7.5.2 General Addition of Two Angular Momenta 498
7.5.3 Actual Construction of States 501
7.6 Two- and Three-Dimensional Systems 504
7.6.1 2- and 3-Dimensional Infinite Wells 505
7.6.2 Two-Dimensional Infinite Circular Well 508
7.6.3 3-Dimensional Finite Well 514
7.6.4 Two-Dimensional Harmonic Oscillator 519
7.6.5 What happens in 3 dimensions? 526
7.6.6 Two-Dimensional Finite Circular Well 528
7.6.7 The 3-Dimensional Delta Function 529
7.6.8 The Hydrogen Atom 532
7.6.9 Algebraic Solution of the Hydrogen Atom 537
7.6.10 The Deuteron 540
7.6.11 The Deuteron - Another Way 542
7.6.12 Linear Potential 544
7.6.13 Modified Linear Potential and Quark Bound States 546
7.7 Problems 548
7.7.1 Position representation wave function 548
7.7.2 Operator identities 549
7.7.3 More operator identities 549
7.7.4 On a circle 549
7.7.5 Rigid rotator 549
7.7.6 A Wave Function 549
7.7.7 $L = 1$ System 550
7.7.8 A Spin-3/2 Particle 550
7.7.9 Arbitrary directions 551
7.7.10 Spin state probabilities 552
7.7.11 A spin operator 552

7.7.12 Simultaneous Measurement 553
7.7.13 Vector Operator 553
7.7.14 Addition of Angular Momentum 553
7.7.15 Spin = 1 system 554
7.7.16 Deuterium Atom 554
7.7.17 Spherical Harmonics 555
7.7.18 Spin in Magnetic Field 555
7.7.19 What happens in the Stern-Gerlach box? 555
7.7.20 Spin = 1 particle in a magnetic field 556
7.7.21 Multiple magnetic fields 556
7.7.22 Neutron interferometer 556
7.7.23 Magnetic Resonance 557
7.7.24 More addition of angular momentum 557
7.7.25 Clebsch-Gordan Coefficients 557
7.7.26 Spin−1/2 and Density Matrices 557
7.7.27 System of N Spin−1/2 Particle 558
7.7.28 In a coulomb field 559
7.7.29 Probabilities 559
7.7.30 What happens? 559
7.7.31 Anisotropic Harmonic Oscillator 560
7.7.32 Exponential potential 560
7.7.33 Bouncing electrons 561
7.7.34 Alkali Atoms 562
7.7.35 Trapped between 562
7.7.36 Logarithmic potential 562
7.7.37 Spherical well 562
7.7.38 In magnetic and electric fields 563
7.7.39 Extra(Hidden) Dimensions 563
7.7.40 Spin−1/2 Particle in a D-State 571
7.7.41 Two Stern-Gerlach Boxes 571
7.7.42 A Triple-Slit experiment with Electrons 572
7.7.43 Cylindrical potential 572
7.7.44 Crazy potentials..... 573
7.7.45 Stern-Gerlach Experiment for a Spin-1 Particle 573
7.7.46 Three Spherical Harmonics 574
7.7.47 Spin operators ala Dirac 574
7.7.48 Another spin = 1 system 574
7.7.49 Properties of an operator 574
7.7.50 Simple Tensor Operators/Operations 575
7.7.51 Rotations and Tensor Operators 575
7.7.52 Spin Projection Operators 575
7.7.53 Two Spins in a magnetic Field 575
7.7.54 Hydrogen d States 576
7.7.55 The Rotation Operator for Spin−1/2 576
7.7.56 The Spin Singlet 576
7.7.57 A One-Dimensional Hydrogen Atom 577

7.7.58 Electron in Hydrogen p–orbital . . . 577
7.7.59 Quadrupole Moment Operators . . . 579
7.7.60 More Clebsch-Gordon Practice . . . 580
7.7.61 Spherical Harmonics Properties . . . 580
7.7.62 Starting Point for Shell Model of Nuclei . . . 580
7.7.63 The Axial-Symmetric Rotor . . . 581
7.7.64 Charged Particle in 2-Dimensions . . . 582
7.7.65 Particle on a Circle Again . . . 583
7.7.66 Density Operators Redux . . . 583
7.7.67 Angular Momentum Redux . . . 584
7.7.68 Wave Function Normalizability . . . 584
7.7.69 Currents . . . 584
7.7.70 Pauli Matrices and the Bloch Vector . . . 585

8 Time-Independent Perturbation Theory 587
8.1 Nondegenerate Case . . . 587
8.1.1 Rayleigh-Schrodinger Perturbation Theory . . . 587
8.2 Degenerate Case . . . 595
8.2.1 More Ideas about Perturbation Methods . . . 604
8.2.2 Thoughts on Degeneracy and Position Representation . . 611
8.3 Spin-Orbit Interaction - Fine Structure . . . 613
8.3.1 Thomas Precession . . . 614
8.4 Another Relativity Correction . . . 615
8.5 External Fields - Zeeman and Stark Effects; Hyperfine Structure 616
8.5.1 Zeeman Effect . . . 616
8.5.2 Stark Effect . . . 618
8.5.3 Hyperfine Structure . . . 619
8.6 Examples . . . 619
8.6.1 Spin-Orbit, Relativity, Zeeman Effect in Hydrogen Atom 619
8.6.2 Spin-Orbit and Arbitrary Magnetic Field . . . 632
8.6.3 Wigner-Eckart Theorem . . . 636
8.6.4 Paschen-Bach Effect . . . 640
8.6.5 Stark Effect . . . 644
8.6.6 Van der Waal's Interaction . . . 647
8.7 Variational Methods . . . 654
8.8 2^{nd}-Order Degenerate Perturbation Theory . . . 656
8.9 Problems . . . 662
8.9.1 Box with a *Sagging Bottom* . . . 662
8.9.2 Perturbing the Infinite Square Well . . . 662
8.9.3 Weird Perturbation of an Oscillator . . . 663
8.9.4 Perturbing the Infinite Square Well Again . . . 663
8.9.5 Perturbing the 2-dimensional Infinite Square Well . . . 663
8.9.6 Not So Simple Pendulum . . . 664
8.9.7 1-Dimensional Anharmonic Oscillator . . . 664
8.9.8 A Relativistic Correction for Harmonic Oscillator . . . 664

8.9.9 Degenerate perturbation theory on a spin = 1 system . . . 665
8.9.10 Perturbation Theory in Two-Dimensional Hilbert Space . 665
8.9.11 Finite Spatial Extent of the Nucleus 666
8.9.12 Spin-Oscillator Coupling 666
8.9.13 Motion in spin-dependent traps 667
8.9.14 Perturbed Oscillator 668
8.9.15 Another Perturbed Oscillator 668
8.9.16 Helium from Hydrogen - 2 Methods 668
8.9.17 Hydrogen atom + xy perturbation 669
8.9.18 Rigid rotator in a magnetic field 669
8.9.19 Another rigid rotator in an electric field 669
8.9.20 A Perturbation with 2 Spins 669
8.9.21 Another Perturbation with 2 Spins 670
8.9.22 Spherical cavity with electric and magnetic fields 670
8.9.23 Hydrogen in electric and magnetic fields 670
8.9.24 $n = 3$ Stark effect in Hydrogen 670
8.9.25 Perturbation of the $n = 3$ level in Hydrogen - Spin-Orbit and Magnetic Field corrections 670
8.9.26 Stark Shift in Hydrogen with Fine Structure 671
8.9.27 2-Particle Ground State Energy 672
8.9.28 1s2s Helium Energies 672
8.9.29 Hyperfine Interaction in the Hydrogen Atom 672
8.9.30 Dipole Matrix Elements 673
8.9.31 Variational Method 1 673
8.9.32 Variational Method 2 675
8.9.33 Variational Method 3 675
8.9.34 Variational Method 4 675
8.9.35 Variation on a linear potential 675
8.9.36 Average Perturbation is Zero 676
8.9.37 3-dimensional oscillator and spin interaction 676
8.9.38 Interacting with the Surface of Liquid Helium 676
8.9.39 Positronium + Hyperfine Interaction 676
8.9.40 Two coupled spins 677
8.9.41 Perturbed Linear Potential 678
8.9.42 The ac-Stark Effect 678
8.9.43 Light-shift for multilevel atoms 680
8.9.44 A Variational Calculation 682
8.9.45 Hyperfine Interaction Redux 682
8.9.46 Find a Variational Trial Function 683
8.9.47 Hydrogen Corrections on 2s and 2p Levels 683
8.9.48 Hyperfine Interaction Again 684
8.9.49 A Perturbation Example 684
8.9.50 More Perturbation Practice 684

Chapter 1

Motivation, Motivation, Motivation .....

In this book we will develop a theory, which is deeply rooted in experiment and can only be understood using a new mathematical language. It not only describes how nature works in all microscopic physical systems, but also in macroscopic physical systems.

It is most important, however, when the characteristic length scale of the physical system is smaller than $10^{-4}\,m$.

This theory is called **quantum mechanics**.

The physical world we will discover in our studies is a strange and fascinating place where nature operates in a way that seems to defy the intuition we have built up living among macroscopic systems. We will not endeavor to explain why nature works in this particular way since it is my strong belief, as we will see in this book, that it is not possible to do so within the context of this theory. We will, however, be able to correctly predict the outcomes of an amazingly wide range of experiments in many different fields of physics, chemistry and biology.

Let me emphasize my strong belief that theories in physics should only endeavor to make predictions about experimental measurements and not attempt to provide reasons for *why* nature works in these particular ways, that is, why we must choose to start with certain postulates.

Feynman put it best.....

> We know how the electrons and light behave. But what can I call it? If I say they behave like particles I give the wrong impression; also if I say they behave like waves. They behave in their own inimitable way, which technically could be called a quantum mechanical way. They behave in a way that is like nothing that you

> have seen before. Your experience with things that you have seen before is incomplete. The behavior of things on a very small scale is simply different. An atom does not behave like a miniature representation of the solar system with little planets going around in orbits. Nor does it appear like a cloud or fog of some sort surrounding the nucleus. It behaves like nothing you have ever seen before.
>
> There is one simplification at least. Electrons behave in this respect exactly the same way as photons; they are both screwy, but in exactly the same way.
>
> The difficulty really is psychological and exists in the perpetual torment that results from your saying to yourself ''but how can it really be like that?'', which is a reflection of an uncontrolled but vain desire to see it in terms of something familiar. I will not describe it in terms of an analogy with something familiar; I will simply describe it...
>
> I am going to tell you what nature behaves like. If you will simply admit that maybe she does behave like this, you will find her a delightful and entrancing thing. Do not keep saying to yourself, if you can avoid it, ''but how can it really be like that?'' because you will get ''down the drain'', into a blind alley from which nobody has yet escaped.
>
> Nobody knows how it can be like that.

and we can add an addendum .....

> and ''nobody knows why it is like that''

We will not be able to reduce the quantum universe to everyday ways of thinking(usually called *common sense*). In fact, in order to understand the ideas and implications of the theory we will have to adjust all of our ways of thinking at the most fundamental level.

Imagine, for a moment, that you are attempting to understand a new culture. If you are serious about it, the first thing you would do is to learn the language appropriate to that culture so that you can put your experiences in the proper context. Understanding the universe of quantum phenomena is much like a understanding a new culture where the appropriate language is mathematics and the experiences we are attempting to put into context are experiments.

As we shall see, we will have to use a mathematical language to describe the

quantum world since ordinary language, which was developed to explain everyday occurrences(experiments on macroscopic objects), will turn out to be totally inadequate.

Since it makes no sense to attempt any understanding of the nature of quantum phenomena without first learning to speak and use the language of the quantum world, we will spend the first several chapters of this book learning the appropriate mathematics, in particular, the subject of linear vector spaces.

The adjustment of our ways of thinking at the fundamental level that will be needed is not simply a mathematical matter, however. The development of the necessary mathematical language will not come into conflict with our everyday modes of thinking in any major way. Although, the mathematics of linear vector spaces is very elegant, you will be able to understand it without much difficulty and without having your basic view of the world changed at any fundamental level.

You will be troubled, however, when we apply the mathematics to physical systems that develop according to quantum ideas. We will attach physical meaning to the mathematical formalism in ways that will conflict with your well-developed views(I will call these classical views) about how the world works.

We will rethink both mathematics and quantum theory using the Dirac language, which, as we shall see, incorporates the very nature of the quantum world in an intrinsic and natural way. Since we are attempting to develop a physical theory, we will link all the mathematical concepts that we introduce to physical concepts as we proceed.

Dirac was able to link the physical structure of quantum mechanics with the mathematical structure in a unique way. His mathematical language incorporates the physical meaning directly into the formalism and the calculational methods. The language explicitly exhibits the physics and clearly exposes the internal logic of quantum mechanics. Once we understand the language, every equation will directly convey its physical meaning without the need for further explanation or any need for inadequate models.

It is very important to understand that the Dirac language is not simply a new notation for quantum mechanics(as many physicists seem to think). It is a way of thinking about quantum mechanics. It will allow us to use the physical ideas of quantum mechanics to develop the appropriate mathematical language rather than the other way around. This will allow the very mathematical quantum theory to be more closely connected to experiment than any other physical theory.

These statements about the importance of understanding the mathematical language appropriate to the physics under consideration do not only apply to the quantum world. It is true for all areas of physics and other sciences. One should

always learn the appropriate language before studying any field that relies on that language for its understanding.

Mathematical Methods

AsI stated, will have to use various mathematical techniques in order to formulate quantum mechanics and in order to solve the resulting equations. Some of these techniques you have may have learned in a Mathematical Methods in Physics course, some you may have learned in Mathematics courses such as Linear Algebra, and Multivariable Calculus and some we will develop in this book.

The techniques we will need are linear operators in Hilbert space (an extension of linear algebra), partial differential equations, special functions of mathematical physics (Legendre polynomials, spherical harmonics, Bessel functions), Fourier transforms, use of Green's functions, integral equations, contour integration in the complex plane, group theory and group representations, etc. We will cover many of these techniques in detail as we proceed.

As we proceed we will cover many applications of quantum theory to physical systems in the areas of atomic and molecular physics.

Chapter 2

The Mathematics of Quantum Physics:

Dirac Language

Because it is so important, I repeat that this study of the mathematics of quantum mechanics assumes that you have a good background in the following areas:

- Calculus
  Differentiation and integration
- Infinite series
  Taylor and power series
- Linear Algebra
  Linear vector spaces and matrices
- Multivariable Calculus
  Partial derivatives
  Gradient, divergence, curl and Laplacian
- Mathematical Methods
  Ordinary and partial differential equations
  Fourier series and Fourier transforms

This study of the mathematical formalism of quantum theory will center around the subject of linear vector spaces. We will present this subject using Dirac language and connect it to the physics as we proceed. At the start of our discussion, we will mix standard mathematical notation and the Dirac language so that the transition to using only the Dirac language will be easier.

Quantum systems cannot be described in terms of our everyday experience. To understand them it is necessary to use the abstract language provided by the algebra of linear operators on Hilbert space.

When we present the mathematical formalism appropriate to a physical theory we have two choices. We can approach the subject abstractly and deal directly

with mathematical quantities of fundamental importance or we can choose a particular a particular representation (abstract coordinate system) and work with the numbers (functions) that correspond to the fundamental quantities in that coordinate representation.

We will follow the abstract approach in most cases since it will allow us to delve more deeply into the fundamental nature of the physical theory and, in addition, enable us to treat the physical laws in a precise and efficient manner.

First let us define *mappings* which are used almost everywhere in the discussion, sometimes explicitly and sometimes implicitly.

2.1 Mappings

Given two sets A and B, let $a \in A$, $b \in B$. A mapping T from A to B:

$$a \mapsto b = Ta \tag{2.1}$$

can be of the following types:

1. T is a mapping of A *into* B if to each $a \in A$ there corresponds a definite element $Ta \in B$(there may be elements of B that do not correspond to any element of A and different elements of A may correspond to the same element of B). The range of the mapping is the subset $TA \subset B$ (TA is a subset set of B but not equal to B) formed by those elements that correspond to some elements of A.

2. T is a mapping of A *onto* B if to each element of A there corresponds a definite element $Ta \in B$ and to every element of B there corresponds at least one element of A (in this case $Ta = B$).

3. A *one-to-one* mapping is where distinct elements of A correspond to distinct elements of B: $Ta \neq Tb$ if $a \neq b$.

4. It follows that if T is a one-to-one mapping from A *onto* B, then there exists an inverse one-to-one mapping T^{-1} from B *onto* A. Such an inverse does not exist if T is a one-to-one mapping A *into* B.

2.2 Linear Vector Spaces

The mathematical formalism of quantum mechanics is based on the theory of linear vector spaces. In this chapter we shall present a complete exposition of the mathematics of linear vector spaces. We shall state theorems without proofs (excellent proofs are available in a variety of texts). Instead of proofs, we shall relate the theorems where appropriate to the physics of quantum mechanics using the Dirac language and provide concrete examples that will help us understand the physical content in our later discussions.

The number of configurations experimental instruments can exhibit is finite. This implies that, in principle, only the language of finite-dimensional vector spaces will be needed to explain experimental results and to understand the theoretical structure of quantum mechanics. However, if we want to embed the theory in a spacetime continuum, then it will be necessary to consider idealized instruments capable of an infinite number of configurations. This will require a description using the language of infinite-dimensional spaces, in particular the use of a vector space called a non-separable or rigged Hilbert space. Because these idealized infinite instruments are approximations of the actual finite ones, physicists usually ignore those properties of the infinite-dimensional Hilbert space that cannot be derived from the properties of finite-dimensional spaces by some, not necessarily unique, physically based limiting procedure.

A working knowledge of the mathematical description that results from the adopted limiting procedure is necessary to understand many of the developments of quantum mechanics. The following mathematical presentation reflects these considerations. The results pertaining to finite-dimensional spaces, necessary for the understanding of the structure of quantum mechanics, are presented with thoroughness. The generalization to infinite-dimensional spaces, which is a very difficult task, is discussed in less detail.

These mathematical details are usually ignored in most textbooks, which I believe makes it very difficult to understand the fundamental ideas underlying quantum mechanics.

In most of our discussions we can assume we are in a finite-dimensional vector space and the results will generalize without change to the Hilbert space case. We will deal with the special problems associated with the infinite dimensionality of Hilbert space as they arise.

2.3 Linear Spaces and Linear Functionals

A vector space is defined with reference to a field. We consider the case where the field is the field C of complex numbers because this is the case of interest in quantum mechanics.

Ket Vectors
The mathematical vectors that will allow us to describe physical states will be called *ket vectors* or *kets* and be represented by the symbol $|\ldots\rangle$ (due to Dirac). We will label different vectors(states) according to their associated physical(measurable) properties and these will be inserted inside the ket symbol $|a, b, ...\rangle$. The ket vectors will form the basis for the Dirac language of the associated vector space defined below.

These simple mathematical objects will turn out to have sufficient mathematical

structure to allow us to represent all of the important features of quantum theory in terms of them. Whether these mathematical objects have any objective physical content themselves must be discussed later as we develop the theory.

At this point, however, let us state a couple of the basic properties of the ket vectors and their connection to physics to help set the stage for the discussion that will follow.

As with any other vectors, kets can be multiplied by complex numbers and can be added together to get another ket vector

$$c_1 |a\rangle + c_2 |b\rangle = |q\rangle \tag{2.2}$$

where c_1 and c_2 are complex numbers.

The crucial assumptions we will make later as we connect the Dirac language, the mathematical formalism and the physics are:

- Each state of a physical system at a given time can be mathematically described by a definite ket vector. There is some correspondence!
- If a physical state is a superposition(sum) of other physical states, its corresponding ket vector is a linear sum(combination) of the ket vectors corresponding to the other states

The state $|q\rangle$ in (2.2) is a superposition of the states $|a\rangle$ and $|b\rangle$ with the mathematical properties of this superposition defined precisely by the two complex numbers c_1 and c_2. For example, the state of a photon passing through a double-slit apparatus might be described by the superposition

$$|photon\ in\ 2-slit\ apparatus\rangle = a\,|slit\ 1\rangle + b\,|slit\ 2\rangle \tag{2.3}$$

where $|slit\ 1\rangle$ is the state of a photon that passes through slit 1.

With these tantalizing thoughts rambling about in your minds, let us turn to the mathematical formalism.

Definition:

A *linear vector space* V is a set of abstract elements, called *vectors*,

$$|1\rangle, |2\rangle, \ldots, |M\rangle, \ldots, |N\rangle, \ldots \tag{2.4}$$

for which there exists two rules:

1. a rule for creating a vector sum or vector addition,

$$|7\rangle + |12\rangle$$

2. a rule for multiplication of a vector by a complex scalar c,

$$c\,|13\rangle$$

The following properties hold:

1. the result of applying either or both of these rules is another vector in the same space; this is called *closure*

$$c_1\,|7\rangle + c_2\,|12\rangle \in V$$

2. scalar multiplication is *distributive* with respect to the vectors

$$c(|7\rangle + |12\rangle) = c\,|7\rangle + c\,|12\rangle$$

3. scalar multiplication is *associative*

$$c_1(c_2\,|7\rangle) = c_1 c_2\,|7\rangle$$

4. addition is *commutative*

$$|7\rangle + |12\rangle = |12\rangle + |7\rangle$$

5. addition is *associative*

$$|6\rangle + (|7\rangle + |12\rangle) = (|6\rangle + |7\rangle) + |12\rangle$$

6. there exists a unique *zero* or *null* vector $|\varnothing\rangle$ such that

$$|7\rangle + |\varnothing\rangle = |7\rangle \text{ and } 0\,|M\rangle = |\varnothing\rangle$$

7. for every vector $|M\rangle$ there exists a unique additive *inverse* vector $|-M\rangle$ where

$$|M\rangle + |-M\rangle = |\varnothing\rangle \text{ and } |-M\rangle = -\,|M\rangle$$

Example: Consider the set of all *3-tuples*, which is a particular example of a finite-dimensional vector space

$$|i\rangle = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} \tag{2.5}$$

Addition is defined by

$$|i\rangle + |i\rangle = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} + \begin{pmatrix} a_j \\ b_j \\ c_j \end{pmatrix} = \begin{pmatrix} a_i + a_j \\ b_i + b_j \\ c_i + c_j \end{pmatrix} \tag{2.6}$$

Multiplication by a scalar q is defined by

$$q\,|i\rangle = \begin{pmatrix} qa_i \\ qb_i \\ qc_i \end{pmatrix} \tag{2.7}$$

The null vector is defined by

$$|\varnothing\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.8}$$

We must emphasize that use of the word "vector" here is not meant to imply that we are talking about a mathematical object that in any way needs to possess properties such as magnitude and direction. These "vectors" are abstract mathematical objects whose properties are defined above.

Other examples are:

1. the set of all functions $f(x)$ of a real variable x for which

   $$\int \; |\, f(x)\,|^2 \; dx < \infty$$

   with addition and multiplication by a scalar defined by

   $$(f+g)(x) = f(x) + g(x) \; and \; (af)(x) = af(x)$$

   This space is called L^2.

2. the set of all infinite sequences of numbers $x = (x_1, x_2, x_3, \ldots, x_i, \ldots)$ such that

   $$x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots, x_k + y_k, \ldots)$$

   with addition and multiplication by a scalar defined by

   $$(f+g)(x) = f(x) + g(x) \qquad (af)(x) = af(x)$$

   $$ax = (ax_1, ax_2, ax_3, \ldots, ax_k, \ldots)$$

3. the set of all $2x2$ matrices

   $$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

   with addition and multiplication by a scalar defined by

   $$A + B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

   $$cA = c\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{pmatrix}$$

These examples should make it clear that it is important, as we have emphasized already, not to think about magnitudes and directions as being the defining properties of "vectors" in any way. We must be very careful not to attribute any specific properties derived from special examples as general properties of vectors in any abstract vector space.

Isomorphism -Two vector spaces U and V (defined over the same field) are said to be *isomorphic* if there exists a one-to-one and onto correspondence$x u \leftrightarrow v$ between every vector $u \in U$ and every other vector $v \in V$ that preserves the linear operations, that is, $\alpha u_1 + \beta u_2 \leftrightarrow \alpha v_1 + \beta v_2$ whenever $u_1 \leftrightarrow v_1$ and $u_2 \leftrightarrow v_2$ for $u_1, u_2 \in U$ and $v_1, v_2 \in V$. Such a correspondence is called an *isomorphism*. If one is only interested in linear operations on vectors, then two isomorphic spaces are indistinguishable from each other.

Definition: A set of vectors is said to be *linearly independent* if a linear relation of the form

$$\sum_{k=1}^{n} c_k \left|k\right\rangle = \left|\varnothing\right\rangle \tag{2.9}$$

implies that $c_k = 0$ for all k; otherwise the set of vectors is *linearly dependent*.

If a set of vectors is linearly dependent, then we can express a member of the set as a linear combination of the other members of the set.

Examples:

1. Consider the set of vectors(3-tuples in this case)

$$\left|1\right\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \left|2\right\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \left|3\right\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

   This set is linearly independent since

$$a_1 \left|1\right\rangle + a_2 \left|2\right\rangle + a_3 \left|3\right\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \left|\varnothing\right\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

   implies that the only solution is

$$a_1 = a_2 = a_3 = 0$$

2. Consider the set of vectors(3-tuples in this case)

$$\left|1\right\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \left|2\right\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad \left|3\right\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ a_1 - a_2 \\ a_3 \end{pmatrix} = |\varnothing\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the only solution is

$$a_1 + a_2 = a_1 - a_2 = a_3 = 0 \qquad \text{or} \qquad a_1 = a_2 = a_3 = 0$$

3. Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 2a_3 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 + 2a_3 \\ a_1 - a_2 \\ a_1 + a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the solution is

$$a_1 + a_2 + 2a_3 = a_1 - a_2 = a_1 + a_3 = 0 \qquad \text{or} \qquad a_2 = a_1, a_3 = -a_1$$

Note: For simplicity and in all cases where no ambiguity will arise, we will simplify the notation for the "null" vector from now on. We will write

$$\sum_{k=1}^{n} c_k |k\rangle = 0 \qquad \text{instead of} \qquad \sum_{k=1}^{n} c_k |k\rangle = |\varnothing\rangle \tag{2.10}$$

We say that an infinite set of vectors is linearly independent if **every finite subset** is linearly independent. Alternatively, we can use this method: if the determinant of the matrix formed by using the vectors as columns is equal to zero, then the vectors are linearly dependent.

Definition: The maximum number of linearly independent vectors in a space V is called the *dimension* of the space $dim(V)$.

Definition: A set of vectors $|k\rangle, k = 1, 2, 3, \ldots, n$ *spans* the space if every vector $|Q\rangle$ in the space can be written as a linear combination of vectors in the set

$$|Q\rangle = \sum_{k=1}^{n} q_k |k\rangle \tag{2.11}$$

This linear combination, which is given by the coefficients $q_k, k = 1, 2, \ldots, n$ is unique.

Definition: A set of vectors is a *basis* for the space if it is a linearly independent set *and* spans the space, that is, if $dim(V) = m$, a set of m linearly independent vectors is called a *basis* on V.

The set of vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.12}$$

is the *maximal* set of linearly independent vectors since any other vector $|g\rangle$ in the space can always be written as a linear combination of them as

$$|g\rangle = a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \tag{2.13}$$

Therefore the dimension of this vector space is 3. This set of vectors is also a basis. The basis is *not unique* since the set of linearly independent vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.14}$$

also spans the space, i.e.,

$$|g\rangle = c_1 |1\rangle + c_2 |2\rangle + c_3 |3\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 \\ c_1 + c_2 \\ c_3 \end{pmatrix} \tag{2.15}$$

implies that

$$2c_1 = a_1 + a_2 \quad 2c_2 = a_1 - a_2 \quad c_3 = a_3 \tag{2.16}$$

and, thus, this set is also a basis. Clearly, a basis spans the whole of V.

Definition: The coefficients in the expansion of an arbitrary vector $|Q\rangle$ in terms of a basis set $|k\rangle, k = 1, 2, 3, \ldots, n$

$$|Q\rangle = \sum_{k=1}^{n} q_k |k\rangle \tag{2.17}$$

are called the *components* of the vector $|Q\rangle$ in that basis.

Example: In the space of 3-tuples, a basis is represented by the three vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.18}$$

so that an arbitrary vector in the space can be written

$$|g\rangle = a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \tag{2.19}$$

so that a_1, a_2 and a_3 are the components.

When we add vectors(must be defined with respect to the same basis) we simply add their components

$$|Q\rangle + |R\rangle = \sum_{k=1}^{n} q_k|k\rangle + \sum_{k=1}^{n} r_k|k\rangle = \sum_{k=1}^{n} (q_k + r_k)|k\rangle \tag{2.20}$$

Subspaces - A subset of a vector space V, which is also a vector space, is a *subspace*, that is, it contains all the linear combinations of any number of its vectors - it said to be *closed*. The smallest subspace that contains the set S of vectors is said to be *spanned* by S. We note that the intersection $M \cap N$ - in the sense of set theory - of two subspaces M and N is a subspace, but, in general, their union $M \cup N$ is not.

2.4 Inner Products

The vector spaces we have been discussing do not need to contain vectors that have a well-defined length or direction in the ordinary sense (remember the example of the $2x2$ matrices). We can, however, define *generalizations* of length and direction in an arbitrary space using a mathematical object called the *inner product*. The inner product is a generalization of the standard dot product. I will first discuss the inner product using *standard* mathematical notation and then return to the *Dirac language*.

Definition: An *inner product* for a linear vector space associates a scalar (f, g) with every *ordered pair* of vectors f, g. It satisfies these properties:

1. $$(f, g) = \text{ complex number}$$

2. $$(f, g) = (g, f)^*$$

3. $$(f, c_1 g_1 + c_2 g_2) = c_1(f, g_1) + c_2(f, g_2)$$

4. $$(f, f) \geq 0 \text{ with equality if and only if } f = \text{ null vector}$$

Now(2) and (3) above imply that

$$(c_1 g_1 + c_2 g_2, f) = c_1^*(f, g_1) + c_2^*(f, g_2) \tag{2.21}$$

Hence, the inner product is said to be *linear* in its *second* argument and *anti-linear* in its *first* argument.

Definition: The non-negative number $\|f\| = (f, f)^{1/2}$ is called the *norm* or *length* of the vector f. Clearly, $\|f\| = 0$ if and only if $f = 0$.

Definition: If the inner product of two vectors is zero, then the vectors are *orthogonal*.

Definition: A set of vectors $\{f_i\}$ is said to be *orthonormal* if the vectors are pairwise orthogonal and of unit norm. We can summarize this property by the equation

$$(f_i, f_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2.22}$$

where the symbol δ_{ij} is called the *Kronecker delta*.

Schwarz's Inequality: scalar products satisfy

$$|\,(f, g)\,|^2 \leq (f, f)(g, g) \tag{2.23}$$

Triangle Inequality: scalar products satisfy

$$\|(f + g)\| \leq \|f\| + \|g\| \tag{2.24}$$

Equality holds in both cases only if one vector is a scalar multiple of the other, i.e., $f = cg$ where the scalar c is real and positive.

An inner product space is simply one in which an inner product is defined. A *unitary* space is a finite-dimensional inner product vector space. In fact, every finite-dimensional space can be made into a unitary space.

Examples:

1. For the case of n-tuples(an n-dimensional vector space)

$$f = (x_1, x_2, x_3, \ldots, x_n), \quad , \quad g = (y_1, y_2, y_3, \ldots, y_n)$$
$$(f, g) = \sum_{k=1}^{n} x_k^* y_k$$

2. For the case of square integrable functions

$$(f, g) = \int f^*(x) g(x) w(x)\, dx$$

where $w(x)$ is a nonnegative weight function.

When a vector Q is expanded as a linear combination

$$Q = \sum_k q_k f_k \tag{2.25}$$

of orthonormal basis vectors f_k, the coefficients or components are given by the inner product, i.e.,

$$(f_j, Q) = \sum_k q_k (f_j, f_k) = \sum_k \delta_{kj} = q_j \tag{2.26}$$

or

$$Q = \sum_k (f_k, Q) f_k \tag{2.27}$$

Of the two basis vector sets for the 3-dimensional vectors we looked at earlier:

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.28}$$

is an orthonormal basis and

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.29}$$

is orthogonal but not orthonormal.

For later use we note the following.

Given two unitary spaces U and V, their *direct sum* $W = U \oplus V$ is the unitary space consisting of the pairs $\{f.g\}$ written $f \oplus g, f \in U, g \in V$ with the operations defined by

$$\alpha_1 f_1 \oplus g_1 + \alpha_2 f_2 \oplus g2 = (\alpha_1 f_1 + \alpha_2 f_2) \oplus (\alpha_1 g_1 + \alpha_2 g_2) \tag{2.30}$$

$$(f_1 \oplus g_1, f_2 \oplus g2) = (f_1, f_2) + (g_1, g_2) \tag{2.31}$$

This is a definition that can clearly be extended to any finite number of spaces.

2.5 Linear Functionals

Linear complex-valued functions of vectors are called *linear functionals* on V. A linear functional Γ assigns a scalar value $\Gamma(f)$ to each vector f in the vector space V, such that linearity holds

$$\Gamma(c_1 f_1 + c_2 f_2) = c_1 \Gamma(f_1) + c_2 \Gamma(f_2) \tag{2.32}$$

for any vectors f_1 and f_2 and any scalars c_1 and c_2.

The set of linear functionals on V form a linear space V' where we define addition and scalar multiplication by

$$(\Gamma_1 + \Gamma_2)(f) = \Gamma_1(f) + \Gamma_2(f) \tag{2.33}$$

$$(c\Gamma)(f) = c\Gamma(f) \tag{2.34}$$

The vector space V' is called the *dual space* of the vector space V.

There is a one-to-one correspondence between linear functionals Γ in V' and the vectors f in V, such that the linear functional associated with f has the form

$$\Gamma_f(g) = (f, g) \tag{2.35}$$

where f is a fixed vector and g is an arbitrary vector. This is called the *Riesz theorem*.

Using the properties of the inner product we can show that

$$\Gamma_f + \Gamma_g = \Gamma_{f+g} \qquad \text{and} \qquad a\Gamma_f = \Gamma_{a^* f} \tag{2.36}$$

or

$$\Gamma_f(h) + \Gamma_g(h) = \Gamma_{(f+g)}(h) \tag{2.37}$$

$$(f, h) + (g, h) = (f + g, h) \tag{2.38}$$

and

$$a\Gamma_f(h) = a(f, h) = (a^* f, h) = \Gamma_{a^* f} \tag{2.39}$$

The scalar product is clearly *antilinear* in Γ and linear in f.

The vector f that corresponds to a given linear functional Γ_f is easily found by direct construction.

Let $\{\alpha^i\}$ be a set of orthonormal basis vectors in V (this means that $(\alpha_i, \alpha_j) = \delta_{ij}$, and let

$$\phi = \sum_n c_n \alpha_n \tag{2.40}$$

be an arbitrary vector in V. Then we have from its definition

$$\Gamma_f(\phi) = \Gamma_f\left(\sum_n c_n \alpha_n\right) = \sum_n c_n \Gamma_f(\alpha_n) \tag{2.41}$$

We now choose

$$f = \sum_n [\Gamma_f(\alpha_n)]^* \alpha_n \tag{2.42}$$

Its inner product with the arbitrary vector ϕ is

$$(f, \phi) = \left(\sum_n [\Gamma_f(\alpha_n)]^* \alpha_n, \sum_m c_m \alpha_m\right) = \sum_{n,m} [\Gamma_f(\alpha_n)] c_m (\alpha_n, \alpha_m) \tag{2.43}$$

$$= \sum_{n,m} [\Gamma_f(\alpha_n)] c_m \delta_{nm} = \sum_n [\Gamma_f(\alpha_n)] c_n = \Gamma_f(\phi) \tag{2.44}$$

So the vector f corresponding to Γ_f is given by

$$f = \sum_n [\Gamma_f(\alpha_n)]^* \alpha_n \tag{2.45}$$

2.6 The Dirac language of Kets and Bras

Let us rewrite everything we have done so far in the Dirac language.

Vectors in the linear space V are called *kets* or *ket vectors* and denoted by $|a\rangle$. Linear functionals Γ_b in the dual linear space V' are called *bras* or *bra vectors* and are denoted by $\langle b|$.

There is a one-to-one correspondence between kets in V and bras in V' or between vectors and linear functionals. We will use the *same label* to denote these corresponding quantities

$$|a\rangle \leftrightarrow \langle a| \tag{2.46}$$

The inner product between two vectors $|a\rangle$ and $|b\rangle$ in V corresponds to the linear functional of the left-hand vector assigning a scalar value to the right-hand vector

$$\Gamma_a(b) = (a, b) = \langle a \,|\, b\rangle = \text{ complex number} \tag{2.47}$$

The orthonormality property of a basis set $\{\alpha_i\}$ is expressed by the linear functional relation

$$\langle \alpha_i \,|\, \alpha_j \rangle = \delta_{ij} \tag{2.48}$$

The expansion of an arbitrary state in an orthonormal basis is given by

$$|q\rangle = \sum_n a_n |\alpha_n\rangle \tag{2.49}$$

and the expansion coefficients are

$$\langle \alpha_m \,|\, q\rangle = \sum_n a_n \langle \alpha_m \,|\, \alpha_n\rangle = \sum_n a_n \delta_{mn} = a_m \tag{2.50}$$

or a_m is the linear functional $\Gamma_{\alpha_m}(q)$ of the state $|q\rangle$ with respect to the corresponding basis vector $|\alpha_m\rangle$.

If we have two vectors $|a\rangle$ and $|b\rangle$ and we expand them in an orthonormal basis set $\{\alpha_i\}$ as

$$|a\rangle = \sum_n a_n |\alpha_n\rangle \text{ and } |b\rangle = \sum_n b_n |\alpha_n\rangle \tag{2.51}$$

$$\rightarrow \langle a \,|\, b\rangle = \sum_n a_n^* b_n = \langle b \,|\, a\rangle^* = \left(\sum_n b_n^* a_n \right)^* \tag{2.52}$$

where we have used the antilinear property of inner products that says that we have the correspondence

$$c\,|a\rangle \leftrightarrow c^*\,\langle a| \tag{2.53}$$

The linear functional itself is directly represented by the bra vector, i.e.,

$$\langle q|\cdots = \Gamma_q(\cdots) = (q,\cdots) \tag{2.54}$$

In Dirac language, there is nothing to distinguish between the value of the linear functional $\langle q|$ for the vector $|p\rangle$ and the inner product of the vectors $|q\rangle$ and $|p\rangle$. They are both $\langle q\,|\,p\rangle$.

2.7 Gram-Schmidt Orthogonalization Process

An orthonormal basis set for an n-dimensional vector space can always be constructed from any set of n linearly independent vectors using the Gram-Schmidt orthogonalization method.

Suppose that we have a set of n linearly independent vectors $|\alpha_i\rangle, i = 1, 2, \ldots, n$ that are not a mutually orthonormal set. We can construct a mutually orthonormal set $|\beta_i\rangle, i = 1, 2, \ldots, n$ using the following steps:

1. let
$$|\beta_1\rangle = |\alpha_1\rangle$$

2. let
$$|\beta_2\rangle = |\alpha_2\rangle + a_1\,|\beta_1\rangle \text{ where we choose } a_1 \text{ such that } \langle\beta_1\,|\,\beta_2\rangle = 0$$

3. this gives
$$\langle\beta_1\,|\,\beta_2\rangle = 0 = \langle\beta_1\,|\,\alpha_2\rangle + a_1\,\langle\beta_1\,|\,\beta_1\rangle$$
$$a_1 = -\frac{\langle\beta_1\,|\,\alpha_2\rangle}{\langle\beta_1\,|\,\beta_1\rangle}$$

Now proceed by induction.

Suppose we have constructed k mutually orthogonal vectors $|\beta_i\rangle, i = 1, 2, \ldots, k$. If we let

$$|\beta_{k+1}\rangle = |\alpha_{k+1}\rangle + \sum_{j=1}^{k} a_j\,|\beta_j\rangle \tag{2.55}$$

with

$$a_j = -\frac{\langle\beta_j\,|\,\alpha_{k+1}\rangle}{\langle\beta_j\,|\,\beta j\rangle} \tag{2.56}$$

then we have $\langle\beta_j\,|\,\beta_{k+1}\rangle = 0$ for $j = 1, 2, \ldots, k$. These steps are repeated until we have n mutually orthogonal vectors. We then normalize them to 1 and create an orthonormal set.

For example, suppose we have the set

$$|\alpha_1\rangle = \begin{pmatrix}1\\1\\0\end{pmatrix} \qquad |\alpha_2\rangle = \begin{pmatrix}0\\1\\1\end{pmatrix} \qquad |\alpha_3\rangle = \begin{pmatrix}1\\0\\1\end{pmatrix} \tag{2.57}$$

These vectors are not orthonormal.

1. let

$$|\beta_1\rangle = |\alpha_1\rangle = \begin{pmatrix}1\\1\\0\end{pmatrix} \qquad \langle\beta_1\,|\,\beta_1\rangle = 2$$

2. let

$$|\beta_2\rangle = |\alpha_2\rangle + a_1\,|\beta_1\rangle$$

with

$$a_1 = -\frac{\langle\beta_1\,|\,\alpha_2\rangle}{\langle\beta_1\,|\,\beta_1\rangle} = -\frac{1}{2}$$

and thus

$$|\beta_2\rangle = |\alpha_2\rangle - \frac{1}{2}\,|\alpha_1\rangle = \frac{1}{2}\begin{pmatrix}-1\\1\\2\end{pmatrix} \qquad \langle\beta_2\,|\,\beta_2\rangle = \frac{3}{2} \qquad \langle\beta_1\,|\,\beta_2\rangle = 0$$

3. let

$$|\beta_3\rangle = |\alpha_3\rangle + a_1\,|\beta_1\rangle + a_2\,|\beta_2\rangle$$

with

$$a_1 = -\frac{\langle\beta_1\,|\,\alpha_3\rangle}{\langle\beta_1\,|\,\beta_1\rangle} = -\frac{1}{2} \qquad a_2 = -\frac{\langle\beta_2\,|\,\alpha_3\rangle}{\langle\beta_2\,|\,\beta_2\rangle} = -\frac{1}{3}$$

and thus

$$|\beta_3\rangle = \frac{2}{3}\begin{pmatrix}1\\-1\\1\end{pmatrix} \qquad \langle\beta_3\,|\,\beta_3\rangle = \frac{4}{3} \qquad \langle\beta_1\,|\,\beta_3\rangle = 0 \qquad \langle\beta_2\,|\,\beta_3\rangle = 0$$

We normalize the vectors by dividing by their respective *norms*,

$$|\gamma_i\rangle = \frac{|\beta_i\rangle}{\|\,|\beta_i\rangle\,\|} = \frac{|\beta_i\rangle}{|\,\langle\beta_i\,|\,\beta_i\rangle\,|^{1/2}}$$

The orthonormal set is then

$$|\gamma_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix} \qquad |\gamma_2\rangle = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\1\\2\end{pmatrix} \qquad |\gamma_3\rangle = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\end{pmatrix}$$

2.8 Linear Operators

In general, an *operator* defined on a vector space maps vectors into vectors $\hat{Q} : V \to V$, i.e., if $\hat{Q}$ = operator and $|k\rangle$ = a vector, then $\hat{Q}\,|k\rangle = |q\rangle$ = a vector(we will use the $\hat{}$ symbol on top to signify operators). We say that $|q\rangle$ is the image of $|k\rangle$ under the mapping $\hat{Q}$. An operator is fully defined if we specify its action on all vectors in its *domain*, which is the set of all vectors for which $\hat{Q}\,|k\rangle$ for all $|k\rangle$ is defined. Its *range* is the set of all vectors of the form $\hat{Q}\,|k\rangle$ for all $|k\rangle$.

In the special case of a *linear operator*, the operator also satisfies the *linearity relation*

$$\hat{Q}(a_1\,|k_1\rangle + a_2\,|k_2\rangle) = \hat{Q}(a_1\,|k_1\rangle) + \hat{Q}(a_2\,|k_2\rangle) \tag{2.58}$$

In general, the operator $\hat{A}$ is linear if, for all $|f\rangle, |g\rangle \in V$, $\hat{A}(|f\rangle+|g\rangle) = \hat{A}\,|f\rangle+\hat{A}\,|g\rangle$, $(\alpha\hat{A})\,|f\rangle = (\alpha\hat{A}\,|f\rangle$ and $\hat{A} = 0$ if and only if $\hat{A}\,|f\rangle = 0$ for all $|f\rangle$. The *norm* $\|\hat{A}\|$ of the linear operator $\hat{A}$ is defined as the maximum of $\|\hat{A}\,|f\rangle\|$ for all $|f\rangle$ such that $\||f\rangle\| \le 1$. The identity operator $\hat{I}$ is defined by $\hat{I}\,|f\rangle = |f\rangle$ for all $|f\rangle$.

Since any vector in the space can be written as a linear combination of a basis set of vectors, we need only define a linear operator on the basis set and then its operation on all vectors is known.

Quantum mechanics works exclusively with linear operators so we will just call them operators from now on.

When are two operators equal?

In general, we say that two operators are *equal*, i.e., $\hat{Q}_1 = \hat{Q}_2$, when $\hat{Q}_1\,|p\rangle = \hat{Q}_2\,|p\rangle$ for *all* vectors $|p\rangle$ in the intersection of the domains of $\hat{Q}_1$ and $\hat{Q}_2$.

Using this equality idea we can define the sum and product of operators by

$$(\hat{Q}_1 + \hat{Q}_2)\,|k\rangle = \hat{Q}_1\,|k\rangle + \hat{Q}_2\,|k\rangle \tag{2.59}$$

$$(\hat{Q}_1\hat{Q}_2)\,|k\rangle = \hat{Q}_1(\hat{Q}_2\,|k\rangle) \tag{2.60}$$

for all vectors $|k\rangle$. It is clear that these relations imply that

$$(\hat{Q}_1\hat{Q}_2\hat{Q}_3)\,|k\rangle = (\hat{Q}_1\hat{Q}_2)\hat{Q}_3\,|k\rangle = \hat{Q}_1(\hat{Q}_2\hat{Q}_3)\,|k\rangle \tag{2.61}$$

or

$$(\hat{Q}_1\hat{Q}_2)\hat{Q}_3 = \hat{Q}_1(\hat{Q}_2\hat{Q}_3) \tag{2.62}$$

which corresponds to *associativity*. Now

$$(\hat{Q}_1\hat{Q}_2)\,|k\rangle = \hat{Q}_1(\hat{Q}_2)\,|k\rangle) = \hat{Q}_1\,|k_2\rangle = |k_{12}\rangle \tag{2.63}$$

$$(\hat{Q}_2\hat{Q}_1)\,|k\rangle = \hat{Q}_2(\hat{Q}_1)\,|k\rangle) = \hat{Q}_2\,|k_1\rangle = |k_{21}\rangle \tag{2.64}$$

does not imply that $|k_{12}\rangle = |k_{21}\rangle$. Therefore, in general we have $\hat{Q}_1\hat{Q}_2 \neq \hat{Q}_2\hat{Q}_1$, which corresponds to the *noncommutivity* of the two operators.

We define the *commutator* of two operators by

$$[\hat{Q}_1, \hat{Q}_2] = \hat{Q}_1\hat{Q}_2 - \hat{Q}_2\hat{Q}_1 \tag{2.65}$$

Two operators are said to *commute* if their corresponding commutator is zero (the null vector).

The commutator will be a fundamental quantity in our discussions of quantum mechanics.

2.9 An Aside : The Connection to Matrices

Now suppose that we have an N-dimensional vector space and we choose an orthonormal basis set

$$\{|q_i\rangle, i = 1, 2, \ldots, N\} \tag{2.66}$$

In an N-dimensional space, we can always represent any vector as an N-element column of numbers or as a column vector.

If we expand some arbitrary state $|\alpha\rangle$ in terms of this basis set and operate on it with an operator $\hat{Q}$, then we have

$$|\beta\rangle = \hat{Q}\,|\alpha\rangle = \hat{Q}\sum_{i=1}^{N} c_i\,|q_i\rangle = \sum_{i=1}^{N} c_i\hat{Q}\,|q_i\rangle \tag{2.67}$$

where $|\beta\rangle$ is another vector in the same space since $\hat{Q}$ is a linear operator.

Now expanding the vector $|\beta\rangle$ in the same basis set as

$$|\beta\rangle = \sum_{i=1}^{N} d_i\,|q_i\rangle \tag{2.68}$$

and constructing the linear functional(inner product) of $|\beta\rangle = \hat{Q}\,|\alpha\rangle$ with respect to $|q_i\rangle$ we get

$$\langle q_k|\,\hat{Q}\,|\alpha\rangle = \langle q_k|\,\hat{Q}\sum_{i=1}^{N} c_i\,|q_i\rangle = \sum_{i=1}^{N} c_i\,\langle q_k|\,\hat{Q}\,|q_i\rangle = \langle q_k\,|\,\beta\rangle = \sum_{j=1}^{N} d_j\,\langle q_k\,|\,q_j\rangle \tag{2.69}$$

Using the orthonormality property of the basis set

$$\langle q_k\,|\,q_j\rangle = \delta_{kj} \tag{2.70}$$

we get

$$\langle q_k|\,\hat{Q}\,|q_i\rangle = \sum_{j=1}^{N} d_j\delta_{kj} = d_k \tag{2.71}$$

If we define the numbers $\hat{Q}_{ki} = \langle q_k | \hat{Q} | q_i \rangle$ we get the relation

$$\sum_{j=1}^{N} \hat{Q}_{ki} c_i = d_k \tag{2.72}$$

Clearly, this has the form of a matrix equation where $\hat{Q}_{ki} = \langle q_k | \hat{Q} | q_i \rangle$ is defined as the (ki)-matrix element of the operator $\hat{Q}$. In this case, we would have the representation

$$(\hat{Q}_{ij}) = N \times N \text{ matrix} \tag{2.73}$$

$$(c_i) = N \times 1 \text{ column matrix} \tag{2.74}$$

$$(d_i) = N \times 1 \text{ column matrix} \tag{2.75}$$

and the matrix equation (using matrix multiplication) represented above is $Qc = d$.

All finite dimensional vector space operator equations can be turned into matrix equations in this manner.

2.10 More about Vectors, Linear Functionals, Operators

We have discussed what happens when an operator acts to the right on a ket vector. What about acting to the left on a bra vector, i.e., what is the meaning of the quantity

$$\langle q | \hat{Q} \tag{2.76}$$

Since the bra vector is really a linear functional we must be careful and look very closely at the mathematical content of the above quantity. Remember that the definition of the linear functional was

$$\langle q | p \rangle = \Gamma_q(p) = (q, p) \tag{2.77}$$

The standard *definition* (in terms of inner products) of the *adjoint* operator $\hat{Q}^\dagger$, of the operator $\hat{Q}$, in a linear vector space is

$$(\hat{Q}^\dagger q, p) = (q, \hat{Q} p) \tag{2.78}$$

The adjoint obeys the following properties:

$$\|\hat{A}^\dagger\| = \|\hat{A}\| \qquad (\hat{A}^\dagger)^\dagger = \hat{A} \qquad (\hat{A} + \hat{B})^\dagger = \hat{A}^\dagger + \hat{B}^\dagger \tag{2.79}$$

$$(\alpha \hat{A})^\dagger = \alpha^* \hat{A}^\dagger \qquad (\hat{A}\hat{B})^\dagger = \hat{B}^\dagger \hat{A}^\dagger \qquad (\hat{A}^{-1})^\dagger = (\hat{A}^\dagger)^{-1} \tag{2.80}$$

If we define a new vector ϕ by the operation $\hat{Q}^\dagger q = \phi$, then using the definition of the adjoint operator we have

$$\Gamma_\phi(p) = (\phi, p) = (\hat{Q}^\dagger q, p) = (q, \hat{Q} p) = \Gamma_q(\hat{Q} p) \tag{2.81}$$

We want the action of an operator on the bra space of linear functionals or bra vectors to create a new linear functional or bra vector in the same way that it does for ket vectors. We can guarantee this by *defining* the operation of an operator on the bra space of linear functionals by

$$\hat{Q}\Gamma_q(p) = \Gamma_q(\hat{Q}p) \tag{2.82}$$

Using this new definition and the definition of a linear functional, we then have

$$\hat{Q}\Gamma_q(c_1p_1 + c_2p_2) = \Gamma_q(\hat{Q}(c_1p_1 + c_2p_2)) \tag{2.83}$$

$$= c_1\Gamma_q(\hat{Q}p_1) + c_2\Gamma_q(\hat{Q}p_2) \tag{2.84}$$

$$= c_1\hat{Q}\Gamma_q(p_1) + c_2\hat{Q}\Gamma_q(p_2) \tag{2.85}$$

which says that $\hat{Q}\Gamma_q(\ldots)$ itself is also a linear functional.

Thus, our definition of how an operator acts on the bra space of linear functionals simply says that it creates a new functional as we desired.

With this definition, we then have

$$\hat{Q}\Gamma_q(p) = \Gamma_\phi(p) = (\phi, p) = (\hat{Q}^\dagger q, p) = (q, \hat{Q}p) = \Gamma_q(\hat{Q}p) \tag{2.86}$$

or, since $(\hat{Q}^\dagger q, p) = \Gamma_{\hat{Q}^\dagger q}(p)$, the relationship among linear functionals is

$$\hat{Q}\Gamma_q(\ldots) = \Gamma_{\hat{Q}^\dagger q}(p) \tag{2.87}$$

In terms of bra and ket vectors, this says that if $\langle q|$ and $|q\rangle$ are *corresponding* bra and ket vectors(remember, the Riesz theorem says there is a unique bra vector for each ket vector), then

$$\hat{Q}\,|q\rangle = |\beta\rangle \quad \text{and} \quad \langle q|\,\hat{Q}^\dagger = \langle\beta| \tag{2.88}$$

should also be emphcorresponding bra and ket vectors.

Since $\langle\beta\,|\,p\rangle^* = \langle p\,|\,\beta\rangle$ we then have that

$$\langle q|\,\hat{Q}^\dagger\,|p\rangle^* = \langle p|\,\hat{Q}\,|q\rangle \tag{2.89}$$

for all p and q. This relation is *equivalent* to the original inner product relation

$$(\hat{Q}^\dagger q, p) = (q, \hat{Q}p) \tag{2.90}$$

The *end result* of this discussion is that $\langle q|\,\hat{Q}^\dagger = \langle\beta|$ is the bra vector(linear functional) corresponding to the ket vector $\hat{Q}\,|q\rangle = |\beta\rangle$. Since the adjoint operator satisfies the relations

$$(c\hat{Q})^\dagger = c^*\hat{Q}^\dagger \qquad (\hat{Q}\hat{R})^\dagger = \hat{R}^\dagger\hat{Q}^\dagger \qquad (\hat{Q} + \hat{R})^\dagger = \hat{Q}^\dagger + \hat{R}^\dagger \tag{2.91}$$

we can define another product among the bra and ket vectors, which is an operator rather than a scalar as in the case of the inner product. It is called the *outer product* and, in the Dirac language, has the form

$$|q\rangle\langle p| \tag{2.92}$$

It is clear that the outer product is an operator since its action on any other vector always results in a vector

$$(|q\rangle\langle p|)\,|s\rangle = |q\rangle\langle p\,|\,s\rangle \tag{2.93}$$

We also have that

$$(\langle q|\,(|q\rangle\langle p|)^\dagger\,|p\rangle)^* = \langle p|\,(|q\rangle\langle p|)\,|q\rangle = \langle p\,|\,q\rangle\langle p\,|\,q\rangle = (\langle q\,|\,p\rangle\langle q\,|\,p\rangle)^* \tag{2.94}$$

or

$$(\langle q|\,(|q\rangle\langle p|)^\dagger\,|p\rangle) = \langle p\,|\,q\rangle\langle p\,|\,q\rangle \tag{2.95}$$

which implies that

$$(|q\rangle\langle p|)^\dagger = |q\rangle\langle p| \tag{2.96}$$

This type of operator will be very important in our later discussions. The special case

$$|p\rangle\langle p| \tag{2.97}$$

is called a *projection operator* because it picks out the component of an arbitrary state vector in the *"direction"* of the vector $|p\rangle$.

Projections operators (and linear combinations) will be the mathematical object with a direct connection to physical states and measurements in our later discussions.

Example in a Finite Dimensional Vector Space

Let us consider the 2-dimensional vector space spanned by the orthonormal basis set

$$|1\rangle = \begin{pmatrix}1\\0\end{pmatrix} \qquad |2\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{2.98}$$

We can define two projection operators as

$$\hat{P}_1 = |1\rangle\langle 1| \qquad \hat{P}_2 = |2\rangle\langle 2| \tag{2.99}$$

The matrix representation of these two projection operator is easily found using $\langle 1\,|\,1\rangle = \langle 2\,|\,2\rangle = 1$ and $\langle 1\,|\,2\rangle = \langle 2\,|\,1\rangle = 0$ and $\hat{Q}_{ki} = \langle k|\,\hat{Q}\,|i\rangle$. We have

$$\begin{aligned}(\hat{P}_1) &= \begin{pmatrix}\langle 1|\,\hat{P}_1\,|1\rangle & \langle 1|\,\hat{P}_1\,|2\rangle\\ \langle 2|\,\hat{P}_1\,|1\rangle & \langle 2|\,\hat{P}_1\,|2\rangle\end{pmatrix} \\ &= \begin{pmatrix}\langle 1\,|\,1\rangle\langle 1\,|\,1\rangle & \langle 1\,|\,1\rangle\langle 1\,|\,2\rangle\\ \langle 2\,|\,1\rangle\langle 1\,|\,1\rangle & \langle 2\,|\,1\rangle\langle 1\,|\,2\rangle\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\end{aligned} \tag{2.100}$$

$$(\hat{P}_2) = \begin{pmatrix} \langle 1|\hat{P}_2|1\rangle & \langle 1|\hat{P}_2|2\rangle \\ \langle 2|\hat{P}_2|1\rangle & \langle 2|\hat{P}_2|2\rangle \end{pmatrix} \tag{2.101}$$
$$= \begin{pmatrix} \langle 1|2\rangle\langle 2|1\rangle & \langle 1|2\rangle\langle 2|2\rangle \\ \langle 2|2\rangle\langle 2|1\rangle & \langle 2|2\rangle\langle 2|2\rangle \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Now consider an arbitrary vector in this space

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1|1\rangle + a_2|2\rangle \tag{2.102}$$

We then have (using both Dirac and matrix languages)

$$\hat{P}_1|a\rangle = a_1|1\rangle\langle 1|1\rangle + a_2|1\rangle\langle 1|2\rangle = a_1|1\rangle \tag{2.103}$$

or

$$(\hat{P}_1)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ 0 \end{pmatrix} = a_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.104}$$

and the projection operator performs as advertised.

We note that (at least in this *special case*)

$$(\hat{P}_1) + (\hat{P}_2) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.105}$$
$$= \hat{I} = \text{identity operator}$$

or

$$(\hat{P}_1) + (\hat{P}_2)|a\rangle = (\langle 1|1\rangle + \langle 2|2\rangle)|a\rangle \tag{2.106}$$
$$= \sum_{j=1}^{2} |j\rangle\langle j|a\rangle = |a\rangle = \hat{I}|a\rangle$$

where we have made use of the expansion formula for an arbitrary state in an orthonormal basis.

Return to Gram-Schmidt

As before, suppose we have a set of linearly independent, but non-orthogonal vectors $|i\rangle$ in an n-dimensional linear vector space, we can construct a set of orthogonal vectors $|\alpha_i\rangle$ as follows:

$$|\alpha_1\rangle = |1\rangle \qquad |\alpha_2\rangle = |2\rangle - |\alpha_1\rangle\langle\alpha_1|2\rangle \tag{2.107}$$

where

$$|\alpha_1\rangle\langle\alpha_1| = \hat{P}_{\alpha_1} = \text{ projection operator} \tag{2.108}$$

Then

$$|\alpha_2\rangle = |2\rangle - \hat{P}_{\alpha_1}|2\rangle \tag{2.109}$$

As earlier, the fact that this type of operator is a *projection operator* is shown clearly by considering its effect on an arbitrary vector

$$|Q\rangle = q_1 |1\rangle + q_2 |2\rangle + q_3 |3\rangle + \ldots = \sum_i q_i |i\rangle \tag{2.110}$$

Then using

$$\hat{P}_{\alpha_n} = |\alpha_n\rangle \langle\alpha_n| \tag{2.111}$$

we get

$$\hat{P}_{\alpha_n} |Q\rangle = \sum_i q_i \hat{P}_{\alpha_n} |i\rangle = \sum_i q_i |\alpha_n\rangle \langle\alpha_n | i\rangle = \sum_i q_i |\alpha_n\rangle \delta_{ni} = q_n |\alpha_n\rangle \tag{2.112}$$

or

$$\sum_n |\alpha_n\rangle \langle\alpha_n| = \hat{I} = \text{ identity operator} \tag{2.113}$$

which makes perfect sense since if you project out all components of a vector you just get the vector back! This generalizes the earlier result we found in the 2-dimensional space. *This Dirac language is very interesting!*

We can then write a general Gram-Schmidt procedure as

$$|\alpha_1\rangle = |1\rangle \qquad |\alpha_2\rangle = (1 - \hat{P}_{\alpha_1}) |2\rangle \qquad |\alpha_3\rangle = (1 - \hat{P}_{\alpha_1} - \hat{P}_{\alpha_2}) |3\rangle$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$|\alpha_n\rangle = \left(1 - \sum_{k=1}^{n-1} \hat{P}_{\alpha_k}\right) |n\rangle$$

which is rather neat. *This shows the power of these projection operators and of the Dirac language.*

Important Point - Looking at the equations

$$\hat{Q} |q\rangle = |\beta\rangle \quad \textit{and} \quad \langle q| \hat{Q}^\dagger = \langle\beta| \tag{2.114}$$

we might be tempted(as is the case in many textbooks) at this point to write

$$(|q\rangle)^\dagger = \langle q| \tag{2.115}$$

This might seem to make sense in a finite dimensional vector space where we can always treat kets as *column vectors* and bras as *row vectors*. However, the adjoint symbol is really only defined for operators and not for vectors, so one should exercise great care before generalizing this result, especially in infinite dimensional spaces!

Finally, we define a property of an operator called the *trace* as

$$Tr\hat{Q} = \sum_j \langle q_j| \hat{Q} |q_j\rangle = \text{ sum of diagonal elements } = \sum_j (\hat{Q})_{jj} \tag{2.116}$$

2.11 Self-Adjoint Operators

If we have an operator $\hat{Q}$ where the adjoint satisfies the relation

$$\langle q|\,\hat{Q}^\dagger\,|p\rangle = (\langle p|\,\hat{Q}\,|q\rangle)^* \tag{2.117}$$

then $\hat{Q}$ is a *Hermitian* or *self-adjoint* operator.

In a finite-dimensional space, we can look at the matrix elements corresponding to a Hermitian operator. Let

$$|p\rangle = \sum_{i=1}^{N} a_i\,|q_i\rangle \tag{2.118}$$

We have

$$\langle q_k|\,\hat{Q}\,|p\rangle^* = \langle p|\,\hat{Q}\,|q_k\rangle \tag{2.119}$$

$$\langle q_k|\,\hat{Q}\sum_{i=1}^{N} a_i\,|q_i\rangle^* = \sum_{i=1}^{N} a_i^*\,\langle q_i|\,\hat{Q}\,|q_k\rangle \tag{2.120}$$

$$\sum_{i=1}^{N} a_i^*\,\langle q_k|\,\hat{Q}\,|q_i\rangle = \sum_{i=1}^{N} a_i^*\,\langle q_i|\,\hat{Q}\,|q_k\rangle \tag{2.121}$$

$$\sum_{i=1}^{N} a_i^*[\langle q_k|\,\hat{Q}\,|q_i\rangle^* - \langle q_i|\,\hat{Q}\,|q_k\rangle] = 0 \tag{2.122}$$

where we have used the *antilinear* property.

This says that the matrix elements of a Hermitian operator satisfy

$$\hat{Q}^*_{ki} = \hat{Q}_{ik} \qquad or \qquad \hat{Q}^\dagger = \hat{Q} \tag{2.123}$$

as we saw above.

If $\hat{H}, \hat{K}$ are Hermitian, then so are $\hat{H} + \hat{K}$, $i[\hat{H}, \hat{K}]$, and $\alpha\hat{H}$ for all real α, and $\hat{A}^\dagger\hat{H}\hat{A}$ for any operator $\hat{A}$, but $\hat{H}\hat{K}$ is Hermitian if and only if the commutator $[\hat{H}, \hat{K}] = 0$.

The Hermitian operator $\hat{H}$ is positive if $(f, \hat{H}f) = \langle f|\,\hat{H}\,|f\rangle \geq 0$ for all $|f\rangle$; note that if the equality holds for all $|f\rangle$, then $\hat{H} = 0$. If $\hat{H}$ is positive, then it obeys the inequality

$$|(f, \hat{H}g)|^2 \geq (f, \hat{H}f)(g, \hat{H}g) \tag{2.124}$$

called the *Schwarz inequality for positive operators.*

If $\hat{K}^\dagger = -\hat{K}$, then the operator $\hat{K}$ is called *antihermitian*; clearly, $i\hat{K}$ is Hermitian. An arbitrary operator $\hat{A}$ can always be written as the sum $\hat{A} = \hat{H}_A + \hat{K}_A$ of its Hermitian part

$$\hat{H}_A = \frac{1}{2}(\hat{A} + \hat{A}^\dagger) \tag{2.125}$$

and its antihermitian part

$$\hat{K}_A = \frac{1}{2}(\hat{A} - \hat{A}^\dagger) \tag{2.126}$$

Hermitian operators will be very important in our discussions because they will represent *measurable quantities or observables.*

2.12 Eigenvalues and Eigenvectors

Suppose that an operator $\hat{Q}$ acting on a particular vector $|\beta\rangle$ returns a scalar multiple of that vector

$$\hat{Q}\,|\beta\rangle = b\,|\beta\rangle \tag{2.127}$$

The vector $|\beta\rangle$ is then called an *eigenvector* and the scalar b an *eigenvalue* of the operator $\hat{Q}$. Using the definition of the adjoint $\hat{Q}^\dagger$ operator and the antilinear correspondence between bras and kets, we then also must have (using any arbitrary vector $|\alpha\rangle$)

$$\langle\alpha|\,\hat{Q}\,|\beta\rangle^* = b^*\,\langle\alpha\,|\,\beta\rangle^* \tag{2.128}$$

$$\langle\beta|\,\hat{Q}^\dagger\,|\alpha\rangle = b^*\,\langle\beta\,|\,\alpha\rangle \tag{2.129}$$

$$\langle\beta|\,\hat{Q}^\dagger = b^*\,\langle\beta| \tag{2.130}$$

Now we assume that $\hat{Q}$ is a Hermitian operator and that $\hat{Q}\,|\beta\rangle = b\,|\beta\rangle$ again. The Hermitian property says that

$$\langle\beta|\,\hat{Q}^\dagger\,|\beta\rangle = \langle\beta|\,\hat{Q}\,|\beta\rangle = \langle\beta|\,\hat{Q}\,|\beta\rangle^* \tag{2.131}$$

$$\langle\beta|\,b\,|\beta\rangle = \langle\beta|\,b\,|\beta\rangle^* \tag{2.132}$$

$$(b - b^*)\,\langle\beta\,|\,\beta\rangle = 0 \tag{2.133}$$

$$b = b^* \tag{2.134}$$

where we have assumed that $\langle\beta\,|\,\beta\rangle \neq 0$. This means that all of the eigenvalues of a Hermitian operator are *real.* Following this up, if $\hat{Q}$ is a Hermitian operator which satisfies

$$\hat{Q}\,|\beta\rangle = b\,|\beta\rangle \quad \textit{with} \quad \hat{Q} = \hat{Q}^\dagger \tag{2.135}$$

then

$$\langle\beta|\,\hat{Q} = b\,\langle\beta| \tag{2.136}$$

or the ket vector and its corresponding bra vector are eigenvectors with the *same* eigenvalue.

Suppose that we have two eigenvectors $|\alpha\rangle$ and $|\beta\rangle$ of a Hermitian operator $\hat{Q}$ with eigenvalues a and b, respectively. We then have

$$\hat{Q}\,|\alpha\rangle = a\,|\alpha\rangle \quad \textit{and} \quad \hat{Q}\,|\beta\rangle = b\,|\beta\rangle \tag{2.137}$$

$$0 = \langle\alpha|\,\hat{Q}\,|\beta\rangle - \langle\beta|\,\hat{Q}^\dagger\,|\alpha\rangle^* = \langle\alpha|\,\hat{Q}\,|\beta\rangle - \langle\beta|\,\hat{Q}\,|\alpha\rangle^* \tag{2.138}$$

$$0 = b\,\langle\alpha\,|\,\beta\rangle - a^*\,\langle\beta\,|\,\alpha\rangle^* = (b-a)\,\langle\alpha\,|\,\beta\rangle \tag{2.139}$$

This implies that eigenvectors corresponding to *distinct(different)* eigenvalues ($a \neq b$) are *orthogonal*, i.e., $\langle\alpha\,|\,\beta\rangle = 0$.

If $a = b$, that is, the two eigenvectors correspond to the same eigenvalue, then they are called *degenerate* eigenvectors. In this case, we have that

$$\hat{Q}\,|\alpha\rangle = a\,|\alpha\rangle \quad \textit{and} \quad \hat{Q}\,|\beta\rangle = a\,|\beta\rangle \tag{2.140}$$

Now, any linear combination of the degenerate eigenvectors is also an eigenvector with the same eigenvalue as can be seen below

$$\begin{aligned}\hat{Q}(c_1\,|\alpha\rangle + c_2\,|\beta\rangle) &= c_1\hat{Q}\,|\alpha\rangle + c_2\hat{Q}\,|\beta\rangle \\ &= c_1 a\,|\alpha\rangle + c_2 a\,|\beta\rangle = a(c_1\,|\alpha\rangle + c_2\,|\beta\rangle)\end{aligned} \tag{2.141}$$

It is, therefore, always possible to replace a nonorthogonal but linearly independent set of degenerate eigenvectors by linear combinations of themselves that are orthogonal (using the Gram-Schmidt process). For the case of two states above, the orthogonal set is easy to find, namely

$$|1\rangle = |\alpha\rangle + |\beta\rangle \qquad |2\rangle = |\alpha\rangle - |\beta\rangle \tag{2.142}$$

The number of distinct vectors corresponding to a given eigenvalue is called the *multiplicity* of that eigenvalue. Non-degenerate eigenvalues are called *simple*.

We will always assume that we have already carried out this orthogonalization process and that the set of eigenvectors(for both non-degenerate and degenerate eigenvalues) of any Hermitian operator is an orthogonal set.

If the norms are all finite, then we can always construct an *orthonormal* set where

$$\langle\alpha_i\,|\,\alpha_j\rangle = \delta_{ij} \tag{2.143}$$

The set of eigenvalues of an operator $\hat{A}$ is called the *spectrum* of $\hat{A}$.

In an m-dimensional space $\mathcal{H}$ choose a matrix representation in which $\hat{A} = (A_{ij}), g = (\gamma_i)$. Written in this representation, the eigenvalue equation $\hat{A}g = \lambda g$ becomes a system of m homogeneous linear equations in the γ_i. The eigenvalues λ are the roots of the algebraic equation of degree m

$$det(A_{ij} - \lambda\delta_{ij}) = 0 \tag{2.144}$$

known as the *secular equation*, which expresses the condition for the homogeneous system of equations to have non-trivial solutions. In an m-dimensional space every operator has at least one and at most m distinct eigenvalues. If all the eigenvalues of an arbitrary operator $\hat{A}$ are simple, then there are m linearly independent eigenvectors of $\hat{A}$ but there may be fewer if $\hat{A}$ has multiple eigenvalues.

2.13 Completeness

A set of orthonormal vectors $\{|\alpha_k\rangle, k = 1, 2, 3, \ldots, N\}$ is *complete* if we can expand an arbitrary vector $|\eta\rangle$ in terms of that set, i.e.,

$$|\eta\rangle = \sum_j a_j |\alpha_j\rangle \tag{2.145}$$

The orthonormality condition then allows us to calculate the expansion coefficients a_j as

$$\langle\alpha_k | \eta\rangle = \sum_j a_j \langle\alpha_k | \alpha_j\rangle = \sum_j a_j \delta_{kj} = a_k \tag{2.146}$$

This implies that(remember the 2-dimensional example earlier)

$$|\eta\rangle = \sum_j |\alpha_j\rangle \langle\alpha_j | \eta\rangle = \left(\sum_j |\alpha_j\rangle \langle\alpha_j| \right) |\eta\rangle \tag{2.147}$$

or

$$\sum_j |\alpha_j\rangle \langle\alpha_j| = \hat{I} = \text{ identity operator} \tag{2.148}$$

This result is quite general.

For any complete set of vectors $\{|q_k\rangle, k = 1, 2, 3, \ldots, N\}$ the sum over all of the *projection* operators $|q_k\rangle \langle q_k|$ is the identity operator. This is one of the most important results we will derive. It will enable us to perform very clever tricks during algebraic manipulations. It is fundamentally linked to the probability interpretation of quantum mechanics as we shall see later.

If a set of vectors $\{|q_k\rangle, k = 1, 2, 3, \ldots, N\}$ are eigenvectors of a Hermitian operator $\hat{Q}$, then we will *always assume that they are a complete set.* Thus, if

$$\hat{Q} |q_k\rangle = q_k |q_k\rangle \tag{2.149}$$

then we can always write

$$\hat{Q} = \hat{Q}\hat{I} = \hat{Q} \sum_j |q_j\rangle \langle q_j| = \sum_j q_j |q_j\rangle \langle q_j| \tag{2.150}$$

which completely expresses an operator *in terms of its eigenvectors and eigenvalues.*

Since we can use the same argument to show that

$$\hat{Q}^n = \hat{Q}^n \hat{I} = \hat{Q}^n \sum_j |q_j\rangle \langle q_j| = \sum_j q_j^n |q_j\rangle \langle q_j| \tag{2.151}$$

we then have for any function $f(x)$ that has a power series expansion of the form

$$f(x) = \sum_k c_k x^k \tag{2.152}$$

that

$$f(\hat{Q}) = f(\hat{Q})\hat{I} = f(\hat{Q})\sum_j |q_j\rangle\langle q_j| = \sum_j\sum_k c_k\hat{Q}^k|q_j\rangle\langle q_j|$$
$$= \sum_j\sum_k c_k q_j^k|q_j\rangle\langle q_j| = \sum_j f(q_j)|q_j\rangle\langle q_j| \tag{2.153}$$

For a *finite dimensional space*, it can be shown that the eigenvectors of a Hermitian operator always form a complete set and all of the above relations hold. For infinite dimensional spaces the proof is not easily carried out in general (it is usually true however).

Before proceeding to a general discussion of the eigenvalue spectrum, the spectral theorem and the problem of continuous eigenvalues let us expand our knowledge of some of the objects we have already defined and add a few definitions that will be useful.

2.14 Expand Our Knowledge - Selected Topics

2.14.1 Hilbert Space

In an infinite dimensional space, we must determine whether the sums involved in many of our definitions *converge*. If they do not converge in some expressions, then the corresponding definitions are not valid. In addition, we must clearly state *what is meant* by a linear combination of an infinite number of vectors?

We assume that an *infinite linear combination*

$$|\alpha\rangle = \sum_{k=1}^{\infty} a_k|q_k\rangle \tag{2.154}$$

is *defined* if the sequence of partial sums

$$|\alpha_n\rangle = \sum_{k=1}^{n} a_k|q_k\rangle \tag{2.155}$$

converges as $n \to \infty$ or, equivalently, that $|\alpha_n\rangle \to |\alpha\rangle$ as $n \to \infty$, where this convergence is defined by the norm relation

$$\| |\alpha\rangle - |\alpha_n\rangle \| \to 0 \, as \, n \to \infty \tag{2.156}$$

The vector $|\alpha\rangle$ is called the *limit vector* of the sequence. A sequence of vectors $|\alpha_n\rangle$ is called a *Cauchy sequence* if

$$\| |\alpha_m\rangle - |\alpha_n\rangle \| \to 0 \, as \, m, n \to \infty \tag{2.157}$$

A space is *complete* if every Cauchy sequence of vectors converges to a limit vector in the space.

If a linear space with an inner product defined is complete, then it is called a *Hilbert space*.

A Hilbert space is *separable* if it has an orthonormal basis consisting of a countable (finite or infinite) number of vectors. Note that every finite-dimensional space is complete(as we assumed earlier).

If a set of n vectors $\{|q_k\rangle$ is such that every sequence of vectors in the set

$$|\alpha_n\rangle = \sum_{k=1}^{n} a_k |q_k\rangle \tag{2.158}$$

has a limit vector also in the set, then the set is *closed*.

Some examples are:

Space ℓ^2: this is the space of infinite sequences

$$(x_1, x_2, x_3, \ldots\ldots)$$

such that

$$\sum_{k=1}^{\infty} |x_k|^2 \; is \; finite$$

This space is a separable Hilbert space. It has an orthonormal basis consisting of

$$|q_1\rangle = (1,0,0,0,\ldots) \qquad |q_2\rangle = (0,1,0,0,\ldots) \qquad |q_3\rangle = (0,0,1,0,\ldots) \qquad \ldots$$

since

$$(x_1, x_2, x_3, \ldots\ldots) = \sum_{k=1}^{\infty} x_k |q_k\rangle$$

Space $\mathbf{L^2(a,b)}$: this is the space of square integrable functions on the interval (a,b). It is a separable Hilbert space. If we choose the interval $(0,1)$, then we have the space $L^2(0,1)$ of square integrable functions $f(x)$ on the interval $0 \le x \le 1$.

Another example comes from the theory of Fourier series, which says that the set of orthonormal vectors (or functions in this case)

$$1, \sqrt{2}\cos 2\pi kx \text{ and } \sin 2\pi kx \; for \; k = 1, 2, 3, \ldots$$

form an orthonormal basis.

If a set of vectors within a larger set is closed and both the larger set and the smaller set both form linear vector spaces, the smaller set is called a *subspace* of the larger set.

A subspace of a separable Hilbert space is also a separable Hilbert space.

If $\mathcal{R}$ is a subspace of a separable Hilbert space, then the set of all vectors which are *orthogonal to every vector* in $\mathcal{R}$ is called the *orthogonal complement* $\mathcal{R}_\perp$ of $\mathcal{R}$. $\mathcal{R}_\perp$is also a subspace.

A Hilbert space *preserves* the *one-to-one correspondence* between the kets in the space and the bras or linear functionals in its dual space.

2.14.2 Bounded Operators

A linear operator $\hat{Q}$ is *continuous* if

$$\hat{Q}\,|\alpha_n\rangle \to \hat{Q}\,|\alpha\rangle \tag{2.159}$$

for any sequence of vectors $|\alpha_n\rangle$ which converge to the limit vector $|\alpha\rangle$.

A linear operator is *bounded* if there exists a positive number a such that

$$\|\hat{Q}\,|\alpha\rangle\,\| \leq a\|\,|\alpha\rangle\,\| \tag{2.160}$$

for every vector $|\alpha\rangle$ in the space. The *smallest* a is called the *norm* $\|\hat{Q}\|$ of $\hat{Q}$.

A linear operator is continuous if and only if it is bounded. Every operator on a finite-dimensional space is bounded. A bounded linear operator on an infinite-dimensional space can be represented by an infinite matrix.

2.14.3 Inverses

A linear operator $\hat{Q}$ has an *inverse* if there exists a linear operator $\hat{M}$ such that

$$\hat{M}\hat{Q} = \hat{I} = \hat{Q}\hat{M} \tag{2.161}$$

We denote the inverse of $\hat{Q}$ by $\hat{M} = \hat{Q}^{-1}$.

In an n-dimensional vector space with a basis set $\{|q_k\rangle, k = 1, 2, 3, \ldots, n\}$, a necessary and sufficient condition for a linear operator $\hat{Q}$ to have an inverse is *any one* of the following (all are equivalent):

1. there is no vector $|\chi\rangle$ (except null vector) such that $\hat{Q}\,|\chi\rangle = 0$.
2. the set of vectors $\{\hat{Q}\,|q_k\rangle, k = 1, 2, 3, \ldots, n\}$ is linearly independent.
3. there exists a linear operator $\hat{M}$ such that $\hat{M}\hat{Q} = \hat{I} = \hat{Q}\hat{M}$
4. the matrix corresponding to $\hat{Q}$ has a nonzero determinant.

We defined the matrix elements with respect to a basis set by $\hat{Q}_{ij} = \langle q_i | \hat{Q} | q_j \rangle$. The determinant of a matrix is defined by

$$det(\hat{Q}) = \sum_{i_1 i_2 \ldots i_n} \epsilon_{i_1 i_2 \ldots i_n} \hat{Q}_{i_1 1} \hat{Q}_{i_2 2} \ldots \hat{Q}_{i_n n} \tag{2.162}$$

where $\epsilon_{i_1 i_2 \ldots i_n}$ is the permutation symbol of order n (n indices) which is defined by

$$\epsilon_{i_1 i_2 \ldots i_n} = \begin{cases} +1 & \text{if } i_1 i_2 \ldots i_n \text{ is an even permutation of } 123 \ldots n \\ -1 & \text{if } i_1 i_2 \ldots i_n \text{ is an odd permutation of } 123 \ldots n \\ 0 & \text{if any index is repeated} \end{cases} \tag{2.163}$$

Example: in the 3×3 case:

$$\begin{aligned} det \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{32} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} &= \sum_{i,j,k=1}^{3} \epsilon_{ijk} A_{i1} A_{j2} A_{k3} \\ &= \epsilon_{123} A_{11} A_{22} A_{33} + \epsilon_{132} A_{11} A_{32} A_{23} + \epsilon_{213} A_{21} A_{12} A_{33} \\ &\quad + \epsilon_{231} A_{21} A_{32} A_{13} + \epsilon_{312} A_{31} A_{12} A_{23} + \epsilon_{321} A_{31} A_{22} A_{13} \\ &= A_{11} A_{22} A_{33} - A_{11} A_{32} A_{23} - A_{21} A_{12} A_{33} \\ &\quad + A_{21} A_{32} A_{13} + A_{31} A_{12} A_{23} - A_{31} A_{22} A_{13} \end{aligned}$$

where we have only included the nonzero terms(no repeated indices).

We note that Rules (1), (2), (3) are *not sufficient conditions* in the infinite-dimensional case.

Finally, if two linear operators have inverses, then the inverse of their product is

$$(\hat{M}\hat{N})^{-1} = \hat{N}^{-1}\hat{M}^{-1} \tag{2.164}$$

2.14.4 Unitary Operators

A linear operator $\hat{G}$ is *unitary* (*orthogonal* if we have a real linear space) if it has an inverse and if $\| \hat{G} | \alpha \rangle \| = \| | \alpha \rangle \|$ for every vector $|\alpha\rangle$, i.e., unitary operators *preserve norms* or lengths of vectors.

If $\hat{G}$ is a unitary operator, then a more general result is

$$|\alpha_1\rangle = \hat{G}|\beta_1\rangle \quad \text{and} \quad |\alpha_2\rangle = \hat{G}|\beta_2\rangle \quad \text{implies} \quad \langle \alpha_1 | \alpha_2 \rangle = \langle \beta_1 | \beta_2 \rangle \tag{2.165}$$

or unitary operators preserve inner products and not just norms. Using the fact that $\langle \alpha_1 | = \langle \beta_1 | \hat{G}$, we then have, for a unitary operator $\hat{G}$

$$\langle \alpha_1 | \alpha_2 \rangle = \langle \beta_1 | \hat{G}^\dagger \hat{G} | \beta_2 \rangle \quad \text{or} \quad \hat{G}^\dagger \hat{G} = \hat{I} \quad \text{or} \quad \hat{G}^\dagger = \hat{G}^{-1} \tag{2.166}$$

i.e., the inverse is the Hermitian conjugate or adjoint.

In addition, the action of a unitary operator on a basis set preserves the fact that it is a basis set.

The evolution of quantum systems in time will be given by a unitary operator. The inner product preserving property will be connected to the probability interpretation of quantum mechanics.

2.14.5 More on Matrix Representations

Let $\{|b_i\rangle\}$ be an orthonormal basis on an m-dimensional space. An arbitrary vector $|g\rangle$ can be written in terms of its components $\gamma_i = \langle b_i \,|\, g\rangle, i = 1, 2, \ldots, m$ as

$$|g\rangle = \sum_i \gamma_i \,|b_i\rangle \tag{2.167}$$

Given an operator $\hat{A}$, the vector $|h\rangle = \hat{A}\,|g\rangle$ has the components

$$\eta_i = \langle b_i|\,(\hat{A}\,|g\rangle) = \langle b_i|\,\hat{A}\,|g\rangle = \sum_i \gamma_j \,\langle b_i|\,\hat{A}\,|b_j\rangle \tag{2.168}$$

One can identify the vectors $|g\rangle$ and $|h\rangle$ with column matrices formed by their components

$$|g\rangle = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdots \\ \gamma_m \end{pmatrix} \qquad |h\rangle = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_m \end{pmatrix} \tag{2.169}$$

In fact, the vector space of column matrices with the usual definitions of addition of matrices and multiplication by a number is isomorphic with $\mathcal{H}$. With these definitions, the operator $\hat{A}$ can be identified with the square matrix whose elements are the m^2 numbers $\hat{A}_{ij} = \langle b_i|\,\hat{A}\,|b_j\rangle$

$$\hat{A} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{23m} \\ \ldots & \ldots & \ldots & \ldots \\ A_{m1} & A_{m2} & \ldots & A_{mm} \end{pmatrix} \tag{2.170}$$

With these identifications, the components η_i can be rewritten as

$$\eta_i = \sum_j A_{ij} b_j \tag{2.171}$$

This is an expression which is identical to the matrix equation

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_m \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{23m} \\ \ldots & \ldots & \ldots & \ldots \\ A_{m1} & A_{m2} & \ldots & A_{mm} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdots \\ \gamma_m \end{pmatrix} \tag{2.172}$$

or more succinctly $(\eta_i) = (A_{ij})(b_j)$.

When the above identifications are made we speak of the *matrix representation of $\mathcal{H}$ with respect to the basis* $\{b_i\}$, or simply the $\{b_i\}$-representation of $\mathcal{H}$.

If in a given representation $\hat{A} \Rightarrow (A_{ij})$ and $\hat{A}^\dagger \Rightarrow (A^\dagger_{ij})$, then $A^\dagger_{ij} = A^*_{ji}$; thus, if $\hat{A}$ is hermitian, $A_{ij} = A^*_{ji}$. The representation of the identity operator with respect to any orthonormal basis is the identity matrix (δ_{ij}). The inverse $\hat{A}^{-1}$ exists if and only if the determinant $det(A_{ij}) \neq 0$; then $A^{-1}_{ij} = cof(A_{ij})/det(A_{ij})$ (cofactors). The matrix elements of a unitary operator $\hat{U}$ satisfy the relation .

Change of Representation - Let $\{|b_i\rangle\}, \{\left|\bar{b}_i\right\rangle\}$ be two bases on $\mathcal{H}$. What is the relationship between the representations of a vector $|g\rangle$ and an operator $\hat{A}$ with respect to the bases $\{|b_i\rangle\}$ and $\{\left|\bar{b}_i\right\rangle\}$?

Consider first the vector $|g\rangle$ and let $\gamma_i = \langle b_i \,|\, g\rangle$ and $\bar{\gamma}_i = \left\langle \bar{b}_i \,\middle|\, g\right\rangle$ be its components along $|b_i\rangle$ and $\left|\bar{b}_i\right\rangle$ respectively. Since

$$\left|\bar{b}_i\right\rangle = \sum_j \left\langle b_j \,\middle|\, \bar{b}_i\right\rangle |b_j\rangle \tag{2.173}$$

we have

$$\bar{\gamma}_i = \left\langle \bar{b}_i \,\middle|\, g\right\rangle = \sum_j \left\langle b_j \,\middle|\, \bar{b}_i\right\rangle \langle b_j \,|\, g\rangle \tag{2.174}$$

Defining the matrix $S \Rightarrow (S_{ij}) = \left\langle b_j \,\middle|\, \bar{b}_i\right\rangle$ we can write $\bar{\gamma}_i = (S_{ij})(\gamma_j)$.

The matrix S is unitary in the sense that

$$\sum_k S_{ik} S^*_{jk} = \delta_{ij} \tag{2.175}$$

Instead of thinking of S as a matrix performing a change of bases we can think of it as a unitary operator that generates a *unitary transformation* of *mathcalH* onto itself given by the correspondence

$$\left|\bar{f}\right\rangle = S\,|f\rangle \qquad \hat{\bar{A}} = \hat{S}\hat{A}\hat{S}^\dagger \tag{2.176}$$

For any vectors $|f\rangle, |g\rangle$ and any operator $\hat{A}$ we then have

$$\left\langle \bar{f} \,\middle|\, \bar{g}\right\rangle = \langle f \,|\, g\rangle \qquad \left\langle \bar{f}\right| \hat{\bar{A}} \left|\bar{g}\right\rangle = \langle f|\, \hat{A}\,|g\rangle \tag{2.177}$$

2.14.6 Projection Operators

Suppose we have a vector $|\alpha\rangle$ in a separable Hilbert space. The associated projection operator $\hat{P}_\alpha$ is a linear operator in Hilbert space whose action on

any other arbitrary vector in the space is to project out the component of that arbitrary vector along the vector $|\alpha\rangle$, i.e.,

$$\hat{P}_\alpha |\beta\rangle = a |\alpha\rangle \tag{2.178}$$

where a is a scalar and $|\beta\rangle$ is any arbitrary vector. This is a projection operator of the entire Hilbert space onto a *one-dimensional subspace*, namely, the single vector $|\alpha\rangle$.

Since, by definition, $\hat{P}_\alpha |\alpha\rangle = |\alpha\rangle$, all projection operators satisfy the following property:

$$\hat{P}_\alpha \hat{P}_\alpha |\beta\rangle = a\hat{P}_\alpha |\alpha\rangle = a |\alpha\rangle = \hat{P}_\alpha |\beta\rangle \tag{2.179}$$

$$(\hat{P}_\alpha^2 - \hat{P}_\alpha) |\beta\rangle = 0 \qquad or \qquad \hat{P}_\alpha^2 = \hat{P}_\alpha \tag{2.180}$$

If $|r\rangle$ is an eigenvector of $\hat{P}_\alpha$, such that $\hat{P}_\alpha |r\rangle = r |r\rangle$, then we have

$$\hat{P}_\alpha^2 |r\rangle = r\hat{P}_\alpha |r\rangle = r^2 |r\rangle \tag{2.181}$$

$$(\hat{P}_\alpha^2 - \hat{P}_\alpha) |r\rangle = (r^2 - r) |r\rangle = 0 \tag{2.182}$$

$$r^2 - r = 0 \quad \rightarrow r = 0, 1 \tag{2.183}$$

The *eigenvalues* of any projection operator are $0, 1$.

In general, any operator that satisfies the relation $\hat{A}^2 = \hat{A}$ is called *idempotent* and has eigenvalues 0,1.

Two projection operators $\hat{P}_{\alpha_1}$ and $\hat{P}_{\alpha_2}$ are orthogonal if, for any arbitrary vector $|\beta\rangle$

$$|\eta\rangle = \hat{P}_{\alpha_1} |\beta\rangle \quad and \quad |\sigma\rangle = \hat{P}_{\alpha_2} |\beta\rangle \quad \text{implies that} \quad \langle \eta | \sigma \rangle = 0 \tag{2.184}$$

The scalar constant in the equation $\hat{P}_\alpha |\beta\rangle = a |\alpha\rangle$, since it is the component of the arbitrary vector along $|\alpha\rangle$, is related to the inner product of the arbitrary vector with $|\alpha\rangle$ by $a = \langle \alpha | \beta \rangle$, which implies that

$$\hat{P}_\alpha |\beta\rangle = \langle \alpha | \beta \rangle |\alpha\rangle = (|\alpha\rangle \langle \alpha|) |\beta\rangle \tag{2.185}$$

or

$$\hat{P}_\alpha = |\alpha\rangle \langle \alpha| \tag{2.186}$$

For an orthonormal basis set $\{|q_i\rangle\}$, where $\langle q_i | q_j \rangle = \delta_{ij}$, we can define a set of projection operators $\{\hat{P}_i\}$, where each projection operator is given by

$$\hat{P}_k = |q_k\rangle \langle q_k| \tag{2.187}$$

We then have

$$\hat{P}_i \hat{P}_j = \delta_{ij} \hat{P}_j \quad \text{or} \quad |q_i\rangle \langle q_i\rangle |q_j \langle q_j| = \delta_{ij} |q_j\rangle \langle q_j| \tag{2.188}$$

so the projection operators are *mutually orthogonal* in this case.

As we stated earlier, the set of projection operators satisfies a *completeness* property, i.e., for any vector we can write

$$|\psi\rangle = \sum_k \langle q_k \,|\, \psi\rangle \, |q_k\rangle = \sum_k |q_k\rangle \, \langle q_k \,|\, \psi\rangle = \sum_k \hat{P}_k \, |\psi\rangle = \left(\sum_k \hat{P}_k \right) |\psi\rangle \tag{2.189}$$

This implies

$$\sum_k \hat{P}_k = \hat{I} = \sum_k |q_k\rangle \, \langle q_k| \tag{2.190}$$

This relation is very powerful and allows us to easily do algebra using the Dirac language by judicious insertion of $\hat{I}$ operators.

Some examples are:

1.
$$\langle \alpha \,|\, \beta\rangle = \langle \alpha| \, \hat{I} \, |\beta\rangle = \sum_k \langle \alpha| \, (|q_k\rangle \, \langle q_k|) \, |\beta\rangle = \sum_k \langle \alpha \,|\, q_k\rangle \, \langle q_k \,|\, \beta\rangle$$

2.
$$\begin{aligned} \langle \alpha| \, \hat{Q} \, |\beta\rangle &= \langle \alpha| \, \hat{I}\hat{Q}\hat{I} \, |\beta\rangle \\ &= \langle \alpha| \left(\sum_k |q_k\rangle \, \langle q_k| \right) \hat{Q} \left(\sum_j |q_j\rangle \, \langle q_j| \right) |\beta\rangle \\ &= \sum_k \sum_j \langle \alpha \,|\, q_k\rangle \, \langle q_k| \, \hat{Q} \, |q_j\rangle \, \langle q_j \,|\, \beta\rangle \end{aligned}$$

3.
$$\hat{Q} \, |\beta\rangle = \hat{I}\hat{Q} \, |\beta\rangle = \left(\sum_k |q_k\rangle \, \langle q_k| \right) \hat{Q} \, |\beta\rangle = \sum_k \langle q_k| \, \hat{Q} \, |\beta\rangle \, |q_k\rangle$$

4.
$$\begin{aligned} \langle \alpha| \, \hat{Q}\hat{R} \, |\beta\rangle &= \langle \alpha| \, \hat{I}\hat{Q}\hat{I}\hat{R}\hat{I} \, |\beta\rangle \\ &= \langle \alpha| \left(\sum_k |q_k\rangle \, \langle q_k| \right) \hat{Q} \left(\sum_k |q_i\rangle \, \langle q_i| \right) \hat{R} \left(\sum_j |q_j\rangle \, \langle q_j| \right) |\beta\rangle \\ &= \sum_k \sum_j \sum_i \langle \alpha \,|\, q_k\rangle \, \langle q_k| \, \hat{Q} \, |q_i\rangle \, \langle q_i| \, \hat{R} \, |q_j\rangle \, \langle q_j \,|\, \beta\rangle \end{aligned}$$

2.14.7 Unbounded Operators

Many of the operators we shall deal with in quantum mechanics are not bounded.

An example is easily constructed in the space $L^2(-\infty, \infty)$ of square-integrable

functions $f(x)$ for $-\infty < x < \infty$. Let $\hat{X}$ be a linear operator defined by the equation

$$\hat{X}f(x) = xf(x) \tag{2.191}$$

where x is real. This is a Hermitian operator since

$$\begin{aligned}(g, \hat{X}f) &= \int_{-\infty}^{\infty} g^*(x)\hat{X}f(x))\,dx = \int_{-\infty}^{\infty} g^*(x)xf(x)\,dx = \int_{-\infty}^{\infty} xg^*(x)f(x))\,dx \\ &= \int_{-\infty}^{\infty} [xg(x)]^* f(x))\,dx = \int_{-\infty}^{\infty} [\hat{X}g(x)]^* f(x))\,dx = (\hat{X}g, f)\end{aligned} \tag{2.192}$$

provided that all the integrals converge. It is not bounded since

$$\|\hat{X}f\|^2 = \int_{-\infty}^{\infty} |xf(x)|^2)\,dx \tag{2.193}$$

is not necessarily finite even if

$$\|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2)\,dx < \infty \tag{2.194}$$

2.15 Eigenvalues and Eigenvectors of Unitary Operators

We need to cover a few more details in this area. If $\hat{U}$ is a linear operator which has an inverse $\hat{U}^{-1}$, then the operators $\hat{U}\hat{Q}\hat{U}^{-1}$ and $\hat{Q}$ have the same eigenvalues, that is, if $\hat{Q}\,|\alpha\rangle = \alpha\,|\alpha\rangle$, then

$$\hat{U}\hat{Q}\hat{U}^{-1}(\hat{U}\,|\alpha\rangle) = \hat{U}\hat{Q}\,|\alpha\rangle = \alpha\hat{U}\,|\alpha\rangle \tag{2.195}$$

which says that $\hat{U}\hat{Q}\hat{U}^{-1}$ has the same eigenvalues as $\hat{Q}$ and its eigenvectors are $\hat{U}\,|\alpha\rangle$.

The eigenvectors and eigenvalues of unitary operators have these properties:

1. the eigenvalues are complex numbers of absolute value one
2. two eigenvectors are orthogonal if they correspond to different eigenvalues
3. in a complex finite-dimensional space, the eigenvectors span the space

2.16 Eigenvalue Decomposition

Let us now look deeper into the representation of operators in terms of their eigenvalues and eigenvectors.

As we stated earlier, for a finite-dimensional vector space, the eigenvectors of a Hermitian operator always form an orthonormal basis set or they always span

the space.

Suppose we have a Hermitian operator $\hat{B}$ with eigenvectors $\{|b_k\rangle, k = 1, 2, 3, \ldots, n\}$ and eigenvalues b_k where

$$\hat{B}\{|b_k\rangle = b_k\{|b_k\rangle \tag{2.196}$$

Labeling the states by the eigenvalues as above will become a standard practice as we get into quantum mechanics.

We showed earlier that we could represent an operator by the expression

$$\hat{B} = \sum_j b_j |b_j\rangle \langle b_j| \tag{2.197}$$

in terms of its eigenvalues and the projection operators constructed from the basis set (its eigenvectors).

2.17 Extension to Infinite-Dimensional Spaces

We now extend the properties we have been discussing to infinite-dimensional spaces.

First, we extend the properties of projection operators. The projection operators we have been considering are a special case of a more general definition. In particular, the projection operator we have been discussing $\hat{P}_\alpha = |\alpha\rangle \langle\alpha|$ projects the vector space onto the 1-dimensional subspace spanned by the vector $|\alpha\rangle$.

We extend the definition by defining a projection operator onto larger subspaces.

Let $\hat{E}_M$ be a projection operator onto the subspace M (not necessarily 1-dimensional). This means that for any vector $|\eta\rangle$ in the space, there are unique vectors $|\eta\rangle_{M_\perp}$ in $M^\perp$ which is called the orthogonal complement of M, such that we can always write

$$|\eta\rangle = |\eta\rangle_M + |\eta\rangle_{M_\perp} \tag{2.198}$$

and

$$\hat{E}_M |\eta\rangle = |\eta\rangle_M \tag{2.199}$$

for every $|\eta\rangle$ in the space. The operator $\hat{P}_\alpha = |\alpha\rangle \langle\alpha|$ is clearly a special case where the subspace M contains only one vector, namely, $|\alpha\rangle$.

The more general projection operators $\hat{E}_M$ satisfy all the same properties listed earlier for the single-state projection operators $\hat{P}_\alpha$.

If $\hat{E}_M$ is the projection on the n-dimensional subspace N, one can select an orthonormal basis $\{|b_i\rangle\}$ on $\mathcal{H}$, n of whose vectors $|b_1\rangle, \ldots, |b_n\rangle$ form a basis on N. In the *corresponding representation* $\hat{E}_N$ has the n diagonal matrix elements

$(\hat{E}_N)_{kk}$, $k = 1, \ldots, n$ equal to one and all the others equal to zero.

Example: Given the 2-dimensional space C spanned by the basis

$$|b_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |b_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The projection operator onto the 1-dimensional space A ($|b_1\rangle$) is

$$\hat{E}_1 = |b_1\rangle \langle b_1| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

The projection operator onto the 1-dimensional space B ($|b_2\rangle$) is

$$\hat{E}_2 = |b_2\rangle \langle b_2| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that $A \oplus B = C$, $B = A_\perp$, $A = B_\perp$.

Before proceeding further let us look at the properties of projections and expand on our earlier discussions.

Let $M \subset \mathcal{H}$ be a subspace and $M^\perp$ its orthogonal complement. Every vector $|h\rangle$ can be written in a unique manner as $|h\rangle = |f\rangle + |g\rangle$, with $|f\rangle \in M, |g\rangle \in M^\perp$. $|f\rangle$ is called the orthogonal projection of $|h\rangle$ on M. The linear operator $\hat{E}_M$ defined by $|f\rangle = \hat{E}_M |h\rangle$ is called the projection operator on M. Its domain is the whole space and its range is the subspace M. We say that $\hat{E}_M$ projects on M. In general, an operator which projects on some subspace is called a projection operator. An operator $\hat{E}$ is a projection operator if and only if $\hat{E}^2 = \hat{E}$ and $\hat{E}^\dagger = \hat{E}$. Let $\hat{E}_M$ and $\hat{E}_N$ be the projections on the subspaces M and N respectively. The product $\hat{E}_M \hat{E}_N$ is a projection if and only if both operators commute, in which case $\hat{E}_M \hat{E}_N$ projects on the intersection of M and N. The sum $\hat{E}_M + \hat{E}_N$ is a projection if and only if $\hat{E}_M \hat{E}_N = 0$, which means that M and N are orthogonal. In that case, $\hat{E}_M + \hat{E}_N$ projects on $M \oplus N$. The difference $\hat{E}_M - \hat{E}_N$ is a projection operator if and only if $\hat{E}_M \hat{E}_N = \hat{E}_N$, which is equivalent to $M \subseteq N$, i.e., N is a subset of M or N is contained in M. In this case $\|\hat{E}_N |f\rangle\| \leq \|\hat{E}_M |f\rangle\|$ for all $|f\rangle \in \mathcal{H}$ and we write $\hat{E}_N \leq \hat{E}_M$.

The dimension of a projection operator $\hat{E}_M$ is the dimension of the range M.

Any two vectors $|f\rangle, |g\rangle$ determine the operator $|f\rangle \langle g|$ defined by $(|f\rangle \langle g|) |h\rangle = \langle g | h\rangle |f\rangle$. We have $(|f\rangle \langle g|)^\dagger = |g\rangle \langle f|$. In particular, $|f\rangle \langle f|$ is Hermitian and, if $|f\rangle$ is normalized, it is the one-dimensional projection whose range is $|f\rangle$.

A subspace M is *invariant* under an operator $\hat{A}$ if for all vectors $|f\rangle \in M$ we have $\hat{A} |f\rangle \in M$. If, furthermore, $\hat{A} |g\rangle \in M^\perp$ for all $|g\rangle \in M^\perp$, that is, if both M and $M^\perp$ are invariant under $\hat{A}$, then the subspace M is said to reduce the

operator $\hat{A}$. The statements "M reduces $\hat{A}$" and "$\hat{A}$ commutes with $\hat{E}_M$" are equivalent.

Let M be invariant under $\hat{A}$. With $|f\rangle$ and $|g\rangle$ given as above, we have

$$(\hat{A}^\dagger |g\rangle)^\dagger |f\rangle) = (\langle g| \hat{A}) |f\rangle) = \langle g| (\hat{A} |f\rangle) = \langle g| \hat{A} |f\rangle = 0 \tag{2.200}$$

Therefore, $\hat{A}^\dagger |g\rangle \in M^\perp$ or, equivalently, $M^\perp$ is invariant under $\hat{A}^\dagger$.

From this result, one sees immediately that if M is invariant under the hermitian operator $\hat{B}$, then M reduces $\hat{B}$. The same is true if $\hat{U}$ is unitary because then, from

$$(\hat{U} |f\rangle)^\dagger (\hat{U} |g\rangle) = \langle f| \hat{U}^\dagger \hat{U} |g\rangle = \langle f | g\rangle = 0 \tag{2.201}$$

we conclude that $\hat{U} |g\rangle$, being orthogonal to $\hat{U} |f\rangle$, must be in $M^\perp$. If no subspaces other than $\mathcal{H}$ (the entire space itself) and $\{0\} = \varnothing$ reduce every member of a set of operators, then the set is called *irreducible*. It follows that a set of operators is irreducible if and only if the only operators that commute with every member of the set are multiples of the identity. If there is a subspace that reduces every operator of the set, then the set is said to be *reducible*. Now let $\hat{W}$ be a hermitian operator with r distinct eigenvalues $\lambda_i, i = 1, \ldots, r$. In addition, let M_i and $\hat{E}_i$ be the eigensubspace and eigenprojection belonging to the eigenvalue λ_i. M_i is invariant under $\mathcal{H}$ and reduces $\hat{W}$ or, equivalently, $\hat{E}_i \hat{W} = \hat{W} \hat{E}_i$. The subspace $M = \oplus_i M_i$ spanned by all the eigenvectors of $\hat{W}$ also reduces $\hat{W}$ and the corresponding projection operator

$$\hat{E} = \sum_i \hat{E}_i \tag{2.202}$$

commutes with $\hat{W}$.

This result has the following important consequence: the *eigenvectors of a Hermitian operator span the whole space*, that is, $\hat{E} = \hat{I}$ or $M =$ entire vector space, which generalizes the ideas we found earlier.

Regarded as an operator on M_i , the operator $\hat{W}$ multiplies every vector by the number λ_i. Therefore, it is equal to $\lambda_i \hat{I}$ and the multiplicity λ_i equals its degeneracy. One can write $\hat{W}$ as the direct sum $\oplus_i \lambda_i \hat{I}_i$ or equivalently, as the sum

$$\sum_i \lambda_i \hat{E}_i \tag{2.203}$$

in terms of the eigenprojections on the vector space.

Collecting together all these results, we have the simplest form of the *spectral theorem* in a unitary space.

To every Hermitian operator $\hat{W}$ on an m-dimensional unitary space there corresponds a unique family of non-zero projection operators, the eigenprojections of the space, $\hat{E}_i$, $i = 1, \ldots, r$, $r \leq m$ with the following properties:

1. The projections $\hat{E}_i$ are pairwise orthogonal

$$\hat{E}_i \hat{E}_j = \hat{E}_i \delta_{ij} \tag{2.204}$$

2. The family of projections is complete

$$\sum_i \hat{E}_i = \hat{I} \tag{2.205}$$

3. There exists a unique set of r distinct real numbers λ_i, the eigenvalues of $\hat{W}$, such that

$$\hat{W} = \sum_{i=1}^{r} \lambda_i \hat{E}_i \tag{2.206}$$

   This expression is the *spectral resolution* of $\hat{W}$.

The range of $\hat{E}_i$ is the eigensubspace M_i belonging to λ_i. Its dimension is the degeneracy or multiplicity s_i of λ_i. On it one can construct an orthonormal basis $\{|b_i^r\rangle,\ r = 1, \ldots, s_i$. It follows from the completeness of the family $\hat{E}_i$ that the union of all those bases $\{|b_i^r\rangle,\ r = 1, \ldots, s_i,\ i = 1, \ldots, m$ constitutes an orthonormal basis on the vector space One often expresses this fact by saying that $\hat{W}$ possesses a *complete set* of eigenvectors.

It is an immediate consequence of (1) and (2) above that for every vector $|f\rangle$ we have $\langle f|\hat{E}_i|f\rangle \geq 0$ and

$$\sum_i \langle f|\hat{E}_i|f\rangle = 1 \tag{2.207}$$

that is, for each vector and each family of projections, the set of numbers $P_i = \langle f|\hat{E}_i|f\rangle$ constitutes a so-called *discrete probability distribution.* This will be of fundamental importance in quantum theory as we shall see later.

Two hermitian operators

$$\sum_i \alpha_i \hat{E}_i \quad \textit{and} \quad \sum_i \beta_i \hat{F}_i \tag{2.208}$$

commute if and only if every pair $\hat{E}_i, \hat{F}_i$ of their eigenprojections commute.

From the orthogonality of the projections in the spectral resolution of $\hat{W}$ we have for every normalized vector $|f\rangle$

$$\|\hat{W}|f\rangle\|^2 = \langle f|\hat{W}^2|f\rangle = \sum_i \lambda_i^2 \langle f|\hat{E}_i|f\rangle \leq \lambda_m^2 \tag{2.209}$$

where λ_m is the largest eigenvalue and the equality holds for $|f\rangle$ in the range of $\hat{E}_m$. It follows that the norm of $\hat{W}$ is $|\lambda_m|$.

Functions of a Hermitian Operator - Using the spectral resolution of a Hermitian operator

$$\hat{Q} = \sum_{i=1}^{r} \lambda_i \hat{E}_i \tag{2.210}$$

one can verify that for a non-negative integer n one has

$$\hat{Q}^n = \sum_{i=1}^{r} \lambda_i^n \hat{E}_i \tag{2.211}$$

(also valid for negative n if all λ_i are non-zero). This property suggests the following *definition* of a function of a Hermitian operator.

> If $F(x)$ is a complex-valued function of the real variable x, then the function $F(\hat{Q})$ of the Hermitian operator $\hat{Q}$ is the operator
>
> $$F(\hat{Q}) = \sum_i F(\lambda_i)\hat{E}_i \tag{2.212}$$

Now, as we saw above, for every vector $|f\rangle$ the family of projections $\hat{E}_i$ belonging to a Hermitian operator determines a probability distribution on a discrete finite sample space with the probabilities given by $P_i = \langle f|\hat{E}_i|f\rangle$.

In probability theory, the passage to a discrete but infinite sample space offers no difficulty; the sums that yield averages are simply replaced by convergent series. The difficulties appear in the case where the sample space is continuous. One cannot then construct the probabilities for every possible set of outcomes from the probabilities of the points of the space (it is in general zero). Instead, one must consider the probabilities of appropriate elementary sets which, in the one-dimensional sample space in which we are interested here, can be taken to be intervals of the real line. To discuss such a case it is convenient to introduce the probability distribution function $D(\alpha)$ defined as the probability that the outcome of a trial is *(*less than) α.

In the discrete case, let P_i be the probability that the outcome of a trial yields the value ξ_i and let us order the indices in such a way $\xi_i < \xi_j$ for $i < j$. Written in terms of the P_i the distribution function is

$$D(\alpha) = \sum_{i=-\infty}^{i_\alpha} P_i \tag{2.213}$$

where i_α is the largest index such that $\xi_i < \alpha$ (we are assuming that the sample space is infinite, but the finite case is included if we put the appropriate $P_i = 0$). The function $D(\alpha)$ is clearly a non-decreasing *ladder function* with the properties $D(-\infty) = 0$, $D(\infty) = 1$. We have chosen the upper limit in the summation so as to satisfy the *convention* that $D(\alpha)$ is continuous on the right.

One can imagine a pure continuous sample space as the limit of a discrete one when the differences $\xi_{i+1} - \xi_i$ tend to zero. The distribution function then becomes the limit of a ladder function with the properties just mentioned. It is, therefore, a continuous non-decreasing function $D(\alpha)$ such that $D(-\infty) = 0$, $D(\infty) = 1$. In the case where $D(\alpha)$ is everywhere differentiable, as is the case

in most physical problems, the probability density $\pi(\alpha)$ is defined by $D'(\alpha) = dD(\alpha)/d\alpha$ and the average of the random variable $f(\alpha)$ is given by

$$\langle f(\alpha)\rangle = \int f(\alpha)\pi(\alpha)\,d\alpha \tag{2.214}$$

In the general case, the sample space is continuous but includes non-zero probability P_i concentrated at a countable number of points α_i. In this case $D(\alpha)$ will have a countable number of discontinuities of size P_i. If $D(\alpha)$ is differentiable everywhere else and $\pi(\alpha)$ is the corresponding probability density, then we have

$$\langle f(\alpha)\rangle = \int_{-\infty}^{\infty} f(\alpha)\pi(\alpha)\,d\alpha + \sum_i f(\alpha_i)P_i \tag{2.215}$$

Alternatively, using the Dirac δ-function we can write

$$\langle f(\alpha)\rangle = \int_{-\infty}^{\infty} f(\alpha)\pi_d(\alpha)\,d\alpha \tag{2.216}$$

where $\pi_d(\alpha)$ is the derivative of $D(\alpha)$, that is,

$$\pi_d(\alpha) = \pi(\alpha) + \sum_i P_i\delta(\alpha - \alpha_i) \tag{2.217}$$

The average can also be conveniently written by using the *Stieltjes integral* defined for a function $g(\alpha)$ with a countable number of discontinuities by

$$\int_a^b f(\alpha)\,dg(\alpha) = lim \sum_{i=1}^{n} f(\alpha_i)[g(\alpha_i) - g(\alpha_{i-1})] \tag{2.218}$$

where the α_i,$(a = \alpha_0 \le \alpha_1 \le \alpha_2 \le \ldots \le \alpha_n = b)$ form a subdivision of the interval (a, b) and the limit is taken over a sequence of ever finer subdivisions. In terms of the Stieltjes integral we then have

$$\langle f(\alpha)\rangle = \int_{-\infty}^{\infty} f(\alpha)\,dD(\alpha) \tag{2.219}$$

Now let us consider a Hermitian operator $\hat{B}$ in an infinite-dimensional space. We label its discrete eigenvalues in order of increasing eigenvalue where we assume

$$b_1 < b_2 < b_3 < \ldots < b_{m-1} < b_m \quad and \quad \hat{B}\,|b_j\rangle = b_j\,|b_j\rangle \tag{2.220}$$

For each real number x we define the operator

$$\hat{E}_x = \sum_{b_j<x} \hat{P}_j = \sum_{b_j<x} |b_j\rangle\,\langle b_j| \tag{2.221}$$

With this definition, $\hat{E}_x$ is the projection operator onto the subspace spanned by all eigenvectors with eigenvalues $b_k < x$. If $x < b_1$ (the smallest eigenvalue), then $\hat{E}_x$ is zero (no terms in the sum) and if $x > b_m$ (the largest eigenvalue), then $\hat{E}_x = \hat{I}$ because of the completeness property

$$\sum_{k=1}^{m} \hat{P}_k = \hat{I} \tag{2.222}$$

$\hat{E}_x$ increases from zero to one as x increases through the *spectrum* of eigenvalues. In fact, $\hat{E}_x$ increases(jumps) by an amount $\hat{P}_k$ when x reaches the eigenvalue b_k.

For each x let us define $d\hat{E}_x = \hat{E}_x - \hat{E}_{x-\epsilon}$ where ϵ is positive and small enough so that there is no eigenvalue b_j such that $(x-\epsilon) \leq b_j < x$. This means that $d\hat{E}_x$ is not zero only when x is an eigenvalue b_k and $d\hat{E}_x = \hat{P}_k$ for $x = b_k$.

Let us say this very important stuff in still another way.

In a unitary space, the family of projections $\hat{E}_i$ belonging to a hermitian operator

$$\hat{A} = \sum_i \xi_i \hat{E}_i \tag{2.223}$$

can be thought of as constituting a probability distribution on a finite sample space of operators. We shall take the eigenvalues to be ordered in ascending order. For every vector $|f\rangle$ the *operator valued probability* $\hat{E}_i$ generates the ordinary probability distribution $P_i = \langle f|\hat{E}_i|f\rangle$, which in quantum mechanics will give the probability of the outcome ξ_i upon measurement of $\hat{A}$ on a system in the state $|f\rangle$ as we shall see.

In analogy with ordinary probability ideas one can construct a corresponding *operator valued probability distribution function*

$$\hat{E}(\alpha) = \sum_{-\infty}^{i_\alpha} \hat{E}_i \tag{2.224}$$

A formula that one would expect to also be applicable in the case of the infinite-dimensional Hilbert space if the operator $\hat{A}$ corresponds to an observable which yields upon measurement a countable set of possible values. From the properties of the family $\hat{E}_i$ it follows that $\hat{E}(\alpha)$ is a projection operator, that $\hat{E}(\alpha)$ and $\hat{E}(\alpha')$ commute, that we have $\hat{E}(\alpha') \geq \hat{E}(\alpha)$ for $\alpha' > \alpha$ and that $\hat{E}(-\infty) = 0, \hat{E}(\infty) = \hat{I}$. $\hat{E}(\alpha)$ is an operator valued *ladder* function whose jumps at the discontinuities are given by $\hat{E}(\xi_i) - \hat{E}(\xi_i-) = \hat{E}_i$.

Quantum mechanics will require that we consider operators on the Hilbert space that will correspond to observables that yield upon measurement a continuous range of possible values. Such operators are associated with *operator valued probability distribution functions* analogous to the continuous distribution functions of ordinary probability, that is, with a family of projections $\hat{E}(\alpha)$ depending continuously on the parameter α such that $\hat{E}(\alpha') \geq \hat{E}(\alpha)$ for $\alpha' > \alpha$ and that $\hat{E}(-\infty) = 0, \hat{E}(\infty) = \hat{I}$. In the most general case corresponding to the continuous sample space with points ξ_i of concentrated probability we would expect the family $\hat{E}(\alpha)$ to have discontinuities at ξ_i of the form $\hat{E}(\xi_i) - \hat{E}(\xi_i-) = \hat{E}_i$ where $\hat{E}_i$ is the *operator valued probability* concentrated at the point ξ_i.

A *spectral family* is a one parameter family of commuting projection operators

$\hat{E}(\alpha)$ depending on the real parameter α that satisfies the following properties:

1. $\hat{E}(\alpha)$ is increasing: $\hat{E}(\alpha') \geq \hat{E}(\alpha)$ for $\alpha' > \alpha$
2. $\hat{E}(\alpha)$ is continuous on the right: $\hat{E}(\alpha') \to \hat{E}(\alpha)$ for $\alpha' \to \alpha$ and $\alpha' > \alpha$
3. $\hat{E}(\alpha)$ is complete: $\hat{E}(-\infty) = 0, \hat{E}(\infty) = \hat{I}$

It follows that $\hat{E}(\alpha)\hat{E}(\alpha') = \hat{E}(\alpha')\hat{E}(\alpha) = \hat{E}(\alpha)$ for $\alpha' > \alpha$.

The tentative conclusions of this intuitive discussion can, in fact, be proven to be correct. We now formulate them precisely.

In place of

$$\sum_{k=1}^{m} = \hat{I} \tag{2.225}$$

we can now *formally* write

$$\int_{-\infty}^{\infty} d\hat{E}_x = \hat{I} \tag{2.226}$$

and in place of

$$\hat{B} = \sum_{j=1}^{m} b_j \left|b_j\right\rangle \left\langle b_j\right| \tag{2.227}$$

we can now *formally* write

$$\hat{B} = \int_{-\infty}^{\infty} x \, d\hat{E}_x \tag{2.228}$$

Additionally, we have

$$\langle \alpha \,|\, \beta \rangle = \int_{-\infty}^{\infty} d \left\langle \alpha\right| \hat{E}_x \left|\beta\right\rangle \qquad \left\langle \alpha\right| \hat{B} \left|\beta\right\rangle = \int_{-\infty}^{\infty} x \, d \left\langle \alpha\right| \hat{E}_x \left|\beta\right\rangle \tag{2.229}$$

We can easily show the validity of these expressions in the case of a finite-dimensional space with discrete eigenvalues. In this case, we can satisfy all of the properties of $\hat{E}_x$ by writing

$$\hat{E}_x = \sum_k \hat{P}_k \theta(x - b_k) \tag{2.230}$$

where

$$\theta(x - b_k) = \begin{cases} +1 & \text{if} \quad x > b_k \\ 0 & \text{if} \quad x < b_k \end{cases} \tag{2.231}$$

This is called the *Heaviside step function.*

Then

$$d\hat{E}_x = \sum_k \hat{P}_k \, d\theta(x - b_k) = \sum_k \hat{P}_k \, \frac{d}{dx} \theta(x - b_k) \, dx = \sum_k \hat{P}_k \delta(x - b_k) \, dx \tag{2.232}$$

where

$$\delta(x - c) = Dirac \ \delta - function = 0 \quad if \quad x \neq c \tag{2.233}$$

and satisfies

$$\int_{-\infty}^{\infty} \delta(x-c)\,dx = 1 \tag{2.234}$$

In addition, we have used

$$\int_{-\infty}^{\infty} g(x)\delta(x-c)\,dx = g(c) \tag{2.235}$$

In the finite-dimensional case, we then have

$$\int_{-\infty}^{\infty} d\hat{E}_x = \int_{-\infty}^{\infty} \sum_k \hat{P}_k \delta(x-b_k)\,dx = \sum_k \hat{P}_k = \hat{I} \tag{2.236}$$

$$\int_{-\infty}^{\infty} x\,d\hat{E}_x = \int_{-\infty}^{\infty} x \sum_k \hat{P}_k \delta(x-b_k)\,dx = \sum_k b_k \hat{P}_k = \hat{B} \tag{2.237}$$

$$\begin{aligned}\int_{-\infty}^{\infty} d\left\langle\alpha\right|\hat{E}_x\left|\beta\right\rangle &= \int_{-\infty}^{\infty} \sum_k \left\langle\alpha\right|\hat{P}_k\left|\beta\right\rangle \delta(x-b_k)\,dx \\ &= \sum_k \left\langle\alpha\right|\hat{P}_k\left|\beta\right\rangle = \left\langle\alpha\right|\left(\sum_k \hat{P}_k\right)\left|\beta\right\rangle = \left\langle\alpha\,\middle|\,\beta\right\rangle\end{aligned} \tag{2.238}$$

$$\begin{aligned}\int_{-\infty}^{\infty} x\,d\left\langle\alpha\right|\hat{E}_x\left|\beta\right\rangle &= \int_{-\infty}^{\infty} x \sum_k \left\langle\alpha\right|\hat{P}_k\left|\beta\right\rangle \delta(x-b_k)\,dx \\ &= \sum_k b_k \left\langle\alpha\right|\hat{P}_k\left|\beta\right\rangle = \left\langle\alpha\right|\left(\sum_k b_k\hat{P}_k\right)\left|\beta\right\rangle = \left\langle\alpha\right|\hat{B}\left|\beta\right\rangle\end{aligned} \tag{2.239}$$

where we have used the fact that $\left\langle\alpha\right|\hat{E}_x\left|\beta\right\rangle$ is a complex function of x which jumps in value by the amount $\left\langle\alpha\right|\hat{P}_k\left|\beta\right\rangle$ at $x = b_k$.

So, we see that result

$$\hat{E}_x = \sum_k \hat{P}_k \theta(x-b_k) \tag{2.240}$$

works in the finite-dimensional case!!!

Unitary operators can be handled in the same manner. We label the eigenvalues (absolute value = 1) of the unitary operator $\hat{U}$ by $u_k = e^{i\theta_k}$ with the θ-values labeled in order

$$0 < \theta_1 < \theta_2 < \ldots < \theta_{m-1} < \theta_m, 2\pi \tag{2.241}$$

As before we define

$$\hat{E}_x = \sum_{\theta_j<x} \hat{P}_j = \sum_{\theta_j<x} \left|u_j\right\rangle\left\langle u_j\right| \tag{2.242}$$

This operator now projects onto the subspace spanned by all eigenvectors for eigenvalues $u_k = e^{i\theta_k}$ with $\theta_k < x$. If $x \le 0$, then $\hat{E}_x = 0$. If $x \ge 2\pi$, then $\hat{E}_x = \hat{I}$. $\hat{E}_x$ increments by $\hat{P}_k$ (the same as for a Hermitian operator) at the eigenvalues $u_k = e^{i\theta_k}$.

We can then write

$$\hat{U} = \sum_{k=1}^{m} u_k \hat{P}_k = \sum_{k=1}^{m} e^{i\theta_k} \hat{P}_k \to \int_0^{2\pi} e^{ix}\, d\hat{E}_x \tag{2.243}$$

and

$$\langle\alpha|\, \hat{U}\, |\beta\rangle = \int_0^{2\pi} e^{ix}\, d\,\langle\alpha|\, \hat{E}_x\, |\beta\rangle \tag{2.244}$$

Summarizing, for every Hermitian operator $\hat{H}$ there corresponds a unique spectral family $\hat{E}_H(\alpha)$ that commutes with $\hat{H}$ such that for every $|f\rangle$ and every $|g\rangle$ in the domain of $\hat{H}$

$$\langle f|\, \hat{H}\, |g\rangle = \int_{-\infty}^{\infty} \alpha\, d\,\langle f|\, \hat{E}_H(\alpha)\, |g\rangle \tag{2.245}$$

where the integral is a Riemann-Stieltjes integral. This expression can be written in short as

$$\hat{H} = \int_{-\infty}^{\infty} \alpha\, d\hat{E}_H(\alpha) \tag{2.246}$$

and is called the *spectral resolution* of $\hat{H}$. We have

$$\hat{E}_H(\alpha) = \int_{-\infty}^{\alpha} d\hat{E}_H(\alpha') \tag{2.247}$$

and therefore

$$\hat{I} = \int_{-\infty}^{\infty} d\hat{E}_H(\alpha) \tag{2.248}$$

Now to every interval $\Delta = (\alpha_1, \alpha_2)$ of the real line, there corresponds a projection operator

$$\hat{E}_H[\Delta] = \int_{\alpha_1}^{\alpha_2} d\hat{E}_H(\alpha) = \hat{E}_H(\alpha_2) - \hat{E}_H(\alpha_1) \tag{2.249}$$

This is a definition that can be extended to a set of intervals. It follows from the properties of the spectral family $\hat{E}_H(\alpha)$ that

$$\hat{E}_H[\Delta]\hat{E}_H[\Delta'] = \hat{E}_H[\Delta \cap \Delta'] \tag{2.250}$$

and

$$\hat{E}_H[\Delta] + \hat{E}_H[\Delta'] = \hat{E}_H[\Delta \cup \Delta'] \tag{2.251}$$

The quantity $d\hat{E}_H(\alpha)$ can be thought of as the projection $\hat{E}_H(d\alpha)$ corresponding to an infinitesimal interval of length $d\alpha$ centered at α.

By definition of the Stieltjes integral one has

$$\hat{H} = \lim_{n\to\infty} \sum_{\Delta_j \in S_n} \alpha_j \hat{E}_H[\Delta_j] \tag{2.252}$$

where $\{S_n\}$ is a sequence of subdivisions of the real line, such that S_n becomes infinitely fine as $n \to \infty$. The sum runs over all the intervals $\Delta_j^{(n)} = \alpha_j^{(n)} - \alpha_{j-1}^{(n)}$

of the subdivision S_n. If the spectral family $\hat{E}_K(\alpha)$ belonging to a Hermitian operator $\hat{K}$ is constant except for a countable number of isolated discontinuities at the points α_i of size $\hat{E}_K(\alpha_i) - \hat{E}_K(\alpha_i -) = \hat{E}_i$, then $\hat{K}$ has a spectral resolution entirely similar to the one in unitary space

$$\hat{K} = \sum_i \alpha_i \hat{E}_i \tag{2.253}$$

although in this case the number of terms in the sum may be infinite.

A general operator $\hat{H}$ can be regarded as the limit of a sequence

$$\hat{H}_n = \sum_{\Delta_j \epsilon S_n} \alpha_j \hat{E}_H[\Delta_j] \tag{2.254}$$

of operators of type $\hat{K}$.

Let $\hat{A}$ and $\hat{B}$ be two operators with spectral families $\hat{E}_A(\alpha), \hat{E}_B(\alpha)$. Suppose that $\hat{A}$ and $\hat{B}$ commute in the sense that $[\hat{A}, \hat{B}]\,|g\rangle = 0$ for every $|g\rangle$ in the domain of $[\hat{A}, \hat{B}]$, then one can show that

$$[\hat{E}_A(\alpha), \hat{E}_B(\alpha)]\,|g\rangle = 0 \tag{2.255}$$

However, unless the domain of $[\hat{A}, \hat{B}]$ coincides with the whole space, this relation does not imply that the spectral families commute. For this reason, when we deal with two operators $\hat{A}$ and $\hat{B}$ that possess a spectral family we shall use a stronger definition of commutativity:

$\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ commute if their spectral families commute.

2.18 Spectral Decomposition - More Details

It turns out that for infinite-dimensional vector spaces there exist Hermitian and unitary operators that have no eigenvectors and eigenvalues.

Consider the eigenvalue equation

$$-i\frac{d}{dx}\psi(x) = \hat{D}\psi(x) = \beta\psi(x) \tag{2.256}$$

This is a differential equation whose solution is

$$\psi(x) = ce^{i\beta x} \qquad c = constant \tag{2.257}$$

Suppose the operator

$$\hat{D} = -i\frac{d}{dx} \tag{2.258}$$

is defined on the interval $a \le x \le b$. Then its adjoint operator $\hat{D}^\dagger$ is defined by the relation

$$\langle\phi|\,\hat{D}^\dagger\,|\psi\rangle = \langle\psi|\,\hat{D}\,|\phi\rangle^* \tag{2.259}$$

or for function spaces

$$\int_a^b \phi^*(x)\hat{D}^\dagger\psi(x)\,dx = \left\{\int_a^b \psi^*(x)\hat{D}\phi(x)\,dx\right\}^* \tag{2.260}$$
$$= \int_a^b \phi^*(x)\hat{D}\psi(x)\,dx + i\left[\psi(x)\phi^*(x)\right]\Big|_a^b$$

where the last step follows from an integration by parts. If boundary conditions are imposed so that the last term(called the *surface* term) vanishes, then $\hat{D}$ will be a Hermitian operator; otherwise it is not Hermitian.

Now let us try to find the eigenvectors (eigenfunctions in this case) within a particular vector space. It turns out that we can define several different vector spaces *depending on the boundary conditions* that we impose.

Case 1: No boundary conditions

In this case, all complex β are eigenvalues and $\hat{D}$ is not Hermitian. In quantum mechanics we will be interested in Hermitian operators, so we are not really interested in this case.

Case 2: $a = -\infty$, $b = +\infty$ with $|\psi(x)|$ bounded as $|x| \to \infty$

All real values of β are eigenvalues. The eigenfunctions $\psi(x)$ are not normalizable since

$$\int_{-\infty}^{\infty} |\psi(x)|^2\,dx = |c|^2 \int_{-\infty}^{\infty} dx = \infty \tag{2.261}$$

They do, however, form a complete set in the sense that an arbitrary function can be represented as a the Fourier integral

$$q(x) = \int_{-\infty}^{\infty} F(\beta)e^{i\beta x}\,d\beta \tag{2.262}$$

which may be regarded as a continuous linear combination of the eigenfunctions. In this case, $F(\beta)$ is the *Fourier transform* of $q(x)$.

Case 3: $a = -\frac{L}{2}$, $b = +\frac{L}{2}$ with periodic boundary conditions $\psi(-\frac{L}{2}) = \psi(\frac{L}{2})$

The eigenvalues form a discrete set, β_n, satisfying

$$e^{-i\beta_n\frac{L}{2}} = e^{i\beta_n\frac{L}{2}} \to e^{i\beta_n L} = 1 \tag{2.263}$$

which implies

$$\beta_n L = 2n\pi \to \beta_n = \frac{2n\pi}{L} \tag{2.264}$$

where n = integers such that $-\infty \le n \le \infty$. These eigenfunctions form a complete orthonormal set (normalize by choosing the constant c appropriately) and $\hat{D}$ is

Hermitian. The completeness of the eigenfunction set follows from the theory of Fourier series.

Case 4: $a = -\infty$, $b = +\infty$ with $\psi(x) \to 0$ as $|x| \to \infty$

In this case, the operator $\hat{D}$ is Hermitian (the surface term vanishes), but it has no eigenfunctions within the space.

So, a Hermitian operator on an infinite-dimensional space may or may not possess a complete set of eigenfunctions, depending on the precise nature of the operator and the vector space.

It turns out, however, that the decomposition into projection operators can be reformulated in a way that does not rely on the existence of eigenfunctions. This alternative formulation uses the integral form of the projection operators derived earlier.

We need, however, to remind ourselves of some ideas we stated above(since it never hurts to repeat important stuff).

Let $\hat{E}_1$ and $\hat{E}_2$ be projection operators onto subspaces M_1 and M_2, respectively. We say that $\hat{E}_1$ and $\hat{E}_2$ are *orthogonal* if M_1 and M_2 are orthogonal (every vector in M_1 is orthogonal to every vector in M_2). We can express this orthogonality, in general, using the relation

$$\hat{E}_j \hat{E}_k = \delta_{jk} \hat{E}_k \tag{2.265}$$

If M_1 is *contained* in M_2, we write $\hat{E}_1 \leq \hat{E}_2$. This means that either $\hat{E}_1 \hat{E}_2 = \hat{E}_1$ or $\hat{E}_2 \hat{E}_1 = \hat{E}_1$. If $\hat{E}_1 \leq \hat{E}_2$, then $\| \hat{E}_1 |\alpha\rangle \| \leq \| \hat{E}_2 |\alpha\rangle \|$ for any vector $|\alpha\rangle$. If $\hat{E}_1 \hat{E}_2 = \hat{E}_2 \hat{E}_1$, then $\hat{E}_1 \hat{E}_2$ is a projection operator that projects onto the subspace which is the *intersection* of M_1 and M_2 , that is, the set of all vectors that are in both M_1 and M_2.

If $\hat{E}_1$ and $\hat{E}_2$ are orthogonal, then $\hat{E}_1 + \hat{E}_2$ is the projection operator onto $M_1 \oplus M_2$ (the direct sum).

If $\hat{E}_1 \leq \hat{E}_2$, $\hat{E}_1 - \hat{E}_2$ is the projection operator onto the subspace which is the orthogonal complement of M_1 in M_2, that is, the set of vectors in M_2 which are orthogonal to every vector in M_1.

Definition: A family of projection operators $\hat{E}_x$ depending on a real parameter x is a *spectral family* if it has the following properties:

1. if $x \leq y$ then $\hat{E}_x \leq \hat{E}_y$ or $\hat{E}_x \hat{E}_y = \hat{E}_x = \hat{E}_y \hat{E}_x$ – this means that $\hat{E}_x$ projects onto the subspace corresponding to eigenvalues $\leq x$.
2. if ϵ is positive, then $\hat{E}_{x+\epsilon} |\eta\rangle \to \hat{E}_x |\eta\rangle$ as $\epsilon \to 0$ for any vector $|\eta\rangle$ and any x.

3. $\hat{E}_x |\eta\rangle \to |0\rangle$ (the null vector) as $x \to -\infty$ and $\hat{E}_x |\eta\rangle \to |\eta\rangle$ as $x \to \infty$ for any vector $|\eta\rangle$.

For each *self-adjoint* operator $\hat{B}$ there is a unique spectral family of projection operators $\hat{E}_x$ such that

$$\langle\gamma| \hat{B} |\eta\rangle = \int_{-\infty}^{\infty} x \, d \langle\gamma| \hat{E}_x |\eta\rangle \tag{2.266}$$

for all vectors $|\eta\rangle$ and $|\gamma\rangle$. We then write

$$\hat{B} = \int_{-\infty}^{\infty} x \, d\hat{E}_x \tag{2.267}$$

This is called the *spectral decomposition* or *spectral resolution* of $\hat{B}$.

The same results hold for unitary operators. For each unitary operator $\hat{U}$ there is a unique spectral family of projection operators $\hat{E}_x$ such that $\hat{E}_x = 0$ for $x \leq 0$ and $\hat{E}_x = 1$ for $x \geq 2\pi$ and

$$\langle\gamma| \hat{U} |\eta\rangle = \int_{0}^{2\pi} e^{ix} \, d \langle\gamma| \hat{E}_x |\eta\rangle \tag{2.268}$$

for all vectors $|\eta\rangle$ and $|\gamma\rangle$. We then write

$$\hat{U} = \int_{0}^{2\pi} e^{ix} \, d\hat{E}_x \tag{2.269}$$

This is called the *spectral decomposition* or *spectral resolution* of $\hat{U}$.

Both of these results generalize for functions of an operator, i.e.,

$$g(\hat{B}) = \int_{-\infty}^{\infty} g(x) \, d\hat{E}_x \tag{2.270}$$

We considered the case of a discrete spectrum of eigenvalues earlier and found that when the operator $\hat{B}$ has the eigenvalue equation

$$\hat{B} |b_k\rangle = b_k |b_k\rangle \tag{2.271}$$

we then have

$$\hat{E}_x = \sum_k \hat{P}_k \theta(x - b_k) = \sum_k |b_k\rangle \langle b_k| \, \theta(x - b_k) \tag{2.272}$$

so that

$$d\hat{E}_x = \sum_k |b_k\rangle \langle b_k| \, \delta(x - b_k) \, dx \tag{2.273}$$

which implies that the only contributions to the integral occur at the eigenvalues b_k.

We can state all of this stuff *even more formally* once again

The Eigenvalue Problem and the Spectrum

Based on our previous discussion, we can now say that the eigenvalue problem $\hat{H}|f\rangle = \lambda|f\rangle$ has solutions if and only if the spectral family $\hat{E}_H(\alpha)$ is discontinuous. The eigenvalues λ_i are the points of discontinuity of $\hat{E}_H(\alpha)$. The eigenprojection $\hat{E}_i$ belonging to λ_i is the discontinuity of $\hat{E}_H(\alpha)$ at λ_i:

$$\hat{E}_i = \hat{E}_H(\lambda_i) - \hat{E}_H(\lambda_i -) \tag{2.274}$$

It follows that the eigenprojections are orthogonal and that (for hermitian $\hat{H}$) the eigenvalues are real.

The *spectrum* of $\hat{H}$ consists of the set Σ of points of the real axis where the spectral family $\hat{E}_H(\alpha)$ is increasing. If the spectral family is constant except for a (necessarily countable) number of discontinuities $\hat{H}$ is said to have a *pure discrete* spectrum. If the spectral family is everywhere continuous $\hat{H}$ is said to have a *pure continuous* spectrum. Otherwise the spectrum is said to be *mixed*. If and only if $\hat{H}$ is bounded (positive) is its spectrum bounded (positive). Since the points of constancy of $\hat{E}_H(\alpha)$ do not contribute to the integral in

$$\hat{H} = \int_{-\infty}^{\infty} \alpha \, d\hat{E}_H(\alpha) \tag{2.275}$$

the region of integration can be restricted to the spectrum and we can write

$$\hat{H} = \int_{\Sigma} \alpha \, d\hat{E}_H(\alpha) \tag{2.276}$$

Suppose that an operator $\hat{H}$ has a mixed spectrum Σ. The subspace M spanned by its eigenvectors reduces $\hat{H}$ so that we can write $\hat{H} = \hat{H}_d \oplus \hat{H}_c$ where $\hat{H}_d$ and $\hat{H}_c$ are operators on M and $M^{\perp}$, respectively. The spectrum Σ_d of $\hat{H}_d$ is pure discrete and is called the *discrete spectrum* of $\hat{H}$ while the spectrum Σ_c of $\hat{H}_c$ is pure continuous and is called the *continuous spectrum* of $\hat{H}$. Note that $\Sigma = \Sigma_d \cup \Sigma_c$, but that Σ_d and Σ_c may have points in common, that is, there may be eigenvalues embedded in the continuous spectrum. Separating out the discrete spectrum, the spectral resolution of $\hat{H}$ can be written

$$\hat{H} = \sum_i \alpha_i \hat{E}_i + \int_{\Sigma_c} \alpha \, d\hat{E}_H(\alpha) \quad , \quad \alpha_i \in \Sigma_d \tag{2.277}$$

Clearly, the terms on the RHS are, respectively $\hat{H}_d$ and $\hat{H}_c$, and when regarded as operators on the vector space.

Now let us consider the case of a *continuous spectrum* of eigenvalues. In particular, consider the multiplicative operator $\hat{Q}$ defined on $L^2(-\infty, \infty)$ by

$$\hat{Q}g(x) = xg(x) \tag{2.278}$$

for all functions $g(x)$ in $L^2(-\infty,\infty)$. This is a Hermitian operator since

$$\int_a^b \phi^*(x)\hat{Q}^\dagger\psi(x)\,dx = \left\{\int_a^b \psi^*(x)\hat{Q}\phi(x)\,dx\right\}^* = \left\{\int_a^b \psi^*(x)x\phi(x)\,dx\right\}^*$$
$$= \int_a^b \phi^*(x)x\psi(x)\,dx = \int_a^b \phi^*(x)\hat{Q}\psi(x)\,dx \qquad (2.279)$$

Now suppose that $\hat{Q}$ has an eigenvalue equation with eigenfunctions $q(x)$ in $L^2(-\infty,\infty)$ of the form

$$\hat{Q}q(x) = \beta q(x) \qquad (2.280)$$

Since all functions in $L^2(-\infty,\infty)$, including $q(x)$, must also satisfy $\hat{Q}g(x) = xg(x)$, we then have

$$\hat{Q}q(x) = xq(x) = \beta q(x) \quad or \quad (x-\beta)q(x) = 0 \qquad for\ all\ x \qquad (2.281)$$

The formal solution to this equation is the Dirac δ-function

$$q(x) = \delta(x-\beta) \qquad (2.282)$$

The spectral theorem still applies to this operator. The projection operators for $\hat{Q}$, in this case, are given by

$$\hat{E}_\beta g(x) = \theta(\beta - x)g(x) \qquad (2.283)$$

This is equal to $g(x)$ if $x < \beta$ and is 0 for $x > \beta$. We then have

$$\hat{Q} = \int_{-\infty}^{\infty} \beta\, d\hat{E}_\beta \qquad (2.284)$$

which can be easily verified by

$$\hat{Q}g(x) = \int_{-\infty}^{\infty} \beta\, d\hat{E}_\beta g(x) = \int_{-\infty}^{\infty} \beta\, d[\theta(\beta-x)g(x)] \qquad (2.285)$$
$$= \int_{-\infty}^{\infty} \beta \frac{d}{d\beta}[\theta(\beta-x)g(x)]\,d\beta = \int_{-\infty}^{\infty} \beta\delta(\beta-x)g(x)]\,d\beta = xg(x)$$

So the decomposition into projection operators can still be defined in the case of a continuous spectrum.

Saying it still another way. Let $\hat{E}_Q(x)$ be its spectral family and let us put $(\hat{E}_Q(\alpha)f)(x) = g(\alpha,x)$. The spectral resolution of $\hat{Q}$ requires that

$$xf(x) = (\hat{Q}f)(x) = \int_{-\infty}^{\infty} \alpha\, d\hat{E}_Q(\alpha)f)(x) = \int_{-\infty}^{\infty} \alpha\, dg(\alpha,x) \qquad (2.286)$$

A solution of this equation is obtained if we set $dg(\alpha,x) = \delta(\alpha-x)f(x)\,d\alpha$. We then get

$$\hat{E}_Q(\alpha)f)(x) = g(\alpha,x) = \int_{-\infty}^{\alpha} dg(\alpha',x) \qquad (2.287)$$
$$= \int_{-\infty}^{\alpha} \delta(\alpha'-x)f(x)\,d\alpha' = \chi(\alpha-x)f(x)$$

where $\chi(\alpha - x)$ is the Heaviside function

$$\chi(x) = \begin{cases} +1 & \text{if} \quad 0 \le x \\ 0 & \text{if} \quad 0 > x \end{cases} \tag{2.288}$$

According to this calculation $\hat{E}_Q(\alpha)$ is the multiplicative operator $\chi(\alpha - x)$ as above. We now need to verify that the solution we have constructed gives the unique spectral family belonging to $\hat{Q}$.

First we show that the solution is a spectral family, that is, that it has the properties(on page 50) required by the definition given earlier. Since the $\hat{E}_Q(\alpha)$ are multiplicative operators, it is clear that they commute. Property (1) is equivalent to

$$\hat{E}(\alpha)\hat{E}(\alpha') = \hat{E}(\alpha) \quad , \quad \alpha \le \alpha' \tag{2.289}$$

a property that is clearly satisfied by the multiplicative operator $\chi(\alpha - x)$. Property (2) holds because $\chi(x)$ was defined to be continuous on the right, while properties (3) are clear.

Next we have to verify that the spectral family found does in fact belong to $\hat{Q}$. According to the spectral theorem we must have

$$\langle f|\,\hat{Q}\,|g\rangle = \int_{-\infty}^{\infty} \alpha\, d\,\langle f|\,\hat{E}_Q(\alpha)\,|g\rangle \tag{2.290}$$

or

$$\int_{-\infty}^{\infty} f^*(x) x g(x)\, dx = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} d\alpha\, f^*(x)\delta(\alpha - x) g(x) \tag{2.291}$$

which is clearly true by definition of the δ-function.

The spectral family $\hat{E}_Q(\alpha)$ is continuous everywhere and increasing in $(-\infty, \infty)$. The spectrum of $\hat{Q}$ consists therefore of the whole of the real axis and is pure continuous. The projection $\hat{E}_Q[\Delta]$ is the *characteristic function* $\chi(\Delta)$ of the interval Δ, that is, a function of x equal to one if $x \in \Delta$ and zero otherwise.

Some final thoughts about these ideas.....

Projection Operators and Continuous Spectra - An Example

In the macroscopic world, if we want to locate the position of an object, we use a calibrated ruler. Formally, the physical position x is a continuous variable. The ruler, however, only has a finite resolution. An outcome anywhere within the j^{th} interval is said to correspond to the value x_j. Thus, effectively, the result of the position measurement is not the original continuous variable x, but rather a staircase function,

$$x' = f(x) = x_j \quad , \quad \forall \quad x_j \le x \le x_{j+1} \tag{2.292}$$

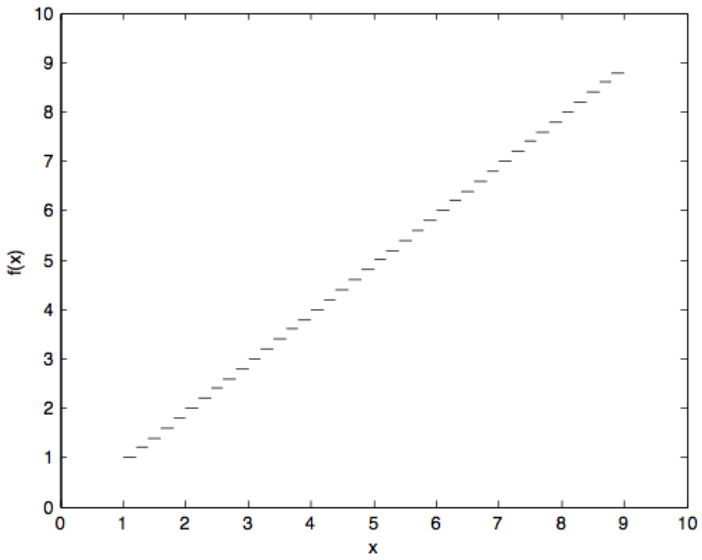


Figure 2.1: Staircase Function

as illustrated in the figure below.

These considerations are easily translated in quantum language.

In the x-representation, an operator x' is defined as multiplication by the staircase function $f(x)$. This operator has a finite number of discrete eigenvalues x_j. Each one of the eigenvalues is infinitely degenerate, that is, any state vector with domain between x_j and x_{j+1} falls entirely within the j^{th} interval of the ruler (see Figure 4.1), and therefore corresponds to the degenerate eigenvalue x_j.

Orthogonal resolution of the Identity

An experimental setup for a quantum test described by the above formalism could have, at its final stage, an array of closely packed detectors, labeled by the real numbers x_j. Such a quantum test thus asks, simultaneously, a set of questions

$$\text{Is} \quad x_j \leq x \leq x_{j+1} \quad ? \tag{2.293}$$

(one question for each j). The answers, *yes* and *no*, can be give numerical values 0 and 1, respectively. Each of these questions therefore corresponds to

the operator $\hat{P}_j$, which is itself a function of x:

$$\hat{P}_j(x) = \begin{cases} 1 & \text{if} \quad x_j \le x \le x_{j+1} \\ 0 & \text{otherwise} \end{cases} \tag{2.294}$$

Clearly, these operators satisfy

$$\hat{P}_j \hat{P}_k = \delta_{jk} \hat{P}_k \quad and \quad \sum_j \hat{P}_j = \hat{I} \tag{2.295}$$

This implies that they are projection operators (or projectors) and the questions will correspond to the *measurement* of the projection operators.

The staircase function $x' = f(x)$ defined above can then be written as

$$x' = \sum_j s_j \hat{P} \tag{2.296}$$

This operator x' approximates the operator x as well as is allowed by the finite resolution of the ruler.

How do we proceed to the continuum limit? Let us define a *spectral family* of operators

$$\hat{E}(x_j) = \sum_{k=0}^{j-1} \hat{P}_k \tag{2.297}$$

They obey the recursion relations

$$\hat{E}(x_{j+1}) = \hat{E}(x_j) + \hat{P}_j \tag{2.298}$$

and the boundary conditions

$$\hat{E}(x_{min}) = 0 \quad , \quad \hat{E}(x_{max}) = \hat{I} \tag{2.299}$$

The physical meaning of the operator $\hat{E}(x_j)$ is the question

$$''Is \quad x \le x_j \quad ?'' \tag{2.300}$$

with answers yes = 1 and no = 0.

We then have

$$\begin{aligned} \hat{E}(x_j)\hat{E}(x_m) &= \sum_{k=0}^{j-1} \hat{P}_k \sum_{n=0}^{m-1} \hat{P}_n = \sum_{k=0}^{j-1} \sum_{n=0}^{m-1} \hat{P}_k \hat{P}_n \\ &= \sum_{k=0}^{j-1} \sum_{n=0}^{m-1} \delta_{kn} \hat{P}_n = \begin{cases} \hat{E}(x_j) & \text{if} \quad x_j \le x_m \\ \hat{E}(x_m) & \text{if} \quad x_m \le x_j \end{cases} \end{aligned} \tag{2.301}$$

so that the $\hat{E}(x_j)$ are projectors.

We can now pass to the continuum limit. We define $\hat{E}(\xi)$ as the projector which represents the question

$$''Is \quad x \leq \xi \quad ?'' \tag{2.302}$$

and which returns, as the answer, a numerical value(yes = 1, no = 0). We then consider two neighboring values, ξ and $\xi + d\xi$, and define an *infinitesimal* projector,

$$d\hat{E}(\xi) = \hat{E}(\xi + d\xi) - \hat{E}(\xi) \tag{2.303}$$

which represents the question "Is $\xi \leq x \leq \xi + d\xi$?". This $d\hat{E}(\xi)$ thus behaves as an infinitesimal increment $\hat{P}_j$ in the equation

$$\hat{E}(x_{j+1}) = \hat{E}(x_j) + \hat{P}_j \tag{2.304}$$

We then have, instead of the staircase approximation, the exact result

$$x = \int_0^1 \xi \, d\hat{E}(\xi) \tag{2.305}$$

Note that the integration limits are actually operators, namely, $\hat{E}(x_{min}) = 0$ and $\hat{E}(x_{max}) = 1$.

This equation is the *spectral decomposition* or *spectral resolution* of the operator x and the operators $\hat{E}(\xi)$ are the *spectral family* (or *resolution of the identity*) generated by x. We can now define any function of the operator x

$$f(x) = \int_0^1 f(\xi) \, d\hat{E}(\xi) \tag{2.306}$$

Note that the right hand sides of the last two equations are Stieltjes integrals.

If we consider a small increment $d\xi \to 0$, then the limit $d\hat{E}(\xi)/d\xi$ exists and the integration step can be taken as the c-number $d\xi$ rather than $d\hat{E}(\xi)$, which is an operator. We then have an operator valued Riemann integral

$$\int_0^1 f(\xi) \, d\hat{E}(\xi) = \int_{x_{min}}^{x_{max}} f(\xi) \frac{d\hat{E}(\xi)}{d\xi} \, d\xi \tag{2.307}$$

This type of spectral decomposition applies not only to operators with continuous spectra, but also to those having discrete spectra, or even mixed spectra.

For a discrete spectrum, $d\hat{E}(\xi) = 0$ if ξ lies between consecutive eigenvalues, and $d\hat{E}(\xi) = \hat{P}_k$, that is, the projector on the k^{th} eigenstate, if the k^{th} eigenvalue lies between ξ and $\xi + d\xi$.

The projector $\hat{E}(\xi)$ is a bounded operator, which depends on the parameter ξ. It may be a discontinuous function of ξ, but it never is infinite, and we never actually need $d\hat{E}(\xi)/d\xi$. This is the advantage of the Stieltjes integral over the more familiar Riemann integral: the left-hand side of the last equation is always

meaningful, even if the the right-hand side is not.

Some Useful Properties

If $f(\xi)$ is a real function, then $f(x)$ is given by

$$f(x) = \int_0^1 f(\xi)\, d\hat{E}(\xi) \tag{2.308}$$

is a self-adjoint operator.

We also have

$$\int f(\xi)\, d\hat{E}(\xi) \int g(\eta)\, d\hat{E}(\eta) = \int f(\zeta) g(\zeta)\, d\hat{E}(\zeta) \tag{2.309}$$

$$e^{if(x)} = \int e^{if(\xi)}\, d\hat{E}(\xi) \tag{2.310}$$

The spectral decomposition of a self-adjoint operator will allow us to give a rigorous definition of the measurement of a continuous variable. It will be equivalent to an infinite set of yes-no questions where each question is represented by a bounded(but infinitely degenerate) projection operator.

Looking ahead to quantum mechanics .....

An operator such as $\hat{Q}$ that has a continuous spectrum is said to have a *formal* eigenvalue equation in Dirac language

$$\hat{Q}\,|q\rangle = q\,|q\rangle \tag{2.311}$$

In the development of the theory, we will make assumptions that will lead to the orthonormality condition for the continuous case taking the form

$$\langle q'\,|\,q''\rangle = \delta(q' - q'') \tag{2.312}$$

Since this implies that $\langle q\,|\,q\rangle = \infty$, these formal eigenvectors have infinite norm. Thus, the Dirac formulation that we will construct for operators with a continuous spectrum will not fit into the mathematical theory of Hilbert space, which admits only vectors of finite norm.

Operators will take the form

$$\hat{Q} = \int_{-\infty}^{\infty} q\,|q\rangle\,\langle q|\, dq \tag{2.313}$$

which is the continuous analog of our earlier expressions.

The projection operator will be formally given by

$$\hat{E}_\beta = \int_{-\infty}^{\beta} |q\rangle\,\langle q|\, dq \tag{2.314}$$

It is well-defined in Hilbert space, but its derivative

$$\frac{d\hat{E}_\beta}{dq} = |q\rangle\langle q| \tag{2.315}$$

does not exist within the Hilbert space framework.

There are two alternative methods for making quantum mechanics fit within a mathematically rigorous Hilbert space framework. The first would be to restrict or revise the formalism to make it fit(still not admit states of infinite norm). The second would be to extend the Hilbert space so that vectors of infinite norm are allowed. We will discuss these ideas later and make appropriate choices.

We saw earlier in the finite-dimensional case, where we have a discrete spectrum of eigenvalues, that the value of the projection operator $\hat{E}_x$ for the operator $\hat{B}$ jumped by $\hat{P}_k = |b_k\rangle\langle b_k|$ as x passed through the k^{th} eigenvalue b_k.

For the infinite dimensional case, where we can have both a discrete and a continuous spectrum of eigenvalues, the projection operator behaves in the same way as we move about the discrete part of the spectrum. In the continuous part of the spectrum, however, it is possible for the projection operator to exhibit a continuous increase in value.

In a more formal manner we state: if $\hat{B}$ is a self-adjoint operator and $\hat{E}_x$ is the projection operator of its spectral decomposition, then

1. the set of points x on which $\hat{E}_x$ increases is called the *spectrum* of $\hat{B}$; alternatively, a point is in the spectrum if it is not in an interval on which $\hat{E}_x$ is constant.

2. the set of points x on which $\hat{E}_x$ jumps is called the point spectrum of $\hat{B}$; the point spectrum is the set of all eigenvalues.

3. the set of points x on which $\hat{E}_x$ increases continuously is called the *continuous spectrum* of $\hat{B}$.

4. the point spectrum and the continuous spectrum comprise the total spectrum.

In quantum mechanics, as we shall see, a real physical, measurable quantity will be represented by a self-adjoint operator. The spectrum of the operator will be the set of real numbers that correspond to the possible measured values of the physical quantity. The projection operators in the spectral decomposition will be used to describe the probability distributions of these values and the state operators. We will get discrete probability distributions over the point spectrum and continuous probability distributions over the continuous spectrum.

All of these new mathematical quantities will have direct and important physical meaning!

2.19 Functions of Operators (general case); Stone's Theorem

We mentioned earlier that functions of operators also have a decomposition into projection operators. In fact, we wrote

$$g(\hat{B}) = \int_{-\infty}^{\infty} g(x)\, d\hat{E}_x \tag{2.316}$$

The ability to deal with functions of operators will be very important in quantum mechanics, so let us spend more time looking at the mathematical details involved.

For any self-adjoint operator we have the spectral decomposition

$$\hat{B} = \int_{-\infty}^{\infty} x\, d\hat{E}_x \tag{2.317}$$

If $f(x)$ is a complex function of a real variable x, then we can *define* the same function of an operator by

$$\langle\alpha|\, f(\hat{B})\, |\beta\rangle = \int_{-\infty}^{\infty} f(x)\, d\,\langle\alpha|\, \hat{E}_x\, |\beta\rangle \tag{2.318}$$

for all vectors $|\alpha\rangle$ and $|\beta\rangle$. Now a self-adjoint operator is bounded if and only if its spectrum is bounded. If $f(\hat{B})$ is a bounded operator on the spectrum of $\hat{B}$, then it turns out that the above equation for all $|\alpha\rangle$ and $|\beta\rangle$ defines $f(\hat{B})\,|\eta\rangle$ for all vectors $|\eta\rangle$.

Is the above definition reasonable? We can see that it is by looking at some properties and some simple examples.

Let $f(x) = x$. This implies that $f(\hat{B}) = \hat{B}$ since we must have

$$\langle\alpha|\, \hat{B}\, |\beta\rangle = \int_{-\infty}^{\infty} x\, d\,\langle\alpha|\, \hat{E}_x\, |\beta\rangle \tag{2.319}$$

$$\langle\alpha|\, \hat{I}\, |\beta\rangle = \int_{-\infty}^{\infty} d\,\langle\alpha|\, \hat{E}_x\, |\beta\rangle = \langle\alpha\,|\,\beta\rangle \tag{2.320}$$

If we let

$$(f+g)(x) = f(x) + g(x) \tag{2.321}$$

$$(cf)(x) = cf(x) \tag{2.322}$$

we then have

$$(f+g)(\hat{B}) = f(\hat{B}) + g(\hat{B}) \tag{2.323}$$

$$(cf)(\hat{B}) = cf(\hat{B}) \tag{2.324}$$

so that sums and multiples of functions of operators are defined in a standard way.

If we let

$$(fg)(x) = f(x)g(x) \tag{2.325}$$

we then have

$$\begin{aligned}\langle\alpha|\, f(\hat{B})g(\hat{B})\,|\beta\rangle &= \int_{-\infty}^{\infty} f(x)\, d\,\langle\alpha|\, \hat{E}_x g(\hat{B})\,|\beta\rangle \\ &= \int_{-\infty}^{\infty} f(x)\, d_x \int_{-\infty}^{\infty} g(y)\, d_y\, \langle\alpha|\, \hat{E}_x \hat{E}_y\,|\beta\rangle \\ &= \int_{-\infty}^{\infty} f(x)\, d \int_{-\infty}^{x} g(y)\, d\,\langle\alpha|\, \hat{E}_y\,|\beta\rangle \\ &= \int_{-\infty}^{\infty} f(x)g(x)\, d\,\langle\alpha|\, \hat{E}_x\,|\beta\rangle \\ &= \int_{-\infty}^{\infty} (fg)(x)\, d\,\langle\alpha|\, \hat{E}_x\,|\beta\rangle = \langle\alpha|\, (fg)(\hat{B})\,|\beta\rangle\end{aligned} \tag{2.326}$$

so that

$$(fg)(\hat{B}) = f(\hat{B})g(\hat{B}) \tag{2.327}$$

as it should. We can also show that $f(\hat{B})g(\hat{B}) = g(\hat{B})f(\hat{B})$ so that all functions of the operator $\hat{B}$ commute.

With these properties, we can then define a polynomial function of an operator by

$$f(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n \tag{2.328}$$

$$f(\hat{B}) = c_0 + c_1 \hat{B} + c_2 \hat{B}^2 + \ldots + c_n \hat{B}^n \tag{2.329}$$

for any vectors $|\alpha\rangle$ and $|\beta\rangle$.

Thus, products of functions of operators are also defined in the standard way.

If we let $(f^*)(x) = f^*(x)$, then for any vectors $|\alpha\rangle$ and $|\beta\rangle$ we have

$$\begin{aligned}\langle\alpha|\left[f(\hat{B})\right]^*|\beta\rangle &= \langle\beta|\, f(\hat{B})\,|\alpha\rangle^* \\ &= \int_{-\infty}^{\infty} f^*(x)\, d\,\langle\alpha|\, \hat{E}_x\,|\beta\rangle = \int_{-\infty}^{\infty} (f^*)(x)\, d\,\langle\alpha|\, \hat{E}_x\,|\beta\rangle\end{aligned} \tag{2.330}$$

or

$$\left[f(\hat{B})\right]^* = (f^*)(\hat{B}) \tag{2.331}$$

If $f(x)$ is a real function, this implies that $f(\hat{B})$ is also a self-adjoint operator.

If $f^* f = 1$, then $f(\hat{B})$ is a unitary operator since

$$\left[f(\hat{B})\right]^\dagger f(\hat{B}) = \hat{I} = f(\hat{B})\left[f(\hat{B})\right]^\dagger \tag{2.332}$$

Now for any vector $|\alpha\rangle$ we have

$$\begin{aligned}\langle\alpha|\, f(\hat{B})\,|\alpha\rangle &= \int_{-\infty}^{\infty} f(x)\, d\,\langle\alpha|\, \hat{E}_x\,|\alpha\rangle \\ &= \int_{-\infty}^{\infty} f(x)\, d\,\langle\alpha|\, \hat{E}_x \hat{E}_x\,|\alpha\rangle = \int_{-\infty}^{\infty} f(x)\, d\|\hat{E}_x\,|\alpha\rangle\,\|^2\end{aligned} \tag{2.333}$$

If we define a self-adjoint operator to be positive if and only if it spectrum is non-negative, then $f(\hat{B})$ is positive if $f(x)$ is non-negative over the spectrum of $\hat{B}$ and $f(\hat{B})$ is bounded if $|f(x)|$ is bounded over the spectrum of $\hat{B}$.

In the special case where $\hat{B}$ has only a point spectrum, we have

$$\hat{B} = \sum_k b_k \hat{P}_k \tag{2.334}$$

and then for any vectors $|\alpha\rangle$ and $|\beta\rangle$

$$\begin{aligned}\langle\alpha| f(\hat{B}) |\beta\rangle = \langle\alpha| f\left(\sum_k b_k \hat{P}_k\right) |\beta\rangle &= \sum_k \langle\alpha| f(b_k \hat{P}_k) |\beta\rangle \\ &= \sum_k f(b_k) \langle\alpha| \hat{P}_k |\beta\rangle = \langle\alpha| \sum_k f(b_k) \hat{P}_k |\beta\rangle\end{aligned} \tag{2.335}$$

or

$$f(\hat{B}) = \sum_k f(b_k) \hat{P}_k \tag{2.336}$$

as we expected.

We define the same properties for unitary operator. Let

$$\hat{U} = \int_0^{2\pi} e^{ix}\, d\hat{E}_x \tag{2.337}$$

and also let

$$\langle\alpha| \hat{B} |\beta\rangle = \int_0^{2\pi} x\, d \langle\alpha| \hat{E}_x |\beta\rangle \tag{2.338}$$

for all vectors in the space.

This defines a bounded self-adjoint operator $\hat{B}$ with spectral decomposition

$$\hat{B} = \int_0^{2\pi} x\, d\hat{E}_x \tag{2.339}$$

We then have

$$\langle\alpha| \hat{U} |\beta\rangle = \int_0^{2\pi} e^{ix}\, d \langle\alpha| \hat{E}_x |\beta\rangle = \int_0^{2\pi} f(x)\, d \langle\alpha| \hat{E}_x |\beta\rangle = \langle\alpha| f(\hat{B}) |\beta\rangle \tag{2.340}$$

which implies that $\hat{U}$, in this case, must be a particular function of $\hat{B}$, namely $\hat{U} = e^{i\hat{B}}$. In addition, any function of $\hat{U}$ is clearly a function of $\hat{B}$.

There exists another relation between Hermitian and unitary operators that is a fundamental property in quantum mechanics. Let $\hat{H}$ be a self-adjoint operator with the spectral decomposition

$$\hat{H} = \int_{-\infty}^{\infty} x\, d\hat{E}_x \tag{2.341}$$

For every real number t let

$$\langle\alpha|\hat{U}_t|\beta\rangle = \int_0^{2\pi} e^{itx}\, d\langle\alpha|\hat{E}_x|\beta\rangle \tag{2.342}$$

This then defines an operator $\hat{U}_t = e^{it\hat{H}}$ which is unitary since $(e^{itx})^* e^{itx} = 1$. We also have $\hat{U}_0 = \hat{I}$.

Now since

$$e^{itx}e^{it'x} = e^{i(t+t')x} \tag{2.343}$$

we must have

$$\hat{U}_t\hat{U}_{t'} = \hat{U}_{t+t'} \tag{2.344}$$

for all real numbers t and t'.

The converse of this property is called *Stone's theorem*:

For each real number t let $\hat{U}_t$ be a unitary operator. Assume that $\langle\alpha|\hat{U}_t|\beta\rangle$ is a continuous function of t for all vectors $|\alpha\rangle$ and $|\beta\rangle$. If $\hat{U}_0 = \hat{I}$ and $\hat{U}_t\hat{U}_{t'} = \hat{U}_{t+t'}$ for all real numbers t and t', then there is a unique self-adjoint operator $\hat{H}$ such that $\hat{U}_t = e^{it\hat{H}}$ for all t. A vector $|\beta\rangle$ is in the domain of $\hat{H}$ if and only if the vectors

$$\frac{1}{it}(\hat{U}_t - \hat{I})|\beta\rangle \tag{2.345}$$

converge to a limit as $t \to 0$. The limit vector is $\hat{H}|\beta\rangle$. If a bounded operator commutes with $\hat{U}_t$, then it commutes with $\hat{H}$.

This theorem leads to the following very important result.

If $\hat{U}_t|\beta\rangle$ is in the domain of $\hat{H}$, then

$$\frac{1}{i\Delta t}(\hat{U}_{\Delta t} - \hat{I})\hat{U}_t|\beta\rangle \to \hat{H}\hat{U}_t|\beta\rangle \tag{2.346}$$

or

$$\frac{1}{i\Delta t}(\hat{U}_{\Delta t}\hat{U}_t - \hat{U}_t)|\beta\rangle = \frac{1}{i\Delta t}(\hat{U}_{t+\Delta t} - \hat{U}_t)|\beta\rangle \to -i\frac{d}{dt}\hat{U}_t|\beta\rangle \tag{2.347}$$

as $\Delta t \to 0$. We can then write

$$-i\frac{d}{dt}\hat{U}_t|\beta\rangle = \hat{H}\hat{U}_t|\beta\rangle \tag{2.348}$$

This equation will eventually tell us how the physical states(ket vectors) and state operators that will represent physical systems evolve in time (it will lead to the Schrodinger equation, which is the time evolution equation in one representation of quantum mechanics).

Examples - Functions of Operators

Suppose that we have the eigenvector/eigenvalue equations for a self-adjoint operator

$$\hat{A}\,|k\rangle = a_k\,|k\rangle \quad , \quad k = 1, 2, \ldots, N \tag{2.349}$$

We then *assume* that

$$f(\hat{A})\,|k\rangle = f(a_k)\,|k\rangle \quad , \quad k = 1, 2, \ldots, N \tag{2.350}$$

for the eigenvectors.

We can show that this works for polynomials and power series as follows:

$$|\psi\rangle = \sum_{k=1}^{N} |k\rangle\,\langle k\,|\,\psi\rangle \tag{2.351}$$

$$\begin{aligned} \hat{A}\,|\psi\rangle &= \hat{A}\sum_{k=1}^{N} |k\rangle\,\langle k\,|\,\psi\rangle = \sum_{k=1}^{N} \hat{A}\,|k\rangle\,\langle k\,|\,\psi\rangle \\ &= \sum_{k=1}^{N} a_k\,|k\rangle\,\langle k\,|\,\psi\rangle = \left(\sum_{k=1}^{N} a_k\,|k\rangle\,\langle k|\right)|\psi\rangle \end{aligned} \tag{2.352}$$

$$\rightarrow \hat{A} = \sum_{k=1}^{N} a_k\,|k\rangle\,\langle k| \rightarrow \text{ spectral resolution of the operator} \tag{2.353}$$

Now define the *projection* operator

$$\hat{P}_k = |k\rangle\,\langle k| \rightarrow \hat{P}_k\hat{P}_j = \hat{P}_k\delta_{kj} \tag{2.354}$$

We then have

$$\hat{A} = \sum_{k=1}^{N} a_k\,|k\rangle\,\langle k| = \sum_{k=1}^{N} a_k\hat{P}_k \tag{2.355}$$

or any operator is represented by a sum over its eigenvalues and corresponding projection operators.

We then have

$$\begin{aligned} \hat{A}^2 &= \left(\sum_{k=1}^{N} a_k\hat{P}_k\right)\left(\sum_{j=1}^{N} a_j\hat{P}_j\right) = \sum_{k,j=1}^{N} a_k a_j \hat{P}_k\hat{P}_j \\ &= \sum_{k,j=1}^{N} a_k a_j \hat{P}_k\delta_{kj} = \sum_{k=1}^{N} a_k^2\hat{P}_k \rightarrow \hat{A}^n = \sum_{k=1}^{N} a_k^n\hat{P}_k \end{aligned} \tag{2.356}$$

Therefore, for

$$f(x) = \sum_{n=1}^{N} q_n x^n \tag{2.357}$$

we have

$$f(\hat{A}) = \sum_{n=1}^{N} q_n \hat{A}^n = \sum_{n=1}^{N} q_n \sum_{k=1}^{N} a_k^n \hat{P}_k \tag{2.358}$$
$$= \sum_{k=1}^{N} \left(\sum_{n=1}^{N} q_n a_k^n \right) \hat{P}_k = \sum_{k=1}^{N} f(a_k) \hat{P}_k$$

This says that, in general, we have

$$f(\hat{A}) |\psi\rangle = f(\hat{A}) \sum_{k=1}^{N} |k\rangle \langle k | \psi\rangle = \sum_{k=1}^{N} f(\hat{A}) |k\rangle \langle k | \psi\rangle \tag{2.359}$$
$$= \sum_{k=1}^{N} f(a_k) |k\rangle \langle k | \psi\rangle = \left(\sum_{k=1}^{N} f(a_k) |k\rangle \langle k| \right) |\psi\rangle$$

$$\rightarrow f(\hat{A}) = \sum_{k=1}^{N} f(a_k)^n |k\rangle \langle k| \rightarrow \text{ spectral resolution of the operator} \tag{2.360}$$

Numerical example: consider the operator

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

which has eigenvalues 7 and 1 with eigenvectors

$$|7\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad , \qquad |1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

This gives

$$\hat{P}_7 = |7\rangle \langle 7| = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad , \qquad \hat{P}_1 = |1\rangle \langle 1| = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and therefore

$$\hat{A} = 7\hat{P}_7 + \hat{P}_1 = \frac{7}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

$$\hat{A}^2 = 7^2 \hat{P}_7 + \hat{P}_1 = \frac{49}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 25 & 24 \\ 24 & 25 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

$$\log(\hat{A}) = \log(7)\hat{P}_7 + \log(1)\hat{P}_1 = \frac{\log(7)}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\sqrt{\hat{A}} = \sqrt{7}\hat{P}_7 + \hat{P}_1 = \frac{\sqrt{7}}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{7}+1 & \sqrt{7}-1 \\ \sqrt{7}-1 & \sqrt{7}+1 \end{pmatrix}$$

Clearly we then have

$$\log(\hat{A}) |7\rangle = \frac{\log(7)}{2\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{\log(7)}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \log(7) |7\rangle$$

$$\log(\hat{A})\,|1\rangle = \frac{\log(7)}{2\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix} = \frac{\log(7)}{2\sqrt{2}}\begin{pmatrix}0\\ 0\end{pmatrix} = 0 = \log(1)\,|1\rangle$$

as expected.

The big question that remains is then

$$\text{Is} \quad \hat{f}(\hat{A}) = f(\hat{A}) \ ? \tag{2.361}$$

no proof exists!

2.20 Commuting Operators

As we stated earlier, the commutator of two operators is given by

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} \tag{2.362}$$

If $\hat{A}$ and $\hat{B}$ are self-adjoint operators, each possessing a complete set of eigenvectors, and if they commute, then there exists a complete set of vectors which are eigenvectors of both $\hat{A}$ and $\hat{B}$, that is, they possess a common set of eigenvectors.

This theorem extends to any number of commuting operators. If we have a set of N mutually commuting operators, then they all have a common set of eigenvectors.

The reverse is also true. If two operators possess a common set of eigenvectors, then they commute.

Let $\{\hat{A}, \hat{B}, \hat{C}, \ldots\}$ be a set of mutually commuting operators that possess a complete set of common eigenvectors. Corresponding to a particular eigenvalue for each operator, there may be more than one eigenvector. If, however, there is no more than one eigenvector for each set of eigenvalues $(a_k, b_k, c_k, \ldots)$, then the operators $\{\hat{A}, \hat{B}, \hat{C}, \ldots\}$ are said to be a complete commuting set of operators.

Any operator that commutes with all members of a complete commuting set must be a function of the operators in that set.

Let us now think about these ideas in terms of projection operators.

Let $\hat{Q}$ be a Hermitian operator with a pure point spectrum so that we can write

$$\hat{Q} = \sum_k q_k \hat{P}_k \tag{2.363}$$

where each q_k is a *different* eigenvalue of $\hat{Q}$ and $\hat{P}_k$ is the projection operator onto the subspace corresponding to eigenvalue q_k.

Let $\hat{R}$ be a bounded Hermitian operator that commutes with $\hat{Q}$. For each k and for any arbitrary vector $|\eta\rangle$ we then have

$$\begin{aligned}\hat{Q}\hat{R}\hat{P}_k|\eta\rangle = \hat{R}\hat{Q}\hat{P}_k|\eta\rangle &= \hat{R}\left(\sum_j q_j\hat{P}_j\right)\hat{P}_k|\eta\rangle \\ &= \hat{R}\sum_j q_j\delta_{jk}\hat{P}_k|\eta\rangle = q_k\hat{R}\hat{P}_k|\eta\rangle\end{aligned} \tag{2.364}$$

where we have used the relation

$$\hat{P}_j\hat{P}_k = \delta_{jk}\hat{P}_k \tag{2.365}$$

for any vector $|\eta\rangle$. Thus, $\hat{R}\hat{P}_k|\eta\rangle$ is an eigenvector of $\hat{Q}$ with eigenvalue q_k. Therefore, $\hat{P}_k\hat{R}\hat{P}_k|\eta\rangle = \hat{R}\hat{P}_k|\eta\rangle$ for all $|\eta\rangle$ and we have

$$\hat{R}\hat{P}_k = \hat{P}_k\hat{R}\hat{P}_k \tag{2.366}$$

Taking the adjoint of both sides we get

$$\left(\hat{R}\hat{P}_k\right)^\dagger = \left(\hat{P}_k\hat{R}\hat{P}_k\right)^\dagger \tag{2.367}$$

$$\hat{R}\hat{P}_k = \hat{P}_k\hat{R}\hat{P}_k \tag{2.368}$$

Thus,

$$\hat{R}\hat{P}_k = \hat{P}_k\hat{R} \tag{2.369}$$

or *each* $\hat{P}_k$ commutes with every bounded Hermitian operator which commutes with $\hat{Q}$.

We can extend this result to operators possessing both a point and continuous spectrum. If $\hat{Q}$ is a self-adjoint operator with the spectral decomposition

$$\hat{Q} = \int_{-\infty}^{\infty} x\, d\hat{E}_x \tag{2.370}$$

and if $\hat{R}$ is a bounded self-adjoint operator that commutes with $\hat{Q}$, then

$$\hat{E}_x\hat{R} = \hat{R}\hat{E}_x \tag{2.371}$$

for every x.

Let us now express the ideas of complete commuting sets in terms of projection operators.

Let $\{\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_N\}$ be a set of mutually commuting Hermitian operators with pure point spectra. For each we then have

$$\hat{B}_r = \sum_k b_k^{(r)}\hat{P}_k^{(r)} \quad , \quad r = 1, 2, \ldots, N \tag{2.372}$$

where each $b_k^{(r)}$ is a different eigenvalue of $\hat{B}_r$ and $\hat{P}_k^{(r)}$ is the projection operator onto the subspace spanned by the eigenvectors of $\hat{B}_r$ corresponding to $b_k^{(r)}$.

By definition then, the projection operators $\hat{P}_k^{(r)}$ commute with each other for all different r and k

$$\hat{P}_j^{(r)} \hat{P}_k^{(s)} = \hat{P}_k^{(s)} \hat{P}_j^{(r)} \tag{2.373}$$

This implies that

$$\hat{P}_i^{(1)} \hat{P}_j^{(2)} \dots \hat{P}_l^{(N)} \tag{2.374}$$

is the projection operator for any $i, j, \dots, N$, that is, it projects onto the subspace of all vectors $|\alpha\rangle$ such that

$$\hat{B}_1 |\alpha\rangle = b_i^{(1)} |\alpha\rangle \ , \ \hat{B}_2 |\alpha\rangle = b_j^{(2)} |\alpha\rangle \ , \ \dots \ , \ \hat{B}_N |\alpha\rangle = b_l^{(N)} |\alpha\rangle \tag{2.375}$$

These projection operators are mutually orthogonal

$$\hat{P}_i^{(1)} \hat{P}_j^{(2)} \dots \hat{P}_l^{(N)} \hat{P}_{i'}^{(1)} \hat{P}_{j'}^{(2)} \dots \hat{P}_{l'}^{(N)} = \delta_{ii'} \delta_{jj'} \dots \delta_{ll'} \hat{P}_i^{(1)} \hat{P}_j^{(2)} \dots \hat{P}_l^{(N)} \tag{2.376}$$

and they have the completeness property that

$$\sum_i \sum_j \dots \sum_l \hat{P}_i^{(1)} \hat{P}_j^{(2)} \dots \hat{P}_l^{(N)} = \hat{I} \tag{2.377}$$

Note that some of these projection operators might be zero.

Suppose that none of them projects onto a subspace of dimension larger than one. In this case, we say that the set of operators $\{\hat{B}_1, \hat{B}_2, \dots, \hat{B}_N\}$ is a *complete set of commuting operators.*

Now let us return to the study of a continuous spectrum.

First, we repeat some earlier material to set the stage.

Let us start with a very simple example so we can figure out how to proceed and then generalize to a more complicated case. We consider the space $L^2(-\infty, \infty)$ and a single Hermitian operator $\hat{Q}$ defined by

$$(\hat{Q}g)(x) = xg(x) \tag{2.378}$$

It turns out that every bounded operator which commutes with $\hat{Q}$ is a function of $\hat{Q}$.

Now, there exists a theorem: Suppose we have a set of mutually commuting operators $\{\hat{A}_i\}$. This is a complete set of commuting operators if an only if every bounded operator which commutes with all $\{\hat{A}_i\}$ is a function of the $\{\hat{A}_i\}$.

In the previous discussion, we had the case of a complete commuting set consisting of a single operator.

The spectrum of $\hat{Q}$ is purely continuous and consists of all real numbers x. Each vector g is a function $g(x)$ on the spectrum of $\hat{Q}$.

We connect this to the case of a complete set of commuting operators with a pure point spectra as follows.

We define two abstract vectors $|x\rangle$ and $|g\rangle$ such that

$$\langle x \,|\, g \rangle = g(x) \tag{2.379}$$

We then have

$$(\hat{Q}g)(x) = \left\langle x \,\middle|\, \hat{Q}g \right\rangle = x \langle x \,|\, g \rangle = xg(x) \tag{2.380}$$

which is the spectral representation of $\hat{Q}$. We can generalize to a function of $\hat{Q}$ with

$$\left\langle x \,\middle|\, f(\hat{Q})g \right\rangle = f(x) \langle x \,|\, g \rangle = f(x)g(x) \tag{2.381}$$

In Dirac language, we write an abstract equation like

$$\hat{Q} \,|x\rangle = x \,|x\rangle \tag{2.382}$$

We then have

$$x^* \langle x| = x \langle x| = \langle x| \hat{Q}^\dagger = \langle x| \hat{Q} \tag{2.383}$$

and

$$\langle x| \hat{Q} \,|g\rangle = x \langle x \,|\, g \rangle = xg(x) \tag{2.384}$$

which again gives the spectral representation of $\hat{Q}$.

Finally we have

$$\langle x| f(\hat{Q}) \,|g\rangle = f(x) \langle x \,|\, g \rangle = f(x)g(x) \tag{2.385}$$

The problem with defining an abstract Hermitian operator $\hat{Q}$ by

$$\hat{Q} \,|x\rangle = x \,|x\rangle \tag{2.386}$$

is that $\hat{Q}$ has no eigenvectors in the Hilbert space L^2 of square-integrable functions. In order for there to be eigenvectors we must have

$$(\hat{Q}g)(x) = \hat{Q}g(x) = xg(x) = ag(x) \tag{2.387}$$

for a real number a. This implies that $g(x)$ is zero for all points $x \neq a$ and

$$\|g\|^2 = \int_{-\infty}^{\infty} |g|^2 = 0 \tag{2.388}$$

because the *standard* integral is not changed by the value of the integrand at a *single point*.

Now we have

$$\langle x| \hat{Q} \,|a\rangle = x \langle x \,|\, a \rangle = a \langle x \,|\, a \rangle \tag{2.389}$$

If we replace the inner product by

$$\langle x \,|\, a \rangle = \langle a \,|\, x \rangle = \delta(x - a) \tag{2.390}$$

or, in general,

$$\langle x \,|\, x' \rangle = \langle x' \,|\, x \rangle = \delta(x - x') \tag{2.391}$$

We then have for each real number a

$$x\delta(x - a) = a\delta(x - a) \tag{2.392}$$

which is a *valid mathematical relation* for delta functions.

Thus, we can formally use Dirac delta functions for eigenfunctions of $\hat{Q}$ as follows: for each real number a we have

$$x\delta(x - a) = a\delta(x - a) \tag{2.393}$$

If we write $|a\rangle$ for $\delta(x - a)$, we then have

$$\hat{Q}\,|a\rangle = a\,|a\rangle \tag{2.394}$$

We must not consider $\delta(x - a)$ as a standard integrable function and we cannot think of $|a\rangle$ as a vector in the Hilbert space L^2. We must do all mathematics using the standard delta function rules.

In this way we have

$$\langle a \,|\, g \rangle = g(a) = \int_{-\infty}^{\infty} \delta(x - a) g(x)\, dx \tag{2.395}$$

as the components of a vector $|g\rangle$ in the spectral representation of $\hat{Q}$. Note the shift from a function $g(x)$ to the ket vector $|g\rangle$ and the relationship between the two mathematical objects. In fact, we can write

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} \delta(x - a) g(a)\, da = \langle x \,|\, g \rangle = \langle x |\, \hat{I}\, |g\rangle \\ &= \langle x | \left(\int_{-\infty}^{\infty} |a\rangle \langle a|\, da \right) |g\rangle = \langle x | \int_{-\infty}^{\infty} \langle a \,|\, g \rangle\, |a\rangle\, da \end{aligned} \tag{2.396}$$

or

$$|g\rangle = \int_{-\infty}^{\infty} \langle a \,|\, g \rangle\, |a\rangle\, da \tag{2.397}$$

In addition, we have (using the properties of projection operators derived earlier)

$$\begin{aligned} \langle a \,|\, g \rangle &= \langle a |\, \hat{I}\, |g\rangle = \langle a | \int |x\rangle \langle x|\, dx\, |g\rangle \\ &= \int \langle a \,|\, x \rangle \langle x \,|\, g \rangle\, dx = \int \delta(x - a) g(x)\, dx \\ &= g(a) \end{aligned} \tag{2.398}$$

as expected.

Thus, formally, we can think of any eigenfunction $g(x)$ as a linear combination of delta functions, where the delta functions are analogous to an orthonormal basis of eigenvectors with the symbol $|x\rangle$. We the have

$$|g\rangle = \int_{-\infty}^{\infty} \langle x\,|\,g\rangle\,|x\rangle\;dx \quad , \quad \hat{I} = \int_{-\infty}^{\infty} |x\rangle\,\langle x|\;dx \tag{2.399}$$

Thus, for each real number a, we define the operator $|a\rangle\,\langle a|$ by

$$(|a\rangle\,\langle a|\,g)(x) = \langle x\,|\,a\rangle\,\langle a\,|\,g\rangle = g(a)\delta(x-a) \tag{2.400}$$

$$|a\rangle\,\langle a|\,g = \langle a\,|\,g\rangle\,|a\rangle = g(a)\,|a\rangle \tag{2.401}$$

In a similar manner, the projection operator $\hat{E}_x$ in the spectral decomposition of $\hat{Q}$ is given by

$$(\hat{E}_x g)(y) = \int_{-\infty}^{x} g(a)\delta(y-a)\,da \tag{2.402}$$

which we can write as

$$(\hat{E}_x g)(y) = \int_{-\infty}^{x} (|a\rangle\,\langle a|\,g)(y)\,da \tag{2.403}$$

This says that

$$\hat{E}_x = \int_{-\infty}^{x} |a\rangle\,\langle a|\;da \tag{2.404}$$

Finally, we have

$$\hat{Q} = \int_{-\infty}^{\infty} x\,d\hat{E}_x \tag{2.405}$$

which we write as

$$\hat{Q} = \int_{-\infty}^{\infty} x\,|x\rangle\,\langle x|\;dx \tag{2.406}$$

This is analogous to a sum of eigenvalues multiplying projection operators onto eigenvector subspaces. We will return to this discussion when we introduce the position operator.

2.21 Another Continuous Spectrum Operator

We already have the relation

$$\langle x\,|\,x'\rangle = \delta(x-x') \tag{2.407}$$

Let us introduce a new operator $\hat{p}$ such that

$$\hat{p}\,|p\rangle = p\,|p\rangle \quad , \quad -\infty \le p \le \infty \quad , \quad p \text{ real} \tag{2.408}$$

Thus $\hat{p}$ is an operator with a continuous spectrum just like $\hat{Q}$. As we found to be true for $\hat{Q}$, we can now also write

$$\hat{p} = \frac{1}{2\pi\hbar}\int p\,|p\rangle\,\langle p|\;dp \qquad , \qquad \hat{I} = \frac{1}{2\pi\hbar}\int |p\rangle\,\langle p|\;dp \tag{2.409}$$

This allows us to write

$$\langle x' \,|\, x \rangle = \delta(x' - x) = \langle x' | \, \hat{I} \, | x \rangle = \frac{1}{2\pi\hbar} \int \langle x' | \, (|p\rangle \langle p|) \, | x \rangle \, dp \tag{2.410}$$
$$= \frac{1}{2\pi\hbar} \int \langle x' \,|\, p \rangle \langle p \,|\, x \rangle \, dp = \frac{1}{2\pi\hbar} \int \langle x' \,|\, p \rangle \langle x \,|\, p \rangle^* \, dp$$

Now, one of the standard representations of the delta function is

$$\delta(x - x') = \frac{1}{2\pi\hbar} \int e^{-ip(x-x')/\hbar} \, dp \tag{2.411}$$

Thus we can write

$$\int \langle x' \,|\, p \rangle \langle x \,|\, p \rangle^* \, dp = \int e^{-ip(x-x')/\hbar} \, dp \tag{2.412}$$

One of the most important solutions to this equation is

$$\langle x \,|\, p \rangle = e^{ipx/\hbar} \rightarrow \langle x \,|\, p \rangle^* = e^{-ipx/\hbar} \quad , \quad \langle x' \,|\, p \rangle = e^{ipx'/\hbar} \tag{2.413}$$

As we shall see later, this choice will correspond to the new operator $\hat{p}$ representing the standard linear momentum.

We can then write

$$\langle x | \, \hat{p} \, | p \rangle = p \langle x \,|\, p \rangle = p e^{ipx/\hbar} = -i\hbar \frac{\partial}{\partial x} e^{ipx/\hbar} = -i\hbar \frac{\partial}{\partial x} \langle x \,|\, p \rangle \tag{2.414}$$

which, in our earlier notation, says

$$(\hat{p}g)(x) = -i\hbar \frac{\partial}{\partial x} g(x) \tag{2.415}$$

In addition we can write

$$\begin{aligned}
\langle p | \, \hat{x} \, | \psi \rangle &= [\langle \psi | \, \hat{x} \, | p \rangle]^* = \left[\langle \psi | \, \hat{x} \left(\int |x'\rangle \langle x'| \, dx' \right) |p\rangle \right]^* \\
&= \left[\int \langle \psi | \, \hat{x} \, |x'\rangle \langle x' \,|\, p \rangle \, dx' \right]^* = \left[\int x' \langle \psi \,|\, x' \rangle \langle x' \,|\, p \rangle \, dx' \right]^* \\
&= \left[\int \langle \psi \,|\, x' \rangle \, x' \langle x' \,|\, p \rangle \, dx' \right]^* = \left[\int \langle \psi \,|\, x' \rangle \left(-i\hbar \frac{\partial}{\partial p} \right) \langle x' \,|\, p \rangle \, dx' \right]^* \\
&= \left[\left(-i\hbar \frac{\partial}{\partial p} \right) \int \langle \psi \,|\, x' \rangle \langle x' \,|\, p \rangle \, dx' \right]^* = i\hbar \frac{\partial}{\partial p} \left[\langle \psi | \left(\int |x'\rangle \langle x'| \, dx' \right) |p\rangle \right]^* \\
&= i\hbar \frac{\partial}{\partial p} [\langle \psi \,|\, p \rangle]^* = i\hbar \frac{\partial}{\partial p} \langle p \,|\, \psi \rangle \\
&\Rightarrow \langle p | \, \hat{x} = i\hbar \frac{\partial}{\partial p} \langle p |
\end{aligned} \tag{2.416}$$

since $|\psi\rangle$ is arbitrary.

Now, since the eigenvectors of $\hat{p}$ form a basis, we can write any arbitrary vector as

$$|g\rangle = \int \langle p\,|\,g\rangle\,|p\rangle\,dp \tag{2.417}$$

which implies

$$\langle x\,|\,g\rangle = \int \langle p\,|\,g\rangle\,\langle x\,|\,p\rangle\,dp = \int \langle p\,|\,g\rangle\,e^{ipx/\hbar}\,dp \tag{2.418}$$

Now the theory of Fourier transforms says that

$$g(x) = \int G(p)e^{ipx/\hbar}\,dp \tag{2.419}$$

where $G(p)$ is the Fourier transform of $g(x)$. Thus, we find that $G(p) = \langle p\,|\,g\rangle$ is the Fourier transform of $g(x)$.

More about Fourier Transforms (in general)

In the space L^2 of square-integrable functions, let us consider a self-adjoint operator defined by the relation we found earlier for the $\hat{p}$ operator

$$(\hat{p}g)(x) = -i\hbar\frac{\partial}{\partial x}g(x) \tag{2.420}$$

As we already have seen there is a direct connection here to Fourier transforms. Let us review some of the mathematical concepts connected with the *Fourier Transform*.

If g is a function(vector) in L^2, then

$$\psi_n(k) = \frac{1}{2\pi}\int_{-n}^{n} e^{-ikx}g(x)\,dx \tag{2.421}$$

defines a sequence of functions(vectors) ψ_n in L^2 which converges as $n \to \infty$ to a limit function(vector) Gg such that $\|Gg|^2 = \|g\|^2$ and

$$\psi(k) = (Gg)(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}g(x)\,dx \tag{2.422}$$

We also have

$$g_n(k) = \frac{1}{2\pi}\int_{-n}^{n} e^{ikx}(Gg)(k)\,dk \tag{2.423}$$

which defines a sequence of functions(vectors) that converges to g as $n \to \infty$ where

$$g(x) = \frac{1}{2\pi}\int (Gg)(k)e^{ikx}\,dk \tag{2.424}$$

Now, this gives

$$(\hat{p}g)(x) = -i\hbar\frac{\partial}{\partial x}g(x) = \frac{\hbar}{2\pi}\int (Gg)(k)ke^{ikx}\,dk \tag{2.425}$$

It is clear from this expression that a vector g is in the domain of $\hat{p}$ if and only if the quantity $k(Gg)(k)$ is square-integrable. We then have

$$\begin{aligned}(G\hat{p}g)(k) &= \frac{1}{2\pi}\int(\hat{p}g)(x)e^{-ikx}\,dx = \frac{1}{2\pi}\int\frac{1}{2\pi}\left(\int(Gg)(k')\hbar k' e^{ik'x}\,dk'\right)e^{-ikx}\,dx \\ &= \frac{1}{2\pi}\int(Gg)(k')\hbar k'\,dk'\,\frac{1}{2\pi}\int e^{i(k'-k)x}\,dx \\ &= \frac{1}{2\pi}\int(Gg)(k')\hbar k'\delta(k'-k)\,dk' = k(Gg)(k) \end{aligned} \tag{2.426}$$

We call G the *Fourier transform* of g. G is a unitary operator on L^2 . Its inverse is given by

$$(G^{-1}h)(x) = \frac{1}{2\pi}\int h(k)e^{ikx}\,dk \tag{2.427}$$

which implies

$$(G^{-1}h)(x) = (Gh)(-x) \tag{2.428}$$

for every h. Since G is unitary, it preserves inner products as well as lengths of vectors so we have

$$\int(Gh)(k)^*(Gg)(k)\,dk = \int h^*(x)g(x)\,dx \tag{2.429}$$

for all vectors h and g.

In terms of the operator $\hat{Q}$ defined by

$$(\hat{Q}g)(x) = xg(x) \tag{2.430}$$

it can be shown that

$$G\hat{p} = \hat{Q}G \tag{2.431}$$

or

$$\hat{p} = G^{-1}\hat{Q}G \tag{2.432}$$

From the spectral decomposition

$$\hat{Q} = \int_{-\infty}^{\infty} y\,d\hat{E}_y \tag{2.433}$$

we can then obtain the spectral decomposition of $\hat{p}$. Since G is unitary and the set of operators $G^{-1}\hat{E}_yG$ is a spectral family of projection operators, the set of operators $\hat{E}_y$ is also a spectral family of projection operators.

Since $G^{-1} = G^\dagger$, we have

$$\begin{aligned}(h,\hat{p}g) &= (h,G^\dagger\hat{Q}Gg) = (Gh,\hat{Q}Gg) \\ &= \int_{-\infty}^{\infty} y\,d(Gh,\hat{E}_yGg) = \int_{-\infty}^{\infty} y\,d(h,G^{-1}\hat{E}_yGg)\end{aligned} \tag{2.434}$$

for any vector h and any vector g in the domain of $\hat{p}$. Thus the spectral decomposition of $\hat{p}$ is

$$\hat{p} = \int_{-\infty}^{\infty} y \, d(G^{-1} \hat{E}_y G) \tag{2.435}$$

Now recall that $\hat{E}_y$ is the projection operator onto the subspace of all vectors g such that $g(x) = 0$ for $x > y$. Therefore, $G^{-1} \hat{E}_y G$ is the projection operator onto the subspace of all vectors g such that $(Gg)(k) = 0$ for $k > y$.

This means that $\hat{p}_r$ has the same spectrum as $\hat{Q}_r$, namely, a purely continuous spectrum consisting of all real numbers as we already assumed at the beginning of our discussion. These results generalize to functions of the operators.

We have been thinking of the Fourier transform as an operator G which takes a vector g to a different vector Gg. We may also think of $g(x)$ and $(Gg)(k)$ as two different ways of representing the same vector g as a function. We can write

$$\langle k \,|\, g \rangle = (Gg)(k) \tag{2.436}$$

provided we are careful not to confuse this with

$$\langle x \,|\, g \rangle = g(x) \tag{2.437}$$

We think of $\langle k \,|\, g \rangle$ as a function on the spectra of $\hat{p}$. We then have

$$\langle k | \, \hat{p} \, | g \rangle = k \, \langle k \,|\, g \rangle \tag{2.438}$$

which is the spectral representation of $\hat{p}$.

For a function f of $\hat{p}$ we have

$$\langle k | \, f(\hat{p}) \, | g \rangle = f(k) \, \langle k \,|\, g \rangle \tag{2.439}$$

The operator $\hat{p}$ has no eigenvectors (as was true earlier for $\hat{Q}$), It does, however, have eigenfunctions which we can use as analogs of eigenvectors as we did earlier for $\hat{Q}$.

If we write $|k\rangle$ for

$$\frac{1}{\sqrt{2\pi}} e^{ikx} \tag{2.440}$$

we have

$$\hat{p} \, |k\rangle = \hbar k \, |k\rangle \tag{2.441}$$

as we assumed at the beginning, since

$$-i\hbar \frac{\partial}{\partial x} \left(\frac{1}{\sqrt{2\pi}} e^{ikx} \right) = \hbar k \left(\frac{1}{\sqrt{2\pi}} e^{ikx} \right) \tag{2.442}$$

For the components of a vector g n the spectral representation of the operators $\hat{p}$ we have

$$\langle k \,|\, g \rangle = (Gg)(k) = \frac{1}{\sqrt{2\pi}} \int g(x) e^{-ikx} \, dx \tag{2.443}$$

We can think of these as the inner products of g with the eigenfunctions $e^{ikx}/\sqrt{2\pi}$. We have also

$$g(x) = \frac{1}{\sqrt{2\pi}} \int (Gg)(k) e^{ikx} \, dk \tag{2.444}$$

which we can write as

$$|k\rangle = \int \langle k \,|\, g \rangle \, |k\rangle \, dk \tag{2.445}$$

so we can think of any vector g as a linear combination of the eigenfunctions.

Thus, the eigenfunctions $e^{ikx}/\sqrt{2\pi}$ are analogous to an orthonormal basis of eigenvectors. They are not vectors in the Hilbert space L^2, however, because they are not square-integrable.

We use them in the same way that we earlier used the delta functions for eigenfunctions of the operators $\hat{Q}$. In fact,

$$\frac{1}{\sqrt{2\pi}} e^{ikx} = \frac{1}{\sqrt{2\pi}} \int \delta(k' - k) e^{ik'x} \, dk' \tag{2.446}$$

is the inverse Fourier transform of the delta function.

be defined by Now let $|k\rangle \langle k|$

$$(|k\rangle \langle k \,|\, g\rangle)(x) = (Gg)k \frac{1}{\sqrt{2\pi}} e^{ikx} \tag{2.447}$$

or

$$|k\rangle \langle k| \, g = \langle k \,|\, g \rangle \, |k\rangle \tag{2.448}$$

Then for the projection operators $G^{-1} \hat{E}_y G$ in the spectral decomposition of $\hat{p}$ we can write

$$(G^{-1} \hat{E}_y G g)(x) = \int_{k \le y} (Gg)k \frac{1}{\sqrt{2\pi}} e^{ikx} \, dk = \int_{k \le y} (|k\rangle \langle k \,|\, g\rangle)(x) \, dk \tag{2.449}$$

or

$$G^{-1} \hat{E}_y G = \int_{k \le y} (|k\rangle \langle k| \, dk \tag{2.450}$$

and for the spectral decomposition of the operators $\hat{p}$ we get

$$\hat{p} = \int \hbar k \, |k\rangle \langle k| \, dk \tag{2.451}$$

which is the same spectral decompositions in terms of eigenvalues and eigenvectors that we saw earlier.

2.22 Problems

2.22.1 Simple Basis Vectors

Given two vectors

$$\vec{A} = 7\hat{e}_1 + 6\hat{e}_2 \quad , \quad \vec{B} = -2\hat{e}_1 + 16\hat{e}_2$$

written in the $\{\hat{e}_1, \hat{e}_2\}$ basis set and given another basis set

$$\hat{e}_q = \frac{1}{2}\hat{e}_1 + \frac{\sqrt{3}}{2}\hat{e}_2 \quad , \quad \hat{e}_p = -\frac{\sqrt{3}}{2}\hat{e}_1 + \frac{1}{2}\hat{e}_2$$

(a) Show that $\hat{e}_q$ and $\hat{e}_p$ are orthonormal.

(b) Determine the new components of $\vec{A}$, $\vec{B}$ in the $\{\hat{e}_q, \hat{e}_p\}$ basis set.

2.22.2 Eigenvalues and Eigenvectors

Find the eigenvalues and normalized eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 0 \\ 5 & 0 & 3 \end{pmatrix}$$

Are the eigenvectors orthogonal? Comment on this.

2.22.3 Orthogonal Basis Vectors

Determine the eigenvalues and eigenstates of the following matrix

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

Using Gram-Schmidt, construct an orthonormal basis set from the eigenvectors of this operator.

2.22.4 Operator Matrix Representation

If the states $\{|1\rangle, |2\rangle |3\rangle\}$ form an orthonormal basis and if the operator $\hat{G}$ has the properties

$$\begin{aligned} \hat{G}|1\rangle &= 2|1\rangle - 4|2\rangle + 7|3\rangle \\ \hat{G}|2\rangle &= -2|1\rangle + 3|3\rangle \\ \hat{G}|3\rangle &= 11|1\rangle + 2|2\rangle - 6|3\rangle \end{aligned}$$

What is the matrix representation of $\hat{G}$ in the $|1\rangle, |2\rangle |3\rangle$ basis?

2.22.5 Matrix Representation and Expectation Value

If the states $\{|1\rangle, |2\rangle |3\rangle\}$ form an orthonormal basis and if the operator $\hat{K}$ has the properties

$$\begin{aligned}\hat{K}|1\rangle &= 2|1\rangle \\ \hat{K}|2\rangle &= 3|2\rangle \\ \hat{K}|3\rangle &= -6|3\rangle\end{aligned}$$

(a) Write an expression for $\hat{K}$ in terms of its eigenvalues and eigenvectors (projection operators). Use this expression to derive the matrix representing $\hat{K}$ in the $|1\rangle, |2\rangle |3\rangle$ basis.

(b) What is the *expectation or average value* of $\hat{K}$, defined as $\langle\alpha|\hat{K}|\alpha\rangle$, in the state

$$|\alpha\rangle = \frac{1}{\sqrt{83}}\left(-3|1\rangle + 5|2\rangle + 7|3\rangle\right)$$

2.22.6 Projection Operator Representation

Let the states $\{|1\rangle, |2\rangle |3\rangle\}$ form an orthonormal basis. We consider the operator given by $\hat{P}_2 = |2\rangle\langle 2|$. What is the matrix representation of this operator? What are its eigenvalues and eigenvectors. For the arbitrary state

$$|A\rangle = \frac{1}{\sqrt{83}}\left(-3|1\rangle + 5|2\rangle + 7|3\rangle\right)$$

What is the result of $\hat{P}_2|A\rangle$?

2.22.7 Operator Algebra

An operator for a two-state system is given by

$$\hat{H} = a\left(|1\rangle\langle 1| - |2\rangle\langle 2| + |1\rangle\langle 2| + |2\rangle\langle 1|\right)$$

where a is a number. Find the eigenvalues and the corresponding eigenkets.

2.22.8 Functions of Operators

Suppose that we have some operator $\hat{Q}$ such that $\hat{Q}|q\rangle = q|q\rangle$, i.e., $|q\rangle$ is an eigenvector of $\hat{Q}$ with eigenvalue q. Show that $|q\rangle$ is also an eigenvector of the operators $\hat{Q}^2$, $\hat{Q}^n$ and $e^{\hat{Q}}$ and determine the corresponding eigenvalues.

2.22.9 A Symmetric Matrix

Let A be a 4×4 symmetric matrix. Assume that the eigenvalues are given by 0, 1, 2, and 3 with the corresponding normalized eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad , \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} \quad , \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

Find the matrix A.

2.22.10 Determinants and Traces

Let A be an $n \times n$ matrix. Show that

$$\det(\exp(A)) = \exp(Tr(A))$$

2.22.11 Function of a Matrix

Let

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

Calculate $\exp(\alpha A)$, α real.

2.22.12 More Gram-Schmidt

Let A be the symmetric matrix

$$A = \begin{pmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{pmatrix}$$

Determine the eigenvalues and eigenvectors of A. Are the eigenvectors orthogonal to each other? If not, find an orthogonal set using the Gram-Schmidt process.

2.22.13 Infinite Dimensions

Let A be a square finite-dimensional matrix (real elements) such that $AA^T = I$.

(a) Show that $A^T A = I$.

(b) Does this result hold for infinite dimensional matrices?

2.22.14 Spectral Decomposition

Find the eigenvalues and eigenvectors of the matrix

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Construct the corresponding projection operators, and verify that the matrix can be written in terms of its eigenvalues and eigenvectors. This is the spectral decomposition for this matrix.

2.22.15 Measurement Results

Given particles in state

$$|\alpha\rangle = \frac{1}{\sqrt{83}}\left(-3|1\rangle + 5|2\rangle + 7|3\rangle\right)$$

where $\{|1\rangle, |2\rangle, |3\rangle\}$ form an orthonormal basis, what are the possible experimental results for a measurement of

$$\hat{Y} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{pmatrix}$$

(written in this basis) and with what probabilities do they occur?

2.22.16 Expectation Values

Let

$$R = \begin{bmatrix} 6 & -2 \\ -2 & 9 \end{bmatrix}$$

represent an observable, and

$$|\Psi\rangle = \begin{bmatrix} a \\ b \end{bmatrix}$$

be an arbitrary state vector(with $|a|^2 + |b|^2 = 1$). Calculate $\langle R^2 \rangle$ in two ways:

(a) Evaluate $\langle R^2 \rangle = \langle \Psi | R^2 | \Psi \rangle$ directly.

(b) Find the eigenvalues(r_1 and r_2) and eigenvectors($|r_1\rangle$ and $|r_2\rangle$) of R^2 or R. Expand the state vector as a linear combination of the eigenvectors and evaluate

$$\langle R^2 \rangle = r_1^2 |c_1|^2 + r_2^2 |c_2|^2$$

2.22.17 Eigenket Properties

Consider a 3–dimensional ket space. If a certain set of orthonormal kets, say $|1\rangle$, $|2\rangle$ and $|3\rangle$ are used as the basis kets, the operators $\hat{A}$ and $\hat{B}$ are represented by

$$\hat{A} \rightarrow \begin{pmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{pmatrix} \quad , \quad \hat{B} \rightarrow \begin{pmatrix} b & 0 & 0 \\ 0 & 0 & -ib \\ 0 & ib & 0 \end{pmatrix}$$

where a and b are both real numbers.

(a) Obviously, $\hat{A}$ has a degenerate spectrum. Does $\hat{B}$ also have a degenerate spectrum?

(b) Show that $\hat{A}$ and $\hat{B}$ commute.

(c) Find a new set of orthonormal kets which are simultaneous eigenkets of both $\hat{A}$ and $\hat{B}$.

2.22.18 The World of Hard/Soft Particles

Let us define a state using a *hardness* basis $\{|h\rangle, |s\rangle\}$, where

$$\hat{O}_{\text{HARDNESS}} |h\rangle = |h\rangle \quad , \quad \hat{O}_{\text{HARDNESS}} |s\rangle = -|s\rangle$$

and the *hardness* operator $\hat{O}_{\text{HARDNESS}}$ is represented by (in this basis) by

$$\hat{O}_{HARDNESS} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Suppose that we are in the state

$$|A\rangle = \cos\theta |h\rangle + e^{i\varphi} \sin\theta |s\rangle$$

(a) Is this state normalized? Show your work. If not, normalize it.

(b) Find the state $|B\rangle$ that is orthogonal to $|A\rangle$. Make sure $|B\rangle$ is normalized.

(c) Express $|h\rangle$ and $|s\rangle$ in the $\{|A\rangle, |B\rangle\}$ basis.

(d) What are the possible outcomes of a hardness measurement on state $|A\rangle$ and with what probability will each occur?

(e) Express the hardness operator in the $\{|A\rangle, |B\rangle\}$ basis.

2.22.19 Things in Hilbert Space

For all parts of this problem, let $\mathcal{H}$ be a Hilbert space spanned by the basis kets $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$, and let a and b be arbitrary complex constants.

(a) Which of the following are Hermitian operators on $\mathcal{H}$?

1. $|0\rangle\langle 1| + i|1\rangle\langle 0|$
2. $|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 3| + |3\rangle\langle 2|$
3. $(a|0\rangle + |1\rangle)^+(a|0\rangle + |1\rangle)$
4. $((a|0\rangle + b^*|1\rangle)^+(b|0\rangle - a^*|1\rangle))|2\rangle\langle 1| + |3\rangle\langle 3|$
5. $|0\rangle\langle 0| + i|1\rangle\langle 0| - i|0\rangle\langle 1| + |1\rangle\langle 1|$

(b) Find the spectral decomposition of the following operator on $\mathcal{H}$:

$$\hat{K} = |0\rangle\langle 0| + 2|1\rangle\langle 2| + 2|2\rangle\langle 1| - |3\rangle\langle 3|$$

(c) Let $|\Psi\rangle$ be a normalized ket in $\mathcal{H}$, and let $\hat{I}$ denote the identity operator on $\mathcal{H}$. Is the operator

$$\hat{B} = \frac{1}{\sqrt{2}}(\hat{I} + |\Psi\rangle\langle\Psi|)$$

a projection operator?

(d) Find the spectral decomposition of the operator $\hat{B}$ from part (c).

2.22.20 A 2-Dimensional Hilbert Space

Consider a 2-dimensional Hilbert space spanned by an orthonormal basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. This corresponds to spin up/down for spin= 1/2 as we will see later in Chapter 9. Let us define the operators

$$\hat{S}_x = \frac{\hbar}{2}(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|) \quad , \quad \hat{S}_y = \frac{\hbar}{2i}(|\uparrow\rangle\langle\downarrow| - |\downarrow\rangle\langle\uparrow|) \quad , \quad \hat{S}_z = \frac{\hbar}{2}(|\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|)$$

(a) Show that each of these operators is Hermitian.

(b) Find the matrix representations of these operators in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis.

(c) Show that $\left[\hat{S}_x, \hat{S}_y\right] = i\hbar\hat{S}_z$, and cyclic permutations. Do this two ways: Using the Dirac notation definitions above and the matrix representations found in (b).

Now let

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle \pm |\downarrow\rangle)$$

(d) Show that these vectors form a new orthonormal basis.

(e) Find the matrix representations of these operators in the $\{|+\rangle, |-\rangle\}$ basis.

(f) The matrices found in (b) and (e) are related through a *similarity transformation* given by a unitary matrix, U, such that

$$\hat{S}_x^{(\uparrow\downarrow)} = U^\dagger \hat{S}_x^{(\pm)} U \quad , \quad \hat{S}_y^{(\uparrow\downarrow)} = U^\dagger \hat{S}_y^{(\pm)} U \quad , \quad \hat{S}_z^{(\uparrow\downarrow)} = U^\dagger \hat{S}_z^{(\pm)} U$$

where the superscript denotes the basis in which the operator is represented. Find U and show that it is unitary.

Now let

$$\hat{S}_{\pm} = \frac{1}{2}\left(\hat{S}_x \pm i\hat{S}_y\right)$$

(g) Express $\hat{S}_{\pm}$ as outer products in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis and show that $\hat{S}_+^{\dagger} = \hat{S}_-$.

(h) Show that

$$\hat{S}_+|\downarrow\rangle = |\uparrow\rangle \ , \ \hat{S}_-|\uparrow\rangle = |\downarrow\rangle \ , \ \hat{S}_-|\downarrow\rangle = 0 \ , \ \hat{S}_+|\uparrow\rangle = 0$$

and find

$$\langle\uparrow|\hat{S}_+ \ , \ \langle\downarrow|\hat{S}_+ \ , \ \langle\uparrow|\hat{S}_- \ , \ \langle\downarrow|\hat{S}_-$$

2.22.21 Find the Eigenvalues

The three matrices M_x, M_y, M_z, each with 256 rows and columns, obey the commutation rules

$$[M_i, M_j] = i\hbar\varepsilon_{ijk}M_k$$

The eigenvalues of M_z are $\pm 2\hbar$ (each once), $\pm 2\hbar$ (each once), $\pm 3\hbar/2$ (each 8 times), $\pm\hbar$ (each 28 times), $\pm\hbar/2$ (each 56 times), and 0 (70 times). State the 256 eigenvalues of the matrix $M^2 = M_x^2 + M_y^2 + M_z^2$.

2.22.22 Operator Properties

(a) If O is a quantum-mechanical operator, what is the definition of the corresponding Hermitian conjugate operator, O^+?

(b) Define what is meant by a Hermitian operator in quantum mechanics.

(c) Show that d/dx is not a Hermitian operator. What is its Hermitian conjugate, $(d/dx)^+$?

(d) Prove that for any two operators A and B, $(AB)^+ = B^+A^+$,

2.22.23 Ehrenfest's Relations

Show that the following relation applies for any operator O that lacks an explicit dependence on time:

$$\frac{\partial}{\partial t}\langle O\rangle = \frac{i}{\hbar}\langle[H, O]\rangle$$

HINT: Remember that the Hamiltonian, H, is a Hermitian operator, and that H appears in the time-dependent Schrodinger equation.

Use this result to derive Ehrenfest's relations, which show that classical mechanics still applies to expectation values:

$$m\frac{\partial}{\partial t}\langle\vec{x}\rangle = \langle\vec{p}\rangle \quad , \quad \frac{\partial}{\partial t}\langle\vec{p}\rangle = -\langle\nabla V\rangle$$

2.22.24 Solution of Coupled Linear ODEs

Consider the set of coupled linear differential equations $\dot{x} = Ax$ where $x = (x_1, x_2, x_3) \in R^3$ and

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

(a) Find the general solution $x(t)$ in terms of $x(0)$ by matrix exponentiation.

(b) Using the results from part (a), write the general solution $x(t)$ by expanding $x(0)$ in eigenvectors of A. That is, write

$$x(t) = e^{\lambda_1} c_1 v_1 + e^{\lambda_2} c_2 v_2 + e^{\lambda_3} c_3 v_3$$

where (λ_i, v_i) are the eigenvalue-eigenvector pairs for A and the c_i are coefficients written in terms of the $x(0)$.

2.22.25 Spectral Decomposition Practice

Find the spectral decomposition of the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix}$$

2.22.26 More on Projection Operators

The basic definition of a projection operator is that it must satisfy $P^2 = P$. If P furthermore satisfies $P = P^+$ we say that P is an orthogonal projector. As we derived in the text, the eigenvalues of an orthogonal projector are all equal to either zero or one.

(a) Show that if P is a projection operator, then so is $I - P$.

(b) Show that for any orthogonal projector P and an normalized state, $0 \leq \langle P \rangle \leq 1$.

(c) Show that the singular values of an orthogonal projector are also equal to zero or one. The singular values of an arbitrary matrix A are given by the square-roots of the eigenvalues of A^+A. It follows that for every singular value σ_i of a matrix A there exist some unit normalized vector u_i such that

$$u_i^+ A^+ A u_i = \sigma_i^2$$

Conclude that the action of an orthogonal projection operator never lengthens a vector (never increases its norm).

For the next two parts we consider the example of a non-orthogonal projection operator

$$N = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix}$$

(d) Find the eigenvalues and eigenvectors of N. Does the usual spectral decomposition work as a representation of N?

(e) Find the singular values of N. Can you interpret this in terms of the action of N on vectors in R^2?

Chapter 3

Probability

3.1 Probability Concepts

Quantum mechanics will necessarily involve probability in order for us to make the connection with experimental measurements.

We will be interested in understanding the quantity

$$P(A|B) = \text{probability of event } A \text{ given that event } B \text{ is true}$$

In essence, event B sets up the conditions or an environment and then we ask about the (conditional) probability of event A given that those conditions exist. All probabilities are conditional in this sense. The | symbol means *given* so that items to the right of this *conditioning* symbol are taken as being true.

In other words, we set up an experimental apparatus, which is expressed by properties B and do a measurement with that apparatus, which is expressed by properties A. We generate numbers (measurements) which we use to give a value to the quantity $P(A|B)$.

3.1.1 Standard Thinking .....

We start with the standard mathematical formalism based on *axioms*. We define these events:

1. A = occurrence of A
   (denotes that proposition A is true)
2. $\sim A$ = NOT A = nonoccurrence of A
   (denotes that proposition A is false)
3. $A \cap B$ = A AND B = occurrence of both A and B
   (denotes that proposition A AND B is true)

4. $A \cup B = A$ OR B =occurrence of at least A and B
   (denotes that proposition A OR B is true)

and standard Boolean logic as shown below:

> Boolean logic uses the basic statements AND, OR, and NOT. Using these and a series of Boolean expressions, the final output would be one TRUE or FALSE statement.

This is illustrated below:

1. If A is true AND B is true, then (A AND B) is true
2. If A is true AND B is false, then (A AND B) is false
3. If A is true OR B is false, then (A OR B) is true
4. If A is false OR B is false, then (A OR B) is false

or written as a *truth table*:

A	B	$(A \cap B)$	$(A \cup B)$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	0	0

Table 3.1: Boolean Logic

where 1 = TRUE and 0 = FALSE.

Then we set up a theory of probability with these axioms:

1. $P(A|A) = 1$
   This is the probability of the occurrence A given the occurrence of A. This represents a certainty and, thus, the probability must = 1. This is clearly an obvious assumption that we must make if our probability ideas are to make any sense at all.

   In other words, if I set the experimental apparatus such that the meter reads A, then it reads A with probability = 1.

2. $0 \leq P(A|B) \leq P(B|B) = 1$
   This just expresses the sensible idea that no probability is greater than the probability of a certainty and it make no sense to have the probability be less than 0.

3. $P(A|B) + P(\sim A|B) = 1$ or $P(\sim A|B) = 1 - P(A|B)$
   This just expresses the fact that the probability of something (anything) happening (A or $\sim A$) given B is a certainty ($= 1$), that is, since the set A or $\sim A$ includes everything that can happen, the total probability that one or the other occurs must be the probability of a certainty and be equal to one.

4. $P(A \cap B|C) = P(A|C)P(B|A \cap C)$
   This says that the probability that 2 events A, B both occur given that C occurs equals the probability of A given C multiplied by the probability of B given $(A \cap C)$, which makes sense which makes sense if you think of them happening in sequence.

All other probability relationships can be derived from these axioms.

The nonoccurrence of A given that A occurs must have probability = 0. This is expressed by

$$P(\sim A|A) = 0 \tag{3.1}$$

This result clearly follows from the axioms since

$$P(A|B) + P(\sim A|B) = 1 \tag{3.2}$$

$$P(A|A) + P(\sim A|A) = 1 \tag{3.3}$$

$$P(\sim A|A) = 1 - P(A|A) = 1 - 1 = 0 \tag{3.4}$$

Example: Let us evaluate $P(X \cap Y|C) + P(X \cap \sim Y|C)$.

We use axiom (4) in the 1^{st} term with and in the 2^{nd} term with to get

$$\begin{aligned} P(X \cap Y|C) + P(X \cap \sim Y|C) & \\ &= P(X|C)P(Y|X \cap C) + P(X|C)P(\sim Y|X \cap C) \\ &= P(X|C)[P(Y|X \cap C) + P(\sim Y|X \cap C)] \\ &= P(X|C)[1] = P(X|C) \end{aligned} \tag{3.5}$$

where we have used axiom (3). Thus we have the result

$$P(X \cap Y|C) + P(X \cap \sim Y|C) = P(X|C) \tag{3.6}$$

Now let us use this result with $X = \sim A, Y = \sim B$. This gives

$$P(\sim A \cap \sim B|C) + P(\sim A \cap \sim\sim B|C) = P(\sim A|C) \tag{3.7}$$

$$P(\sim A \cap \sim B|C) + P(\sim A \cap B|C) = 1 - P(A|C) \tag{3.8}$$

$$P(\sim A \cap \sim B|C) = 1 - P(A|C) - P(\sim A \cap B|C) \tag{3.9}$$

Then use the result again with $X = B, Y = \sim A$. This gives

$$P(B \cap \sim A|C) + P(B \cap A|C) = P(B|C) \tag{3.10}$$

or

$$P(\sim A \cap B|C) = P(B|C) - P(A \cap B|C) \tag{3.11}$$

which gives

$$P(\sim A \cap \sim B|C) = 1 - P(A|C) - P(B|C) + P(A \cap B|C) \tag{3.12}$$

Now

$$P(A \cup B|C) = 1 - P(\sim (A \cup B)|C) = 1 - P((\sim A \cap \sim B)|C) \tag{3.13}$$

since

$$\sim (A \cup B) = (\sim A \cap \sim B) \tag{3.14}$$

i.e., we can construct a *truth table* as shown below, which illustrates the equality directly

A	B	$(\sim (A \cup B))$	$(\sim A \cap \sim B)$
1	1	0	0
1	0	0	0
0	1	0	1
0	0	1	1

Table 3.2: Equivalent Expressions

We finally get

$$P(A \cup B) = P(A|C) + P(B|C) - P(A \cap B|C) \tag{3.15}$$

which is a very important and useful result.

If we have $P(A \cap B|C)=0$, then events A and B are said to be *mutually exclusive* given that C is true and the relation then reduces to

$$P(A \cup B) = P(A|C) + P(B|C) \tag{3.16}$$

This is the rule of *addition of probabilities for exclusive events.*

Some other important results are:

$$If\ A \cap B = B \cap A, |; then\ P(A|C)P(B|A \cap C) = P(B|C)P(A|B \cap C) \tag{3.17}$$

$$If\ P(A|C) \neq 0,\ then\ P(B|A \cap C) = P(A|B \cap C)\frac{P(B|C)}{P(A|C)} \tag{3.18}$$

which is *Baye's theorem.* It relates the probability of B given A to the probability of A given B.

When we say that B is *independent* of A, we will mean

$$P(B|A \cap C) = P(B|C) \tag{3.19}$$

or the occurrence of A has *NO influence* on the probability of B given C. Using axiom (4) we then have the result:

If A and B are independent given C, then

$$P(A \cap B|C) = P(A|C)P(B|C) \tag{3.20}$$

This is called *statistical* or *stochastic* independence. The result generalizes to a set of events $\{A_i,\ i = 1, 2, \ldots, n\}$. All these events are independent if and only if

$$P(A_1 \cap A_2 \cap \cdots \cap A_m|C) = P(A_1|C)P(A_2|C)\ldots P(A_m|C) \tag{3.21}$$

for all $m \leq n$.

Now let us think about these ideas in another way that has fundamental importance in modern approaches to quantum theory. The fundamental result in this view will turn out to be the Bayes formula and its relationship to measurements.

3.1.2 Bayesian Thinking .....

Two Different Axioms

1. If we specify how much we believe something is true, then we must have implicitly specified how much we believe it is false.
2. If we first specify how much we believe that proposition Y is true, and then state how much we believe X is true given that Y is true, then we must implicitly have specified how much we believe that both X and Y are true.

We assign *real* numbers to each proposition in a manner so that the larger the numerical value associated with a proposition, the more we believe it.

Only using the rules of *Boolean logic*, ordinary algebra, and the constraint that if there are several different ways of using the same information, then we should always arrive at the same conclusions independent of the particular analysis-path chosen, it is then found that this consistency could only be guaranteed if the real numbers we had attached to our beliefs in the various propositions could be *mapped* (or transformed) to another set of *real* positive numbers which obeyed the usual rules of probability theory:

$$prob(X|I) + prob(\sim X|I) = 1 \quad \text{(same as axiom (3))} \tag{3.22}$$

$$prob(X \cap Y|I) = prob(X|Y \cap I) \times prob(Y|I) \text{ (same as axiom (4))} \tag{3.23}$$

The first of these equations is called the sum rule and states (as earlier) that the probability that X is true plus the probability that X is false is equal to

one.

The second of these equations is called the product rule. It states (as earlier) that the probability that both X and Y are true is equal to the probability that X is true given that Y is true times the probability that Y is true (independent of X).

Note all the probabilities are conditional on proposition(s) or conditioning(s) I, which denotes the relevant background information on hand. It is important to understand that there is no such thing as an absolute probability (one without prior information).

Baye's Theorem and Marginalization

As before, we can use the sum and product rules to derive other results.

First, starting with the product rule we have

$$prob(X \cap Y|I) = prob(X|Y \cap I) \times prob(Y|I) \tag{3.24}$$

We can rewrite this equation with X and Y interchanged

$$prob(Y \cap X|I) = prob(Y|X \cap I) \times prob(X|I) \tag{3.25}$$

Since the probability that both X and Y are true must be logically the same as the probability that both Y and X are true we must also have

$$prob(Y \cap X|I) = prob(X \cap Y|I) \tag{3.26}$$

or

$$prob(X|Y \cap I) \times prob(Y|I) = prob(Y|X \cap I) \times prob(X|I) \tag{3.27}$$

or

$$prob(X|Y \cap I) = \frac{prob(Y|X \cap I) \times prob(X|I)}{prob(Y|I)} \tag{3.28}$$

which is *Bayes theorem* (as derived earlier).

Most standard treatments of probability do not attach much importance to Bayes' rule.

This rule, which relates $prob(A|B \cap C)$ to $prob(B|A \cap C)$, allows us to turn things around with respect to the conditioning symbol, which leads to a reorientation of our thinking about probability.

The fundamental importance of this property to data analysis becomes apparent if we replace A and B by *hypothesis* and *data*:

$$prob(A|B \cap C) \propto prob(B|A \cap C) \times prob(A|C) \tag{3.29}$$

$$prob(hypothesis|data \cap C) \propto prob(data|hypothesis \cap C) \times prob(hypothesis|C)$$

Note that the equality has been replaced with a proportionality because the term $prob(data|I) = evidence$ has been omitted. The proportionality constant can be found from the normalization requirement that the sum of the probabilities for something happening must equal 1.

The power of Bayes' theorem lies in the fact that it relates the quantity of interest, the probability that the hypothesis is true given the data, to the term that we have a better chance of being able to assign, the probability that we would have obtained the measured data if the hypothesis was true.

The various terms in Bayes' theorem have formal names.

The term $prob(hypothesis|C) = prior$ probability represents our state of knowledge(or ignorance) about the truth of the hypothesis before we have analyzed the current data. This is modified by the experimental measurements through the term $prob(data|hypothesis \cap C) = likelihood$ function. This product gives $prob(hypothesis|data \cap C) = posterior$ probability representing our state of knowledge about the truth of the hypothesis in the light of the data(after measurements).

In some sense, Bayes' theorem encapsulates the process of learning, as we shall see later.

Second, consider the following results from the product rule

$$prob(X \cap Y|I) = prob(Y \cap X|I) = prob(Y|X \cap I) \times prob(X|I) \tag{3.30}$$

$$prob(X \cap \sim Y|I) = prob(\sim Y \cap X|I) = prob(\sim Y|X \cap I) \times prob(X|I) \tag{3.31}$$

Adding these equations we get

$$\begin{aligned} prob(X \cap Y|I) + prob(X \cap \sim Y|I) & \\ = (prob(Y|X \cap I) + prob(\sim Y|X \cap I))\, prob(X|I) & \end{aligned} \tag{3.32}$$

Since $prob(Y|X \cap I) + prob(\sim Y|X \cap I) = 1$ we have

$$prob(X \cap Y|I) + prob(X \cap \sim Y|I) = prob(X|I) \tag{3.33}$$

which, again, is the same result as earlier. If, on the other hand, $Y \rightarrow \{Y_k, k = 1, 2, \ldots, M\}$ representing a set of M alternative possibilities, then we generalize the two-state result above as

$$\sum_{k=1}^{M} prob(X \cap Y_k|I) = prob(X|I) \tag{3.34}$$

We can derive this result as follows

$$prob(X \cap Y_1|I) = prob(Y_1 \cap X|I) = prob(Y_1|X \cap I) \times prob(X|I) \tag{3.35}$$

$$prob(X \cap Y_2|I) = prob(Y_2 \cap X|I) = prob(Y_2|X \cap I) \times prob(X|I) \tag{3.36}$$

$$\ldots\ldots\ldots\ldots \tag{3.37}$$

$$prob(X \cap Y_M|I) = prob(Y_M \cap X|I) = prob(Y_M|X \cap I) \times prob(X|I) \tag{3.38}$$

Adding these equations we get

$$\sum_{k=1}^{M} prob(X \cap Y_k|I) = prob(X|I)\left(\sum_{k=1}^{M} prob(Y_k \cap X|I)\right) \tag{3.39}$$

If we assume that the $\{Y_k\}$ form a mutually exclusive and exhaustive set of possibilities, that is, if one of the $Y_k's$ is true, then all the others must be false, we then get

$$\sum_{k=1}^{M} prob(Y_k \cap X|I) = \hat{I} \tag{3.40}$$

which is a normalization condition. This completes the derivation.

If we go to the continuum limit where we consider an arbitrarily large number of propositions about some result (the range in which a given result might lie), then as long as we choose the intervals in a contiguous fashion, and cover a big enough range of values, we will have a mutually exclusive and exhaustive set of possibilities. In the limit of $M \to \infty$, we obtain

$$prob(X|I) = \int_{-\infty}^{\infty} prob(x \cap Y|I)\, dY \tag{3.41}$$

which is the *marginalization* equation. The integrand here is technically a *probability density function(pdf)* rather than a probability. It is defined by

$$pdf(X \cap Y = y|I) = \lim_{\delta y \to 0} \frac{prob(X \cap y \le Y \le y + \delta y|I)}{\delta y} \tag{3.42}$$

and the probability that the value of Y lies in a finite range between y_1 and y_2 (and X is also true) is given by

$$prob(X \cap y_1 \le Y \le y_2|I) = \int_{y_1}^{y_2} pdf(X \cap Y|I)\, dY \tag{3.43}$$

which leads directly to the marginalization equation.

In this continuum limit the normalization condition takes the form

$$1 = \int_{-\infty}^{\infty} pdf(Y|X \cap I)\, dY \tag{3.44}$$

Marginalization is a very powerful device in data analysis because it enables us to deal with *nuisance parameters*, that is, quantities which necessarily enter the analysis but are of no intrinsic interest. The unwanted background signal present in many experimental measurements, and instrumental parameters which are difficult to calibrate, are examples of nuisance parameters.

3.2 Probability Interpretation

In the standard way of thinking about probability in relation to experiments, measured results are related to probabilities using the concept of a *limit frequency*. The limit frequency is linked to probability by this definition:

> If C can lead to either A or $\sim A$ and if in n repetitions, A occurs m times, then

$$P(A|C) = \lim_{n\to\infty} \frac{m}{n} \tag{3.45}$$

We must now connect the mathematical formalism with this limit frequency concept so that we can use the formalism to make predictions for experiments in real physical systems.

This approach depends on whether we can prove that the limit makes sense for real physical systems. Let us see how we can understand the real meaning of the above interpretation of probability and thus learn how to use it in quantum mechanics, where probability will be the dominant property.

Suppose that we have an experimental measurement, M, that can yield either A or $\sim A$ as results, with a probability for result A given by

$$P(A|M) = p \tag{3.46}$$

In general, we let any sequence of n independent measurements be labeled as event M^n and we define n_A as the number of times A occurs, where $0 \le n_A \le n$.

Now imagine we carry out a sequence of n independent measurements and we find that A occurs r times. The probability for a sequence of results that includes result A occurring r times and $\sim A$ occurring $(n-r)$ times (independent of their order in the sequence) is given by

$$p^r q^{n-r} \tag{3.47}$$

where

$$q = P(\sim A|M) = 1 - P(A|M) = 1 - p \tag{3.48}$$

The different sequence orderings are mutually exclusive events and thus we have

$$P(n_A = r|M^n) = \sum_{all\ possible\ orderings} p^r q^{n-r} \tag{3.49}$$

The sum

$$\sum_{all\ possible\ orderings} \tag{3.50}$$

just counts the number of ways to distribute r occurances of A and (n-r) occurances of $\sim A$, where all the terms contain the common factor $p^r q^{n-r}$. This

result is given by the Binomial probability distribution (more about this later) as

$$\frac{n!}{r!(n-r)!} \tag{3.51}$$

so that

$$P(n_A = r|M^n) = \frac{n!}{r!(n-r)!}p^r q^{n-r} \tag{3.52}$$

Now to get to the heart of the problem. The frequency of A in M^n is given by

$$f_n = \frac{n_A}{n} \tag{3.53}$$

This is *not necessarily* $= p$ in any set of measurements.

What is the relationship between them? Consider the following:

$$\begin{aligned} \langle n_A \rangle &= \text{ average or expectation value} \\ &= \text{ sum over [possible values times probability of that value} \\ &= \sum_{r=0}^{n} rP(n_A = r|M^n) = \sum_{r=0}^{n} r\frac{n!}{r!(n-r)!}p^r q^{n-r} \end{aligned} \tag{3.54}$$

We now use a clever mathematical trick to evaluate this sum. For the moment consider p and q to be two arbitrary independent variables. At the end of the calculation we will let $q = 1 - p$ as is appropriate for a real physical system.

From the Binomial expansion formula, we have, in general,

$$\sum_{r=0}^{n} \frac{n!}{r!(n-r)!}p^r q^{n-r} = (p+q)^n \tag{3.55}$$

We then have

$$p\frac{\partial}{\partial p}\sum_{r=0}^{n} \frac{n!}{r!(n-r)!}p^r q^{n-r} = p\frac{\partial}{\partial p}(p+q)^n \tag{3.56}$$

so that

$$\sum_{r=0}^{n} r\frac{n!}{r!(n-r)!}p^r q^{n-r} = np(p+q)^{n-1} \tag{3.57}$$

This gives

$$\sum_{r=0}^{n} rP(n_A = r|M^n) = np(p+q)^{n-1} \tag{3.58}$$

or

$$\langle n_A \rangle = np(p+q)^{n-1} \tag{3.59}$$

In a real physical system, we must have $p + q = 1$, so that we end up with the result

$$\langle n_A \rangle) = np \tag{3.60}$$

and

$$\langle f_n \rangle = \frac{\langle n_A \rangle}{n} = p \tag{3.61}$$

This says that p = the average frequency.

This does not say, however, that f_n is close to p.

Now consider a more general experiment where the outcome of a measurement is the value of some continuous variable Q, with probability density (for its continuous spectrum) given by

$$P(q < Q < q + dq|M) = h(q)dq \tag{3.62}$$

If we let $h(q)$ contain delta-functions, then this derivation is also valid for the discrete part of the spectrum. We can now derive the following useful result. If Q is a nonnegative variable, which means that $h(q) = 0$ for $q < 0$, then for any $\epsilon > 0$

$$\langle Q \rangle = \int_0^\infty h(q) q \, dq \geq \int_\epsilon^\infty h(q) q \, dq \geq \epsilon \int_\epsilon^\infty h(q) \, dq = \epsilon P(Q \geq \epsilon|M) \tag{3.63}$$

This implies that

$$P(Q \geq \epsilon|M) \leq \frac{\langle Q \rangle}{\epsilon} \tag{3.64}$$

Now we apply this result to the nonnegative variable $|Q - c|^\alpha$ where $\alpha > 0$ and c = number, to obtain

$$P(|Q - c| \geq \epsilon|M) = P(|Q - c|^\alpha \geq \epsilon^\alpha|M) \leq \frac{\langle |Q - c|^\alpha \rangle}{\epsilon^\alpha} \tag{3.65}$$

which is called *Chebyshev's inequality.*

In the special case where $\alpha = 2, c = \langle Q \rangle$=mean of distribution we have

$$\langle |Q - c|^2 \rangle = \langle |Q - \langle Q \rangle|^2 \rangle = -\langle Q^2 \rangle - \langle Q^2 \rangle^2 = \sigma^2 = \text{ variance} \tag{3.66}$$

so that letting $\epsilon = k\sigma$ we get

$$P(|Q - \langle Q \rangle| \geq k\sigma|M) \leq \frac{1}{k^2} \tag{3.67}$$

or, the probability of Q being k or more standard deviations from the mean is no greater than $1/k^2$ (*independent* of the form of the probability distribution).

In a similar manner, it can also be shown that

$$P(|f_n - p| \geq \delta|M) \leq \frac{1}{n\delta^2} \tag{3.68}$$

which implies that the probability of f_n (the relative frequency of A in n independent repetitions of M) being more than δ away from p converges to 0 as

$n \to \infty$. This is an example of the *law of large numbers* in action. This *DOES NOT* say $f_n = p$ at any time or that f_n remains close to p as $n \to \infty$.

It DOES say that the deviation of f_n from p becomes more and more improbable or that the probability of any deviation approaches 0 as $n \to \infty$.

It is in this sense that one uses the limit frequency from experiment to compare with theoretical probability predictions in physics. *From probability theory one derives only statements of probability, not of necessity.*

3.3 First hints of "subversive" or "Bayesian" thinking.....

How do we reason in situations where it is not possible to argue with certainty? In other words, is there a way to use the techniques of deductive logic to study the inference problem arising when using inductive logic? *No matter what scientists say, this is what they are actually doing most of the time.*

The answer to this last question resides in the Bayes' rule.

To Bayes(along with Bernoulli and Laplace), a probability represented a "degree-of-belief" or "plausibility", that is, how much one thinks that something is true, *based* on the *evidence on hand.*

The developers of standard probability theory(Fisher, Neyman and Pearson) thought this seemed too vague and subjective a set of ideas to be the basis of a "rigorous" mathematical theory. Therefore, they *defined* probability as the long-run relative frequency with which an event occurred, given *infinitely many repeated* experimental trials. Since such probabilities can be measured, probability was then thought to be an objective tool for dealing with random phenomena.

This frequency definition certainly *seems* to be more objective, but it turns out that its *range of validity* is far more limited.

In this Bayesian view, probability represents a *state of knowledge.* The conditional probabilities represent *logical* connections rather than *causal* ones.

Example:

Consider an urn that contains 5 red balls and 7 green balls.

If a ball is selected at *random*, then we would all agree that the probability of picking a red ball would be 5/12 and of picking a green ball would be 7/12.

If the ball is not returned to the urn, then it seems reasonable that the probability of picking a red or green ball must depend on the outcome of the first pick (because there will be one less red or green ball in the urn).

Now suppose that we are not told the outcome of the first pick, but are given the result of the second pick. Does the probability of the first pick being red or green change with the knowledge of the second pick?

Initially, many observers would probably say *no*, that is, at the time of the first draw, there were still 5 red balls and 7 green balls in the urn, so the probabilities for picking red and green should still be 5/12 and 7/12 independent of the outcome of the second pick. The error in this argument becomes clear if we consider the extreme example of an urn containing only 1 red and 1 green ball. Although, the second pick cannot affect the first pick in a physical sense, a knowledge of the second result does influence what we can infer about the outcome of the first pick, that is, if the second ball was green, then the first ball must have been red, and vice versa.

We can calculate the result as shown below:

Y = pick is GREEN (2^{nd} pick)

X = pick is RED (1^{st} pick)

I = initial number of RED/GREEN balls = (n, m)

A Bayesian would say:

$$prob(X|Y \cap I) = \frac{prob(Y|X \cap I) \times prob(X|I)}{prob(Y|I)} \tag{3.69}$$

$$\begin{aligned} prob(X|Y \cap \{n,m\}) &= \frac{prob(Y|X \cap \{n,m\}) \times \frac{n}{n+m}}{\frac{n}{n+m}\frac{m}{n+m-1} + \frac{m}{n+m}\frac{m-1}{n+m-1}} \\ &= \frac{\frac{m}{n+m-1} \times n}{\frac{nm}{n+m-1} + \frac{m(m-1)}{n+m-1}} = \frac{n}{n+m-1} \end{aligned} \tag{3.70}$$

$$n = m = 1 \Rightarrow prob(X|Y \cap \{1,1\}) = \frac{1}{1+1-1} = 1 \tag{3.71}$$

$$n = 5, m = 7 \Rightarrow prob(X|Y \cap \{5,7\}) = \frac{5}{5+7-1} = \frac{5}{11} = 0.456 \tag{3.72}$$

Non-Bayesian says:

$$prob(X|\{5,7\}) = \frac{5}{12} = 0.417 \tag{3.73}$$

Clearly, the Bayesian and Non-Bayesian disagree.

However, the non-Bayesian is just *assuming* that the calculated result 0.417 is correct, whereas, the Bayesian is using the rules of probability (Bayes' Rule) to infer the result 0.456 *correctly*.

The concerns about the subjectivity of the Bayesian view of probability are understandable. I think that the presumed shortcomings of the Bayesian approach merely reflect a confusion between subjectivity and the difficult technical question of how probabilities(especially prior probabilities) should be assigned.

The popular argument is that if a probability represents a degree-of-belief, then it must be subjective, because my belief could be different from yours. The Bayesian view is that a probability does indeed represent how much we believe that something is true, but that this belief should be based on all the relevant information available(all prior probabilities).

While this makes the assignment of probabilities an open-ended question, because the information available to me may not be the same as that available to you, it is not the same as subjectivity. It simply means that probabilities are *always conditional*, and this conditioning must be stated *explicitly*.

Objectivity demands only that two people having the same information should assign the same probability.

Cox looked at the question of plausible reasoning from the perspective of logical consistency. He found that the only rules that worked were those of probability theory! Although the sum and product rules of probability are straightforward to prove for frequencies (using Venn diagrams), Cox showed that their range of validity goes much further. Rather than being restricted to frequencies, he showed that probability theory constitutes the basic calculus for logical and consistent plausible reasoning, which means scientific inference!

Another Example - Is this a fair coin?

We consider a simple coin-tossing experiment. Suppose that I had found this coin and we observed 4 heads in 11 flips.

If by the word *fair* we mean that we would be prepared to make a $50:50$ bet on the outcome of a flip being a head or a tail, then do you think that it is a fair coin?

If we ascribe fairness to the coin, then we naturally ask how sure are we that this was so or if it was not fair, how unfair do we think it was?

A way of formulating this problem is to consider a large number of *contiguous* hypotheses about the range in which the *bias-weighting* of the coin might lie. If we denote bias-weighting by H, then $H = 0$ and $H = 1$ can represent a coin which produces a tail(not a head!) or a head on every flip, respectively. There

is a continuum of possibilities for the value of H between these limits, with $H = 1/2$ indicating a fair coin. The hypotheses might then be, for example

$$\begin{gathered}\text{(a) } 0.00 \leq H \leq 0.01 \\ \text{(b) } 0.01 \leq H \leq 0.03 \\ \text{(c) } 0.02 \leq H \leq 0.03 \\ \text{and so on}\end{gathered}$$

Our state of knowledge about the fairness, or the degree of unfairness, of the coin is then completely summarized by specifying how much we believe these various hypotheses to be true. If we assign a high probability to one (or a closely grouped few) of these hypotheses, compared to others, then this indicates that we are confident in our estimate of the bias-weighting. If there was no such distinction, then it would reflect a high level of ignorance about the nature of the coin.

In this case, our inference about the fairness of the data is summarized by the conditional pdf $prob(H|\{data\}H\cap I)$. This is just a representation of the limiting case of a continuum of hypotheses for the value of H, that is, the probability that H lies in an infinitesimally narrow range between h and $h + \delta h$ is given by $prob(H = h|\{data\} \cap I)\, dH$. To estimate this posterior pdf, we need to use Baye's theorem, which relates the pdf of interest to two others that are easier to assign:

$$prob(H|\{data\} \cap I) \propto prob(\{data\}|H \cap I) \times prob(H|I) \tag{3.74}$$

We have omitted the denominator $prob(\{data\}|I)$ since it does not involve bias-weighting explicitly and replaced the equality by a proportionality. The omitted constant can be determined by normalization

$$\int_0^1 prob(H|\{data\} \cap I)\, dH = 1 \tag{3.75}$$

The prior pdf, $prob(H|I)$, on the right side represents what we know about the coin given only that I found the coin. This means that we should keep an open mind about the nature of the coin. A simple probability assignment which reflects this is a uniform pdf

$$prob(H|I) = \begin{cases} 1 & 0 \leq H \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.76}$$

This prior state of knowledge (or ignorance) is modified by the data through the likelihood function, $prob(\{data\}|H \cap I)$, which is a measure of the chance that we would have obtained the data we actually observed if the value of the bias-weighting H was given (as known). If, in the conditioning information I, we assume that the flips of the coin were independent events, so that the outcome of one did not influence that of another, then the probability of obtaining the data R heads in N tosses is given by the binomial distribution

$$prob(\{data\}|H \cap I) \propto H^R(1 - H)^{N-R} \tag{3.77}$$

The product of these last two results then gives the posterior pdf that we require.

It represents our state of knowledge about the nature of the coin in light of the data.

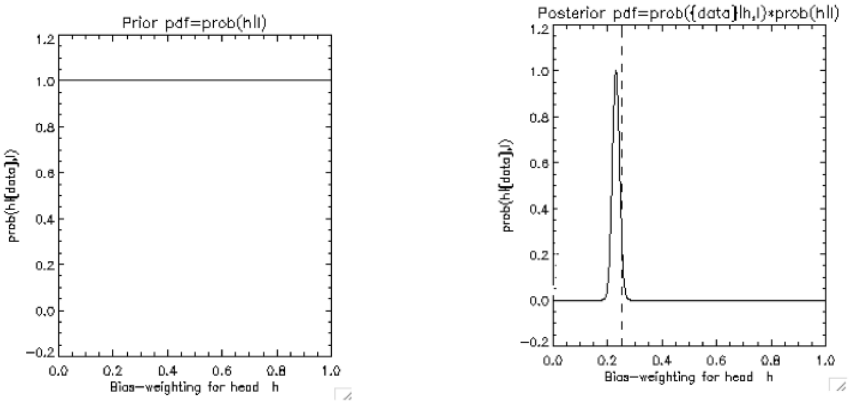
It is instructive to see how this pdf evolves as we obtain more and more data pertaining to the coin. A computer simulation is shown below allows us to demonstrate what happens in some typical cases.

The simulation allows for three distinct and very different prior probabilities:
(1) Uniform distribution
(2) Gaussian distribution centered around 0.5 with some spread
(3) Sum of two Gaussians with different centers

These prior probabilities represent very different initial knowledge:
(1) total ignorance-we have no idea if it is fair
(2) knowledge that mean is 0.5[with spread]-we think it is fair
(3) knowledge that it is unfair (either all tails or all heads) [with spreads]

In the simulation we can choose the true mean value (h_0), which is then reflected in the simulated coin tosses (the data).

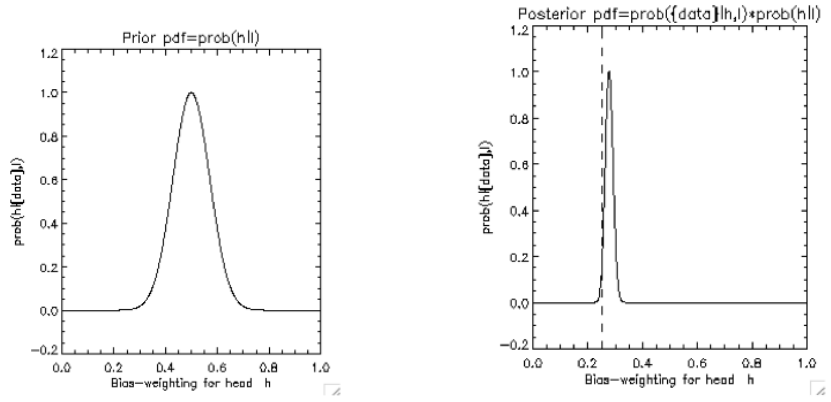
As can be seen from the images below, the only effect that different prior probabilities have is to change the period of time evolution to the final posterior pdf (which is the same eventually in all cases)!
(1) total ignorance - we have no idea if it is fair



(a) Prior (b) Posterior

Figure 3.1: Total Ignorance

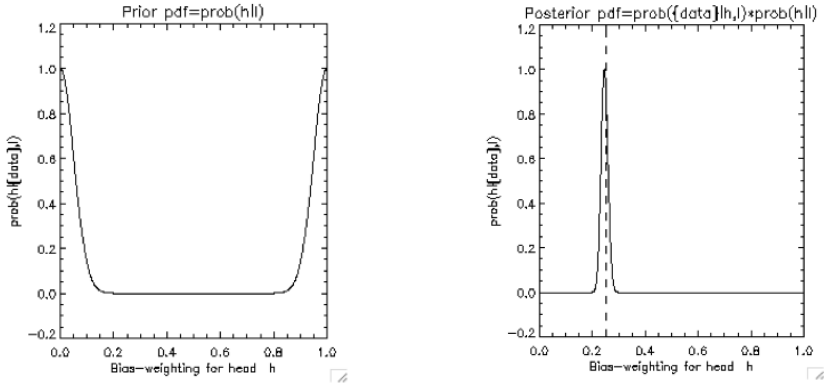
(2) knowledge that mean is 0.5 [with spread] - we think it is fair



(a) Prior (b) Posterior

Figure 3.2: Knowledge that Mean is 0.5

(3) knowledge that it is unfair (either all tails or all heads) [with spreads]



(a) Prior (b) Posterior

Figure 3.3: Knowledge that it is Unfair

In each case, the first figure shows the posterior pdf for H given no data (it is the same as the prior pdf) and the second figure shows the posterior pdf after 1000 tosses and they clearly indicate that no matter what our initial knowledge, the final posterior pdf will be the same, that is, the posterior pdf is dominated by the likelihood function(the actual data) and is independent of the prior pdf.

3.3.1 The Problem of Prior Probabilities

We are now faced with the most difficult question. How do we assign probabilities based on prior information?

The oldest idea was devised by Bernoulli - the *principle of insufficient reason* or the *principle of indifference*. It states that if we determine a set of basic, mutually exclusive, possibilities, and we have no reason to believe that any one of them is more likely to be true than another, then we must assign the same *probability* to each of them. Clearly, this makes sense. Think of flipping a coin with two possibilities, heads and tails. If it is a legitimate coin, then we have no reason to favor heads over tails and we must assign equal probability to each possibility, that is,

$$prob(heads|I) = prob(tails|I) = \frac{1}{2} \tag{3.78}$$

Let us elaborate on the idea of *not having any reason to believe.....* Suppose we had ten possibilities labeled by $X_i, i = 1, 2, \ldots, 10$ and we had no reason to think any was more likely than any other. We would then have

$$prob(X_1|I) = prob(X_2|I) = \ldots = prob(X_{10}|I) = \frac{1}{10} \tag{3.79}$$

Suppose that we relabel or reorder the possibilities. If the conditioning on I truly represents gross ignorance about any details of the situation, then such a reordering should not make any difference in the probability assignments. Any other statement has to mean that we have other important information besides the simple ordering of the possibilities. For example, imagine that you called a certain side of the coin heads and therefore the other side tails. Nothing changes if your friend switches the meaning of heads and tails. This justification of the Bernoulli principle led Jaynes to suggest that we think of it as a consequence of the *requirement of consistency*.

This principle of insufficient reason can only be applied to a limited set of problems involving games of chance. It leads, however, to some very familiar and very important results if combined with the product and sum rules of probability theory.

Example 1: Assume W white balls and R red balls in an urn. We now pick the balls out of the urn randomly. The principle of indifference says that we should assign a uniform prior probability (actually a pdf)

$$prob(j|I) = \frac{1}{R+W} \quad , \quad j = 1, 2, 3, \ldots, R+W \tag{3.80}$$

for the proposition that any particular ball, denoted by index j, will be picked. Using the marginalization idea from earlier

$$prob(X|I) = \int_{-\infty}^{\infty} prob(X \cap Y|I)\, dY \tag{3.81}$$

we have

$$prob(red|I) = \sum_{j=1}^{R+W} prob(red \cap j|I) \tag{3.82}$$
$$= \sum_{j=1}^{R+W} prob(j|I)prob(red|j \cap I) = \frac{1}{R+W} \sum_{j=1}^{R+W} prob(red|j \cap I)$$

where we have used the product rule. The term $prob(red|j \cap I)$ is one if the j^{th} ball is red and zero if it is white. Therefore the summation equals the number of red balls R and we get

$$prob(red|I) = \frac{1}{R+W} \sum_{j=1}^{R+W} prob(red|j \cap I) = \frac{R}{R+W} \tag{3.83}$$

as expected. However, we have derived this result from the principle of indifference and the product rule. It also follows from the basic notion of probability, that is,

$$prob(red|I) = \frac{\text{number of cases favorable to red}}{\text{total number of equally possible cases}} = \frac{R}{R+W} \tag{3.84}$$

We now assume that after each pick the ball is returned to the urn and we ask the question: what is the probability that N such picks (trials) will result in r red balls?

Using marginalization and the product rule we can write

$$prob(r|N \cap I) = \sum_k prob(r \cap S_k|N \cap I) = \sum_k prob(r|S_k \cap N \cap I)prob(S_k|N \cap I) \tag{3.85}$$

where the summation is over the 2^N possible sequences of red-white outcomes $\{S_k\}$ of N picks. The term $prob(r|S_k \cap N \cap I)$ equals one if S_k contains exactly r red balls and is zero otherwise, so that we need only consider those sequences which have exactly r red outcomes for $prob(S_k|N \cap I)$.

Now we have

$$prob(S_k|N \cap I) = [prob(red|I)]^r [prob(white|I)]^{N-r} = \frac{R^r W^{N-r}}{(R+W)^N} \tag{3.86}$$

Hence,

$$prob(r|N \cap I) = \frac{R^r W^{N-r}}{(R+W)^N} \sum_k prob(r|S_k \cap N \cap I) \tag{3.87}$$

for those S_k that matter, i.e., we are only considering those S_k which contain exactly r red balls. In this case we have

$$prob(r|N \cap I) = \frac{R^r W^{N-r}}{(R+W)^N} \frac{N!}{r!(N-r)!} \tag{3.88}$$

where the last factor just corresponds to the number of sequences (permutations) containing r red balls. Thus,

$$prob(r|N \cap I) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \tag{3.89}$$

where

$$p = \frac{R}{R+W} = \text{probability of picking a red ball} \tag{3.90}$$

and

$$q = 1 - p = \frac{W}{R+W} = \text{probability of picking a white ball} \tag{3.91}$$

Note that $p+q = 1$ as it should since red and white balls are the only possibilities.

We can then compute the frequency r/N with which we expect to observe red balls. We have

$$\begin{aligned} \langle \frac{r}{N} \rangle &= \sum_{r=0}^{N} \frac{r}{N} prob(r|N \cap I) = \sum_{r=0}^{N} \frac{r}{N} \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\ &= \sum_{r=1}^{N} \frac{(N-1)!}{(r-1)!(N-r)!} p^r (1-p)^{N-r} = p \sum_{j=0}^{N-1} \frac{(N-1)!}{j!(N-r)!} p^j (1-p)^{N-1-j} \\ &= p(p+q)^{N-1} = p = \frac{R}{R+W} \end{aligned} \tag{3.92}$$

as the *expected* or *anticipated* result. Thus, the expected frequency of red balls, in repetitions of the urn experiment, is equal to the probability of picking one red ball in a single trial.

A similar calculation for the mean-square deviation gives the result

$$\langle \left(\frac{r}{N} - \langle \frac{r}{N} \rangle \right)^2 \rangle = \langle \left(\frac{r}{N} - p \right)^2 \rangle = \frac{pq}{N} \tag{3.93}$$

Since this becomes zero in the limit of large N, it agrees with the result we derived earlier. It also verifies that Bernoulli's famous theorem or law of large numbers is valid:

$$\lim_{N \to \infty} \left(\frac{r}{N} \right) = prob(red|I) \tag{3.94}$$

This relationship, which allows prediction of the long-run frequency of occurrence from the probability assignment, goes in a direction opposite to the one we want, that is, we would like to be able to determine the probability of obtaining a red ball, in a single pick, given a finite number of observed outcomes. This is, in fact, exactly what Bayes theorem allows us to do!

How do we generalize Bernoulli's principle of insufficient reason to the case of continuous parameters, that is, when the quantity of interest is not restricted to certain discrete values (heads/tails)?

Suppose we have a variable X which represents the position of some object. We then define a probability as follows. Given the information I, the probability that X lies in the infinitesimal range between x and $x + dx$ is

$$prob(X = x|I) = \lim_{\delta x \to 0} prob(x \leq X \leq x + \delta x|I) \tag{3.95}$$

so that we are treating continuous pdfs as the limiting case of discrete ones. Although it is still awkward to enumerate the possibilities in this case, we can still make use of the principle of consistency which underlies the principle of indifference.

Examples:

A Location Parameter

Suppose that we are unsure about the actual location of the origin. Should this make any difference to the pdf assigned for X? Since I represents gross ignorance about any details of the situation other than the knowledge that X pertains to a location, the answer must be no; otherwise we must already have information regarding the position of the object. Consistency then demands that the pdf for X should not change with the location of the origin or any offset in the position values. Mathematically, we say

$$prob(X|I)\, dX = prob(X + x_0|I)\, d(X + x_0) \tag{3.96}$$

Since x_0 is a constant, $d(X + x_0) = dX$ so that we have

$$prob(X|I) = prob(X + x_0|I) = \text{constant} \tag{3.97}$$

so that the complete ignorance about a location parameter is represented by the assignment of a *uniform* pdf.

A Scale Parameter

Suppose that we have another parameter that tells us about size or magnitude, a so-called *scale* parameter. If we are interested in the size L of some object and we have no idea about the length scale involved, then the pdf should be invariant with respect to shrinking or stretching the length scale. Mathematically, the requirement of consistency can be written

$$prob(L|I)\, dL = prob(\beta L|I)\, d(\beta L) \tag{3.98}$$

where β is a positive constant. Then since $d(\beta L) = \beta\, dL$ we must have

$$prob(L|I) = \beta prob(\beta L|I) \tag{3.99}$$

which can only be satisfied if

$$prob(L|I) \propto \frac{1}{L} \tag{3.100}$$

which is called *Jeffrey's prior*. It represents complete ignorance about the value of a scale parameter.

Now we must have

$$prob(L|I)\,dL = prob(f(L)|I)\,df(L) \tag{3.101}$$

since we are looking at the same domain of values in each case. We then have

$$prob(\log L||)\,d(\log L) = prob(L|I)\,dL \tag{3.102}$$

$$prob(\log L||)\,\frac{dL}{L} = prob(L|I)\,dL \tag{3.103}$$

$$prob(\log L||) = Lprob(L|I) = \text{constant} \tag{3.104}$$

So that assignment of a *uniform* pdf for $\log L$ is the way to represent complete ignorance about a scale parameter.

3.4 Testable Information: The Principle of Maximum Entropy

Clearly, some pdfs can be assigned given only the nature of the quantities involved (as we saw above). The methods employed hinge on the use of consistency arguments along with *transformation groups*, which characterize the ignorance for a given situation.

For a set of discrete probabilities(finite) the associated pdf must be invariant with respect to any permutation of the propositions (permutation group). In the continuous parameter case, the associated transformations are translation (origin shift) and dilation (shrink/stretch), which are also group transformations.

Let us move on to a situation where we do not have total ignorance.

Suppose that a die, with the usual six faces, was rolled a very large number of times and we are only told that the average result was 4.5. What probability should we assign for the various outcomes $\{X_i\}$ that the face on top had i dots?

The information or condition I provided by the experiment is written as a simple *constraint* equation

$$\sum_{i=1}^{n} iprob(X_i|I) = 4.5 \tag{3.105}$$

If we had assumed a uniform pdf, then we would have predicted a different average

$$\sum_{i=1}^{n} iprob(X_i|I) = \frac{1}{6}\sum_{i=1}^{n} i = 3.5 \tag{3.106}$$

which means the uniform pdf is not a valid assignment.

There are many pdfs that are consistent with the experimental results. Which one is the best?

The constraint equation above is called *testable information*.

With such a condition, we can either accept or reject any proposed pdf. Jaynes(one of the most brilliant theoretical physicists) proposed that, in this situation, we should make the assignment by using the principle of maximum entropy (**MaxEnt**), that is, we should choose that pdf which has the most entropy S while satisfying the available constraints.

Explicitly, for case in the die experiment above, we need to maximize

$$S = -\sum_{i=1}^{6} p_i \log_e(p_i) \tag{3.107}$$

where $p_i = prob(X_i|I)$ subject to the conditions:

(1) normalization constraint

$$\sum_{i=1}^{6} p_i = 1 \tag{3.108}$$

and

(2) testable information constraint

$$\sum_{i=1}^{6} i p_i = 4.5 \tag{3.109}$$

Such a constrained optimization is done using the method of *Lagrange multipliers* as shown below.

Define the functions

$$f(p_i) = \sum_{i=1}^{6} p_i - 1 = 0 \Rightarrow \frac{\partial f}{\partial p_j} \equiv 1 \tag{3.110}$$

$$g(p_i) = \sum_{i=1}^{6} i p_i - 4.5 = 0 \Rightarrow \frac{\partial g}{\partial p_j} \equiv j \tag{3.111}$$

The maximization problem can then be written in the following way. Instead we maximize the quantity

$$S + \lambda_f f + \lambda_g g = S \text{ (by definition)} \tag{3.112}$$

where the *constants* are called undetermined or Lagrange multipliers. Thus the maximization equation becomes

$$\frac{\partial S}{\partial p_j} + \lambda_f \frac{\partial f}{\partial p_j} + \lambda_g \frac{\partial g}{\partial p_j} = 0 \quad j = 1, 2, 3, 4, 5, 6 \tag{3.113}$$

We get the equations

$$-\log_e(p_j) - 1 + \lambda_f + j\lambda_g = 0 \quad j = 1, 2, 3, 4, 5, 6 \tag{3.114}$$

and we obtain

$$-\log_e(p_{j+1}) - 1 + \lambda_f + (j+1)\lambda_g = -\log_e(p_j) - 1 + \lambda_f + j\lambda_g \tag{3.115}$$

This implies that

$$\log_e \frac{p_{j+1}}{p_j} = \lambda_g \Rightarrow \frac{p_{j+1}}{p_j} = \beta = \text{constant} \tag{3.116}$$

This gives

$$-\log_e(p_1) - 1 + \lambda_f + j\log\beta = 0 \Rightarrow \lambda_f = 1 + \log_e \frac{p_1}{\beta} \tag{3.117}$$

Therefore

$$\sum_{i=1}^{6} p_i = 1 = p_1(1 + \beta + \beta^2 + \beta^3 + \beta^4 + \beta^5) \tag{3.118}$$

$$\sum_{i=1}^{6} i p_i = 4.5 = p_1(1 + 2\beta + 3\beta^2 + 4\beta^3 + 5\beta^4 + 6\beta^5) \tag{3.119}$$

or dividing to get rid of p_1 we have

$$\frac{1 + 2\beta + 3\beta^2 + 4\beta^3 + 5\beta^4 + 6\beta^5}{1 + \beta + \beta^2 + \beta^3 + \beta^4 + \beta^5} = 4.5 \tag{3.120}$$

which gives

$$1.5\beta^5 + 0.5\beta^4 - 0.5\beta^3 - 1.5\beta^2 - 2.5\beta - 3.5 = 0 \tag{3.121}$$

Solving numerically for β we get 1.449255 so that

$$\begin{aligned}
p_1 &= \frac{1}{1 + \beta + \beta^2 + \beta^3 + \beta^4 + \beta^5} = 0.05435 \\
p_2 &= \beta p_1 = 0.07877 \\
p_3 &= \beta p_2 = 0.11416 \\
p_4 &= \beta p_3 = 0.16545 \\
p_5 &= \beta p_4 = 0.23977 \\
p_6 &= \beta p_5 = 0.34749
\end{aligned} \tag{3.122}$$

is the MaxEnt assignment for the pdf for the outcomes of the die roll, given only that it has the usual six faces and yields an average result of 4.5.

Why should the entropy function

$$S = -\sum_{i=1}^{6} p_i \log_e(p_i) \tag{3.123}$$

specified above be the choice for a selection criterion?

Let us look at two examples that suggest this criterion is highly desirable and probably correct.

Kangaroo Problem(Gull and Skilling)

The kangaroo problem is as follows:

Information: 1/3 of all kangaroos have blue eyes and 1/3 of all kangaroos are left-handed

Question: On the basis of this information alone, what proportion of kangaroos are both blue-eyed and left-handed?

For any particular kangaroo, there are four distinct possibilities, namely, that it is

(1) blue-eyed and left-handed

(2) blue-eyed and right-handed

(3) not blue-eyed but left-handed

(4) not blue-eyed but right-handed

Bernoulli's law of large numbers says that the expected values of the fraction of kangeroos with characteristics (1)-(4) will be equal to the probabilities (p_1, p_2, p_3, p_4) we assign to each of these propositions.

This is represented by a 2×2 *truth or contingency table* as shown below:

	Left-handed True	Left-handed False
Blue-Eyed True	p_1	p_2
Blue-Eyed False	p_3	p_4

Table 3.3: GeneralTruth Table

Although there are four possible combinations of eye-color and handedness to be considered, the related probabilities are not completely independent of each other. We have the standard normalization requirement

$$\sum_i p_i = 1 \tag{3.124}$$

In addition, we also have two conditions on the so-called *marginal* probabilities

$$p_1 + p_2 = prob(blue \cap left|I) + prob(blue \cap right|I) = 1/3 \tag{3.125}$$

$$p_1 + p_3 = prob(blue \cap left|I) + prob(not - blue \cap left|I) = 1/3 \tag{3.126}$$

Since any $p_i \geq 0$, these imply that $0 \leq p_1 \leq 1/3$. Using this result we can characterize the contingency table by a single variable $x = p_1$ as in the table below:

	Left-handed True	Left-handed False
Blue-Eyed True	$0 \leq x \leq 1/3$	$1/3 - x$
Blue-Eyed False	$1/3 - x$	$1/3 + x$

Table 3.4: For Kangaroo Problem

where we have used

$$x = p_1 \tag{3.127}$$

$$p_1 + p_2 = \frac{1}{3} \rightarrow p_2 = \frac{1}{3} - x \tag{3.128}$$

$$p_1 + p_3 = \frac{1}{3} \rightarrow p_3 = \frac{1}{3} - x \tag{3.129}$$

$$p_1 + p_2 + p_3 + p_4 = 1 \rightarrow p_4 = \frac{1}{3} + x \tag{3.130}$$

All such solutions, where $0 \leq x \leq 1/3$, satisfy the constraints of the testable information that is available. Which one is best?

Common sense leads us towards the assignment based on independence of these two traits, that is, any other assignment would indicate a knowledge of kangaroo eye-color told us something about its handedness. Since we have no information to determine even the sign of any potential correlation, let alone its magnitude, any choice other than independence is not justified.

The independence choice says that

$$x = p_1 = prob(blue \cap left|I) = prob(blue|I)prob(left|I) = \frac{1}{9} \tag{3.131}$$

In this particular example it was straightforward to decide the most sensible pdf assignment in the face of the inadequate information.

We now ask whether there is some function of the $\{p_i\}$ which, when maximized subject to the known constraints, yields the *independence* solution. The importance of finding an answer to this question is that it would become a good candidate for a general *variational principle* that could be used in situations that were too complicated for our common sense.

Skilling has shown that the only function which gives $x = 1/9$ is the entropy S as specified above or

$$\begin{aligned} S &= -\sum_{i=1}^{4} p_i \log_e(p_i) \\ &= -x\log_e(x) - 2\left(\frac{1}{3} - x\right)\log_e\left(\frac{1}{3} - x\right) - \left(\frac{1}{3} + x\right)\log_e\left(\frac{1}{3} + x\right) \end{aligned} \tag{3.132}$$

The results of Skilling's investigations, including three proposed alternatives,

$$\begin{aligned} S1 &= -\sum_{i=1}^{4} p_i \log_e(p_i) \Rightarrow \text{MaxEnt} \\ S2 &= -\sum_{i=1}^{4} p_i^2 \qquad S3 = -\sum_{i=1}^{4} \log_e(p_i) \qquad S4 = -\sum_{i=1}^{4} \sqrt{p_i} \end{aligned} \tag{3.133}$$

is shown in the table below:

Function	Optimal x	Implied Correlation
S1	0.1111	None
S2	0.0833	Negative
S3	0.1301	Positive
S4	0.1218	Positive

Table 3.5: Skilling Results

Clearly, only the MaxEnt assumption leads to an optimal value with no correlations as expected.

Let us look at another example that lends further support to the MaxEnt principle.

The Team of Monkeys

Suppose there are M distinct possibilities $\{X_i\}$ to be considered. How can we assign truth tables ($prob(X_i|I) = p_i$) to these possibilities given some testable information I (experimental results).

What is the most honest and fair procedure?

Imagine playing the following game.

The various propositions are represented by different boxes all of the same size into which pennies are thrown at random. The tossing job is often assigned to a team of monkeys under the assumption that this will not introduce any underlying bias into the process.

After a very large number of coins have been distributed into the boxes, the fraction found in each of the boxes gives a possible assignment of the probability for the corresponding $\{X_i\}$.

The resulting pdf may not be consistent with the constraints of I, of course, in which case it must be rejected as a potential candidate. If it is in agreement, then it is a viable option.

The process is then repeated by the monkeys many times. After many such trials, some distributions will be found to come up more often than others. The one that occurs most frequently (and satisfies I) would be a sensible choice for $prob(\{X_i\}|I)$.

This is so because the team of monkeys has no axe to grind (no underlying bias) and thus the most frequent solution can be regarded as the one that best represents our state of knowledge. It agrees with all the testable information available while being as indifferent as possible to everything else.

Does this correspond to the pdf to the greatest value of $S = -\sum p_i \log_e(p_i)$?

After the monkeys have tossed all the pennies given to them, suppose that we find n_1 in the first box, n_2 in the second box, and so on. We then have

$$N = \sum_{i=1}^{M} n_i = \text{total number of coins} \tag{3.134}$$

which will be assumed to be very large and also much greater than the number of boxes M.

This distribution gives rise to the candidate pdf $\{p_i\}$ for the possibilities $\{X_i\}$:

$$p_i = \frac{n_i}{N} \quad , \quad i = 1, 2, \ldots, M \tag{3.135}$$

Since every penny can land in any of the boxes there are M^N number of different ways of tossing the coins among the boxes. Each way, by assumption of randomness and no underlying bias by the monkeys, is equally likely to occur. All of the basic sequences, however, are not distinct, since many yield the same distribution $\{n_i\}$. The expected frequency F with which a set $\{p_i\}$ will arise, is given by

$$F\left(\{p_i\}\right) = \frac{\text{number of ways of obtaining} \{n_i\}}{M^N} \tag{3.136}$$

The numerator is just the number of ways to distribute N coins in a distribution $\{n_i\}$ which is given by

$$\text{number of ways of obtaining} \{n_i\} = \frac{N!}{n_1! n_2! \ldots n_M!} \tag{3.137}$$

Putting everything together we have

$$F(\{p_i\}) = \frac{\text{number of ways of obtaining}\{n_i\}}{M^N} = \frac{\frac{N!}{n_1!n_2!\ldots n_M!}}{M^N} \tag{3.138}$$

$$\log(F) = -N\log(M) + \log(N!) - \sum_{i=1}^{M}\log(n_i!) \tag{3.139}$$

Using Stirling's approximation $\log(n_i!) \approx n\log(n) - n$ for large n, we find

$$\begin{aligned}\log(F) &= -N\log(M) + N\log(N) - \sum_{i=1}^{M} n_i\log(n_i) - N + \sum_{i=1}^{M} n_i \\ &= -N\log(M) + N\log(N) - \sum_{i=1}^{M} n_i\log(n_i)\end{aligned} \tag{3.140}$$

and thus

$$\begin{aligned}\log(F) &= -N\log(M) + N\log(N) - \sum_{i=1}^{M} p_i N\log(p_i N) \\ &= -N\log(M) + N\log(N) - \sum_{i=1}^{M} p_i N(\log(p_i) + \log(N)) \\ &= -N\log(M) + N\log(N) - N\sum_{i=1}^{M} p_i\log(p_i) - N\log(N)\sum_{i=1}^{M} p_i \\ &= -N\log(M) + N\log(N) - N\sum_{i=1}^{M} p_i\log(p_i) - N\log(N) \\ &= -N\log(M) - N\sum_{i=1}^{M} p_i\log(p_i)\end{aligned} \tag{3.141}$$

Maximizing the $\log(F)$ is equivalent to maximizing F, which is the expected frequency with which the monkeys will come up with the candidate pdf $\{p_i\}$, that is, maximizing $\log(F)$ will give us the assignment $prob(\{X_i\}|I)$ which best represents our state of knowledge consistent with the testable information I. Since M and N are constants, this is equivalent to the constrained maximization of the entropy function

$$S = -\sum p_i\log_e(p_i) \tag{3.142}$$

and so we recover the MaxEnt procedure once again.

3.5 Discussion

In discussions of Bayesian methods, opponents often use the words *subjective probabilities* to say that the methods are not as valid as normal *objective* probability theory.

These opponents are misguided.

The main point of concern centers around the choice of the prior pdf, that is, what should we do if it is not known?

This is actually a very strange question. It is usually posed this way by opponents of the Bayesian methods in an attempt to prove its subjective nature.

No probability, whether prior, likelihood or whatever, is ever *known*. It is simply an assignment which reflects the relevant information that is available. Thus, $prob(x|I_1) \neq prob(x|I_2)$, in general, where the conditioning statements I_1 and I_2 are different.

Nevertheless, objectivity can, and must, be introduced by demanding the two people with the same information I should assign the same pdf. I think that this consistency requirement is the most important idea of all.

Invariance arguments, under transformation groups, can be used to uniquely determine a pdf when given only the nature of the quantities involved. MaxEnt provides a powerful extension when we have testable constraints.

While we may yet be far from knowing how to convert every piece of vague information into a concrete probability assignment, we can deal with a wide variety of problems with these ideas.

The important point is that nowhere in our discussion have we explicitly differentiated between a prior and a likelihood. We have only considered how to assign $prob(X|I)$ for different types of I. If X pertains to data, then we call $prob(X|I)$ a likelihood. If neither X nor I refers to (new) measurements, then we may say it is a prior.

The distinction between the two cases is one of nomenclature and not of objectivity or subjectivity. If it appears otherwise, then this is because we are usually prepared to state conditioning assumptions for the likelihood function but shy away from doing likewise for the prior pdf.

The use of Bayesian methods in quantum mechanics presents a very different view of quantum probability than normally appears in quantum theory textbooks. It is becoming increasingly important in discussions of measurement.

3.6 Problems

3.6.1 Simple Probability Concepts

There are 14 short problems in this section. If you have not studied any probability ideas before using this book, then these are all new to you and doing them should enable you to learn the basic ideas of probability methods. If you have studied probability ideas before, these should all be straightforward.

(a) Two dice are rolled, one after the other. Let A be the event that the second number if greater than the first. Find $P(A)$.

(b) Three dice are rolled and their scores added. Are you more likely to get 9 than 10, or vice versa?

(c) Which of these two events is more likely?

1. four rolls of a die yield at least one six
2. twenty-four rolls of two dice yield at least one double six

(d) From meteorological records it is known that for a certain island at its winter solstice, it is wet with probability 30%, windy with probability 40% and both wet and windy with probability 20%. Find

(1) $Prob$(dry)

(2) $Prob$(dry AND windy)

(3) $Prob$(wet OR windy)

(e) A kitchen contains two fire alarms; one is activated by smoke and the other by heat. Experiment has shown that the probability of the smoke alarm sounding within one minute of a fire starting is 0.95, the probability of the heat alarm sounding within one minute of a fire starting is 0.91, and the probability of both alarms sounding within one minute is 0.88. What is the probability of at least one alarm sounding within a minute?

(f) Suppose you are about to roll two dice, one from each hand. What is the probability that your right-hand die shows a larger number than your left-hand die? Now suppose you roll the left-hand die first and it shows 5. What is the probability that the right-hand die shows a larger number?

(g) A coin is flipped three times. Let A be the event that the first flip gives a head and B be the event that there are exactly two heads overall. Determine

(1) $P(A|B)$

(2) $P(B|A)$

(h) A box contains a double-headed coin, a double-tailed coin and a conventional coin. A coin is picked at random and flipped. It shows a head. What is the probability that it is the double-headed coin?

(i) A box contains 5 red socks and 3 blue socks. If you remove 2 socks at random, what is the probability that you are holding a blue pair?

(j) An inexpensive electronic toy made by Acme Gadgets Inc. is defective with probability 0.001. These toys are so popular that they are copied and sold illegally but cheaply. Pirate versions capture 10% of the market and any pirated copy is defective with probability 0.5. If you buy a toy, what is the chance that it is defective?

(k) Patients may be treated with any one of a number of drugs, each of which may give rise to side effects. A certain drug C has a 99% success rate in the absence of side effects and side effects only arise in 5% of cases. However, if they do arise, then C only has a 30% success rate. If C is used, what is the probability of the event A that a cure is effected?

(l) Suppose a multiple choice question has c available choices. A student either knows the answer with probability p or guesses at random with probability $1 - p$. Given that the answer selected is correct, what is the probability that the student knew the answer?

(m) Common PINs do not begin with zero. They have four digits. A computer assigns you a PIN at random. What is the probability that all four digits are different?

(n) You are dealt a hand of 5 cards from a conventional deck(52 cards). A full house comprises 3 cards of one value and 2 of another value. If that hand has 4 cards of one value, this is called four of a kind. Which is more likely?

3.6.2 Playing Cards

Two cards are drawn at random from a shuffled deck and laid aside without being examined. Then a third card is drawn. Show that the probability that the third card is a spade is 1/4 just as it was for the first card. *HINT*: Consider all the (mutually exclusive) possibilities (two discarded cards spades, third card spade or not spade, etc).

3.6.3 Birthdays

What is the probability that you and a friend have different birthdays? (for simplicity let a year have 365 days). What is the probability that three people have different birthdays? Show that the probability that n people have different birthdays is

$$p = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\left(1 - \frac{3}{365}\right).....\left(1 - \frac{n-1}{365}\right)$$

Estimate this for $n << 365$ by calculating $\log(p)$ (use the fact that $\log(1+x) \approx x$ for $x \ll 1$). Find the smallest integer N for which $p < 1/2$. Hence show that for a group of N people or more, the probability is greater than $1/2$ that two of them have the same birthday.

3.6.4 Is there life?

The number of stars in our galaxy is about $N = 10^{11}$. Assume that the probability that a star has planets is $p = 10^{-2}$, the probability that the conditions on the planet are suitable for life is $q = 10^{-2}$, and the probability of life evolving, given suitable conditions, is $r = 10^{-2}$. These numbers are rather arbitrary.

(a) What is the probability of life existing in an arbitrary solar system (a star and planets, if any)?

(b) What is the probability that life exists in at least one solar system?

3.6.5 Law of large Numbers

This problem illustrates the law of large numbers.

(a) Assuming the probability of obtaining *heads* in a coin toss is 0.5, compare the probability of obtaining *heads* in 5 out of 10 tosses with the probability of obtaining *heads* in 50 out of 100 tosses and with the probability of obtaining *heads* in 5000 out of 10000 tosses. What is happening?

(b) For a set of 10 tosses, a set of 100 tosses and a set of 10000*tosses*, calculate the probability that the fraction of *heads* will be between 0.445 and 0.555. What is happening?

3.6.6 Bayes

Suppose that you have 3 nickels and 4 dimes in your right pocket and 2 nickels and a quarter in your left pocket. You pick a pocket at random and from it select a coin at random. If it is a nickel, what is the probability that it came from your right pocket? Use Baye's formula.

3.6.7 Psychological Tests

Two psychologists reported on tests in which subjects were given the *prior information*:

I = In a certain city, 85% of the taxicabs
are blue and 15% are green

and the *data*:

D = A witness to a crash who is 80% reliable (i.e.,

who in the lighting conditions prevailing can distinguish between green and blue 80% of the time) reports that the taxicab involved in the crash was green

The subjects were then asked to judge the probability that the taxicab was actually blue. What is the correct answer?

3.6.8 Bayes Rules, Gaussians and Learning

Let us consider a classical problem(no quantum uncertainty). Suppose we are trying to measure the position of a particle and we assign a prior probability function,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x-x_0)^2/2\sigma_0^2}$$

Our measuring device is not perfect. Due to noise it can only measure with a resolution Δ, i.e., when I measure the position, I must assume error bars of this size. Thus, if my detector registers the position as y, I assign the likelihood that the position was x by a Gaussian,

$$p(y|x) = \frac{1}{\sqrt{2\pi\Delta^2}} e^{-(y-x)^2/2\Delta^2}$$

Use Bayes theorem to show that, given the new data, I must now update my probability assignment of the position to a new Gaussian,

$$p(x|y) = \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-(x-x')^2/2\sigma'^2}$$

where

$$x' = x_0 + K_1(y - x_0) \ , \ \sigma'^2 = K_2\sigma_0^2 \ , \ K_1 = \frac{\sigma_0^2}{\sigma_0^2 + \Delta^2} \ , \ K_2 = \frac{\Delta^2}{\sigma_0^2 + \Delta^2}$$

Comment on the behavior as the measurement resolution improves. How does the learning process work?

3.6.9 Berger's Burgers-Maximum Entropy Ideas

A fast food restaurant offers three meals: burger, chicken, and fish. The price, Calorie count, and probability of each meal being delivered cold are listed below in Table 3.6:

Item	Entree	Cost	Calories	Prob(hot)	Prob(cold)
Meal 1	burger	\$1.00	1000	0.5	0.5
Meal 2	chicken	\$2.00	600	0.8	0.2
Meal 3	fish	\$3.00	400	0.9	0.1

Table 3.6: Berger's Burgers Details

We want to identify the state of the system, i.e., the values of

$$\begin{aligned} Prob(burger) &= P(B) \\ Prob(chicken) &= P(C) \\ Prob(fish) &= P(F) \end{aligned}$$

Even though the problem has now been set up, we do not know which state the actual state of the system. To express what we do know despite this ignorance, or uncertainty, we assume that each of the possible states A_i has some probability of occupancy $P(A_i)$, where i is an index running over the possible states. As stated above, for the restaurant model, we have three such possibilities, which we have labeled $P(B)$, $P(C)$, and $P(F)$.

A probability distribution $P(A_i)$ has the property that each of the probabilities is in the range $0 \leq P(A_i) \leq 1$ and since the events are mutually exclusive and exhaustive, the sum of all the probabilities is given by

$$1 = \sum_i P(A_i) \tag{3.143}$$

Since probabilities are used to cope with our lack of knowledge and since one person may have more knowledge than another, it follows that two observers may, because of their different knowledge, use different probability distributions. In this sense probability, and all quantities that are based on probabilities are *subjective*.

Our uncertainty is expressed quantitatively by the *information* which we do not have about the state occupied. This *information* is

$$S = \sum_i P(A_i) \log_2 \left(\frac{1}{P(A_i)} \right) \tag{3.144}$$

This *information* is measured in bits because we are using logarithms to base 2.

In physical systems, this uncertainty is known as the *entropy*. Note that the entropy, because it is expressed in terms of probabilities, depends on the observer.

The principle of maximum entropy (**MaxEnt**) is used to discover the probability distribution which leads to the largest value of the entropy (a maximum),

thereby assuring that no information is inadvertently assumed.

If one of the probabilities is equal to 1, the all the other probabilities are equal to 0, and the entropy is equal to 0.

It is a property of the above entropy formula that it has its maximum when all the probabilities are equal (for a finite number of states), which the state of *maximum ignorance*.

If we have no additional information about the system, then such a result seems reasonable. However, if we have additional information, then we should be able to find a probability distribution which is better in the sense that it has less uncertainty.

In this problem we will impose only one constraint. The particular constraint is the known average price for a meal at Berger's Burgers, namely \$1.75. This constraint is an example of an expected value.

(a) Express the constraint in terms of the unknown probabilities and the prices for the three types of meals.

(b) Using this constraint and the total probability equal to 1 rule find possible ranges for the three probabilities in the form

$$a \leq P(B) \leq b$$
$$c \leq P(C) \leq d$$
$$e \leq P(F) \leq f$$

(c) Using this constraint, the total probability equal to 1 rule, the entropy formula and the MaxEnt rule, find the values of $P(B)$, $P(C)$, and $P(F)$ which maximize S.

(d) For this state determine the expected value of Calories and the expected number of meals served cold.

In finding the state which maximizes the entropy, we found the probability distribution that is consistent with the constraints and has the largest uncertainty. Thus, we have not inadvertently introduced any biases into the probability estimation.

3.6.10 Extended Menu at Berger's Burgers

Suppose now that Berger's extends its menu to include a Tofu option as shown in Table 3.7 below:

Entree	Cost	Calories	Prob(hot)	Prob(cold)
burger	\$1.00	1000	0.5	0.5
chicken	\$2.00	600	0.8	0.2
fish	\$3.00	400	0.9	0.1
tofu	\$8.00	200	0.6	0.4

Table 3.7: Extended Berger's Burgers Menu Details

Suppose you are now told that the average meal price is \$2.50.

Use the method of Lagrange multipliers to determine the state of the system (i.e., $P(B)$, $P(C)$, $P(F)$ and $P(T)$).

You will need to solve some equations numerically.

3.6.11 The Poisson Probability Distribution

The arrival time of rain drops on the roof or photons from a laser beam on a detector are completely random, with no correlation from count to count. If we count for a certain time interval we won't always get the same number - it will fluctuate from shot-to-shot. This kind of noise is sometimes known as *shot noise* or *counting statistics*.

Suppose the particles arrive at an average rate R. In a small time interval $\Delta t \ll 1/R$ no more than one particle can arrive. We seek the probability for n particles to arrive after a time t, $P(n,t)$.

(a) Show that the probability to detect zero particles exponentially decays, $P(0,t) = e^{-Rt}$.

(b) Obtain a differential equation as a recursion relation

$$\frac{d}{dt}P(n,t) + RP(n,t) = RP(n-1,t)$$

(c) Solve this to find the Poisson distribution

$$P(n,t) = \frac{(Rt)^n}{n!}e^{-Rt}$$

Plot a histogram for $Rt = 0.1, 1.0, 10.0$.

(d) Show that the mean and standard deviation in number of counts are:

$$\langle n \rangle = Rt \quad , \quad \sigma_n = \sqrt{Rt} = \sqrt{\langle n \rangle}$$

[HINT: To find the variance consider $\langle n(n-1) \rangle$].

Fluctuations going as the square root of the mean are characteristic of counting statistics.

(e) An alternative way to derive the Poisson distribution is to note that the count in each small time interval is a Bernoulli trial(find out what this is), with probability $p = R\Delta t$ to detect a particle and $1 - p$ for no detection. The total number of counts is thus the binomial distribution. We need to take the limit as $\Delta t \to 0$ (thus $p \to 0$) but Rt remains finite (this is just calculus). To do this let the total number of intervals $N = t/\Delta t \to \infty$ while $Np = Rt$ remains finite. Take this limit to get the Poisson distribution.

3.6.12 Modeling Dice: Observables and Expectation Values

Suppose we have a pair of six-sided dice. If we roll them, we get a pair of results

$$a \in \{1,2,3,4,5,6\} \quad , \quad b \in \{1,2,3,4,5,6\}$$

where a is an observable corresponding to the number of spots on the top face of the first die and b is an observable corresponding to the number of spots on the top face of the second die. If the dice are fair, then the probabilities for the roll are

$$Pr(a=1) = Pr(a=2) = Pr(a=3) = Pr(a=4) = Pr(a=5) = Pr(a=6) = 1/6$$
$$Pr(b=1) = Pr(b=2) = Pr(b=3) = Pr(b=4) = Pr(b=5) = Pr(b=6) = 1/6$$

Thus, the expectation values of a and b are

$$\langle a \rangle = \sum_{i=1}^{6} iPr(a=i) = \frac{1+2+3+4+5+6}{6} = 7/2$$
$$\langle b \rangle = \sum_{i=1}^{6} iPr(b=i) = \frac{1+2+3+4+5+6}{6} = 7/2$$

Let us define two new observables in terms of a and b:

$$s = a + b \quad , \quad p = ab$$

Note that the possible values of s range from 2 to 12 and the possible values of p range from 1 to 36. Perform an explicit computation of the expectation values of s and p by writing out

$$\langle s \rangle = \sum_{i=2}^{12} iPr(s=i)$$

and

$$\langle p \rangle = \sum_{i=1}^{36} iPr(p=i)$$

Do this by explicitly computing all the probabilities $Pr(s=i)$ and $Pr(p=i)$. You should find that $\langle s \rangle = \langle a \rangle + \langle b \rangle$ and $\langle p \rangle = \langle a \rangle \langle b \rangle$. Why are these results not surprising?

3.6.13 Conditional Probabilities for Dice

Use the results of Problem 5.12. You should be able to intuit the correct answers for this problems by straightforward probabilistic reasoning; if not you can use Baye's Rule

$$Pr(x|y) = \frac{Pr(y|x)Pr(x)}{Pr(y)}$$

to calculate the results. Here $Pr(x|y)$ represents *the probability of x given y*, where x and y should be propositions of equations (for example, $Pr(a = 2|s = 8)$ is the probability that $a = 2$ given the $s = 8$).

(a) Suppose your friend rolls a pair of dice and, without showing you the result, tells you that $s = 8$. What is your conditional probability distribution for a?

(b) Suppose your friend rolls a pair of dice and, without showing you the result, tells you that $p = 12$. What is your conditional expectation value for s?

3.6.14 Matrix Observables for Classical Probability

Suppose we have a biased coin, which has probability p_h of landing heads-up and probability p_t of landing tails-up. Say we flip the biased coin but do not look at the result. Just for fun, let us represent this preparation procedure by a *classical state vector*

$$x_0 = \begin{pmatrix} \sqrt{p_h} \\ \sqrt{p_t} \end{pmatrix}$$

(a) Define an observable (random variable) r that takes value +1 if the coin is heads-up and −1 if the coin is tails-up. Find a matrix R such that

$$x_0^T R x_0 = \langle r \rangle$$

where $\langle r \rangle$ denotes the mean, or expectation value, of our observable.

(b) Now find a matrix F such that the *dynamics* corresponding to turning the coin over (after having flipped it, but still without looking at the result) is represented by

$$x_0 \mapsto F x_0$$

and

$$\langle r \rangle \mapsto x_0^T F^T R F x_0$$

Does $U = F^T R F$ make sense as an observable? If so explain what values it takes for a coin-flip result of heads or tails. What about RF or $F^T R$?

(c) Let us now define the algebra of flipped-coin observables to be the set V of all matrices of the form

$$v = aR + bR^2 \quad , \quad a, b \in R$$

Show that this set is closed under matrix multiplication and that it is commutative. In other words, for any $v_1, v_2 \in V$, show that

$$v_1, v_2 \in V \quad , \quad v_1 v_2 = v_2 v_1$$

Is U in this set? How should we interpret the observable represented by an arbitrary element $v \in V$?

Chapter 4

The Formulation of Quantum Mechanics

4.1 Introduction

A *physical theory* is

1. a set of basic physical concepts
2. a mathematical formalism
3. a set of connections between the physical concepts and the mathematical objects that represent them

The *process* of doing theoretical physics involves:

1. constructing a mathematical formalism that allows us to express a physical problem or experiment in terms of mathematical objects
2. solving the mathematical system representing the physical problem – this a pure mathematics at this point
3. translating the mathematical solution to the physical world using the set of connections

A *description* of a physical system will involve three ingredients:

1. variables or measurable quantities that characterize the system
2. states that describe values of the variables as a function of time
3. equations of motion that determine how the states variables change in time

In classical mechanics, the position of a particle, which is a physical concept, is connected to a set of real numbers, which is a mathematical object. This does not, in general, present us with any conceptual difficulties because we are familiar with both positions and numbers from our everyday experiences.

In quantum mechanics, however, the mathematical formalism that we will be discussing, is very abstract. In addition, we lack any intuition based on experience concerning quantum phenomena. The everyday world that we live in does not appear to consist of Hermitian operators and infinite-dimensional vectors.

Throughout the last century, many experiments have shown that the various dynamical variables, which seem to have a continuous spectrum in classical mechanics, generally have completely discrete or quantized spectra or, at least, a spectrum that is both discrete and continuous. In quantum mechanics, this property will lead to most of the so-called non-classical results and force us to work in a world of probabilities.

We will take an approach to the theoretical formalism of quantum mechanics presented in terms of postulates.

4.2 Two Postulates: Presentation, Rationalization and Meaning

Postulate 1

> For each *dynamical variable* or *observable*, which is a physical concept, there corresponds a *Hermitian, linear operator*, which is a mathematical object.
>
> The *possible values of any measurement* of the observable are restricted to the *eigenvalues* of the corresponding operator.

It is the nature of a postulate that we can only make ourselves feel good about it or, in other words, make ourselves feel that it *makes sense* in some way. A priori, we cannot justify it in any way and we certainly cannot prove it is true or we would have assumed a more fundamental postulate(the one we used to prove the truth). A posteriori, we justify postulates by their effect on our predictions about experimental results. My philosophy of theoretical physics is based on this statement

> If the predictions agree with experiment,
> on certain quantum systems, then the
> postulates are valid for that class of systems.

So what can we say about the 1st postulate? We already know that linear operators possess both discrete and continuous spectra. In addition, we have a vast storehouse of mathematical knowledge available about linear operators. So making this choice certainly is a sensible(and the simplest) place to start.

Note that the postulate does not give us any rules for assigning operators to observables.

Singling out the eigenvalues is also sensible since they represent a special connection to the properties of the associated linear operator. Choosing Hermitian operators also makes sense since this guarantees that we have only real eigenvalues representing measured quantities.

Now we need to deal with states. The mathematical object that we connect to a physical state must allow us to calculate the probability distributions for all observables.

Before stating the 2nd postulate, we need to present some new mathematical objects and ideas.

$$Tr\hat{W} = 1 = \sum_k W_{kk} = \sum_k \langle \phi_k | \hat{W} | \phi_k \rangle \tag{4.1}$$

where W_{kk} is the diagonal matrix element(in the basis) of the density operator $\hat{W}$.

Some Properties of the Trace:

$$Tr(AB) = Tr(BA) \tag{4.2}$$

$$Tr(cB) = cTr(B) \tag{4.3}$$

$$Tr(c(A+B)) = Tr(cA) + Tr(cB) = cTr(A) + cTr(B) \tag{4.4}$$

If we denote the eigenvalues of $\hat{W}$ by w_k and the corresponding eigenvectors by $|w_k\rangle$ so that

$$\hat{W} |w_k\rangle = w_k |w_k\rangle \tag{4.5}$$

then, since $\hat{W}$ has a pure point spectrum, we can write $\hat{W}$ in terms of its eigenvalues and eigenvectors (spectral representation) as

$$\hat{W} = \sum_k w_k |w_k\rangle \langle w_k| \tag{4.6}$$

Since $\hat{W}$ is self-adjoint, its eigenvectors form an orthonormal basis where

$$\langle w_k | w_j \rangle = \delta_{kj} \tag{4.7}$$

We now derive some properties of this density operator mathematical object.

The spectrum of $\hat{W}$ is the set of numbers $\{w_k\}$. We then have

$$\begin{aligned} Tr\hat{W} = 1 &= \sum_j \langle w_j | \hat{W} | w_j \rangle \\ &= \sum_j \langle w_j | w_j | w_j \rangle = \sum_j w_j \langle w_j | w_j \rangle = \sum_j w_j \end{aligned} \tag{4.8}$$

So,

$$\sum_j w_j = 1 \tag{4.9}$$

$\hat{W}$ is a bounded, Hermitian(self-adjoint) operator. Since $\hat{W}$ is self-adjoint, this means that $\hat{W} = \hat{W}^\dagger$, which implies that the eigenvalues are real, i.e., $w_k = w_k^*$. Using the fact that $\hat{W}$ is defined to be a positive operator, we then have

$$\begin{aligned}\langle a|\hat{W}|a\rangle &= \langle a|\left(\sum_k w_k |w_k\rangle\langle w_k|\right)|a\rangle = \sum_k w_k \langle a|w_k\rangle\langle w_k|a\rangle \\ &= \sum_k w_k |\langle a|w_k\rangle|^2 \geq 0\end{aligned} \tag{4.10}$$

for any vector $|a\rangle$. This can only be true, in general, if

$$w_k \geq 0 \text{ for all } k \tag{4.11}$$

The results

$$w_k \geq 0 \text{ for all } k \text{ and } \sum_j w_j = 1 \tag{4.12}$$

$$0 \leq w_k \leq 1 \tag{4.13}$$

Finally, for any other operator $\hat{B}$

$$\hat{W}\hat{B} = \sum_k w_k |w_k\rangle\langle w_k|\hat{B} \tag{4.14}$$

We then have

$$\begin{aligned}Tr(\hat{W}\hat{B}) &= \sum_j \langle w_j|\left(\sum_k w_k |w_k\rangle\langle w_k|\hat{B}\right)|w_j\rangle \\ &= \sum_k w_k \sum_j \langle w_j|w_k\rangle\langle w_k|\hat{B}|w_j\rangle = \sum_k w_k \sum_j \delta_{jk}\langle w_k|\hat{B}|w_j\rangle \\ &= \sum_k w_k \langle w_k|\hat{B}|w_k\rangle\end{aligned} \tag{4.15}$$

which is a weighted sum (where the eigenvalues of the density operator are the weight factors) of the diagonal matrix elements of the operator $\hat{B}$ in a basis consisting of eigenvectors of $\hat{W}$.

Before we can show how this connects to probabilities, we must discuss the concept of a state and its representation as a mathematical object.

In classical mechanics, a state refers to the set of coordinates and momenta of the particles in the system.

In quantum mechanics, we say that a state describes the values of a set of measurable quantities. The state description must be probabilistic so that we

can only know the probabilities that these values might occur.

In essence, we assume (this is the standard approach to quantum mechanics) that the quantum state description refers to an *ensemble* of similarly prepared systems(so that we can make probability statements) and NOT to any individual system in the ensemble. We note that there is some dispute about this point.

We are, thus, *identifying* the state directly with a set of probability distributions.

Our earlier discussion(Chapter 4) of the experimental limit frequency and its connection to probability indicated that if we know the average (or mean) value or the *expectation* value, which has the symbol $\langle \ldots \rangle$, for all of the observables in our physical system, we have then specified the state of the system as *exactly as is possible.*

Some properties that we require of expectation values in order that they make physical sense as average values are:

1. if $\hat{B}$ is self-adjoint, then $\langle \hat{B} \rangle$ is real
2. if $\hat{B}$ is positive, then $\langle \hat{B} \rangle$ is nonnegative
3. if c = complex number, then $\langle c\hat{B} \rangle = c\langle \hat{B} \rangle$
4. $\langle \hat{A} + \hat{B} \rangle = \langle \hat{A} \rangle + \langle \hat{B} \rangle$
5. $\langle \hat{I} \rangle = 1$

Postulate 2

(a) A density operator exists for every real physical system.

This rather innocuous looking postulate is at the heart of quantum theory. Its full meaning will only become clear as we learn its connection to the probability interpretation of quantum mechanics.

(b) The expectation value of an operator $\hat{B}$ is given by

$$\langle \hat{B} \rangle = Tr(\hat{W}\hat{B}) \tag{4.16}$$

If we assume that every bounded, Hermitian operator represents some measurable quantity for a physical system, then each state is represented by a unique density operator $\hat{W}$.

Let us choose a simple example of a density operator to get some handle on what this postulate is saying. In particular, let us choose as our density operator $\hat{W}$ the projection operator onto a 1-dimensional subspace spanned by the vector $|\alpha\rangle$, namely

$$\hat{W} = |\alpha\rangle \langle \alpha| \tag{4.17}$$

This is an idempotent operator since $\hat{W}^2 = \hat{W}$ and thus it has eigenvalues $w_k = 0, 1$ only. The eigenvector corresponding to eigenvalue 1 is $|\alpha\rangle$. We also have

$$\sum_k w_k = 1 = Tr(\hat{W}) \tag{4.18}$$

$$\langle a|\hat{W}|a\rangle = |\langle a|\alpha\rangle|^2 \geq 0 \tag{4.19}$$

so that all required properties for being a density operator are satisfied.

Since,

$$0 \leq w_k \leq 1 \text{ and } \sum_k w_k = 1 \tag{4.20}$$

we can think of

$$\hat{W} = \sum_k w_k |w_k\rangle\langle w_k| \tag{4.21}$$

as representing a *statistical mixture* of the states represented by the vectors $|w_k\rangle$ where the probability is w_k that we have the state $|w_k\rangle$ present.

In this simple case, we then have

$$\langle\hat{B}\rangle = Tr(\hat{W}\hat{B}) = \langle\alpha|\hat{B}|\alpha\rangle \tag{4.22}$$

which is the expectation value of $\hat{B}$ in the state $|\alpha\rangle$.

Proof: Let $|w_1\rangle = |\beta\rangle$, $|w_2\rangle = |\alpha\rangle$, $\langle\alpha|\beta\rangle = 0$. Then

$$\begin{aligned} Tr(\hat{W}\hat{B}) &= \sum_{m=1}^{2} \langle w_m|\hat{W}\hat{B}|w_m\rangle = \sum_{m=1}^{2} \langle w_m|\alpha\rangle\langle\alpha|\hat{B}|w_m\rangle \\ &= \sum_{m=1}^{2} \langle w_m|w_2\rangle\langle w_2|\hat{B}|w_m\rangle = \sum_{m=1}^{2} \delta_{m2}\langle w_2|\hat{B}|w_m\rangle \\ &= \langle w_2|\hat{B}|w_2\rangle = \langle\alpha|\hat{B}|\alpha\rangle \end{aligned} \tag{4.23}$$

Since the important quantities for connection to experiment will be expectation values, we see that the state represented by $\hat{W}$, in this case, is *equally well represented* by the state vector $|\alpha\rangle$; the density operator and the state vector are *equivalent* ways of representing a physical system *in this simple case.*

In general, when

$$\hat{W} = \sum_k w_k |w_k\rangle\langle w_k| \tag{4.24}$$

we saw earlier that

$$Tr(\hat{W}\hat{B}) = \sum_k w_k \langle w_k|\hat{B}|w_k\rangle \tag{4.25}$$

Postulate #2 then says that the average or expectation value of $\hat{W}$, $\langle\hat{B}\rangle$, equals the weighted sum of the expectation value in each of the eigenstates $|w_k\rangle$ of $\hat{W}$

with a weighting factor equal to the corresponding eigenvalue w_k.

Thus, if we know the numbers $\langle w_k| \hat{B} |w_k\rangle$ we can find the expectation value of the corresponding operator, which is the maximum amount of information we can know.

4.3 States and Probabilities

4.3.1 States

Among the set of all possible states there exists a special group called *pure states*. The density operator for a pure state is given by

$$\hat{W} = |\psi\rangle \langle\psi| \tag{4.26}$$

where the vector $|\psi\rangle$ is of unit norm and is called the *state vector*. We discussed this density operator above and found that the expectation value for an observable represented by $\hat{Q}$ now in this pure state $|\psi\rangle$, is given by

$$\langle\hat{Q}\rangle = Tr(\hat{W}\hat{Q}) = Tr(|\psi\rangle \langle\psi| \hat{Q}) = \langle\psi| \hat{Q} |\psi\rangle \tag{4.27}$$

Note that the density operator is independent of this arbitrary phase.

As we saw earlier the density operator $\hat{W} = |\psi\rangle \langle\psi|$ for the pure state is idempotent and thus the only possible eigenvalues are $w_k = 0, 1$ and therefore, the form chosen for $\hat{W}$ agrees with the expansion in eigenvectors and eigenvalues.

Another property of a pure state goes as follows: we have, in the general case,

$$0 \leq w_k \leq 1 \text{ and } \sum_k w_k = 1 \tag{4.28}$$

which implies that

$$w_k^2 \leq w_k \tag{4.29}$$

so that

$$Tr(\hat{W}^2) = \sum_k w_k^2 \leq \sum_k w_k = 1 \tag{4.30}$$

or

$$Tr(\hat{W}^2) \leq 1 \tag{4.31}$$

in this case. For a pure state, however, where $w_k^2 = w_k$ because $\hat{W}$ is an idempotent projection operator, we then have

$$Tr(\hat{W}^2) = 1 \tag{4.32}$$

The most important way of distinguishing whether a state is pure or not follows from this property of density operators:

The density operator for a *pure state* cannot be written as a linear combination of the density operators of other states, but the density operator for a *nonpure* state can always be so written.

We can see that this is true by assuming the opposite and getting a contradiction. So let the density operator for a pure state be written as a linear combination of density operators for other states

$$\hat{W} = \sum_j c_j \hat{W}^{(j)} \quad , \quad 0 \le c_j \le 1 \quad , \quad \sum_j c_j = 1 \tag{4.33}$$

Now

$$Tr(\hat{W}^2) = \sum_i \sum_j c_i c_j Tr(\hat{W}^{(i)} \hat{W}^{(j)}) \tag{4.34}$$

Now each density operator $\hat{W}^{(i)}$ has a spectral representation of the form

$$\hat{W}^{(j)} = \sum_n w_k^{(j)} \left| w_n^{(j)} \right\rangle \left\langle w_n^{(j)} \right| \tag{4.35}$$

Substitution gives

$$\begin{aligned} Tr(\hat{W}^{(i)} \hat{W}^{(j)}) &= \sum_n \sum_m w_n^{(i)} w_m^{(j)} Tr(\left| w_n^{(i)} \right\rangle \left\langle w_n^{(i)} \right| w_m^{(j)} \rangle \left\langle w_m^{(j)} \right| & (4.36) \\ &= \sum_n \sum_m w_n^{(i)} w_m^{(j)} |\left\langle w_n^{(i)} \right| w_m^{(j)} \rangle|^2 \\ &\le \sum_n w_n^{(i)} \sum_m w_m^{(j)} = 1 & (4.37) \end{aligned}$$

This then gives

$$Tr(\hat{W}^2) = \sum_i \sum_j c_i c_j Tr(\hat{W}^{(i)} \hat{W}^{(j)}) \le \sum_i c_i \sum_j c_j = 1 \tag{4.38}$$

But we assumed that $\hat{W}$ represents a pure state and the equality must hold. The only way for this to be true is as follows:

$$|\left\langle w_n^{(i)} \right| w_m^{(j)} \rangle| = 1 \text{ for all } m, n \text{ such that } w_n^{(i)} w_m^{(j)} \neq 0 \tag{4.39}$$

The Schwarz inequality $|\langle a|b\rangle|^2 \le \langle a|a\rangle \langle b|b\rangle = 1$ then says that $\left| w_n^{(i)} \right\rangle$ and $\left| w_m^{(j)} \right\rangle$ can differ at most by a phase factor. But by assumption, the eigenvectors for each separate density operator are orthogonal. The only conclusion then is that they are all the same operator, i.e.,

$$\hat{W}^{(i)} = \hat{W}^{(j)} \text{ for all } i \text{ and } j \tag{4.40}$$

This contradicts the assumption that we could write a pure state density operator as a linear combination of other density operators. In fact all the density

operators must be the same.

Are pure states more fundamental than nonpure states?

Can we regard nonpure states as some kind of mixture of pure states?

The answer depends on whether we require uniqueness. It turns out that we cannot write nonpure states uniquely in terms of pure states. For example let us consider the density operator given by

$$\hat{W}_{(a)} = a\,|u\rangle\,\langle u| + (1-a)\,|v\rangle\,\langle v| \tag{4.41}$$

with $0 < a < 1$ and

$$\langle u\,|\,u\rangle = \langle v\,|\,v\rangle = 1 \text{ and } \langle u\,|\,v\rangle = \langle v\,|\,u\rangle = 0 \tag{4.42}$$

Now define two other vectors $|x\rangle$ and $|y\rangle$ as linear combinations of $|u\rangle$ and $|v\rangle$.

$$|x\rangle = \sqrt{a}\,|u\rangle + \sqrt{1-a}\,|v\rangle \tag{4.43}$$

$$|y\rangle = \sqrt{a}\,|u\rangle - \sqrt{1-a}\,|v\rangle \tag{4.44}$$

$$\langle u\,|\,u\rangle = \langle v\,|\,v\rangle = 1 \text{ and } \langle u\,|\,v\rangle = \langle v\,|\,u\rangle = 0 \tag{4.45}$$

such that we still have

$$\langle x\,|\,x\rangle = \langle y\,|\,y\rangle = 1 \text{ and } \langle x\,|\,y\rangle = \langle y\,|\,x\rangle = 0 \tag{4.46}$$

We then have

$$\hat{W}_{(a)} = \frac{1}{2}\,|x\rangle\,\langle x| + \frac{1}{2}\,|y\rangle\,\langle y| \tag{4.47}$$

so the linear combination is not unique. In fact, there are an infinite number of ways to do it.

We will assume that both pure and nonpure states are fundamental and we will see physical systems later on that are described by each type. Due to the lack of uniqueness, we will not call the nonpure state a *mixture* or *mixed state* but just stick with the name *nonpure*.

4.3.2 Probabilities

Postulate #2 says that the average value of an observable represented by $\hat{Q}$ in the state corresponding to density operator $\hat{W}$ is given by

$$\langle\hat{Q}\rangle = Tr(\hat{W}\hat{Q}) \tag{4.48}$$

Consider a function $F(\hat{Q})$ of the operator $\hat{Q}$. Let

$h(q)\,dq$ = probability that the measured value of the observable

represented by $\hat{Q}$ lies between q and $q + dq$

Thus,

$$h(q) = \text{probability density function} \tag{4.49}$$

Now, the general definition of an average value says that

$$\langle F(\hat{Q})\rangle = \int_{-\infty}^{\infty} F(q')h(q')\,dq' \tag{4.50}$$

Postulate #2, however, says

$$\langle F(\hat{Q})\rangle = Tr(\hat{W}F(\hat{Q})) \tag{4.51}$$

Taken together, these two relations will allow us to extract the probability density function $h(q)$.

Case 1 : Discrete Spectrum

Let $\hat{Q}$ be a self-adjoint operator representing the observable Q. Assume that it has a pure discrete spectrum with eigenvectors $|q_n\rangle$ and corresponding eigenvalues q_n such that

$$\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle \tag{4.52}$$

We can write its spectral representation (express the operator in terms of its eigenvectors and eigenvalues) as

$$\hat{Q} = \sum_n q_n\,|q_n\rangle\,\langle q_n| \tag{4.53}$$

Now consider the function

$$F(Q) = \theta(q - Q) \tag{4.54}$$

where

$$\theta(q - Q) = \begin{cases} 1 & q > Q \\ 0 & q < Q \end{cases} \tag{4.55}$$

The expectation value of this function is then

$$\begin{aligned} \langle F(\hat{Q})\rangle = \langle\theta(q - \hat{Q})\rangle &= \int_{-\infty}^{\infty} F(q')h(q')\,dq' = \int_{-\infty}^{\infty} \theta(q - q')h(q')\,dq' \\ &= \int_{-\infty}^{q} h(q')\,dq' = Prob(Q < q|\hat{W}) \\ &= \text{probability that } Q < q \text{ given } \hat{W} \end{aligned} \tag{4.56}$$

However, we can also write

$$\begin{aligned}\langle\theta(q-\hat{Q})\rangle &= Tr(\hat{W}\theta(q-\hat{Q})) = Tr\left[\hat{W}\sum_n \theta(q-q_n)\,|q_n\rangle\langle q_n|\right] \\ &= \sum_m \langle q_m|\left(\hat{W}\sum_n \theta(q-q_n)\,|q_n\rangle\langle q_n|\right)|q_m\rangle \\ &= \sum_m\sum_n \langle q_m|\hat{W}|q_n\rangle\,\theta(q-q_n)\,\langle q_n\,|\,q_m\rangle \\ &= \sum_m\sum_n \langle q_m|\hat{W}|q_n\rangle\,\theta(q-q_n)\delta_{nm} = \sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q-q_n) \\ &= Prob(Q<q|\hat{W})\end{aligned} \tag{4.57}$$

Now we have

$$\begin{aligned}h(q) &= \frac{\partial}{\partial q}\int_{-\infty}^{q} h(q')\,dq' = \frac{\partial}{\partial q}Prob(Q<q|\hat{W}) \\ &= \frac{\partial}{\partial q}\sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q-q_n) = \sum_n \langle q_n|\hat{W}|q_n\rangle\,\delta(q-q_n)\end{aligned} \tag{4.58}$$

This implies that the probability density $h(q) = 0$ unless $q =$ an eigenvalue. In other words, the probability $= 0$ that a dynamical variable or observable will take on a value other than an eigenvalue of the corresponding operator. This makes it clear the postulate #2 is consistent with the last part of postulate #1, where we assumed this to be the case.

Now

$Prob(Q = q|\hat{W}) =$ probability that the observable represented by $\hat{Q}$ will have the discrete value q in the ensemble characterized by $\hat{W}$

We calculate this as follows:

$$\begin{aligned}Prob(Q=q|\hat{W}) &= \lim_{\epsilon\to 0}(Prob(Q<q+\epsilon|\hat{W}) - Prob(Q<q-\epsilon|\hat{W})) \\ &= \lim_{\epsilon\to 0}\left(\sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q+\epsilon-q_n) - \sum_n \langle q_n|\hat{W}|q_n\rangle\,\theta(q-\epsilon-q_n)\right) \\ &= \sum_n \langle q_n|\hat{W}|q_n\rangle \lim_{\epsilon\to 0}(\theta(q+\epsilon+q_n) - \theta(q+\epsilon-q_n)) \\ &= \sum_n \langle q_n|\hat{W}|q_n\rangle \lim_{\epsilon\to 0}(\delta(q-q_n)2\epsilon) = \sum_n \langle q_n|\hat{W}|q_n\rangle\,\delta_{q,q_n}\end{aligned} \tag{4.59}$$

If we now define the *projection operator*

$$\hat{P}(q) = \sum_n |q_n\rangle\langle q_n|\,\delta_{q,q_n} \tag{4.60}$$

which projects onto the subspace spanned by all the eigenvectors (degenerate) with eigenvalue $q = q_n$, we then have the general result that

$$\begin{aligned} Tr(\hat{W}\hat{P}(q)) &= \sum_m \langle q_m | \hat{W}\hat{P}(q) | q_m \rangle \\ &= \sum_m \langle q_m | \hat{W} \left(\sum_n |q_n\rangle \langle q_n| \, \delta_{q,q_n} \right) |q_m\rangle \\ &= \sum_{m,n} \langle q_m | \hat{W} | q_n \rangle \, \delta_{q,q_n} \langle q_n | q_m \rangle \\ &= \sum_{m,n} \langle q_m | \hat{W} | q_n \rangle \, \delta_{q,q_n} \delta_{nm} \\ &= \sum_n \langle q_n | \hat{W} | q_n \rangle \, \delta_{q,q_n} = Prob(Q = q|\hat{W}) \end{aligned} \tag{4.61}$$

If we have a pure state, then

$$\hat{W} = |\psi\rangle \langle\psi| \tag{4.62}$$

and if q_n is a non-degenerate eigenvalue,

$$\begin{aligned} Prob(Q = q_n|\hat{W}) &= Tr(\hat{W}\hat{P}(q_n)) = Tr((|\psi\rangle\langle\psi|)\hat{P}(q_n)) \\ &= \sum_m \langle q_m | (|\psi\rangle\langle\psi|)\hat{P}(q_n) | q_m \rangle \\ &= \sum_m \langle q_m | (|\psi\rangle\langle\psi|)(|q_n\rangle\langle q_n|) | q_m \rangle \\ &= \sum_m \langle q_m | \psi \rangle \langle \psi | q_n \rangle \langle q_n | q_m \rangle \\ &= \sum_m \langle q_m | \psi \rangle \langle \psi | q_n \rangle \delta_{nm} \\ &= \langle q_n | \psi \rangle \langle \psi | q_n \rangle = |\langle q_n | \psi \rangle|^2 \end{aligned} \tag{4.63}$$

We note the following alternate and useful form of this result. We have

$$\begin{aligned} Prob(Q = q_n|\hat{W}) &= |\langle q_n | \psi \rangle|^2 = \langle q_n | \psi \rangle \langle \psi | q_n \rangle \\ &= \langle q_n | \hat{W} | q_n \rangle = \sum_m \langle q_m | \hat{P}_{q_n} \hat{W} | q_m \rangle \\ &= Tr(\hat{P}_{q_n}\hat{W}) \end{aligned} \tag{4.64}$$

Thus, for a physical system that is characterized by a density operator $\hat{W} = |\psi\rangle\langle\psi|$ *or* represented by a pure state characterized by a state vector $|\psi\rangle$, the probability of measuring a particular eigenvalue q_n of an observable Q represented by the operator $\hat{Q}$ is given by

$$|\langle q_n | \psi \rangle|^2 \tag{4.65}$$

To see that this all makes sense relative to the standard definition of an average value we consider the quantity

$$\langle \alpha | \hat{Q} | \alpha \rangle = \text{ average value in the state } |\alpha\rangle \tag{4.66}$$

Now using the eigenvector/eigenvalue relation

$$\hat{Q}\,|q_k\rangle = q_k\,|q_k\rangle \tag{4.67}$$

which says that we can write

$$\hat{Q} = \sum_k q_k\,|q_k\rangle\,\langle q_k| \tag{4.68}$$

we get

$$\langle\alpha|\,\hat{Q}\,|\alpha\rangle = \sum_k q_k\,\langle\alpha\,|\,q_k\rangle\,\langle q_k\,|\,\alpha\rangle = \sum_k q_k|\,\langle q_k\,|\,\alpha\rangle\,|^2 \tag{4.69}$$

which says that the average value in the state $|\alpha\rangle$ is equal to the sum over all possible values (the eigenvalues) of the value(eigenvalue) times the probability of finding measuring that eigenvalue when in the state $|\alpha\rangle$, which is clearly given by

$$|\,\langle q_k\,|\,\alpha\rangle\,|^2 \tag{4.70}$$

This corresponds to the standard definition of the average value.

Thus, the connection between the mathematical object $|\,\langle q_k\,|\,\alpha\rangle\,|^2$ and the probability of measuring the eigenvalue q_n when in the state $|\alpha\rangle$ is now clear and *derivable* from the earlier postulates.

We do not need to assume this as another postulate as is done in many texts.

A Special Case: In general, any observable will exhibit a nonzero statistical dispersion in its measured value for most states. For the case of a discrete spectrum, however, it is possible for all of the probability to reside in a single value.

Suppose the state involved is an eigenvector. The observable represented by the operator $\hat{Q}$ takes on a unique value, say q_0 (a non-degenerate eigenvalue) with probability = 1 in this state.

This means that

$$Prob(Q = q_0|\hat{W}) = \sum_n \langle q_n|\,\hat{W}\,|q_n\rangle\,\delta_{q_0,q_n} = \langle q_0|\,\hat{W}\,|q_0\rangle = 1 \tag{4.71}$$

Now, any $\hat{W}$ satisfies

$$Tr(\hat{W}^2) \leq 1 \tag{4.72}$$

or

$$\begin{aligned} Tr(\hat{W}^2) &= \sum_n \langle q_n|\,\hat{W}^2\,|q_n\rangle = \sum_n \langle q_n|\,\hat{W}\hat{I}\hat{W}\,|q_n\rangle \\ &= \sum_n \sum_m \langle q_n|\,\hat{W}\,|q_m\rangle\,\langle q_m|\,\hat{W}\,|q_n\rangle \\ &= \sum_n \sum_m |\,\langle q_n|\,\hat{W}\,|q_m\rangle\,|^2 \\ &= \langle q_0|\,\hat{W}\,|q_0\rangle + \text{ rest of terms } = 1 + \text{ rest of terms } \leq 1 \end{aligned} \tag{4.73}$$

which says that

$$\langle q_n | \hat{W} | q_m \rangle = \delta_{n0}\delta_{m0} \tag{4.74}$$

or all the other diagonal and non-diagonal matrix elements of $\hat{W}$ must vanish.

Thus, the only state for which the observable takes on the value q_0 (non-degenerate) with probability = 1 is the pure state represented by the density operator $\hat{W} = |q_0\rangle \langle q_0|$.

This state, whether described by the density operator $\hat{W}$ or the state vector $|q_0\rangle$ is called an eigenstate of $\hat{Q}$. The observable has a definite value (probability = 1) in an eigenstate and *ONLY* in an eigenstate.

Case 2: Continuous Spectrum

Now we let $\hat{Q}$ be a self-adjoint operator with a purely continuous spectrum, so that we can write

$$\hat{Q} = \int q' |q'\rangle \langle q'| \, dq' \text{ where } \hat{Q} |q'\rangle = q' |q'\rangle \tag{4.75}$$

The eigenvectors are normalized using the Dirac delta-function, as we showed earlier

$$\langle q' | q'' \rangle = \delta(q' - q'') \tag{4.76}$$

and we let

$h(q)\, dq$ = probability that the measured value of the observable represented by $\hat{Q}$ lies between q and $q + dq$

Then, as in the discrete case, we have

$$\langle \theta(q - \hat{Q}) \rangle = \int_{-\infty}^{q} h(q')\, dq' = Prob(Q < q|\hat{W}) \tag{4.77}$$

which gives the probability that $Q < q$. We also have

$$\langle \theta(q - \hat{Q}) \rangle = Tr(\hat{W}\theta(q - \hat{Q})) \tag{4.78}$$

and using the standard expansion of an operator

$$\theta(q - \hat{Q}) = \int_{-\infty}^{\infty} \theta(q - q') |q'\rangle \langle q'| \, dq' \tag{4.79}$$

we get

$$\begin{aligned}
\theta(q-\hat{Q}) &= Tr\left(\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'|\,dq'\right) \qquad (4.80)\\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\left(\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'||q''\rangle\,dq'\right)\\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\langle q'\,|\,q''\rangle\,dq'\\
&= \int_{-\infty}^{\infty} dq''\,\langle q''|\,\hat{W}\int_{-\infty}^{\infty}\theta(q-q')\,|q'\rangle\,\delta(q'-q'')\,dq'\\
&= \int_{-\infty}^{\infty}\theta(q-q')\,\langle q'|\,\hat{W}\,|q'\rangle\,dq'\\
&= \int_{-\infty}^{q}\langle q'|\,\hat{W}\,|q'\rangle\,dq' = Prob(Q<q|\hat{W})
\end{aligned}$$

This says that the probability density for the observable represented by $\hat{Q}$ within the ensemble characterized by $\hat{W}$ is

$$h(q) = \frac{\partial}{\partial q}Prob(Q<q|\hat{W}) = \langle q|\,\hat{W}\,|q\rangle \qquad (4.81)$$

in general. For the special case of a pure state where $\hat{W} = |\psi\rangle\,\langle\psi|$ we have

$$h(q) = |\,\langle q\,|\,\psi\rangle\,|^2 \qquad (4.82)$$

Later on we shall find that this is equivalent to a probability statement for the famous Schrodinger equation wave function.

We now turn to the topic of quantum dynamics.

As we have seen :

States or density operators describe values or probabilities for measurable quantities at given times.

We want to find :

Equations of motion that determine how these values or probabilities of measurable quantities change with time.

It is clear from our discussion, that, in the theory of quantum mechanics, a state only needs to be able specify how to calculate expectations values of measurable quantities. Therefore, we will assume that

Equations of motion, at a minimum, only need to specify how expectation values change in time.

4.4 Quantum Pictures

There are three ways that are commonly used in quantum mechanics to make expectation values depend on time:

Schrodinger Picture

1. states are represented by ket vectors that depend on time, $|\psi(t)\rangle$
2. operators representing observables or measurable quantities are independent of time, $\hat{Q}$

We then get a time-dependent expectation value of the form

$$\langle \hat{Q}(t) \rangle = \langle \psi(t) | \hat{Q} | \psi(t) \rangle \tag{4.83}$$

Heisenberg Picture

1. operators representing observables or measurable quantities are dependent of time, $\hat{Q}(t)$
2. states are represented by ket vectors that do not depend on time, $|\psi\rangle$

We then get a time-dependent expectation value of the form

$$\langle \hat{Q}(t) \rangle = \langle \psi | \hat{Q}(t) | \psi \rangle \tag{4.84}$$

Interaction Picture

It is a mixture of the Schrodinger and Heisenberg pictures that is appropriate for a very important class of problems. We will discuss it later after presenting the Schrodinger and Heisenberg pictures.

All of these pictures must agree in the sense that they must all give the same time dependence for $\langle \hat{Q}(t) \rangle$. There is, after all, a unique real world out there!!!

We will discuss these *pictures* in terms of state vectors first, mixing in state operator aspects as we proceed.

We have been discussing the so-called *formal structure* of quantum mechanics. This structure is a fundamental part of the theory behind quantum mechanics, but it has very little physical content on its own.

We cannot solve any physical problem with the formalism as it stands. We must first develop connections or correspondence rules that tell us the specific operators that actually represent particular dynamical variables.

The fundamental dynamical variables that we will be working with are *position, linear momentum, angular momentum* and *energy*. All such quantities are related to space-time symmetry transformations. As we proceed with our discussions, we will introduce any needed aspects of symmetry transformations and discuss further aspects of the subject later.

4.5 Transformations of States and Observables The way it must be.....

Experimental evidence leads us to believe that the laws of nature are invariant under certain space-time symmetry transformations, including displacements in space and time, rotations and Lorentz boosts (relative velocity changes).

For each such symmetry, both the state vectors and the observables must have transformations, i.e.,

$$|\psi\rangle \to |\psi'\rangle \qquad , \qquad \hat{Q} \to \hat{Q}' \tag{4.85}$$

We will only use pure states for most of our discussions in this development since nothing new appears in more complex cases.

What must be preserved in these transformations?

1. If $\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle$, then $\hat{Q}'\,|q'_n\rangle = q_n\,|q'_n\rangle$
   Here we are assuming that the eigenvalues are unchanged since $\hat{Q}$ and $\hat{Q}'$ are the *SAME* observable represented in two frames of reference. Mathematically operators representing the same dynamical variable must have the *same spectrum*. For example, in both frames the *position* operator will have identical continuous spectra in the range $[-\infty, \infty]$ in each frame.
2. If $|\psi\rangle = \sum_n c_n\,|q_n\rangle$ where $\hat{Q}\,|q_n\rangle = q_n\,|q_n\rangle$, then $|\psi'\rangle = \sum_n c'_n\,|q'_n\rangle$ where $\hat{Q}'\,|q'_n\rangle = q_n\,|q'_n\rangle$. This actually follows from (1).

Now, equivalent events observed in each frame must have the same probability. If this were not true, then some event might occur more often in one frame than it does in another frame, which makes no physical sense.

This means that probabilities are equal or that

$$|c_n|^2 = |c'_n|^2 \text{ or } |\,\langle q_n\,|\,\psi\rangle\,|^2 = |\,\langle q'_n\,|\,\psi'\rangle\,|^2 \tag{4.86}$$

We now present the mathematical formalism that characterizes this type of transformation. But first, a digression to cover two new mathematical topics.

4.5.1 Antiunitary/Antilinear Operators

If, for any vector $|\psi\rangle$, an operator $\hat{T}$ satisfies

$$\hat{T}(|\psi\rangle + |\phi\rangle) = \hat{T}\,|\psi\rangle + \hat{T}\,|\phi\rangle \text{ and } \hat{T}c\,|\psi\rangle = c^*\hat{T}\,|\psi\rangle \tag{4.87}$$

then this type of operator is called an *antilinear* operator. For an antilinear operator $\hat{T}$ to have an inverse $\hat{T}^{-1}$ such that

$$\hat{T}^{-1}\hat{T} = \hat{I} = \hat{T}\hat{T}^{-1} \tag{4.88}$$

it is necessary and sufficient that for each vector $|\psi\rangle$ there is one and only one vector $|\phi\rangle$ such that $|\psi\rangle = \hat{T}|\phi\rangle$. This implies that $\hat{T}^{-1}$ is unique and antilinear.

If an antilinear operator $\hat{T}$ has an inverse $\hat{T}^{-1}$ and if $\|\hat{T}|\psi\rangle\| = \||\psi\rangle\|$ (preserves norm) for all $|\psi\rangle$, then $\hat{T}$ is *antiunitary*.

Assume $\hat{T}$ is antiunitary. Therefore, if

$$\left|\tilde{\phi}\right\rangle = \hat{T}|\phi\rangle \text{ and } \left|\tilde{\psi}\right\rangle = \hat{T}|\psi\rangle \tag{4.89}$$

then we have $\left\langle\tilde{\phi}\middle|\tilde{\psi}\right\rangle = \langle\phi|\psi\rangle^*$. Now if $\hat{T}$ is antilinear, then $\hat{T}^2$ is a linear operator and if $\hat{T}$ is antiunitary, then $\hat{T}^2$ is a unitary operator.

4.5.2 Wigner's Theorem

Any mapping of the vector space onto itself

$$|\psi\rangle \to |\psi\rangle' = \hat{U}|\psi\rangle \text{ and } |\phi\rangle \to |\phi\rangle' = \hat{U}|\phi\rangle \tag{4.90}$$

that preserves $|\langle\phi|\psi\rangle|$ can be implemented by an operator $\hat{U}$ that is unitary (linear) when $\langle\phi'|\psi'\rangle = \langle\phi|\psi\rangle$ or antiunitary(antilinear) when $\langle\phi'|\psi'\rangle = \langle\phi|\psi\rangle^*$.

We can show that all such transformation operators of interest in quantum mechanics are linear operators. For example, let $\hat{U}(\ell)$ describe a displacement through a distance ℓ. We know from experiment that this can be done as a series of two displacements of size $\ell/2$ and thus, we must have $\hat{U}(\ell) = \hat{U}(\ell/2)\hat{U}(\ell/2)$. Now the product of two antilinear operators is linear. Therefore, regardless of whether $\hat{U}(\ell/2)$ is linear or antilinear, $\hat{U}(\ell)$ is linear. There is nothing special, however, about the value ℓ, i.e., we must also have $\hat{U}(\ell/2) = \hat{U}(\ell/4)\hat{U}(\ell/4)$ which implies that $\hat{U}(\ell/2)$ is linear and so on.

Operators in quantum mechanics are continuous, which means that they cannot change discontinuously from linear to antilinear as a function of ℓ. This means that we need only consider *continuous* linear transformations in quantum mechanics.

Antilinear operators will appear later when we discuss *discrete* symmetries.

Now, if the transformation rule for state vectors is $|q_n'\rangle = \hat{U}|q_n\rangle$ where

$$\hat{Q}|q_n\rangle = q_n|q_n\rangle \text{ and } \hat{Q}'|q_n'\rangle = q_n|q_n'\rangle \tag{4.91}$$

then we must have

$$\hat{Q}'\hat{U}\,|q_n\rangle = q_n\hat{U}\,|q_n\rangle \text{ or } \hat{U}^{-1}\hat{Q}'\hat{U}\,|q_n\rangle = q_n\,|q_n\rangle \tag{4.92}$$

Therefore,

$$(\hat{Q} - \hat{U}^{-1}\hat{Q}'\hat{U})\,|q_n\rangle = (q_n - q_n)\,|q_n\rangle = 0 \tag{4.93}$$

for all $|q_n\rangle$. Since the set $\{|q_n\rangle\}$ is complete (the eigenvectors of a Hermitian operator), this result holds for any vector $|\psi\rangle$, which can be constructed from the set $\{|q_n\rangle\}$. Therefore, we must have

$$\hat{Q} - \hat{U}^{-1}\hat{Q}'\hat{U} = 0 \tag{4.94}$$

or

$$\hat{Q} \to \hat{Q}' = \hat{U}\hat{Q}'\hat{U}^{-1} \tag{4.95}$$

is the corresponding transformation rule for linear operators.

4.5.3 The Transformation Operator and its Generator

Let t be a continuous parameter. We consider a family of unitary operators $\hat{U}(t)$, with the properties

$$\hat{U}(0) = \hat{I} \text{ and } \hat{U}(t_1 + t_2) = \hat{U}(t_1)\hat{U}(t_2) \tag{4.96}$$

Transformations such as displacements, rotations and Lorentz boosts clearly satisfy these properties and so it make sense to require them in general.

Now we consider infinitesimal t. We can then write the infinitesimal version of the unitary transformation as

$$\hat{U}(t) = \hat{I} + \left.\frac{d\hat{U}}{dt}\right|_{t=0} t + O(t^2) \tag{4.97}$$

Now all unitary operators must satisfy the unitarity condition

$$\hat{U}\hat{U}^\dagger = \hat{I} \text{ for all } t \tag{4.98}$$

Therefore, we have to *first order* in t (an infinitesimal)

$$\begin{aligned}\hat{U}\hat{U}^\dagger = \hat{I} &= \left[\hat{I} + \left.\frac{d\hat{U}(t)}{dt}\right|_{t=0} t + \ldots\right]\left[\hat{I} + \left.\frac{d\hat{U}^\dagger(t)}{dt}\right|_{t=0} t + \ldots\right] \\ &= \hat{I} + \left.\left[\frac{d\hat{U}(t)}{dt} + \frac{d\hat{U}^\dagger(t)}{dt}\right]\right|_{t=0} t + \ldots\end{aligned} \tag{4.99}$$

which implies that

$$\left.\left[\frac{d\hat{U}(t)}{dt} + \frac{d\hat{U}^\dagger(t)}{dt}\right]\right|_{t=0} = 0 \tag{4.100}$$

If we let

$$\left.\frac{d\hat{U}(t)}{dt}\right|_{t=0} = -i\hat{H} \tag{4.101}$$

then the condition (4.99) becomes

$$-i\hat{H} + (i\hat{H})^\dagger = 0 \text{ or } \hat{H} = \hat{H}^\dagger \tag{4.102}$$

which says that $\hat{H}$ is a Hermitian operator. It is called the *generator* of the family of transformations $\hat{U}(t)$ because it determines these operators uniquely.

Now consider the property $\hat{U}(t_1 + t_2) = \hat{U}(t_1)\hat{U}(t_2)$. Taking the appropriate partial derivative we have

$$\left.\frac{\partial}{\partial t_1}\hat{U}(t_1 + t_2))\right|_{t_1=0} = \left.\left(\frac{d}{dt_1}\hat{U}(t_1)\right)\right|_{t_1=0} \hat{U}(t_2) \tag{4.103}$$

or

$$\left.\frac{d}{dt}\hat{U}(t))\right|_{t=t_2} = -i\hat{H}\hat{U}(t_2) \tag{4.104}$$

which can be written for arbitrary t as

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}\hat{U}(t) \tag{4.105}$$

If $\hat{H}$ is *not explicitly dependent on time*, then this equation is satisfied by the unique solution

$$\hat{U}(t) = e^{-i\hat{H}t} \tag{4.106}$$

Thus, the generator $\hat{H}$ of the infinitesimal transformation, determines the operator $\hat{U}(t) = e^{-i\hat{H}t}$ for a finite transformation.

This is just *Stone's theorem*, which we discussed earlier, but now derived in an alternative way.

We will now approach the study of various pictures using simple methods and then repeat the process using symmetry operations, which will clearly show the power of using the latter approach and give us a deeper understanding about what is happening.

4.6 The Schrodinger Picture

The Schrodinger picture follows directly from the previous discussion of the $\hat{U}(t)$ operator. Suppose we have some physical system that is represented by the state vector $|\psi(0)\rangle$ at time $t = 0$ and represented by the state vector $|\psi(t)\rangle$ at time t.

We ask this question. How are these state vectors related to each other?

We make the following assumptions:

1. *Every vector* $|\psi(0)\rangle$ such that $\langle\psi(0)\,|\,\psi(0)\rangle = \||\psi(0)\rangle\| = 1$ represents a *possible state* at time $t = 0$.

2. *Every vector* $|\psi(t)\rangle$ such that $\langle\psi(t)\,|\,\psi(t)\rangle = \||\psi(t)\rangle\| = 1$ represents a *possible state* at time $t = 0$.

3. *Every bounded Hermitian operator* represents an *observable* or *measurable* quantity.

4. The properties of the physical system determine the state vectors to *within a phase factor* since $\|e^{i\alpha}|\psi\rangle\| = \||\psi\rangle\|$.

5. $|\psi(t)\rangle$ is *determined* by $|\psi(0)\rangle$. Now if $|\psi(0)\rangle$ and $|\phi(0)\rangle$ represent two possible states at $t = 0$ and $|\psi(t)\rangle$ and $|\phi(t)\rangle$ represent the *corresponding* states at time t, then $|\langle\phi(0)\,|\,\psi(0)\rangle|^2$ equals the probability of finding the system in the state represented by $|\phi(0)\rangle$ given that the system is in the state $|\psi(0)\rangle$ at $t = 0$ and $|\langle\phi(t)\,|\,\psi(t)\rangle|^2$ equals the probability of finding the system in the state represented by $|\phi(t)\rangle$ given that the system is in the state $|\psi(t)\rangle$ at time t.

6. It makes physical sense to assume that these two probabilities should be the same
$$|\langle\phi(0)\,|\,\psi(0)\rangle|^2 = |\langle\phi(t)\,|\,\psi(t)\rangle|^2$$

Wigner's theorem then says that there exists a unitary, linear operator $\hat{U}(t)$ such that

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle \tag{4.107}$$

and an expression of the form

$$|\langle\alpha|\,\hat{U}(t)\,|\beta\rangle|^2 \tag{4.108}$$

gives the probability that the system is in state $|\alpha\rangle$ at time t given that it was in state $|\beta\rangle$ at time t = 0.

This clearly agrees with our earlier assumption of the existence of such an operator and strengthens our belief that item (6) above is a valid assumption.

We assume that this expression is a continuous function of t. As we have already showed (4.104), we then have $\hat{U}(t)$ satisfying the equation

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}\hat{U}(t) \tag{4.109}$$

or

$$\hat{U}(t) = e^{-i\hat{H}t} \tag{4.110}$$

and thus

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle = e^{-i\hat{H}t}\,|\psi(0)\rangle \tag{4.111}$$

which implies the following equation of motion for the state vector

$$i\frac{d\hat{U}(t)}{dt}\left|\psi(0)\right\rangle = \hat{H}\hat{U}(t)\left|\psi(0)\right\rangle \tag{4.112}$$

or

$$i\frac{d}{dt}\left|\psi(t)\right\rangle = \hat{H}\left|\psi(t)\right\rangle \tag{4.113}$$

which is the *abstract form* of the *famous Schrodinger equation*. We will derive the standard form of this equation later. The operator $\hat{U}(t) = e^{-i\hat{H}t}$ is called the *time evolution operator* for reasons that will become clear shortly.

Finally, we can write a time-dependent expectation value as

$$\left|\psi(t)\right\rangle = \hat{U}(t)\left|\psi(0)\right\rangle = e^{-i\hat{H}t}\left|\psi(0)\right\rangle \tag{4.114}$$

$$\langle\hat{Q}(t)\rangle = \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle \tag{4.115}$$

This is the *Schrodinger picture* where *state vectors change with time* and *operators are constant in time*.

As we saw in the discussion above, using the Schrodinger picture depends on a full knowledge of the Hamiltonian operator $\hat{H}$. However, in the Schrodinger picture, where we need to know $\hat{H}$ to solve the equation of motion for $|\psi(t)\rangle$, the equation of motion is such that we seem to need to know the complete solution for all time, $|\psi(t)\rangle$ to deduce $\hat{H}$. We are trapped in a *circle*. Put another way, the Schrodinger equation has no physical content unless we have an *independent* way to choose the Hamiltonian operator $\hat{H}$. Before deriving a way to choose $\hat{H}$ (we will use a symmetry approach), we will look at the other pictures.

We note that the Schrodinger picture is *not the same* as the Schrodinger equation. The Schrodinger equation involves a mathematical object called the *wave function* which is *one particular representation* of the state vector, namely the *position representation*, as we shall see later. Thus, the Schrodinger equation is *applicable* only to Hamiltonians that describe operators dependent on external degrees of freedom like position and momentum. The Schrodinger picture, on the other hand, works with *both internal and external degrees of freedom* and can handle a much wider class of physical systems, as we will see.

Digression: Alternative Approach to Unitary Translation Operators

We now consider *active translations* of the state vector in space-time. Since the length of a state vector cannot change (always normalized to 1) in the standard probability interpretation of quantum mechanics, active translations must be represented by *unitary* operators and correspond to rotations(no change of length) of the vectors in the Hilbert space.

We use the quantities

$$\psi_\alpha(\vec{r}) = \langle \vec{r} \,|\, \alpha \rangle \tag{4.116}$$
$$= \text{ probability amplitude for the state } |\alpha\rangle \text{ to be found in the state } |\vec{r}\rangle$$

for this discussion.

First, we consider *translations in space*. For the shifted amplitude we write

$$\psi_{\alpha'}(\vec{r}) = \psi_\alpha(\vec{r} - \vec{\rho}) = \hat{U}_r(\vec{\rho})\psi_\alpha(\vec{r}) \tag{4.117}$$

where

α, α' label the state vectors

$\vec{r}$ indicates spatial translations involved

$\vec{\rho}$ is the displacement vector

The relationship above follows from Figure 4.1 below (it gives the meaning of translation in space).

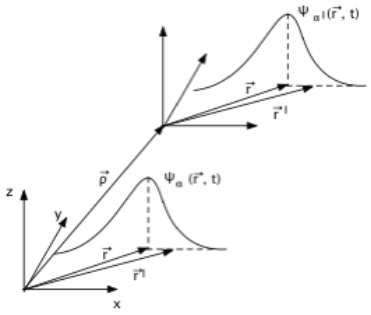


Figure 4.1: Space Translation

which implies that

$$\psi_{\alpha'}(\vec{r} + \rho\hat{x}, t) = \psi_\alpha(\vec{r}, t) \tag{4.118}$$

To determine the operator $\hat{U}_r(\vec{\rho})$ explicitly, we have oriented the translation vector $\vec{\rho}$ parallel to the x-axis. We get

$$\begin{aligned} \psi_\alpha(\vec{r} - \rho\hat{x}) &= \psi_\alpha(x - \rho, y, z) \\ &= \psi_\alpha(x, y, z) - \rho\frac{\partial}{\partial x}\psi_\alpha(x, y, z) + \frac{\rho^2}{2!}\frac{\partial^2}{\partial x^2}\psi_\alpha(x, y, z) - \dots \\ &= e^{-\rho\frac{\partial}{\partial x}}\psi_\alpha(x, y, z) \end{aligned} \tag{4.119}$$

where we have used a Taylor expansion and *defined* the exponential operator by

$$e^{-\rho\frac{\partial}{\partial x}} = 1 - \rho\frac{\partial}{\partial x} + \frac{\rho^2}{2!}\frac{\partial^2}{\partial x^2} - \dots \tag{4.120}$$

For a translation in an arbitrary direction $\vec{\rho}$, the generalization of this result is accomplished by the replacement

$$\rho \frac{\partial}{\partial x} \rightarrow \vec{\rho} \cdot \nabla \tag{4.121}$$

so that

$$\psi_\alpha(\vec{r} - \vec{\rho}) = e^{-\vec{\rho} \cdot \nabla} \psi_\alpha(\vec{r}) = e^{-i\vec{\rho} \cdot \hat{p}/\hbar} \psi_\alpha(\vec{r}) \tag{4.122}$$

where we have used

$$\hat{p} = -i\hbar\nabla \tag{4.123}$$

Later we shall find that $\hat{p}$ is the *linear momentum operator*. Thus, we find that the *spatial translation operator* is given by

$$\hat{U}_r(\vec{\rho}) = e^{-i\vec{\rho} \cdot \hat{p}/\hbar} \tag{4.124}$$

We will derive this result from first principles using symmetry arguments shortly.

Time Displacements in Quantum Mechanics

We now investigate the time displacement of a state function $\psi_\alpha(\vec{r}, t)$ by a time interval τ as shown in Figure 4.2 below.

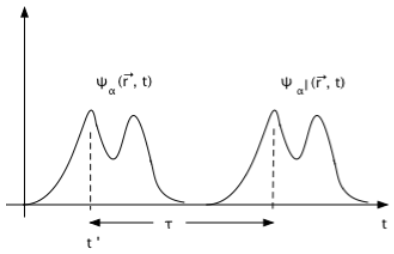


Figure 4.2: Time Translation

As before, we have

$$\psi_{\alpha'}(\vec{r}, t + \tau) = \psi_\alpha(\vec{r}, t) \tag{4.125}$$

We now represent this transformation with a unitary operator $\hat{U}_t(\tau)$ such that

$$\psi_{\alpha'}(\vec{r}, t) = \hat{U}_t(\tau)\psi_\alpha(\vec{r}, t) = \psi_\alpha(\vec{r}, t - \tau) \tag{4.126}$$

We again make a Taylor expansion to get

$$\begin{aligned} \hat{U}_t(\tau)\psi_\alpha(\vec{r}, t) &= \psi_\alpha(\vec{r}, t) + (-\tau)\frac{\partial}{\partial t}\psi_\alpha(\vec{r}, t) + \frac{(-\tau)^2}{2!}\frac{\partial^2}{\partial t^2}\psi_\alpha(\vec{r}, t) - \ldots \\ &= e^{-\tau \frac{\partial}{\partial t}} \psi_\alpha(\vec{r}, t) \end{aligned} \tag{4.127}$$

It follows that

$$\hat{U}_t(\tau) = e^{-\tau \frac{\partial}{\partial t}} = e^{i\tau \hat{E}/\hbar} = e^{i\tau \hat{H}/\hbar} \tag{4.128}$$

where we have used

$$\hat{E} = \hat{H} = i\hbar \frac{\partial}{\partial t} \tag{4.129}$$

Later we shall find that $\hat{E} = \hat{H}$ is the energy or Hamiltonian operator.

We find that the time evolution operator for state vectors(kets) is given as follows

$$|\vec{r}, t, \alpha'\rangle = e^{i\hat{H}\tau/\hbar} |\vec{r}, t, \alpha\rangle = |\vec{r}, t - \tau, \alpha\rangle \tag{4.130}$$

$$e^{i\hat{H}\tau/\hbar} |\vec{r}, t + \tau, \alpha\rangle = |\vec{r}, t, \alpha\rangle \tag{4.131}$$

$$|\vec{r}, t + \tau, \alpha\rangle = e^{-i\hat{H}\tau/\hbar} |\vec{r}, t, \alpha\rangle \tag{4.132}$$

$$|\vec{r}, \tau, \alpha\rangle = e^{-i\hat{H}\tau/\hbar} |\vec{r}, 0, \alpha\rangle \tag{4.133}$$

$$|\vec{r}, t, \alpha\rangle = e^{-i\hat{H}t/\hbar} |\vec{r}, 0, \alpha\rangle \tag{4.134}$$

A result identical to our earlier derivation (4.110).

4.7 The Heisenberg Picture

As we saw earlier, we can think of the expectation value in two different ways, namely,

$$|\psi(t)\rangle = \hat{U}(t) |\psi(0)\rangle = e^{-i\hat{H}t} |\psi(0)\rangle \tag{4.135}$$

$$\begin{aligned} \langle \hat{Q}(t) \rangle &= \langle \psi(t)| \hat{Q} |\psi(t)\rangle = \langle \psi(0)| \hat{U}^\dagger(t) \hat{Q} \hat{U}(t) |\psi(0)\rangle \\ &= \langle \psi(0)| \hat{Q}(t) |\psi(0)\rangle \end{aligned} \tag{4.136}$$

where

$$\hat{Q}(t) = \hat{U}(t)^\dagger \hat{Q} \hat{U}(t) = \hat{U}^\dagger(t) \hat{Q}(0) \hat{U}(t) = e^{i\hat{H}t} \hat{Q} e^{-i\hat{H}t} \tag{4.137}$$

This implies that in the Heisenberg picture operators change with time and states are constant in time.

Now $\hat{U}(t)$ is a unitary, linear operator. This means that the transformation preserves these properties:

1. $\hat{Q}(0)$ bounded $\rightarrow$ $\hat{Q}(t)$ bounded
2. $\hat{Q}(0)$ Hermitian $\rightarrow$ $\hat{Q}(t)$ Hermitian
3. $\hat{Q}(0)$ positive $\rightarrow$ $\hat{Q}(t)$ positive
4. $\hat{Q}(0)$ and $\hat{Q}(t)$ have the same spectrum

In addition, if the spectral decomposition of $\hat{Q}(0)$ give the projection operators $\hat{E}_x(0)$, then the spectral decomposition of $\hat{Q}(0)$ give the projection operators $\hat{E}_x(t) = e^{i\hat{H}t}\hat{E}_x(0)e^{-i\hat{H}t}$.

This follows from:

$$\begin{aligned}\langle\phi|\,\hat{Q}(t)\,|\psi\rangle &= \int_{-\infty}^{\infty} x\,d\,\langle\phi|\,\hat{E}_x(t)\,|\psi\rangle \\ &= \langle\phi|\,e^{i\hat{H}t}\hat{Q}(0)e^{-i\hat{H}t}\,|\psi\rangle = \langle\phi(t)|\,\hat{Q}(0)\,|\psi(t)\rangle \\ &= \int_{-\infty}^{\infty} x\,d\,\langle\phi(t)|\,\hat{E}_x(0)\,|\psi(t)\rangle = \int_{-\infty}^{\infty} x\,d\,\langle\phi|\,e^{i\hat{H}t}\hat{E}_x(0)e^{-i\hat{H}t}\,|\psi\rangle\end{aligned} \tag{4.138}$$

For a function F of $\hat{Q}(t)$ we have

$$\begin{aligned}\langle\phi|\,F(\hat{Q}(t))\,|\psi\rangle &= \int_{-\infty}^{\infty} F(x)\,d\,\langle\phi|\,\hat{E}_x(t)\,|\psi\rangle \\ &= \int_{-\infty}^{\infty} F(x)\,d\,\langle\phi|\,e^{i\hat{H}t}\hat{E}_x(0)e^{-i\hat{H}t}\,|\psi\rangle \\ &= \int_{-\infty}^{\infty} F(x)\,d\,\langle\phi(t)|\,\hat{E}_x(0)\,|\psi(t)\rangle \\ &= \langle\phi(t)|\,F(\hat{Q}(0))\,|\psi(t)\rangle = \langle\phi|\,e^{i\hat{H}t}F(\hat{Q}(0))e^{-i\hat{H}t}\,|\psi\rangle\end{aligned} \tag{4.139}$$

or

$$F(\hat{Q}(t)) = e^{i\hat{H}t}F(\hat{Q}(0))e^{-i\hat{H}t} \tag{4.140}$$

If $\{\hat{Q}_i(0)\}$ represents a complete set of mutually commuting Hermitian operators, then $\{\hat{Q}_i(t)\}$ is also a complete set of mutually commuting Hermitian operators. In addition,

$$F(\{\hat{Q}_i(t)\}) = e^{i\hat{H}t}F(\{\hat{Q}_i(0)\})e^{-i\hat{H}t} \tag{4.141}$$

and all algebraic relations between non-commuting operators are preserved in time.

Since $[f(\hat{Q},\hat{Q}] = 0$ we have

$$\hat{U}^\dagger(t)\hat{H}\hat{U}(t) = e^{i\hat{H}t}\hat{H}e^{-i\hat{H}t} = \hat{H} \tag{4.142}$$

which implies that $\hat{H}$, the Hamiltonian operator is a *constant of the motion (in time)*.

4.8 Interaction Picture

We have derived the equation

$$i\frac{d}{dt}\,|\psi(t)\rangle = \hat{H}(t)\,|\psi(t)\rangle \tag{4.143}$$

Now let us assume that $\hat{H}$ can be split up as follows:

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t) \tag{4.144}$$

where $\hat{H}_0$ is independent of t. This will be possible in many real physical systems. Note that the equation

$$i\frac{d\hat{U}(t)}{dt} = \hat{H}(t)\hat{U}(t) \tag{4.145}$$

does not have the simple solution

$$\hat{U}(t) = e^{-i\hat{H}t} \tag{4.146}$$

in this case.

We now define a new state vector by the relation

$$|\psi_I(t)\rangle = e^{i\hat{H}_0 t}\,|\psi(t)\rangle \tag{4.147}$$

Taking derivatives we get

$$\begin{aligned} i\frac{d}{dt}\,|\psi_I(t)\rangle &= i\frac{d}{dt}(e^{i\hat{H}_0 t}\,|\psi(t)\rangle) = -e^{i\hat{H}_0 t}\hat{H}_0\,|\psi_I(t)\rangle + e^{i\hat{H}_0 t}i\frac{d}{dt}\,|\psi(t)\rangle \\ &= -e^{i\hat{H}_0 t}\hat{H}_0\,|\psi_I(t)\rangle + e^{i\hat{H}_0 t}\hat{H}\,|\psi(t)\rangle = e^{i\hat{H}_0 t}\hat{V}\,|\psi(t)\rangle \\ &= e^{i\hat{H}_0 t}\hat{V}e^{-i\hat{H}_0 t}e^{i\hat{H}_0 t}\,|\psi(t)\rangle \end{aligned} \tag{4.148}$$

or

$$i\frac{d}{dt}\,|\psi_I(t)\rangle = \hat{V}_I(t)\,|\psi_I(t)\rangle \tag{4.149}$$

where we have defined

$$\hat{V}_I(t) = e^{i\hat{H}_0 t}\hat{V}e^{-i\hat{H}_0 t} \tag{4.150}$$

We then have

$$\begin{aligned} \langle\hat{Q}(t)\rangle &= \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle = \langle\psi_I(t)|\,e^{i\hat{H}_0 t}\hat{Q}e^{-i\hat{H}_0 t}\,|\psi_I(t)\rangle \\ &= \langle\psi_I(t)|\,\hat{Q}_I(t)\,|\psi_I(t)\rangle \end{aligned} \tag{4.151}$$

This says that in the interaction picture both the state vectors and the operators are dependent on time. Their time development, however, depends on different parts of $\hat{H}$. The state vector time development depends on $\hat{V}_I(t)$ and the operator time development depends on $\hat{H}_0$. It is, in some sense, intermediate between the Schrodinger and Heisenberg picture.

The question still remains, however, how do we find $\hat{H}$?

We now turn to a more general approach based on symmetries to get a handle on how to deal with this problem.

4.9 Symmetries of Space-Time

Since we are only considering non-relativistic quantum mechanics at this stage, we will restrict our attention to velocities that are small compared to the speed of light.

In this case, the set of all displacements in space-time, rotations in space and Galilean boosts(Lorentz boosts in the low velocity limit) can be represented by transformations where

$$\vec{x} \to \vec{x}' = \mathbb{R}\vec{x} + \vec{a} + \vec{v}t \quad , \quad t' = t + s \tag{4.152}$$

where $\mathbb{R}$ is a rotation of the real 3-vectors $\vec{x}$ in 3-dimensional space, $\vec{a}$ is a real 3-vector that specifies space translations, $\vec{v}$ is a real 3-vector that represents the velocity of a moving coordinate system and specifies the Galilean transformations, and the real number s specifies the time translation.

$\mathbb{R}$ can be thought of as a 3×3 matrix such that under a pure rotation

$$\vec{x}' = \mathbb{R}\vec{x} \text{ or } x'_j = \sum_i R_{ji} x_i \tag{4.153}$$

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{4.154}$$

For example, a rotation by angle θ about the x_3(or z)-axis is given by

$$\mathbb{R}_3(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.155}$$

or

$$\begin{aligned} x'_1 &= x_1 \cos\theta + x_2 \sin\theta \\ x'_2 &= -x_1 \sin\theta + x_2 \cos\theta \\ x'_3 &= x_3 \end{aligned} \tag{4.156}$$

which corresponds to Figure 6.3 below.

If we let T_1 and T_2 be two such transformations, then $T_1 T_2$ is the transformation corresponding to T_2 followed by T_1.

A set of transformations forms a *group* when:

1. The product of two transformation in the group is also a transformation in the group.
2. The product is associative $T_3(T_2 T_1) = (T_3 T_2 T_1)$.

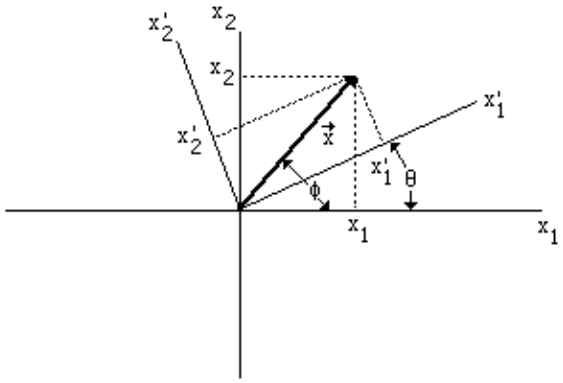


Figure 4.3: Rotation about z-axis

3. An identity transformation T_0 exists such that $\vec{x} \to \vec{x}$ and $t \to t$ or $T_0 T = T = T T_0$ for all transformations T.

4. An inverse transformation T^{-1} exists for every transformation such that $T^{-1}T = T_0 = TT^{-1}$.

A subset of transformations $T(\tau)$ depending on a real parameter τ is called a *one-parameter subgroup* if $T_0 =$ the identity and $T(\tau_1 + \tau_2) = T(\tau_1)T(\tau_2)$. Rotations about a *fixed axis*, as we saw above, form a one-parameter subgroup where the parameter is the angle of rotation.

Since products of the transformations from the three one-parameter subgroups, which correspond to rotations about the 1–, 2–, and 3–axes separately, include all possible rotations, there are three independent parameters(three angles) describing all rotations (remember the Euler angles from Classical Mechanics).

The total group of transformations has 10 parameters

1. rotations = 3 (3 angles)

2. space translations = 3 (3 components of $\vec{a}$)

3. Galilean boosts = 3 (3 components of $\vec{v}$)

4. time translation = 1 (parameter s)

Each of these ten parameters corresponds to a one-parameter subgroup. A ten-parameter group transformation is then defined by the product of ten one-parameter subgroup transformations.

This means that in our discussions we need only consider the properties of one-parameter subgroup transformations in order to understand general group transformations.

In our earlier discussion we have already covered part of this topic when we derived the time evolution operator $\hat{U}(t)$, which clearly corresponds to the time translation transformation.

Our earlier results tell us that for each of the ten transformations there exists a linear, unitary operator $\hat{U}(\tau)$ on the state vectors and observables such that

$$|\psi\rangle \to |\psi'\rangle = \hat{U}(\tau)\,|\psi\rangle \text{ and } \hat{Q} \to \hat{Q}' = \hat{U}(\tau)\hat{Q}\hat{U}^{-1}(\tau) \tag{4.157}$$

where $\hat{U}(\tau)$ takes the general form

$$\hat{U}(\tau) = e^{i\tau\hat{G}} \tag{4.158}$$

and $\hat{G}$ = a *Hermitian* operator = the *generator* of the transformation.

The time evolution operator we derived earlier is a good example:

if $|\psi(t)\rangle$ = state vector at time t, then for $t' = t + s$

$$|\psi(t')\rangle = e^{i\tau\hat{H}}\,|\psi(t)\rangle \tag{4.159}$$

4.10 Generators of the Group Transformations

In general we can write

$$\hat{U}(s_\mu) = \prod_{\mu=1}^{10} e^{is_\mu\hat{K}_\mu} \tag{4.160}$$

where the different s_μ represent the ten parameters defining the group transformation represented by the operator $\hat{U}(s_\mu)$ and

$$\hat{K}_\mu = \hat{K}_\mu^\dagger = \text{ the Hermitian generators (there are 10)} \tag{4.161}$$

Now, let all the parameters become infinitesimally small so that we get infinitesimal unitary operators of the form (expanding the exponentials)

$$\hat{U}(s_\mu) = \prod_{\mu=1}^{10} e^{is_\mu\hat{K}_\mu} = \prod_{\mu=1}^{10}(\hat{I} + is_\mu\hat{K}_\mu) \tag{4.162}$$

or to first order in the parameters s_μ,

$$\hat{U} = \hat{I} + i\sum_{\mu=1}^{10} s_\mu\hat{K}_\mu \tag{4.163}$$

Note that the inverse transformation corresponding to $e^{is_\mu\hat{K}_\mu}$ is $e^{-is_\mu\hat{K}_\mu}$.

We now construct the product transformation consisting of two infinitesimal transformations followed by their inverses to get

$$e^{i\epsilon\hat{K}_\mu}e^{i\epsilon\hat{K}_\nu}e^{-i\epsilon\hat{K}_\mu}e^{-i\epsilon\hat{K}_\nu} = \hat{I} + \epsilon^2[\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3) \tag{4.164}$$

where

$$[\hat{K}_\nu, \hat{K}_\mu] = \hat{K}_\nu \hat{K}_\mu - \hat{K}_\mu \hat{K}_\nu = \text{ commutator} \tag{4.165}$$

The algebraic steps involved are shown below.

$$\begin{aligned}
& e^{i\epsilon \hat{K}_\mu} e^{i\epsilon \hat{K}_\nu} e^{-i\epsilon \hat{K}_\mu} e^{-i\epsilon \hat{K}_\nu} \\
&\quad = \left(\hat{I} + i\epsilon \hat{K}_\mu - \frac{1}{2}\epsilon^2 \hat{K}_\mu^2\right)\left(\hat{I} + i\epsilon \hat{K}_\nu - \frac{1}{2}\epsilon^2 \hat{K}_\nu^2\right) \times \\
&\qquad \left(\hat{I} - i\epsilon \hat{K}_\mu - \frac{1}{2}\epsilon^2 \hat{K}_\mu^2\right)\left(\hat{I} - i\epsilon \hat{K}_\nu - \frac{1}{2}\epsilon^2 \hat{K}_\nu^2\right) \\
&\quad = \left(\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2 \hat{K}_\mu \hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right) \times \\
&\qquad \left(\hat{I} - i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2 \hat{K}_\mu \hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2)\right) \\
&\quad = (\hat{I} + i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2 \hat{K}_\mu \hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2) \\
&\qquad -i\epsilon(\hat{K}_\mu + \hat{K}_\nu) - \epsilon^2 \hat{K}_\mu \hat{K}_\nu - \frac{1}{2}\epsilon^2(\hat{K}_\mu^2 + \hat{K}_\nu^2) \\
&\qquad +(i\epsilon(\hat{K}_\mu + \hat{K}_\nu))(-i\epsilon(\hat{K}_\mu + \hat{K}_\nu)) \\
&\quad = \hat{I} - \epsilon^2 \hat{K}_\mu^2 - \epsilon^2 \hat{K}_\nu^2 - 2\epsilon^2 \hat{K}_\mu \hat{K}_\nu + \epsilon^2 \hat{K}_\mu^2 + \epsilon^2 \hat{K}_\nu^2 + \epsilon^2 \hat{K}_\mu \hat{K}_\nu + \epsilon^2 \hat{K}_\nu \hat{K}_\mu \\
&\quad = \hat{I} + \epsilon^2 \hat{K}_\nu \hat{K}_\mu - \epsilon^2 \hat{K}_\mu \hat{K}_\nu \\
&\quad = \hat{I} + \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] + O(\epsilon^3)
\end{aligned}$$

We have expanded to 2^{nd} order in ϵ in this derivation to explicitly see all the higher order terms cancel out. In general, this is not necessary.

But remember that these transformations are part of a group and therefore the product of the four transformations must also be a transformation $\hat{W}$ in the same group. Actually, it only needs to be a member of the group to within an arbitrary phase factor of the form $e^{i\alpha}$ (remember that the unitary transformations only need to preserve $|\langle\phi|\psi\rangle|$). We then have

$$e^{i\alpha}\hat{W} = \hat{I} + \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] \tag{4.166}$$

$$(1 + i\alpha)\left(\hat{I} + i\sum_\mu s_\mu \hat{K}_\mu\right) = \hat{I} + \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] \tag{4.167}$$

$$i\alpha\hat{I} + i\sum_\mu s_\mu \hat{K}_\mu = \epsilon^2 [\hat{K}_\nu, \hat{K}_\mu] \tag{4.168}$$

where we have used

$$\hat{W} = e^{i\sum_\mu s_\mu \hat{K}_\mu} \tag{4.169}$$

and expanded all the exponentials on the left-hand side of the equation to first order.

Therefore, the most general mathematical statement that we can make based on this result is that the commutator must take the form

$$[\hat{K}_\nu, \hat{K}_\mu] = i \sum_\lambda c^\lambda_{\mu\nu} \hat{K}_\mu + i b_{\mu\nu} \hat{I} \tag{4.170}$$

where the real numbers $c^\lambda_{\mu\nu}$ = the *structure constants* (of the group) and the term involving the identity operator just corresponds to the existence of an arbitrary phase factor. The structure factors and the b's are completely determined by the group algebra as we shall see.

By *convention*, we define the transformations as follows:

1. *Rotation* about the α-axis ($\alpha = 1, 2, 3$)

$$\vec{x} \to \mathbb{R}_\alpha(\theta_\alpha)\vec{x} \tag{4.171}$$

corresponds to the group operator

$$\hat{U} = e^{-i\theta_\alpha \hat{J}_\alpha} \tag{4.172}$$

where $\hat{J}_\alpha$ = generators ($\alpha = 1, 2, 3$)

2. *Displacement* along the α-axis ($\alpha = 1, 2, 3$)

$$x_\alpha \to x_\alpha + a_\alpha \tag{4.173}$$

corresponds to the group operator

$$\hat{U} = e^{-ia_\alpha \hat{P}_\alpha} \tag{4.174}$$

where $\hat{P}_\alpha$ = generators ($\alpha = 1, 2, 3$)

3. *Velocity boost* along the α-axis ($\alpha = 1, 2, 3$)

$$x_\alpha \to x_\alpha + v_\alpha t \tag{4.175}$$

corresponds to the group operator

$$\hat{U} = e^{iv_\alpha \hat{G}_\alpha} \tag{4.176}$$

where $\hat{G}_\alpha$ = generators ($\alpha = 1, 2, 3$)

4. *Time displacement*

$$t \to t + s \tag{4.177}$$

corresponds to the group operator

$$\hat{U} = e^{is\hat{H}} \tag{4.178}$$

where $\hat{H}$ = generator = Hamiltonian

4.11 Commutators and Identities

Initially, we will ignore the extra $\hat{I}$ term in the equation(4.169) below

$$[\hat{K}_\nu, \hat{K}_\mu] = i \sum_\lambda c^\lambda_{\mu\nu} \hat{K}_\mu + i b_{\mu\nu} \hat{I} \tag{4.179}$$

in our discussions and then include its effect(if any) later on.

We can determine some of the commutators using physical arguments as follows:

1. space displacements along different axes are independent of each other which implies that
$$[\hat{P}_\alpha, \hat{P}_\beta] = 0 \tag{4.180}$$

2. space displacements are independent of time displacements which implies that
$$[\hat{P}_\alpha, \hat{H}] = 0 \tag{4.181}$$

3. velocity boosts along different axes are independent of each other which implies that
$$[\hat{G}_\alpha, \hat{G}_\beta] = 0 \tag{4.182}$$

4. rotations are independent of time displacements which implies that
$$[\hat{J}_\alpha, \hat{H}] = 0 \tag{4.183}$$

5. space displacements and velocity boosts along a given axis are independent of rotations about that axis
$$[\hat{J}_\alpha, \hat{P}_\alpha] = 0 = [\hat{J}_\alpha, \hat{G}_\alpha] \tag{4.184}$$

6. obviously
$$[\hat{H}, \hat{H}] = 0 \tag{4.185}$$

This leaves us to consider these remaining unknown commutators:

$$[\hat{J}_\alpha, \hat{P}_\beta], [\hat{G}_\alpha, \hat{H}], [\hat{J}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{J}_\beta], [\hat{G}_\alpha, \hat{P}_\beta] \tag{4.186}$$

Let us consider $[\hat{G}_1, \hat{H}]$ first.

The general procedure is as follows. We write down a product of four operators consisting of the product of two operators representing a velocity boost in the 1-direction and a time translation and their inverses (as we did earlier).

Now these four successive transformation correspond to these changes of the coordinates:

$$\begin{aligned}
(x_1, x_2, x_3, t) &\to (x_1 - \epsilon t, x_2, x_3, t) && \text{Lorentz boost at time t} \\
&\to (x_1 - \epsilon t, x_2, x_3, t - \epsilon) && \text{Time translation - only affects t} \\
&\to (x_1 - \epsilon t + \epsilon(t - \epsilon), x_2, x_3, t - \epsilon) && \text{Lorentz boost at time t-}\epsilon \\
&\to (x_1 - \epsilon^2, x_2, x_3, t) && \text{Time translation - only affects t}
\end{aligned}$$

This last result just corresponds to a space displacement $-\epsilon^2$ along the 1-axis, so equating the product of four transformations to a space translation, we have the result

$$\begin{aligned}
& e^{i\epsilon\hat{H}} e^{i\epsilon\hat{G}_1} e^{-i\epsilon\hat{H}} e^{-i\epsilon\hat{G}_1} \\
&\qquad = (\hat{I} + i\epsilon\hat{H})(\hat{I} + i\epsilon\hat{G}_1)(\hat{I} - i\epsilon\hat{H})(\hat{I} - i\epsilon\hat{G}_1) \\
&\qquad = \hat{I} + \epsilon^2[\hat{G}_1, \hat{H}] + O(\epsilon^3) \\
&\qquad = e^{-i(-\epsilon^2)\hat{P}_1} = \hat{I} + i\epsilon^2\hat{P}_1
\end{aligned}$$

so that we find the result for the commutator

$$[\hat{G}_1, \hat{H}] = i\hat{P}_1 \tag{4.187}$$

In general, we get (using this same procedure)

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha \tag{4.188}$$

So we have determined one of the unknown commutators.

Now let us determine $[\hat{J}_\alpha, \hat{J}_\beta]$ using the same type of procedure. For a rotation we saw that

$$x_j \to \sum_k R_{jk} x_k \tag{4.189}$$

where

$$\mathbb{R}_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{4.190}$$

$$\mathbb{R}_2(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \tag{4.191}$$

$$\mathbb{R}_3(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.192}$$

For small θ we have

$$\mathbb{R}_\alpha(\theta) = \hat{I} - i\theta M_\alpha \tag{4.193}$$

where M_α is determined by expanding the exponential to 1^{st} order as

$$M_\alpha = i \left.\frac{d\mathbb{R}_\alpha}{d\theta}\right|_{\theta=0} \tag{4.194}$$

$$M_1(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix} \tag{4.195}$$

$$M_2(\theta) = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \tag{4.196}$$

$$M_3(\theta) = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.197}$$

Then we have

$$\begin{aligned} &\mathbb{R}_2(-\epsilon)\mathbb{R}_1(-\epsilon)\mathbb{R}_2(\epsilon)\mathbb{R}_1(\epsilon) \\ &\quad = \hat{I} + \epsilon^2[M_1, M_2] = \hat{I} + i\epsilon^2 M_3 \\ &\quad = \mathbb{R}_3(-\epsilon^2) \end{aligned} \tag{4.198}$$

or the product of four rotations is equivalent to a single rotation, which implies that

$$[\hat{J}_1, \hat{J}_2] = i\hat{J}_3 \tag{4.199}$$

and in general

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma \tag{4.200}$$

where we are using the Einstein summation convention for repeated indices and $\epsilon_{\alpha\beta\gamma}$ is the antisymmetric permutation symbol we introduced earlier with the properties

$$\epsilon_{\alpha\beta\gamma} = \begin{cases} 1 & \text{if } \alpha\beta\gamma \text{ is an even permutation of 123} \\ -1 & \text{if } \alpha\beta\gamma \text{ is an odd permutation of 123} \\ 0 & \text{if any two indices are the same} \end{cases} \tag{4.201}$$

Finally, we consider

$$\begin{aligned} &e^{i\epsilon\hat{G}_2}e^{i\epsilon\hat{J}_1}e^{-i\epsilon\hat{G}_2}e^{-i\epsilon\hat{J}_1} \\ &\quad = (\hat{I} + i\epsilon G_2)(\hat{I} + i\epsilon\hat{G}_1)(\hat{I} - i\epsilon G_2)(\hat{I} - i\epsilon\hat{G}_1) \\ &\quad = \hat{I} + \epsilon^2[\hat{J}_1, \hat{G}_2] + O(\epsilon^3) \end{aligned} \tag{4.202}$$

This involves a rotation by ϵ about the 1-axis and a velocity boost of ϵ along the 2-axis. This product transformation changes the coordinates as follows:

$$\begin{aligned} (x_1, x_2, x_3) &\to (x_1, x_2\cos\epsilon + x_3\sin\epsilon, -x_2\sin\epsilon + x_3\cos\epsilon) \\ &\to (x_1, x_2\cos\epsilon + x_3\sin\epsilon - \epsilon t, -x_2\sin\epsilon + x_3\cos\epsilon) \\ &\to (x_1, x_2 + \epsilon t\cos\epsilon, x_3 + \epsilon t sin\epsilon) \\ &\to (x_1, x_2, x_3 + \epsilon^2 t) \text{ to } 2^{nd} \text{ order in } \epsilon \end{aligned}$$

This is the same as

$$e^{i\epsilon^2\hat{G}_3} = \hat{I} + i\epsilon^2\hat{G}_3 \tag{4.203}$$

Thus, we have

$$[\hat{J}_1, \hat{G}_2] = i\hat{G}_3 \tag{4.204}$$

or in general,

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma \tag{4.205}$$

In a similar way,

$$[\hat{J}_\alpha, \hat{P}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{P}_\gamma \tag{4.206}$$

Summarizing so far we have

$$[\hat{P}_\alpha, \hat{P}_\beta] = 0\ ,\ [\hat{P}_\alpha, \hat{H}] = 0\ ,\ [\hat{G}_\alpha, \hat{G}_\beta] = 0\ ,\ [\hat{J}_\alpha, \hat{H}] = 0 \tag{4.207}$$

$$[\hat{J}_\alpha, \hat{P}_\alpha] = 0\ ,\ [\hat{J}_\alpha, \hat{G}_\alpha] = 0\ ,\ [\hat{H}, \hat{H}] = 0 \tag{4.208}$$

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha\ ,\ [\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma \tag{4.209}$$

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma\ ,\ [\hat{J}_\alpha, \hat{P}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{P}_\gamma \tag{4.210}$$

Before figuring out the last unknown commutator $[\hat{G}_\alpha, \hat{P}_\beta]$, we need to see whether the additional $\hat{I}$ term has an effect on any of the commutators we have already determined.

There are two relations, which are true for all commutators, that we will need to use.

Commutators are antisymmetric so that

$$[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}] \tag{4.211}$$

and they satisfy *Jacobi's identity* which is

$$[[\hat{A}, \hat{B}], \hat{C}] = [[\hat{C}, \hat{B}], \hat{A}] + [[\hat{A}, \hat{C}], \hat{B}] \tag{4.212}$$

These identities, as we shall see, limit the possible multiples of $\hat{I}$ that can be present. So we now assume that each commutator has an additional multiple of $\hat{I}$.

Thus, using $[\hat{J}_1, \hat{P}_2] = i\hat{P}_3 - c\hat{I}$, where we have added a multiple of $\hat{I}$. We then have

$$\begin{aligned} i[\hat{P}_3, \hat{P}_1] &= [[\hat{J}_1, \hat{P}_2], \hat{P}_1] + [c\hat{I}, \hat{P}_1] = [[\hat{J}_1, \hat{P}_2], \hat{P}_1] \\ &= [[\hat{P}_1, \hat{P}_2], \hat{J}_1] + [[\hat{J}_1, \hat{P}_1], \hat{P}_2] \end{aligned}$$

Now, $[\hat{P}_1, \hat{P}_2] = 0$ and $[\hat{J}_1, \hat{P}_1] = 0$, therefore $[\hat{P}_3, \hat{P}_1] = 0$ and no additional multiple of $\hat{I}$ is needed in this commutator. Everything is consistent without it!

In a similar manner we can show that

$$[\hat{P}_\alpha, \hat{P}_\beta] = 0 \ , \ [\hat{P}_\alpha, \hat{H}] = 0 \ , \ [\hat{G}_\alpha, \hat{G}_\beta] = 0 \ , \ [\hat{J}_\alpha, \hat{H}] = 0 \tag{4.213}$$

so that no additional multiple of $\hat{I}$ is needed in any of these commutators.

Since $[\hat{J}_\alpha, \hat{J}_\beta] = -[\hat{J}_\beta, \hat{J}_\alpha]$, if the commutator $[\hat{J}_\alpha, \hat{J}_\beta]$ is going to contain an extra multiple of $\hat{I}$, then the constant must be antisymmetric also. Therefore we must have

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma + i\epsilon_{\alpha\beta\gamma}b_\gamma\hat{I} \tag{4.214}$$

If we redefine $\hat{J}_\alpha \to \hat{J}_\alpha + b_\alpha\hat{I}$, then we get the original commutator back

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma \tag{4.215}$$

This change of definition implies that the transformation operator becomes

$$\hat{U}_\alpha(\theta) = e^{-i\theta\hat{J}_\alpha} \to e^{-i\theta\hat{J}_\alpha}e^{-i\theta b_\alpha} \tag{4.216}$$

and thus $|\psi'\rangle = \hat{U}\,|\psi\rangle$ changes to $e^{i\theta b_\alpha}\,|\psi'\rangle = \hat{U}\,|\psi\rangle$. Since overall phase factors for the state vector do not change any physics we can ignore the extra $\hat{I}$ terms in this case. They do not change any real physics content!

In a similar manner, we show that

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha \ , \ [\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma \ , \ [\hat{J}_\alpha, \hat{P}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{P}_\gamma \tag{4.217}$$

so that no additional multiple of $\hat{I}$ is needed in any of these commutators.

Finally, we are left to consider the unknown commutator $[\hat{G}_\alpha, \hat{P}_\beta]$. Now using the fact that $[c\hat{I}, \hat{P}_1] = 0$ we have

$$\begin{aligned} i[\hat{G}_3, \hat{P}_1] &= [[\hat{J}_1, \hat{G}_2], \hat{P}_1] + [c\hat{I}, \hat{P}_1] = [[\hat{J}_1, \hat{G}_2], \hat{P}_1] \\ &= [[\hat{P}_1, \hat{G}_2], \hat{J}_1] + [[\hat{J}_1, \hat{P}_1], \hat{G}_2] = [[\hat{P}_1, \hat{G}_2], \hat{J}_1] \end{aligned} \tag{4.218}$$

which has the solution

$$[\hat{G}_\alpha, \hat{P}_\beta] = 0 \quad \alpha \neq \beta \tag{4.219}$$

In addition, we have

$$\begin{aligned} i[\hat{G}_3, \hat{P}_3] &= [[\hat{J}_1, \hat{G}_2], \hat{P}_3] + [c\hat{I}, \hat{P}_3] = [[\hat{J}_1, \hat{G}_2], \hat{P}_3] \\ &= [[\hat{P}_3, \hat{G}_2], \hat{J}_1] + [[\hat{J}_1, \hat{P}_3], \hat{G}_2] \\ &= -i[\hat{P}_2, \hat{G}_2] \end{aligned} \tag{4.220}$$

or

$$[\hat{G}_3, \hat{P}_3] = [\hat{G}_2, \hat{P}_2] \tag{4.221}$$

and, in general

$$[\hat{G}_\alpha, \hat{P}_\alpha] = [\hat{G}_\beta, \hat{P}_\beta] \tag{4.222}$$

The only way to satisfy all of these commutators is to have the result

$$[\hat{G}_\alpha, \hat{P}_\beta] = \delta_{\alpha\beta} M \hat{I} \tag{4.223}$$

The value of M is undetermined. It cannot be eliminated by including multiples of $\hat{I}$ in any of the other commutators. It must have some real physical significance and we will identify it shortly.

4.12 Identification of Operators with Observables

Operators and Observables We now use the dynamics of a free particle, which is a physical system that is invariant under the Galilei group of space-time transformations, to identify the operators representing the dynamical variables or observables in that case. This section follows and expands on the work of Jordan(1975).

We assume that there exists a *position operator* (as we discussed earlier)

$$\hat{\mathbf{Q}} = (\hat{Q}_1, \hat{Q}_2, \hat{Q}_3) \tag{4.224}$$

(**boldface** = multi-component or vector operator) where

$$\hat{Q}_\alpha |\vec{x}\rangle = x_\alpha |\vec{x}\rangle \quad , \quad \alpha = 1, 2, 3 \tag{4.225}$$

The position operator has an unbounded ($-\infty < x_\alpha < \infty$), continuous spectrum. The three operators $\hat{Q}_\alpha$ have a common set of eigenvectors and, thus, they commute

$$[\hat{Q}_\alpha, \hat{Q}_\beta] = 0 \tag{4.226}$$

We now assume that there also exists a *velocity operator*

$$\hat{\mathbf{V}} = (\hat{V}_1, \hat{V}_2, \hat{V}_3) \tag{4.227}$$

such that we can make the following statement about expectation values (*this is an assumption*)

$$\frac{d}{dt}\langle \hat{\mathbf{Q}} \rangle = \langle \hat{\mathbf{V}} \rangle \tag{4.228}$$

for any state vector.

It is important to note that each treatment of quantum mechanics must make some assumption at this point. Although, they might outwardly look like different assumptions, they clearly must be equivalent since they result in the same theory with the same predictions.

For a pure state $|\psi(t)\rangle$ we would then have

$$\begin{aligned} \langle \psi(t)| \hat{\mathbf{V}} |\psi(t)\rangle &= \frac{d}{dt}\left(\langle \psi(t)| \hat{\mathbf{Q}} |\psi(t)\rangle \right) \\ &= \left[\frac{d}{dt} \langle \psi(t)| \right] \hat{\mathbf{Q}} |\psi(t)\rangle + \langle \psi(t)| \hat{\mathbf{Q}} \left[\frac{d}{dt} |\psi(t)\rangle \right] \end{aligned} \tag{4.229}$$

Now a time displacement corresponds to $t \to t' = t + s$. To use this transformation, we first have to figure out a rule for the transformation of the ket vector argument. When we represent the abstract vector $|\psi\rangle$ as a function of space-time we must be very careful about defining its properties. We derived this result earlier, but it is so important that we do it again here.

Given $|\psi\rangle$, we have seen that

$$\psi(\vec{x}, t) = \langle \vec{x}, t \,|\, \psi \rangle \tag{4.230}$$

Operating on the abstract vector with the time translation operator $e^{is\hat{H}} |\psi\rangle$ then gives

$$\langle \vec{x}, t | \, e^{is\hat{H}} \, |\psi\rangle \tag{4.231}$$

Now

$$e^{-is\hat{H}} \, |\vec{x}, t\rangle = |\vec{x}, t - s\rangle \tag{4.232}$$

and therefore we get

$$\langle \vec{x}, t | \, e^{is\hat{H}} \, |\psi\rangle = \langle \vec{x}, t - s \,|\, \psi \rangle = \psi(\vec{x}, t - s) \tag{4.233}$$

or

$$|\psi(t)\rangle \to e^{is\hat{H}} \, |\psi(t)\rangle = |\psi(t - s)\rangle \tag{4.234}$$

Here we have taken the so-called *active point of view* that the state is translated relative to a fixed coordinate system.

Now, we let $s = t$ to get

$$|\psi(t)\rangle \to e^{it\hat{H}} \, |\psi(t)\rangle = |\psi(0)\rangle \tag{4.235}$$

or

$$|\psi(t)\rangle = e^{-it\hat{H}} \, |\psi(0)\rangle \tag{4.236}$$

and

$$\frac{d}{dt} \, |\psi(t)\rangle = -i\hat{H} \, |\psi(t)\rangle \tag{4.237}$$

as we had already found in an earlier discussion. Using this result in (4.228) we then find

$$\begin{aligned} \langle \psi(t) | \, \hat{\mathbf{V}} \, |\psi(t)\rangle &= i \, \langle \psi(t) | \, \hat{H}\hat{\mathbf{Q}} \, |\psi(t)\rangle - i \, \langle \psi(t) | \, \hat{\mathbf{Q}}\hat{H} \, |\psi(t)\rangle \\ &= i \, \langle \psi(t) | \, [\hat{H}, \hat{\mathbf{Q}}] \, |\psi(t)\rangle \end{aligned} \tag{4.238}$$

which says that

$$\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}] \tag{4.239}$$

is a *valid* velocity operator for a free particle.

However, since we *do not know* $\hat{H}$, we are *still stuck at the starting line.*

Now a space displacement $\vec{x} \to \vec{x}' = \vec{x} + \vec{a}$ corresponds to

$$|\vec{x}\rangle \to |\vec{x}'\rangle = e^{-i\sum_\alpha a_\alpha \hat{P}_\alpha} |\vec{x}\rangle = |\vec{x} + \vec{a}\rangle \tag{4.240}$$

Again, in this active point of view, the state is displaced relative to a fixed coordinate system.

We then have, from our earlier discussion

$$\hat{\mathbf{Q}} \to \hat{\mathbf{Q}}' = e^{-i\sum_\alpha a_\alpha \hat{P}_\alpha} \hat{\mathbf{Q}} e^{i\sum_\alpha a_\alpha \hat{P}_\alpha} \tag{4.241}$$

where

$$\hat{Q}'_\alpha |\vec{x}\rangle' = x_\alpha |\vec{x}\rangle' \text{ or } \hat{Q}'_\alpha |\vec{x} + \vec{a}\rangle = x_\alpha |\vec{x} + \vec{a}\rangle \tag{4.242}$$

Since

$$\hat{Q}_\alpha |\vec{x}\rangle = x_\alpha |\vec{x}\rangle \text{ or } \hat{Q}_\alpha |\vec{x} + \vec{a}\rangle = (x_\alpha + a_\alpha) |\vec{x} + \vec{a}\rangle \tag{4.243}$$

we must have

$$(\hat{Q}_\alpha - a_\alpha \hat{I}) |\vec{x} + \vec{a}\rangle = x_\alpha |\vec{x} + \vec{a}\rangle \tag{4.244}$$

which says that

$$\hat{Q}'_\alpha = \hat{Q}_\alpha - a_\alpha \hat{I} \text{ or } \hat{\mathbf{Q}}' = \hat{\mathbf{Q}} - \vec{a}\hat{I} \tag{4.245}$$

Now we need to work out how to use an operator of the form

$$e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} \tag{4.246}$$

There are two ways to do this. The first way recognizes that it is the solution of the 1^{st}-order linear differential equation

$$\begin{aligned} i\frac{d}{du} e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} &= -e^{iu\hat{A}} \hat{A} \hat{B} e^{-iu\hat{A}} + e^{iu\hat{A}} \hat{B} \hat{A} e^{-iu\hat{A}} \\ &= e^{iu\hat{A}} [\hat{A}, \hat{B}] e^{-iu\hat{A}} = [e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}}, \hat{A}] \end{aligned} \tag{4.247}$$

with boundary condition

$$e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} = \hat{B} \text{ at } u = 0 \tag{4.248}$$

A second way is to expand the exponentials in power series and regroup the terms. In both cases, we get

$$e^{iu\hat{A}} \hat{B} e^{-iu\hat{A}} = \hat{B} - iu[\hat{B}, \hat{A}] - \frac{u^2}{2!}[[\hat{B}, \hat{A}], \hat{A}] - \ldots \tag{4.249}$$

Using this last relation we can find useful results like

$$\begin{aligned} e^{iu\hat{J}_1}\hat{J}_3 e^{-iu\hat{J}_1} &= \hat{J}_3 - iu[\hat{J}_3,\hat{J}_1] - \frac{u^2}{2!}[[\hat{J}_3,\hat{J}_1],\hat{J}_1] - \dots \\ &= \hat{J}_3 - iu(i\hat{J}_2) - \frac{u^2}{2!}[i\hat{J}_2,\hat{J}_1] - \dots \\ &= \hat{J}_3 + u\hat{J}_2 - \frac{u^2}{2!}\hat{J}_3 - \frac{u^3}{3!}\hat{J}_2 + \frac{u^4}{4!}\hat{J}_3 - \dots \\ &= \hat{J}_3\left(1 - \frac{u^2}{2!} + \frac{u^4}{4!} - \dots\right) + \hat{J}_2\left(u - \frac{u^3}{3!} + \frac{u^5}{5!} - \dots\right) \\ &= \hat{J}_3\cos u + \hat{J}_2\sin u \end{aligned} \tag{4.250}$$

Now we finally arrive at the result we are looking for

$$\begin{aligned} \hat{Q}'_\alpha = \hat{Q}_\alpha - a_\alpha\hat{I} &= e^{-i\sum_\beta a_\beta\hat{P}_\beta}\hat{Q}_\alpha e^{i\sum_\beta a_\beta\hat{P}_\beta} \\ &= \hat{Q}_\alpha + i\sum_\beta a_\beta[\hat{Q}_\alpha,\hat{P}_\beta] + \dots \end{aligned} \tag{4.251}$$

which implies that

$$[\hat{Q}_\alpha,\hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I} \tag{4.252}$$

This solution, not only satisfies the equation to 1^{st}-order as above, but implies that all higher order terms are explicitly equal to zero. Thus, the solution is exact.

This equation is one of the most important results in the theory of quantum mechanics.

Now we continue the derivation of that last elusive commutator.

A rotation through an infinitesimal angle θ about an axis along the unit vector $\hat{\mathbf{n}}$ has the effect

$$\vec{x} \to \vec{x}' = \vec{x} + \theta\hat{\mathbf{n}}\times\vec{x} \tag{4.253}$$

The corresponding transformation of the position eigenvectors is

$$|\vec{x}\rangle \to |\vec{x}'\rangle = e^{-i\theta\hat{\mathbf{n}}\cdot\hat{\mathbf{J}}}|\vec{x}\rangle \tag{4.254}$$

and for the position operator

$$\begin{aligned} \hat{Q}_\alpha \to \hat{Q}'_\alpha &= e^{-i\theta\hat{\mathbf{n}}\cdot\hat{\mathbf{J}}}\hat{Q}_\alpha e^{i\theta\hat{\mathbf{n}}\cdot\hat{\mathbf{J}}} \\ &= \hat{Q}_\alpha - i\theta[\hat{\mathbf{n}}\cdot\hat{\mathbf{J}},\hat{Q}_\alpha] + O(\theta^2) \end{aligned} \tag{4.255}$$

We note that $|\vec{x}'\rangle$ = an eigenvector of $\hat{\mathbf{Q}}'$ and $|\vec{x}\rangle$ = an eigenvector of $\hat{\mathbf{Q}}$ (different eigenvalues), but they are the *same* vector! (think about rotations of the axes).

Now as before (4.241) and (4.242),

$$\hat{Q}'_\alpha |\vec{x}\rangle' = x_\alpha |\vec{x}\rangle' \tag{4.256}$$

$$\begin{aligned} \hat{Q}_\alpha |\vec{x}'\rangle = x'_\alpha |\vec{x}'\rangle &= (\vec{x} + \theta\hat{\mathbf{n}} \times \vec{x})_\alpha |\vec{x}'\rangle \\ &= (\hat{\mathbf{Q}}' + \theta\hat{\mathbf{n}} \times \hat{\mathbf{Q}}')_\alpha |\vec{x}\rangle' \end{aligned} \tag{4.257}$$

Now since the vectors $|\vec{x}'\rangle = |\vec{x}\rangle'$ are a complete set, we must have

$$\hat{\mathbf{Q}} = \hat{\mathbf{Q}}' + \theta\hat{\mathbf{n}} \times \hat{\mathbf{Q}}' \tag{4.258}$$

or inverting this expression to 1^{st}-order we have

$$\hat{\mathbf{Q}}' = \hat{\mathbf{Q}} - \theta\hat{\mathbf{n}} \times \hat{\mathbf{Q}} \tag{4.259}$$

Therefore, we find

$$[\hat{\mathbf{n}} \cdot \hat{\mathbf{J}}, \hat{Q}_\alpha] = -i\hat{\mathbf{n}} \times \hat{\mathbf{Q}} \tag{4.260}$$

For an arbitrary unit vector $\hat{\mathbf{u}}$ this says that

$$[\hat{\mathbf{n}} \cdot \hat{\mathbf{J}}, \hat{\mathbf{u}} \cdot \hat{Q}_\alpha] = -i\hat{\mathbf{u}} \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{Q}}) = i(\hat{\mathbf{n}} \times \hat{\mathbf{u}}) \cdot \hat{\mathbf{Q}} \tag{4.261}$$

or

$$[\hat{J}_\alpha, \hat{Q}_\beta] = i(\hat{e}_\alpha \times \hat{e}_\beta) \cdot \hat{\mathbf{Q}} = i(\epsilon_{\alpha\beta\gamma}\hat{e}_\gamma) \cdot \hat{\mathbf{Q}} = i\epsilon_{\alpha\beta\gamma}\hat{Q}_\gamma \tag{4.262}$$

We note that this result is not only true for the components of the position operator, but it is true for the components of *any* vector operator $\hat{\mathbf{A}}$.

$$[\hat{J}_\alpha, \hat{A}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{A}_\gamma \tag{4.263}$$

In a similar manner, since $\hat{\mathbf{G}}$ generates a *displacement* in velocity space and we have

$$\hat{\mathbf{V}}' = \hat{\mathbf{V}} - \vec{v}\hat{I} = e^{i\vec{v}\cdot\hat{\mathbf{G}}}\hat{\mathbf{V}}e^{-i\vec{v}\cdot\hat{\mathbf{G}}} \tag{4.264}$$

This is the same way the $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$ operators behaved earlier. Now the unitary operator $\hat{U}(\vec{v}) = e^{i\vec{v}\cdot\hat{\mathbf{G}}}$ describes the instantaneous ($t = 0$) effect of a transformation to a frame of reference moving at velocity $\vec{v}$ with respect to the original frame. It affects the $\hat{\mathbf{V}}$ operator as in (4.263). Due to its *instantaneous* nature we must also have

$$\hat{U}\hat{\mathbf{Q}}\hat{U}^{-1} = \hat{\mathbf{Q}} \text{ or } [\hat{G}_\alpha, \hat{Q}_\beta] = 0 \tag{4.265}$$

Now, $\hat{\mathbf{Q}}$, the position operator, is clearly identified with an observable of the physical state. After much work we have determined the commutators of $\hat{\mathbf{Q}}$ with all the symmetry generators of the Galilei group. We now have

$$[\hat{G}_\alpha, \hat{Q}_\beta] = 0 \quad , \quad [\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I} \quad , \quad [\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I} \tag{4.266}$$

A *possible solution* of these equations is

$$\hat{G}_\alpha = M\hat{Q}_\alpha \tag{4.267}$$

We cannot show this is a unique solution.

We note, however, at this point that it is certainly true (using the commutators in (4.265)) that

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0 \tag{4.268}$$

We must now turn to special cases in order to learn more about the physical meaning of the generators.

Before discussing the special cases, we must again digress to review some mathematics discussed earlier and also cover some new mathematical ideas that we will need.

A subspace M *reduces* a linear operator $\hat{A}$ if $\hat{A}|\psi\rangle$ is in M for every $|\psi\rangle$ in M and $\hat{A}|\phi\rangle$ is in $M^\perp$ for every $|\phi\rangle$ in $M^\perp$.

A set of operators is *reducible* if there is a subspace, other than the whole space or the subspace containing only the null vector, which *reduces* every operator in the set. Otherwise, we say that the set is *irreducible*.

A subspace M is *invariant* under a set of operators if $\hat{A}|\psi\rangle$ is in M for every operator in the set and every vector in M.

Thus, a subspace M reduces a set of operators *if and only if* M and $M^\perp$ are invariant under the set of operators.

A set of operators is *symmetric* if $\hat{A}^\dagger$ is in the set for every operator $\hat{A}$ in the set.

If a subspace is invariant under a symmetric set of operators, then it reduces the set of operators.

Schur's Lemma

A symmetric set of bounded or Hermitian operators is *irreducible* if and only if multiples of $\hat{I}$ are the *only* bounded operators which commute with all operators in the set.

Example 0: The commutator $[\hat{Q}_\alpha, \hat{Q}_\beta] = 0$ says that the set of operators $\{\hat{Q}_\alpha\ ,\ \alpha = 1, 2, 3\}$ is a complete set of commuting operators. Since $[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I}$ any function of $\hat{Q}_\alpha$ that commutes with the $\hat{P}_\alpha$ must be a multiple of $\hat{I}$. Therefore, the set $\{\hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \hat{P}_1, \hat{P}_2, \hat{P}_3\}$ is irreducible.

In other words, if an operator commutes with the $\hat{Q}_\alpha$, then it is not a function of $\hat{P}_\alpha$ since $[\hat{Q}_\alpha, \hat{P}_\beta] \neq 0$. If an operator commutes with the $\hat{P}_\alpha$, then it is not

a function of $\hat{Q}_\alpha$, for the same reason. If an operator is independent of the set $\{\hat{Q}_\alpha, \hat{P}_\beta\}$ and there are no internal(not dependent on $\{\hat{Q}_\alpha, \hat{P}_\beta\}$) degrees of freedom, then it must be a multiple of $\hat{I}$.

Example 1 : Free Particle - no internal degrees of freedom

Now, as we stated earlier (4.267) that

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0 \tag{4.269}$$

and therefore

$$\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = \text{ multiple of } \hat{I} \tag{4.270}$$

or

$$\hat{G}_\alpha = M\hat{Q}_\alpha + c_\alpha \hat{I} \tag{4.271}$$

But $\hat{G}_\alpha$ is a component of a vector operator and therefore it must satisfy

$$[\hat{J}_\alpha, \hat{G}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{G}_\gamma \tag{4.272}$$

The first term $M\hat{Q}_\alpha$ satisfies this relation, but the $c_\alpha \hat{I}$ term cannot unless we choose $c_\alpha = 0$.

Therefore, we do not have any extra multiples of $\hat{I}$ and we find that

$$\hat{G}_\alpha = M\hat{Q}_\alpha \tag{4.273}$$

when there are no internal degrees of freedom.

In a similar manner, we can show that

$$[\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}}, \hat{\mathbf{Q}}] = 0 \tag{4.274}$$

which then implies that

$$\hat{\mathbf{J}} - \hat{\mathbf{Q}} \times \hat{\mathbf{P}} = \vec{c}\hat{I} \tag{4.275}$$

But since

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma \tag{4.276}$$

again we are forced to choose $\vec{c} = 0$ and thus

$$\hat{\mathbf{J}} = \hat{\mathbf{Q}} \times \hat{\mathbf{P}} \tag{4.277}$$

when there are no internal degrees of freedom.

The remaining generator we need to identify is $\hat{H}$. It must satisfy

$$[\hat{G}_\alpha, \hat{H}] = i\hat{P}_\alpha \rightarrow [M\hat{Q}_\alpha, \hat{H}] = i\hat{P}_\alpha \rightarrow [\hat{Q}_\alpha, \hat{H}] = i\frac{\hat{P}_\alpha}{M} \tag{4.278}$$

A solution of this equation is given by

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} = \frac{\hat{\mathbf{P}}^2}{2M} \tag{4.279}$$

as can be seen below

$$\begin{aligned}
\left[\hat{Q}_\alpha, \frac{1}{2M}\hat{\mathbf{P}}^2\right] &= \left[\hat{Q}_\alpha, \frac{1}{2M}\sum_\beta \hat{P}_\beta^2\right] = \frac{1}{2M}\sum_\beta [\hat{Q}_\alpha, \hat{P}_\beta^2] \qquad (4.280)\\
&= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta^2 - \hat{P}_\beta^2 \hat{Q}_\alpha) = \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta \hat{P}_\beta \hat{Q}_\alpha)\\
&= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta (\hat{Q}_\alpha \hat{P}_\beta - i\delta_{\alpha\beta}\hat{I}))\\
&= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - \hat{P}_\beta \hat{Q}_\alpha \hat{P}_\beta + i\delta_{\alpha\beta}\hat{P}_\beta)\\
&= \frac{1}{2M}\sum_\beta (\hat{Q}_\alpha \hat{P}_\beta \hat{P}_\beta - (\hat{Q}_\alpha \hat{P}_\beta - i\delta_{\alpha\beta}\hat{I})\hat{P}_\beta + i\delta_{\alpha\beta}\hat{P}_\beta)\\
&= \frac{1}{2M}\sum_\beta 2i\delta_{\alpha\beta}\hat{P}_\beta = \frac{i\hat{P}_\alpha}{M}
\end{aligned}$$

This result implies that $\hat{H} - \hat{\mathbf{P}} \cdot \hat{\mathbf{P}}/2M$ commutes with $\hat{\mathbf{Q}}$ and since $[\hat{H}, \hat{\mathbf{P}}] = 0$, it also commutes with $\hat{\mathbf{P}}$. Therefore, it is a multiple of $\hat{I}$, and we find

$$\hat{H} - \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} = E_0 \hat{I} \quad , \quad E_0 = \text{ constant} \tag{4.281}$$

or

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} + E_0 \hat{I} \tag{4.282}$$

Now, earlier we found that

$$[\hat{Q}_\alpha, \hat{H}] = i\hat{V}_\alpha \tag{4.283}$$

which now implies that

$$[\hat{Q}_\alpha, \hat{H}] = \left[\hat{Q}_\alpha, \frac{1}{2M}\hat{\mathbf{P}}^2\right] = \frac{i\hat{P}_\alpha}{M} = i\hat{V}_\alpha \tag{4.284}$$

Thus,

$$\hat{V}_\alpha = \frac{\hat{P}_\alpha}{M} \text{ or } \hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M} \tag{4.285}$$

Summarizing, we have found that

$$\hat{\mathbf{P}} = M\hat{\mathbf{V}} \quad , \quad \hat{H} = \frac{1}{2}\hat{\mathbf{V}} \cdot \hat{\mathbf{V}} + E_0 \quad , \quad \hat{\mathbf{J}} = \hat{\mathbf{Q}} \times M\hat{\mathbf{V}} \tag{4.286}$$

where

$$\hat{\mathbf{V}} = \text{the velocity operator and } \hat{\mathbf{Q}} = \text{the position operator} \tag{4.287}$$

This implies, since we are talking about a *free particle*, that

M must be *proportional* to the mass of the free particle

i.e., suppose that

$$M = \text{constant} \times \text{mass} \; = \beta m \tag{4.288}$$

then we must have

$$\begin{aligned} \hat{P}_\alpha &= \text{constant} \times \text{linear momentum} \; = \beta \hat{p}_\alpha \\ \hat{H} &= \; \text{constant} \times \text{energy} \; = \beta \hat{E} \\ \hat{J}_\alpha &= \text{constant} \; \times \text{angular momentum} \; = \beta \hat{j}_\alpha \end{aligned} \tag{4.289}$$

and all the relations (4.285)

$$\hat{\mathbf{P}} = M\hat{\mathbf{V}} \quad , \quad \hat{H} = \frac{1}{2}\hat{\mathbf{V}} \cdot \hat{\mathbf{V}} + E_0 \quad , \quad \hat{\mathbf{J}} = \hat{\mathbf{Q}} \times M\hat{\mathbf{V}} \tag{4.290}$$

still hold.

Now also remember that

$$[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I} \tag{4.291}$$

which then implies that

$$[\hat{Q}_\alpha, \hat{p}_\beta] = i\frac{1}{\beta}\delta_{\alpha\beta}\hat{I} \tag{4.292}$$

This says that $1/\beta$ must have the units Joule-Sec. With hindsight as to later developments, we will now choose

$$\hbar = \frac{1}{\beta} \tag{4.293}$$

and thus we finally obtain

$$[\hat{Q}_\alpha, \hat{p}_\beta] = i\hbar\delta_{\alpha\beta}\hat{I} \tag{4.294}$$

At this point, we do not know the numerical value of $\hbar$ (any value will do at this stage of our development). *It can only be determined by experiment.* Later we shall see that it is

$$\hbar = \frac{h}{2\pi} \tag{4.295}$$

where h = Planck's constant = 6.62×10^{-34} Joule-Sec. $\hbar$ must be determined from experiment since it sets the scale of all quantum phenomena and the scale factors in physical theories cannot be known a priori (even though Kant thought just the opposite was true).

Example 2 : Free Particle - with Spin

Internal degrees of freedom are, by definition, independent of the center of mass

degrees of freedom(Example 1). This means they are represented by operators that are independent of both $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$, or that they are represented by operators that commute with both $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$.

The set of operators $\{\hat{\mathbf{Q}}, \hat{\mathbf{P}}\}$ is *not* irreducible in this case since an operator that commutes with the set may still be a function of the operators corresponding to the internal degrees of freedom.

The *spin* $\hat{\mathbf{S}}$ is defined to be an internal contribution to the total angular momentum of the system. Therefore, we must modify the operator representing angular momentum to be

$$\hat{\mathbf{J}} = \hat{\mathbf{Q}} \times \hat{\mathbf{P}} + \hat{\mathbf{S}} \tag{4.296}$$

with $[\hat{\mathbf{Q}}, \hat{\mathbf{S}}] = 0 = [\hat{\mathbf{P}}, \hat{\mathbf{S}}]$.

We will study spin in great detail later in this book.

For now, let us see what we can say just based on the fact that spin is an angular momentum. This means that the $\hat{S}_\alpha$ must have the same commutators among themselves as the $\hat{J}_\alpha$.

$$[\hat{S}_\alpha, \hat{S}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{S}_\gamma \tag{4.297}$$

Earlier we found the equation

$$[\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I} \tag{4.298}$$

is satisfied by $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$ and

$$[\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{P}}] = 0 \text{ and } [\hat{\mathbf{G}} - M\hat{\mathbf{Q}}, \hat{\mathbf{Q}}] = 0 \tag{4.299}$$

implied that $\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = \vec{a}\hat{I}$ and we decided that we must have $\vec{a} = 0$.

Now, however, we could also have terms involving $\hat{\mathbf{S}}$ of the form

$$\hat{\mathbf{G}} - M\hat{\mathbf{Q}} = c\hat{\mathbf{S}} \tag{4.300}$$

Higher order powers of $\hat{\mathbf{S}}$ do not contribute any new terms because the commutator, which we can express as

$$\hat{\mathbf{S}} \times \hat{\mathbf{S}} = i\hat{\mathbf{S}} \tag{4.301}$$

indicates that they will reduce to a first-order term. So this result is general.

Now we found earlier that $[\hat{G}_\alpha, \hat{G}_\beta] = 0$, which implies that $c = 0$ and thus, we still have $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$.

The previous argument to get

$$\hat{H} = \frac{\hat{\mathbf{P}} \cdot \hat{\mathbf{P}}}{2M} + E_0\hat{I} \tag{4.302}$$

is identical to before except now E_0 can be a function of $\hat{\mathbf{S}}$. Now we must have

$$[\hat{\mathbf{J}}, \hat{H}] = 0 \rightarrow [\hat{\mathbf{S}}, E_0] = 0 \tag{4.303}$$

which implies that $E_0 = c\hat{\mathbf{S}} \cdot \hat{\mathbf{S}}$ is the only possibility. This has no effect on the $\hat{\mathbf{V}}$ operator given by

$$\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}] \tag{4.304}$$

since $[E_0, \hat{\mathbf{Q}}] = 0$. Therefore, the relation

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M} \tag{4.305}$$

remains valid.

So everything is the same as in Example 1 except that E_0 now corresponds to an *internal (spin dependent) contribution* to the energy.

Example 3 - A Particle Interacting with External Fields

We will only consider a *spinless* particle.

Interactions change the time evolution operator and as a consequence, the probability distributions for observables.

We *assume* that the equation of motion for the state vector *retains its form*

$$\frac{d}{dt}|\psi(t)\rangle = -i\hat{H}|\psi(t)\rangle \tag{4.306}$$

but the generator $\hat{H}$ changes to include the interactions. We single out $\hat{H}$ as the only generator that changes in the presence of interactions because it generates *dynamical* evolution in time and interactions only change that property. All the other generators imply purely *geometric* transformations which are *not dynamical*.

We also retain the definition $\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}]$, but note that interactions (a change in $\hat{H}$) will affect its value. So we assume that $\hat{H}$ still satisfies

$$\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}] \tag{4.307}$$

If we shift to a different frame of reference moving uniformly with respect to the original frame, we saw earlier that $\hat{\mathbf{V}}$ transforms as

$$e^{i\vec{v}\cdot\hat{\mathbf{G}}}\hat{\mathbf{V}}e^{-i\vec{v}\cdot\hat{\mathbf{G}}} = \hat{\mathbf{V}} - \vec{v}\hat{I} \tag{4.308}$$

Expanding the left-hand side to 1^{st}-order in $\vec{v}$ gives

$$i[\vec{v}\cdot\hat{\mathbf{G}}, \hat{\mathbf{V}}] = -\vec{v}\hat{I} \text{ and } [\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I} \tag{4.309}$$

Now, we still have the relation $\hat{G}_\alpha = M\hat{Q}_\alpha$, since its derivation does not use any commutators involving $\hat{H}$.

Our earlier solution for $\hat{\mathbf{V}}$ (no external fields) was

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}}}{M} \tag{4.310}$$

Now the commutators

$$[\hat{G}_\alpha, \hat{V}_\beta] = i\delta_{\alpha\beta}\hat{I} \text{ and } [\hat{G}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}M\hat{I} \tag{4.311}$$

imply that $\hat{\mathbf{V}} = \hat{\mathbf{P}}/M$ commutes with $\hat{\mathbf{G}}$. Since $\hat{\mathbf{G}} = M\hat{\mathbf{Q}}$, we must have

$$\left[\hat{\mathbf{V}} - \frac{\hat{\mathbf{P}}}{M}, \hat{\mathbf{Q}}\right] = 0 \tag{4.312}$$

With no internal degrees of freedom present, the set of operators $\{\hat{Q}_\alpha\}$ is a complete commuting set. This means that we must have

$$\hat{\mathbf{V}} - \frac{\hat{\mathbf{P}}}{M} = -\frac{\vec{A}(\hat{\mathbf{Q}})}{M} = \text{a function only of } \hat{\mathbf{Q}} \tag{4.313}$$

or

$$\hat{\mathbf{V}} = \frac{\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})}{M} \tag{4.314}$$

We now need to solve $\hat{\mathbf{V}} = i[\hat{H}, \hat{\mathbf{Q}}]$ for $\hat{H}$. We have the result

$$\frac{\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})}{M} = i[\hat{H}, \hat{\mathbf{Q}}] \tag{4.315}$$

A possible solution is

$$\hat{H}_0 = \frac{\left(\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})\right)^2}{2M} \tag{4.316}$$

as can be seen from the derivation below. We have

$$\begin{aligned}[\hat{H}_0, \hat{Q}_\alpha] &= \frac{1}{2M}\sum_\beta \left[(\hat{P}_\beta^2 - A_\beta\hat{P}_\beta - \hat{P}_\beta A_\beta - A_\beta^2), \hat{Q}_\alpha\right] \\ &= \frac{1}{2M}\sum_\beta \left([\hat{P}_\beta^2, \hat{Q}_\alpha] - [A_\beta\hat{P}_\beta, \hat{Q}_\alpha] - [\hat{P}_\beta A_\beta, \hat{Q}_\alpha] - [A_\beta^2, \hat{Q}_\alpha]\right)\end{aligned} \tag{4.317}$$

Now

$$[\hat{Q}_\alpha, \hat{P}_\beta] = i\delta_{\alpha\beta}\hat{I} \text{ and } [A_\beta, \hat{Q}_\alpha] = 0 \tag{4.318}$$

which gives

$$\begin{aligned}
&[A_\beta\hat{P}_\beta, \hat{Q}_\alpha] = A_\beta[\hat{P}_\beta, \hat{Q}_\alpha] = -iA_\beta\delta_{\alpha\beta}\hat{I} \\
&[\hat{P}_\beta A_\beta, \hat{Q}_\alpha] = [\hat{P}_\beta, \hat{Q}_\alpha]A_\beta = -iA_\beta\delta_{\alpha\beta}\hat{I} \\
&[\hat{P}_\beta^2, \hat{Q}_\alpha] = \hat{P}_\beta\hat{P}_\beta\hat{Q}_\alpha - \hat{Q}_\alpha\hat{P}_\beta\hat{P}_\beta = \hat{P}_\beta[\hat{P}_\beta, \hat{Q}_\alpha] - i\delta_{\alpha\beta}\hat{P}_\beta = -2i\delta_{\alpha\beta}\hat{P}_\beta \\
&[A_\beta^2, \hat{Q}_\alpha] = 0
\end{aligned}$$

We then have the final step in the proof

$$[\hat{H}_0, \hat{Q}_\alpha] = \frac{1}{2M} \sum_\beta 2i\delta_{\alpha\beta}(\hat{P}_\beta - A_\beta) = \frac{\hat{P}_\alpha - A_\alpha}{M} \tag{4.319}$$

That completes the proof. Finally, we can then say

$$[(\hat{H} - \hat{H}_0), \hat{Q}_\alpha] = 0 \tag{4.320}$$

which implies that at most $\hat{H}$ can differ from $\hat{H}_0$ only by a function of $\hat{\mathbf{Q}}$

$$\hat{H} - \hat{H}_0 = W(\hat{\mathbf{Q}}) = \text{ function of } \hat{\mathbf{Q}} \tag{4.321}$$

or, in general

$$\hat{H} = \frac{\left(\hat{\mathbf{P}} - \vec{A}(\hat{\mathbf{Q}})\right)^2}{2M} + W(\hat{\mathbf{Q}}) \tag{4.322}$$

This is the only form of $\hat{H}$ *consistent with invariance under the Galilei group of transformations.*

The two new functions of $\hat{\mathbf{Q}}$ are called

$$\vec{A}(\hat{\mathbf{Q}}) = \text{vector potential and } W(\hat{\mathbf{Q}}) = \text{scalar potential} \tag{4.323}$$

Both of the functions can be time-dependent. As operators they are functions only of $\hat{\mathbf{Q}}$ and not of $\hat{\mathbf{P}}$.

This form certainly includes the classical electromagnetic interaction. However, we cannot identify $\vec{A}(\hat{\mathbf{Q}})$ and $W(\hat{\mathbf{Q}})$ with the electromagnetic potential because nothing in the derivation implies that they need to satisfy Maxwell's equations.

This method can be generalized to cover the case of more than one particle interacting with external fields.

An Aside

The conventional notation involving the generators and the constant $\hbar$ is to use

$$\hat{\mathbf{P}} \to \frac{\hat{\mathbf{P}}}{\hbar} \quad , \quad \hat{\mathbf{J}} \to \frac{\hat{\mathbf{J}}}{\hbar} \quad , \quad M \to \frac{M}{\hbar} \quad , \quad \hat{H} \to \frac{\hat{H}}{\hbar} \tag{4.324}$$

leading to changed transformation operators of the form

$$e^{-i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} \quad , \quad e^{-i\theta\mathbf{n}\cdot\hat{\mathbf{J}}/\hbar} \quad , \quad e^{-it\hat{H}/\hbar} \tag{4.325}$$

or we could just let $\hbar = 1$ and continue to use all previous results unchanged.

4.13 Multiparticle Systems

How do we generalize these single particle results to systems with more than one particle?

We will deal with a special case that illustrates the general procedure without adding any extra complexity.

Consider two particles forming a composite system where

$$\hat{Q}^{(1)} = \text{operator representing an observable of particle 1}$$
$$\hat{T}^{(2)} = \text{operator representing an observable of particle 2}$$

We *assume* that the two particles in the composite system can be separated physically so that they would not have any interaction with each other. This means that when they are separated, the composite system must reduce to two independent one-particle systems which we can describe using all of the one-particle results we have already derived.

Thus, when we find a description of the two-particle system, it must include the separate one-particle descriptions in some way. This means that it must be possible to prepare the one-particle states as separate, independent entities in the laboratory.

Now, earlier we proved that there exists a state of the system where an observable represented by an operator $\hat{Q}^{(1)}$ has a definite value (probability = 1). This state is any one of the eigenvectors of $\hat{Q}^{(1)}$ and the definite value is the corresponding eigenvalue. A similar result holds for the operator $\hat{T}^{(2)}$.

Now we have assumed that the properties of the two particles can be measured independently. This means that a two-particle state vector for the composite system must exist such that it is a common eigenvector for all operators representing observables of both particles. This says that if

$$\hat{Q}^{(1)} |q_m\rangle_1 = q_m |q_m\rangle_1 \text{ and } \hat{T}^{(2)} |t_n\rangle_2 = t_n |t_n\rangle_2 \tag{4.326}$$

then for every m and n there exists a two-particle state vector $|q_m, t_n\rangle$ for the composite system with the properties

$$\hat{Q}^{(1)} |q_m, t_n\rangle = q_m |q_m, t_n\rangle \text{ and } \hat{T}^{(2)} |q_m, t_n\rangle = t_n |q_m, t_n\rangle \tag{4.327}$$

The way to satisfy these conditions is to represent the two-particle state vector as a mathematical object known as the *Kronecker* or *direct product*. We write this (symbolically) as

$$|q_m, t_n\rangle = |q_m\rangle_1 |t_n\rangle_2 \text{ or } |q_m, t_n\rangle = |q_m\rangle_1 \otimes |t_n\rangle_2 \tag{4.328}$$

If the set of vectors $\{|q_m\rangle_1\}$ spans an M-dimensional vector space and the set of vectors $\{|t_n\rangle_2\}$ spans an N-dimensional vector space, then the set of all direct product vectors spans an $M \times N$ dimensional vector space.

Suppose we have two sets of operators, $\{\hat{Q}_i^{(1)}\}$ for the system consisting of particle 1 and $\{\hat{T}_j^{(2)}\}$ for the system consisting of particle 2. We define how these operators act on the direct product states by the rules

$$\begin{aligned} \hat{Q}_i^{(1)} |q_m, t_n\rangle &= (\hat{Q}_i^{(1)} |q_m\rangle_1) \otimes |t_n\rangle_2 \\ \hat{T}_j^{(2)} |q_m, t_n\rangle &= |q_m\rangle_1 \otimes (\hat{T}_j^{(2)} |t_n\rangle_2) \end{aligned} \tag{4.329}$$

and we define a direct product between operators in the two sets by the relation

$$(\hat{Q}_i^{(1)} \otimes \hat{T}_j^{(2)}) |q_m, t_n\rangle = (\hat{Q}_i^{(1)} |q_m\rangle_1) \otimes (\hat{T}_j^{(2)} |t_n\rangle_2) \tag{4.330}$$

When we write $\hat{Q}_i^{(1)}$ or $\hat{T}_j^{(2)}$ alone, we really mean the following

$$\begin{aligned} \hat{Q}_i^{(1)} &\to \hat{Q}_i^{(1)} \otimes \hat{I}^{(2)} \\ \hat{T}_j^{(2)} &\to \hat{I}^{(1)} \otimes \hat{T}_j^{(2)} \end{aligned} \tag{4.331}$$

This definition of the direct product operators does not include all relevant physical operators that we use in quantum mechanics. When the particles are interacting, there must exist interaction operators that act on both sets of states. Although this means that an individual direct product state vector for a composite system cannot directly represent interacting particles, we will be able to use the set of all such states(which is complete) as a basis for representing states of interacting particles.

Since the common set of states $\{|q_m, t_n\rangle\}$ comprise a complete basis set for all of the $\hat{Q}_i^{(1)}$ and $\hat{T}_j^{(2)}$ operators, these operators must form a set of mutually commuting operators or

$$\left[\hat{Q}_i^{(1)}, \hat{T}_j^{(2)}\right] = 0 \text{ for all } i, j \tag{4.332}$$

An Example

Imagine we are in a fictitious world in which the single-particle Hilbert space is 2-dimensional. Let us denote the corresponding basis vectors by $|+\rangle$ and $|-\rangle$. In addition, let arbitrary operators $\beta_1^{(1)}$ and $\beta_1^{(2)}$ be represented by (in their respective spaces) by

$$\begin{aligned} \beta_1^{(1)} &= \begin{pmatrix} {}_1\langle +| \beta_1^{(1)} |+\rangle_1 & {}_1\langle +| \beta_1^{(1)} |-\rangle_1 \\ {}_1\langle -| \beta_1^{(1)} |+\rangle_1 & {}_1\langle -| \beta_1^{(1)} |-\rangle_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ \beta_1^{(2)} &= \begin{pmatrix} {}_2\langle +| \beta_1^{(2)} |+\rangle_2 & {}_2\langle +| \beta_1^{(2)} |-\rangle_2 \\ {}_2\langle -| \beta_1^{(2)} |+\rangle_2 & {}_2\langle -| \beta_1^{(2)} |-\rangle_2 \end{pmatrix} = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \end{aligned} \tag{4.333}$$

These are operators in the space 1 (V_1) and space 2 (V_2), respectively. The space $V_1 \otimes V_2$ is therefore spanned by 4 vectors. They are

$$\begin{aligned} |+\rangle_1 |+\rangle_2 &= |+\rangle_1 \otimes |+\rangle_2 = |{+}{+}\rangle \ , \ |+\rangle_1 |-\rangle_2 = |+\rangle_1 \otimes |-\rangle_2 = |{+}{-}\rangle \\ |-\rangle_1 |+\rangle_2 &= |-\rangle_1 \otimes |+\rangle_2 = |{-}{+}\rangle \ , \ |-\rangle_1 |-\rangle_2 = |-\rangle_1 \otimes |-\rangle_2 = |{-}{-}\rangle \end{aligned} \tag{4.334}$$

We define the general operator (acts in both spaces)

$$\begin{aligned} \hat{O}_{12} &= \beta_1^{(1)} \otimes \hat{I}^{(2)} \\ &= \begin{pmatrix} \langle{+}{+}|\hat{O}_{12}|{+}{+}\rangle & \langle{+}{+}|\hat{O}_{12}|{+}{-}\rangle & \langle{+}{+}|\hat{O}_{12}|{-}{+}\rangle & \langle{+}{+}|\hat{O}_{12}|{-}{-}\rangle \\ \langle{+}{-}|\hat{O}_{12}|{+}{+}\rangle & \langle{+}{-}|\hat{O}_{12}|{+}{-}\rangle & \langle{+}{-}|\hat{O}_{12}|{-}{+}\rangle & \langle{+}{-}|\hat{O}_{12}|{-}{-}\rangle \\ \langle{-}{+}|\hat{O}_{12}|{+}{+}\rangle & \langle{-}{+}|\hat{O}_{12}|{+}{-}\rangle & \langle{-}{+}|\hat{O}_{12}|{-}{+}\rangle & \langle{-}{+}|\hat{O}_{12}|{-}{-}\rangle \\ \langle{-}{-}|\hat{O}_{12}|{+}{+}\rangle & \langle{-}{-}|\hat{O}_{12}|{+}{-}\rangle & \langle{-}{-}|\hat{O}_{12}|{-}{+}\rangle & \langle{-}|\hat{O}_{12}|{-}{-}\rangle \end{pmatrix} \\ &= \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix} \end{aligned} \tag{4.335}$$

where we have used the type of calculation below to determine the matrix elements.

$$\begin{aligned} \langle{+}{+}|\hat{O}_{12}|{+}{+}\rangle &= \langle{+}{+}|\beta_1^{(1)} \otimes \hat{I}^{(2)}|{+}{+}\rangle \\ &= {}_1\langle{+}|\beta_1^{(1)}|{+}\rangle_{1\,2}\langle{+}|\hat{I}^{(2)}|{+}\rangle_2 \\ &= {}_1\langle{+}|\beta_1^{(1)}|{+}\rangle_{1\,2}\langle{+}|{+}\rangle_2 \\ &= {}_1\langle{+}|\beta_1^{(1)}|{+}\rangle_1 = a \end{aligned}$$

Similarly, we have

$$\hat{I}^{(1)} \otimes \beta_1^{(2)} = \begin{pmatrix} e & f & 0 & 0 \\ g & h & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & h \end{pmatrix} \tag{4.336}$$

and

$$\beta_1^{(1)} \otimes \beta_1^{(2)} = \begin{pmatrix} ae & af & be & bf \\ ag & ah & bg & bh \\ ce & cf & de & df \\ cg & ch & dg & dh \end{pmatrix} \tag{4.337}$$

Therefore, using the *Pauli spin operators* defined (in each space) by

$$\hat{\sigma}_1|{\pm}\rangle = |{\mp}\rangle, \hat{\sigma}_2|{\pm}\rangle = {\mp}i|{\mp}\rangle, \hat{\sigma}_3|{\pm}\rangle = {\pm}|{\mp}\rangle \tag{4.338}$$

so that these operators have the matrix representations (in each space)

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.339}$$

We then have

$$\hat{\sigma}_1^{(1)} \otimes \hat{\sigma}_1^{(2)} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{4.340}$$

$$\hat{\sigma}_2^{(1)} \otimes \hat{\sigma}_2^{(2)} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{4.341}$$

$$\hat{\sigma}_3^{(1)} \otimes \hat{\sigma}_3^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.342}$$

Thus,

$$\hat{\sigma}_1^{(1)} \cdot \hat{\sigma}_2^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.343}$$

and

$$\hat{f} = a\hat{I} + b\hat{\sigma}_1^{(1)} \cdot \hat{\sigma}_2^{(2)} = \begin{pmatrix} a+b & 0 & 0 & 0 \\ 0 & a-b & 2b & 0 \\ 0 & 2b & a-b & 0 \\ 0 & 0 & 0 & a+b \end{pmatrix} \tag{4.344}$$

The states

$$\begin{aligned} |+\rangle_1 |+\rangle_2 &= |{+}{+}\rangle \, , |+\rangle_1 |-\rangle_2 = |{+}{-}\rangle \\ |-\rangle_1 |+\rangle_2 &= |{-}{+}\rangle \, , |-\rangle_1 |-\rangle_2 = |{-}{-}\rangle \end{aligned} \tag{4.345}$$

are given by

$$|{+}{+}\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \, , \; |{+}{-}\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \, , \; |{-}{+}\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \, , \; |{-}{-}\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.346}$$

and we have

$$\hat{\sigma}_1^{(1)} |{+}{+}\rangle = \hat{\sigma}_1^{(1)} \otimes \hat{I}^{(2)} |{+}{+}\rangle = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = |{-}{+}\rangle \tag{4.347}$$

as expected.

An alternative way to think about the outer product is shown below:

$$\hat{\sigma}_1^{(1)} \otimes \hat{\sigma}_1^{(2)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \hat{\sigma}_1^{(2)} = \begin{pmatrix} 0 & \hat{\sigma}_1^{(2)} \\ \hat{\sigma}_1^{(2)} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{4.348}$$

Extending This Idea

This procedure for constructing composite state vectors using the direct product also works when dealing with certain dynamical properties of a one-particle state.

Any one-particle state must represent the various degrees of freedom for the one-particle system. These degrees of freedom, as we have seen, are related to observables and their associated operators and eigenvalues. Thus, quantities like $\hat{Q}_\alpha$, $\hat{P}_\beta$, and $\hat{S}_\gamma$ all represent different degrees of freedom of the physical system or state vector.

In the case of certain one-particle system degrees of freedom, which are said to be independent, we can write one-particle state vectors and operators using direct products. For instance, as we saw in an earlier example, both $\hat{Q}_\alpha$ and $\hat{P}_\beta$ are independent of an internal degree of freedom that we called spin $\hat{S}_\gamma$. In fact, we defined an internal degree of freedom as one which was independent of the center of mass degrees of freedom. We defined this independence via commutators by $[\hat{Q}_\alpha, \hat{S}_\gamma] = 0 = [\hat{P}_\beta, \hat{S}_\gamma]$.

Another example of independence of degrees of freedom was our assumption that $[\hat{Q}_\alpha, \hat{Q}_\beta] = 0$, which says that the three components of the position operator are independent degrees of freedom. This should not surprise us since it just reflects the physical assumption that it is possible to prepare a state where a particle is localized arbitrarily close to a single point in 3-dimensional position space.

Similarly, our assumption that $[\hat{P}_\alpha, \hat{P}_\beta] = 0$ means it is possible to prepare a state where a particle is localized arbitrarily close to a single point in 3-dimensional momentum space.

It is also clear, since $[\hat{Q}_\alpha, \hat{P}_\beta] \neq 0$ that we will not be able to prepare a state in which both $\hat{Q}_\alpha$ and $\hat{P}_\alpha$ (components along the same axis) are simultaneously localized arbitrarily close to single point in phase space and since $[\hat{J}_\beta, \hat{J}_\gamma] \neq 0$ we will not be able to prepare a state in which two different components of the angular momentum have definite values.

This means that we can write single particle states as direct products. For example

$$|\vec{x}\rangle = |x_1\rangle \otimes |x_2\rangle \otimes |x_3\rangle \tag{4.349}$$

$$\hat{X}_1\hat{X}_2\hat{X}_3 = \hat{X}_1 \otimes \hat{X}_2 \otimes \hat{X}_3 \tag{4.350}$$

$$\hat{X}_1\hat{X}_2\hat{X}_3 |\vec{x}\rangle = (\hat{X}_1 |x_1\rangle) \otimes (\hat{X}_2 |x_2\rangle) \otimes (\hat{X}_3 |x_3\rangle) \tag{4.351}$$

or

$$|\vec{x}' s_z\rangle = |x_1\rangle \otimes |x_2\rangle \otimes |x_3\rangle \otimes |s_z\rangle \tag{4.352}$$

$$\hat{X}_1\hat{X}_2\hat{X}_3\hat{S}_z = \hat{X}_1 \otimes \hat{X}_2 \otimes \hat{X}_3 \otimes \hat{S}_z \tag{4.353}$$

$$\hat{X}_1\hat{X}_2\hat{X}_3\hat{S}_z |\vec{x}' s_z\rangle = (\hat{X}_1 |X\rangle_1) \otimes (\hat{X}_2 |X\rangle_2) \otimes (\hat{X}_3 |X\rangle_3) \otimes (\hat{S}_z |\vec{x}' s_z\rangle) \tag{4.354}$$

We can use words as follows:

$$|\vec{x}\rangle = |x_1\rangle \otimes |x_2\rangle \otimes |x_3\rangle \rightarrow \text{ particle at } (x_1, x_2, x_3) \tag{4.355}$$

or

$$|\vec{x}' s_z\rangle = |x_1\rangle \otimes |x_2\rangle \otimes |x_3\rangle \otimes |s_z\rangle \rightarrow \text{ particle at } (x_1, x_2, x_3) \text{ with spin } s_z \tag{4.356}$$

If we can construct state vectors for two-particle systems, then it must be possible to define appropriate probability distributions. These are called joint probability distributions.

If we prepare both particles into a direct product state such that their individual preparations are independent and they do not interact with each other, then their joint probability distribution for the observables Q and T corresponding to the operators $\hat{Q}^{(1)}$ and $\hat{T}^{(2)}$ should obey the statistical independence condition we defined earlier for events A, B, and C, namely,

$$Prob(A \cap B|C) = Prob(A|C)Prob(B|C) \tag{4.357}$$

For the direct product state

$$|\psi\rangle = |\alpha\rangle_1 \otimes |\beta\rangle_2 \tag{4.358}$$

this says that the joint probability distribution of Q and T is

$$\begin{aligned} Prob((Q = q_m) \cap (T = t_n)|\psi) &= |\langle q_m, t_n |\psi\rangle|^2 \\ &= Prob(Q = q_m|\alpha)Prob(T = t_n|\beta) \\ &= |\langle q_m |\alpha\rangle_1|^2 |\langle t_n |\beta\rangle_2|^2 \end{aligned} \tag{4.359}$$

The state does not have to be a pure state for this factorization to occur. It only needs to be represented by a density operator of the form

$$\hat{W} = \hat{W}^{(1)} \otimes \hat{W}^{(2)} \tag{4.360}$$

Further discussion of this topic follows in section(17).

Later, we will use these direct product procedures to construct operators and basis states for multiparticle systems like atoms. Just how we will include interactions is not clear yet.

4.14 Equations of Motion Revisited and Finished

We discussed time dependence and the time evolution operator earlier. This is the most important topic in quantum mechanics and will eventually enable us to make predictions about the behavior of real physical systems. Let us now review our earlier discussion from a more general point of view in light of some of the new ideas we have introduced.

We derived a differential equation of motion for the state vector of the form

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}(t)|\psi(t)\rangle \tag{4.361}$$

If, initially (at $t = t_0$) we are in the state represented by the ket vector $|\psi(t_0)\rangle = |\psi_0\rangle$, then we wrote the formal solution of the differential equation in terms of the time development operator $\hat{U}(t, t_0)$ as

$$|\psi(t)\rangle = \hat{U}(t, t_0)|\psi(t_0)\rangle \tag{4.362}$$

We also showed that $\hat{U}(t, t_0)$ satisfies the same differential equation as $|\psi(t)\rangle$

$$\frac{\partial}{\partial t}\hat{U}(t, t_0) = -\frac{i}{\hbar}\hat{H}(t)\hat{U}(t, t_0) \tag{4.363}$$

with the boundary condition $\hat{U}(t_0, t_0) = \hat{I}$. This implies, using the relation $(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger$ that

$$\frac{\partial}{\partial t}\hat{U}^\dagger(t, t_0) = \frac{i}{\hbar}\hat{U}^\dagger(t, t_0)\hat{H}(t) \tag{4.364}$$

We then have

$$\begin{aligned}\frac{\partial}{\partial t}(\hat{U}^\dagger\hat{U}) &= \hat{U}^\dagger\frac{\partial\hat{U}}{\partial t} + \frac{\partial\hat{U}^\dagger}{\partial t}\hat{U} \\ &= -\frac{i}{\hbar}\hat{U}^\dagger\hat{H}\hat{U} + \frac{i}{\hbar}\hat{U}^\dagger\hat{H}^\dagger\hat{U} \\ &= \frac{i}{\hbar}\hat{U}^\dagger(\hat{H} - \hat{H}^\dagger)\hat{U}\end{aligned} \tag{4.365}$$

If $\hat{H}$ is Hermitian, then $\hat{H} = \hat{H}^\dagger$ and

$$\frac{\partial}{\partial t}(\hat{U}^\dagger\hat{U}) = 0 \tag{4.366}$$

which implies that

$$\hat{U}^\dagger(t, t_0)\hat{U}(t, t_0) = c\hat{I} \text{ where } c = \text{constant} \tag{4.367}$$

But the boundary condition implies that $\hat{U}^\dagger(t_0, t_0)\hat{U}(t_0, t_0) = \hat{I}$ which implies that $c = 1$. Therefore, in general we have $\hat{U}^\dagger(t, t_0)\hat{U}(t, t_0) = \hat{I}$, which is the *unitarity condition* and implies $\hat{U}^\dagger = \hat{U}^{-1}$ or that $\hat{U}$ is unitary when $\hat{H}$ is Hermitian.

This agrees with our earlier results.

Now, if $\hat{H}(t)$ is independent of t, then the differential equation has a simple solution

$$\hat{U}(t,t_0) = e^{-i(t-t_0)\frac{\hat{H}}{\hbar}} \tag{4.368}$$

which is the form we assumed for the time evolution operator during the discussion of symmetry transformations.

If $\hat{H}(t)$ is not independent of t, then *no simple closed form solution* can be given for $\hat{U}(t,t_0)$. We must use perturbation theory as we shall see later when we develop time-dependent perturbation theory.

We can now derive the *equation of motion for the density operator*. We use a pure state for simplicity. We have

$$\begin{aligned}\hat{W}(t) &= |\psi(t)\rangle\langle\psi(t)| = \hat{U}(t,t_0)|\psi(t_0)\rangle\langle\psi(t_0)|\hat{U}^\dagger(t,t_0) \\ &= \hat{U}(t,t_0)\hat{W}(t_0)\hat{U}^\dagger(t,t_0)\end{aligned} \tag{4.369}$$

Now, using $\hat{W}(t_0) = \hat{W}_0$, differentiating with respect to t gives

$$\begin{aligned}\frac{d\hat{W}(t)}{dt} &= \frac{d\hat{U}(t,t_0)}{dt}\hat{W}_0\hat{U}^\dagger(t,t_0) + \hat{U}(t,t_0)\hat{W}_0\frac{d\hat{U}^\dagger(t,t_0)}{dt} \\ &= -\frac{i}{\hbar}\hat{H}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0) + \frac{i}{\hbar}\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{H} \\ &= -\frac{i}{\hbar}[\hat{H}(t),\hat{W}(t)]\end{aligned} \tag{4.370}$$

We will assume that this equation is also true for general states.

Now, we have stated earlier that no physical significance can be attached in quantum mechanics to operators and vectors. The *only physically significant* objects are the probability distributions of observables or their expectation values. Earlier we derived the result

$$\langle\hat{Q}\rangle = Tr(\hat{W}\hat{Q}) \tag{4.371}$$

We now assume that this result carries over to the time-dependent case and we have

$$\langle\hat{Q}\rangle_t = Tr(\hat{W}(t)\hat{Q}) \tag{4.372}$$

Using the expression (4.368) for $\hat{W}(t)$ and the fact that the trace is invariant under cyclic permutation, i.e.,

$$Tr(\hat{A}\hat{B}\hat{C}) = Tr(\hat{C}\hat{A}\hat{B}) = Tr(\hat{B}\hat{C}\hat{A}) \tag{4.373}$$

we get

$$\begin{aligned}\langle\hat{Q}\rangle_t &= Tr(\hat{W}\hat{Q}) = Tr(\hat{U}(t,t_0)\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{Q}) \\ &= Tr(\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{Q}\hat{U}(t,t_0))\end{aligned} \tag{4.374}$$

This result is the formal basis of the Schrodinger and Heisenberg pictures we discussed earlier.

If we leave the time dependence in the density operator, then

$$\langle \hat{Q} \rangle_t = Tr(\hat{W}(t)\hat{Q}) \tag{4.375}$$

where

$$\frac{d\hat{W}(t)}{dt} = -\frac{i}{\hbar}[\hat{H}(t), \hat{W}(t)] \text{ and } \frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}(t)|\psi(t)\rangle \tag{4.376}$$

$\hat{Q}$ is independent of time is these equations. This is the *Schrodinger picture*.

On the other hand, if we write

$$\langle \hat{Q} \rangle_t = Tr(\hat{W}_0\hat{U}^\dagger(t,t_0)\hat{Q}\hat{U}(t,t_0)) = Tr(\hat{W}_0\hat{Q}_H(t)) \tag{4.377}$$

where $\hat{Q}_H(t) = \hat{U}^\dagger(t,t_0)\hat{Q}\hat{U}(t,t_0)$, then the operator is time dependent and the density operator(and hence the state vectors) are independent of time. This is the *Heisenberg picture*.

We can derive the equation of motion of the time-dependent operators in the Heisenberg picture as follows:

$$\begin{aligned}
\frac{d}{dt}\hat{Q}_H(t) &= \frac{\partial \hat{U}^\dagger}{\partial t}\hat{Q}\hat{U} + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U} + \hat{U}^\dagger\hat{Q}\frac{\partial \hat{U}}{\partial t} \\
&= \frac{i}{\hbar}(\hat{U}^\dagger\hat{H}\hat{Q}\hat{U} - \hat{U}^\dagger\hat{Q}\hat{H}\hat{U}) + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U} \\
&= \frac{i}{\hbar}(\hat{U}^\dagger\hat{H}\hat{U}\hat{U}^\dagger\hat{Q}\hat{U} - \hat{U}^\dagger\hat{Q}\hat{U}\hat{U}^\dagger\hat{H}\hat{U}) + \hat{U}^\dagger\frac{\partial \hat{Q}}{\partial t}\hat{U} \\
&= \frac{i}{\hbar}[\hat{H}_H(t), \hat{Q}_H(t)] + \left(\frac{\partial \hat{Q}}{\partial t}\right)_H
\end{aligned} \tag{4.378}$$

where we have included the possibility that $\hat{Q}$ has some explicit time dependence and we have used the definition

$$\hat{A}_H(t) = \hat{U}^\dagger(t,t_0)\hat{A}\hat{U}(t,t_0) \tag{4.379}$$

for any operator. Note the *change in sign* in front of the commutator in this equation for the operator from that of the density operator in the Schrodinger picture.

The two pictures are clearly *equivalent* as mathematical formalisms, since they are derived from the same expectation value formula by an internal rearrangement of terms. Another way to say this is that the two pictures are equivalent because the only physically significant quantity $\langle \hat{Q} \rangle_t$ depends only on the relative motion in time of $\hat{W}$ and $\hat{Q}$, which is the same in both cases.

In the Schrodinger picture $\hat{W}(t)$ moves *forward* in time (term of the form $\hat{U}\hat{W}_0\hat{U}^\dagger = \hat{U}\hat{W}_0\hat{U}^{-1}$) and in the Heisenberg picture $\hat{Q}(t)$ moves *backward* in time (term of the form $\hat{U}^\dagger\hat{W}_0\hat{U} = \hat{U}^{-1}\hat{W}_0\hat{U}$). The opposite senses of motion in time produce the sign difference of the commutator terms in the equations of motion. These two picture are mutually exclusive and cannot be used together.

Finally, we determine expressions for $d\langle\hat{Q}\rangle_t/dt$ in each picture.

In the Schrodinger picture

$$\begin{aligned}\frac{d\langle\hat{Q}\rangle_t}{dt} &= \frac{d}{dt}Tr(\hat{W}(t)\hat{Q}) = Tr\left[\frac{d\hat{W}}{dt}\hat{Q} + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right] \\ &= Tr\left[-\frac{i}{\hbar}(\hat{H}\hat{W}\hat{Q} - \hat{W}\hat{H}\hat{Q}) + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right] \\ &= Tr\left[-\frac{i}{\hbar}(\hat{W}\hat{Q}\hat{H} - \hat{W}\hat{H}\hat{Q}) + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right] \\ &= Tr\left[\frac{i}{\hbar}\hat{W}(t)[\hat{H},\hat{Q}] + \hat{W}\frac{\partial\hat{Q}}{\partial t}\right]\end{aligned} \tag{4.380}$$

In the Heisenberg picture

$$\frac{d\langle\hat{Q}\rangle_t}{dt} = Tr(\hat{W}_0\frac{d\hat{Q}_H(t)}{dt}) = Tr\left[\frac{i}{\hbar}\hat{W}_0[\hat{H},\hat{Q}_H(t)] + \hat{W}_0\left(\frac{\partial\hat{Q}}{\partial t}\right)_H\right] \tag{4.381}$$

For a pure state, we can rewrite these results in terms of the state vectors instead of the density operator. We have in the Schrodinger picture

$$\langle\hat{Q}\rangle_t = \langle\psi(t)|\,\hat{Q}\,|\psi(t)\rangle \text{ where } |\psi(t)\rangle = \hat{U}(t,t_0)\,|\psi_0\rangle \tag{4.382}$$

and in the Heisenberg picture

$$\langle\hat{Q}\rangle_t = \langle\psi_0|\,\hat{Q}_H(t)\,|\psi_0\rangle \text{ where } \hat{Q}_H(t) = \hat{U}^\dagger(t,t_0)\hat{Q}\hat{U}(t,t_0) \tag{4.383}$$

as we saw in our earlier discussions.

4.15 Symmetries, Conservation Laws and Stationary States

Let $\hat{T}(s) = e^{is\hat{K}}$ represent a continuous unitary transformation with a Hermitian generator $\hat{K} = \hat{K}^\dagger$. Another operator $\hat{A}$ representing some observable is *invariant* under this transformation if

$$\hat{T}(s)\hat{A}\hat{T}^{-1}(s) = \hat{A} \tag{4.384}$$

or

$$\hat{A}\hat{T}(s) - \hat{T}(s)\hat{A} = [\hat{A},\hat{T}(s)] = 0 \tag{4.385}$$

If s is an infinitesimal, then we can write

$$(\hat{I} + is\hat{K})\hat{A}(\hat{I} - is\hat{K}) = \hat{A} \rightarrow \hat{A} + is[\hat{K}, \hat{A}] = \hat{A} \tag{4.386}$$

or

$$[\hat{K}, \hat{A}] = 0 \tag{4.387}$$

In words, the invariance of $\hat{A}$ under the continuous transformation $\hat{T}(s) = e^{is\hat{K}}$ for all s implies it is true for infinitesimal s and thus leads to the *commutator condition* for invariance, which says that the operator commutes with the Hermitian generator of the transformation.

It works both ways:

$$\begin{aligned}[\hat{T}, \hat{A}] = 0 \rightarrow &\text{ invariance under finite transformation} \\ \rightarrow &\text{ invariance under infinitesimal transformation} \\ \rightarrow &[\hat{K}, \hat{A}] = 0\end{aligned}$$

or

$$\begin{aligned}[\hat{K}, \hat{A}] = 0 \rightarrow &\text{ invariance under infinitesimal transformation} \\ \rightarrow &\text{ invariance under finite transformation} \\ \rightarrow &[\hat{T}, \hat{A}] = 0\end{aligned}$$

If $\hat{K}$ depends on t, then the commutators $[\hat{K}(t), \hat{A}] = 0$ and $[\hat{T}(t), \hat{A}] = 0$ must hold for all t.

Now, the Hermitian generators of the symmetry transformation, as we have seen, correspond to dynamical variables of a physical system.

$$\begin{aligned}\text{space displacements} &\Leftrightarrow \hat{\mathbf{P}} \\ \text{rotations} &\Leftrightarrow \hat{\mathbf{J}} \\ t \text{ displacements} &\Leftrightarrow \hat{H}\end{aligned}$$

These symmetry generators have no explicit time dependence, i.e., $\partial\hat{K}/\partial t = 0$. Therefore,

$$\frac{d\langle\hat{K}\rangle_t}{dt} = Tr\left[\frac{i}{\hbar}\hat{W}(t)[\hat{H}, \hat{K}]\right] = 0 \tag{4.388}$$

if $\hat{H}$ is invariant under the corresponding symmetry transformation, i.e., $[\hat{H}, \hat{K}] = 0$. Now $[\hat{H}, \hat{K}] = 0 \rightarrow [\hat{H}, f(\hat{K})] = 0$. This says that we must have

$$[\hat{H}, \theta(x - \hat{K})] = 0 \tag{4.389}$$

But in our probability discussions, we showed that

$$\begin{aligned}\langle\theta(x - \hat{K})\rangle &= Prob(K < x|\hat{W}) \\ &= \text{ probability that observable } \hat{K} \text{ has a value } < x \text{ given } \hat{W}\end{aligned}$$

This then says that $Prob(K < x|\hat{W})$ is independent of t no matter what initial state we start with. In this case, observable $K = a$ is a constant of the motion.

Examples:

$[\hat{H}, \hat{P}_\alpha] = 0$ implies invariance under a space displacement along the α-axis. This implies that $\hat{P}_\alpha$= constant of the motion (is conserved) = linear momentum along the α-axis.

$[\hat{H}, \hat{J}_\alpha] = 0$ implies invariance under a rotation about the α-axis. This implies that $\hat{J}_\alpha$= constant of the motion (is conserved) = angular momentum about the α-axis.

If $\hat{H}$ is not an explicit function of t, then since $[\hat{H}, \hat{H}] = 0$, $\hat{H}$ is invariant under time translations = constant of the motion = energy of the system.

Now suppose that $\hat{H}$ is independent of t and that

$$|\psi(0)\rangle = [\text{ eigenvector of } \hat{H}] = |E_n\rangle \text{ such that } \hat{H}|E_n\rangle = E_n|E_n\rangle \tag{4.390}$$

Then we have, using $[\hat{H}, \hat{U}] = [\hat{H}, f(\hat{U})] = 0$,

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)|\psi(0)\rangle \tag{4.391}$$

$$= -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)|E_n\rangle = -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)\hat{H}|E_n\rangle$$

$$= -\frac{i}{\hbar}\hat{H}\hat{U}(t,0)E_n|E_n\rangle = -\frac{i}{\hbar}E_n\hat{H}\hat{U}(t,0)|E_n\rangle$$

$$= -\frac{i}{\hbar}E_n|\psi(t)\rangle \tag{4.392}$$

which has the solution

$$|\psi(t)\rangle = e^{-i\frac{E_n}{\hbar}t}|E_n\rangle \tag{4.393}$$

In this case, we then have for the expectation value of any observable represented by the operator $\hat{R}$

$$\langle\hat{R}\rangle = \langle\psi(t)|\hat{R}|\psi(t)\rangle = \langle E_n|e^{i\frac{E_n}{\hbar}t}\hat{R}e^{-i\frac{E_n}{\hbar}t}|E_n\rangle = \langle E_n|\hat{R}|E_n\rangle \tag{4.394}$$

or $\langle\hat{R}\rangle$ is independent of t (for this state). This implies that $\langle f(\hat{R})\rangle$ is also independent of t, which finally, implies that $\langle\theta(x-\hat{R})\rangle$ is independent of t. This means that, in this state,

$$Prob(x < R|\psi) \text{ is independent of } t \tag{4.395}$$

This kind of state is called a *stationary state*. In a stationary state the expectation values and probabilities of all observables are independent of time. If an observable is a constant of the motion, however, then this would be true for

any state and not just a stationary state. *So these are very different physical concepts.*

Now, if $[\hat{K}, \hat{H}] = 0$, then $\hat{K}$ and $\hat{H}$ have a common set of eigenvectors (it is a complete set). But the eigenvectors of $\hat{H}$ are stationary states. This means that we can prepare systems in stationary states where *both* the energy and the observable represented by $\hat{K}$ have *definite* values (no dispersion).

Suppose we have a set of mutually commuting observables and that they all also commute with $\hat{H}$. Then they have *a* common set of eigenvectors.

We will use the eigenvalues of $\hat{H}$ and all the eigenvalues of this mutually commuting set of observables to label state vectors. We will call the labels *quantum numbers*. They will designate *all we know* about a state vector.

4.16 The Collapse or Reduction Postulate?

Based on unitary time evolution postulate, a system consisting of a quantum system (Q-system) and a measurement system (M-system), would necessarily evolve in this way

$$\begin{aligned} |initial\rangle &= (a\,|+\rangle_Q + b\,|-\rangle_Q)\,|0\rangle_M \\ &\rightarrow |final\rangle = \rightarrow a\,|+\rangle_Q\,|+1\rangle_M + b\,|-\rangle_Q\,|-1\rangle_M \end{aligned} \tag{4.396}$$

which is a superposition of Q-states and M-states. We assume that the M-states represent *macroscopic* pointer locations on some meter.

This says that *measurement*, within the framework of the standard postulates, *CORRELATES* or *ENTANGLES* the *dynamical variable* (Q-system) to be measured and the *macroscopic* (M-system) indicator that can be *directly* (macroscopically) observed.

Derivation: Suppose that the meter has eigenvectors (labeled by eigenvalues)

$$\begin{aligned} |+\rangle_M &\Rightarrow \text{ meter on: reading } = +1 \\ |-\rangle_M &\Rightarrow \text{ meter on: reading } = -1 \\ |0\rangle_M &\Rightarrow \text{ meter off: reading } = 0 \end{aligned}$$

and the system has eigenvectors (labeled by eigenvalues)

$$\begin{aligned} |+\rangle_Q &\Rightarrow \text{ value } = +1 \\ |-\rangle_Q &\Rightarrow \text{ value } = -1 \end{aligned}$$

The initial state is

$$|initial\rangle = (a\,|+\rangle_Q + b\,|-\rangle_Q)\,|0\rangle_M \tag{4.397}$$

which represents the quantum system in a superposition and the meter off.

We are interested in the evolution of this state according to quantum mechanics.

If, instead of the above initial state, we started with the initial state

$$|A\rangle = |+\rangle_Q |0\rangle_M \tag{4.398}$$

and then turn on the meter, this state must evolve into

$$|A'\rangle = |+\rangle_Q |+1\rangle_M \tag{4.399}$$

indicating that the meter has measured the appropriate value (that is the definition of a "good" meter).

Similarly, if, instead of the above initial state, we started with the initial state

$$|B\rangle = |-\rangle_Q |0\rangle_M \tag{4.400}$$

and then turn on the meter, this state must evolve into

$$|B'\rangle = |-\rangle_Q |-1\rangle_M \tag{4.401}$$

indicating that the meter has measured the appropriate value (again, that is the definition of a "good" meter).

If the system is in the initial state corresponding to a superposition of these two special states, however, then the linearity of quantum mechanics says that it must evolve into

$$|final\rangle = a |+\rangle_Q |+1\rangle_M + b |-\rangle_Q |-1\rangle_M \tag{4.402}$$

as we assumed above(4.395).

Interpreting the state vector: Two models....

1. Pure state $|\psi\rangle$ implies a *complete* description of an individual Q-system. This corresponds to the statement that a dynamical variable $\hat{P}$ has the value p in the state $|\psi\rangle$ if and only if $\hat{P}|\psi\rangle = p|\psi\rangle$.
2. Pure state $|\psi\rangle$ implies *statistical* properties of an *ensemble of similarly prepared* systems.

Interpretation (1) is the standard interpretation espoused by 90% of all physicists. It assumes that, because the state vector plays the most important role in the mathematical formalism of QM, it must have an equally important role in the interpretation of QM, so that

$$\text{Properties of world} \Leftrightarrow \text{Properties of } |\psi\rangle \tag{4.403}$$

Interpretation (1) by itself is not consistent with the unitary evolution postulate, that is, the state $|final\rangle$ as defined in (4.401) is not equal to an eigenvector of any indicator (macroscopic pointer) variable. This means that the pointer (of the meter) will *flutter* since the $|\pm\rangle$ states could be macroscopically separated. Since we never observe this flutter, any interpretation of $|final\rangle$ as a description of an individual system cannot be reconciled with both observation and unitary time evolution.

Interpretation (2) has no such difficulties. $|\psi\rangle$ is just an abstract mathematical object which implies the probability distributions of the dynamical variables of an ensemble. It represents a *state of knowledge*.

Physicists that believe interpretation (1) are forced to introduce a *new postulate* at this point to remove these difficulties. This is the so-called *reduction/collapse* of the state vector postulate, which says that during any measurement we have a *new real process* which causes the *transition*

$$|final\rangle \to a\,|+\rangle_Q\,|+1\rangle_M \;\text{ or }\; \to b\,|-\rangle_Q\,|-1\rangle_M \tag{4.404}$$

so that we end up with an eigenvector of the indicator variable and thus there will be no flutter.

Various *reasons* are put forth for making this assumption, i.e.,

measurements are repeatable

Since this experiment(where the repeated measurement takes place immediately after the first measurement) has never been realized in the laboratory, I do not know what to make of a requirement like this one. In addition, in many experiments (like those involving photons), the system is destroyed by the measurement (photon is absorbed) making it silly to talk about a repeatable measurement.

The fact that the *reduction* process has never been observed in the laboratory makes it hard to understand in what sense it can it be thought of as a real physical process.

It is important to note that this difficulty only arises for interpretation (1) where statements are made about state vectors representing individual systems.

Some Proposed Mechanisms for the Reduction

1. **The reduction process is caused by an unpredictable and uncontrollable disturbance of the object by the measuring apparatus (a non-unitary process).**

   This means that the Hamiltonian of the system must take the form

   $$\hat{H} = \hat{H}_Q + \hat{H}_M + \hat{H}_{QM} \text{ where } \hat{H}_{QM} \rightarrow \text{ disturbance} \tag{4.405}$$

   This means, however, that it is already built into the standard unitary time evolution via $\hat{U} = e^{-i\hat{H}t/\hbar}$ and, thus, the disturbance terms can only lead to a final state that is still a superposition of indicator variable states. *IT DOES NOT WORK* unless we are not told what is meant by *unpredictable and uncontrollable disturbance!*

2. **The observer causes the reduction process when she reads the result of the measurement from the apparatus**.

   This is just a variation of (1). Here, the observer is just another indicator device. The new final state becomes

   $$|final\rangle = a\,|+\rangle_Q\,|+1\rangle_M\,|sees\ +1\rangle_O + b\,|-\rangle_Q\,|-1\rangle_M\,|sees\ -1\rangle_O \tag{4.406}$$

   which is still a superposition and thus is *NO HELP*. It also introduces *consciousness* into QM and *that, in my opinion, is just silly!*

3. **The reduction is caused by the environment (called decoherence), where by environment is meant the rest of the universe other than the Q-system and the M-system**.

   In this model, the environment is a very large system with an enormous number of degrees of freedom. We do not have any information about most of the degrees of freedom and thus must *average* over them. This causes pure states to *change* into nonpure or mixed states in a *non-unitary* process as we will see later in the book.

Why do many physicists think an individual Q-system must have its own state vector or wave function and then assume the collapse postulate?

IT WORKS for doing calculations!

This view has survived so long because it does not lead to any serious errors in most situations. Why?

In general, predictions in quantum mechanics are derived from $|\psi\rangle$ which gives the wave function and which, in turn, gives the probabilities. The operational significance of a probability is a relative frequency so that the experimentalist

has to invoke an ensemble of similar systems to make any comparisons with theory that is independent of any particular interpretation of the wave function. So that interpretation (2) is being used in the end anyway.

Does this mean that we should stop worrying about the interpretation of the wave function? NO!

But that is the subject of another book.....

In this book, we will not be dealing with such questions, that is, we do not ask questions that require the collapse postulate and use a different mathematical formalism for quantum mechanics.

What about interpretation (2)? It says that

> **A pure state describes the statistical properties of an ensemble of similarly prepared systems.**

This means that in many situations we must use the density operator $\hat{W}$ or $\hat{\rho}$ as the fundamental mathematical object of quantum mechanics instead of the state vector.

It turns out that some systems only have a density operator $\hat{\rho}$ and do not have a *legitimate* state vector $|\psi\rangle$.

For example, consider a box containing a very large number of electrons, each having spin = 1/2. As we shall see later, this means the spin can have a measurable component = $\pm 1/2$ along any direction. An oriented Stern-Gerlach device measures these spin components as we will see later.

Now, suppose the box has a hole so that electrons can get out and go into a Stern-Gerlach device oriented to measure z-components (an arbitrary choice). We will find the results

$$+\frac{1}{2} \quad 50\% \text{ of the time and } -\frac{1}{2} \quad 50\% \text{ of the time} \tag{4.407}$$

We then ask the question - what are the properties of the electrons in the box?

There are two possibilities, namely,

1. *Each individual* electron has the same state vector

$$|\psi\rangle_Q = \frac{1}{\sqrt{2}}|z = +1/2\rangle + \frac{1}{\sqrt{2}}|z = -1/2\rangle = |\psi\rangle_{box} \tag{4.408}$$

which is a superposition.

2. 1/2 of the electrons have $z = +1/2$ and 1/2 of the electrons have $z = -1/2$ so that

$$|\psi\rangle_Q = |z = +1/2\rangle \text{ OR } |z = -1/2\rangle \tag{4.409}$$

so that

$$|\psi\rangle_{BOX} = \frac{1}{\sqrt{2}}|z = +1/2\rangle + \frac{1}{\sqrt{2}}|z = -1/2\rangle \tag{4.410}$$

which seems to be the same state $|\psi\rangle_{box}$ as in (1), but it really *NOT* a superposition state in this case.

Therefore, it seems that we will not be able to tell which possibility is the correct one!

However, it will turn out that

$$|x - component = +1/2\rangle = \frac{1}{\sqrt{2}}|z = +1/2\rangle + \frac{1}{\sqrt{2}}|z = -1/2\rangle \tag{4.411}$$

so that, in case (1), if we orient the Stern-Gerlach device to measure x-components we would find all the electrons are in the same state $|x - component = +1/2\rangle$, that is, they are all the same!

On the other hand, in case (2) since (as we will see later)

$$|z = \pm 1/2\rangle = \frac{1}{\sqrt{2}}|x = +1/2\rangle \pm \frac{1}{\sqrt{2}}|x - 1/2\rangle \tag{4.412}$$

we would find that

$$+\frac{1}{2} \text{ give the } |x = +1/2\rangle \text{ result and } +\frac{1}{2} \text{ give the } |x = -1/2\rangle \text{ result} \tag{4.413}$$

Therefore, the *states are not the same!* If we try to write a state vector for case (2) we have to write

$$|\psi\rangle_Q = \frac{1}{\sqrt{2}}|z = +1/2\rangle + \frac{e^{i\alpha}}{\sqrt{2}}|z = -1/2\rangle \tag{4.414}$$

instead of

$$|\psi\rangle_{BOX} = \frac{1}{\sqrt{2}}|z = +1/2\rangle + \frac{1}{\sqrt{2}}|z = -1/2\rangle \tag{4.415}$$

where α is a *completely unknown relative phase factor*, which must be averaged over during any calculations since it is different for each separate measurement (each member of the ensemble). With that property for α, this is not a legitimate state vector in my opinion. We note that in a true superposition, the relative phase factors between components is known *exactly*!

If we use density matrices we have a different story. For a pure state we can always write $\hat{\rho} = |\psi\rangle\langle\psi|$ for some state vector $|\psi\rangle$.

In fact, case (1) gives

$$\hat{\rho} = \frac{1}{2}\left(|1/2\rangle\langle 1/2| + |1/2\rangle\langle -1/2| + |-1/2\rangle\langle 1/2| + |-1/2\rangle\langle -1/2|\right)$$
$$\Rightarrow \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{4.416}$$

where, as we saw earlier, the diagonal matrix elements represent probabilities. The existence of the off-diagonal matrix elements implies that we will observe quantum interference effects in this system.

Clearly, any pure state density operator cannot be written as the sum of pure state projection operators as we proved earlier.

In case (2), however, we have

$$\hat{\rho} = \frac{1}{2}\left(|1/2\rangle\langle 1/2| + |-1/2\rangle\langle -1/2|\right) \Rightarrow \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.417}$$

which clearly is the sum of pure state projection operators. This corresponds to a *nonpure or mixed state*. Note that the off-diagonals are zero so that this density operator cannot lead to any quantum interference effects as we might expect.

If we treat case(2) as a pure state with the extra relative phase factor we would obtain

$$\hat{\rho} = \frac{1}{2}\left(|1/2\rangle\langle 1/2| + e^{-i\alpha}|1/2\rangle\langle -1/2| + e^{i\alpha}|-1/2\rangle\langle 1/2| + |-1/2\rangle\langle -1/2|\right)$$
$$\Rightarrow \frac{1}{2}\begin{pmatrix} 1 & e^{-i\alpha} \\ e^{i\alpha} & 1 \end{pmatrix} \tag{4.418}$$

which becomes

$$\hat{\rho} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.419}$$

when we average over α. The decoherence process has this effect on a very short time scale.

We will return to this topic in Chapter 10 where we will repeat some of the material in section 4.16 and then work out a complete description of the so-called "measurement problem" and propose solutions which do not require a collapse postulate!

4.17 Putting Some of These Ideas Together

4.17.1 Composite Quantum Systems; Tensor Product

Let us now look at composite systems again but now in the context of a special kind of state called an "entangled" state, which illustrates some of the more dramatic features of quantum physics.

Most of our discussions so far apply easily to quantum systems comprised of only one part, i.e., a single particle. As we will see shortly, it will be straightforward to use this formalism to deal with a single particle evolving in the presence of external fields. In these cases, the external fields are treated as ordinary classical fields.

We did not, however, attempt to solve the system at the level where the particle is interacting(quantum mechanically) with the other particles that are actually generating the external fields. In this case all parts of the system must be dealt with using quantum mechanics.

We also indicated earlier in this chapter how we might set up a such a multiparticle system, without, however, indicating how this formalism might be used.

We now redo the multiparticle formalism and expand our discussion in several directions with the goal of describing a system where all the particles are interacting quantum mechanically.

Hilbert Space for Individual Quantum Systems

If we have a quantum system, then we can describe it as a vector

$$|\psi\rangle = c_1 |\phi_1\rangle + c_2 |\phi_2\rangle + \ldots + c_N |\phi_N\rangle \tag{4.420}$$

with respect to a set of basis vectors $|\phi_i\rangle$. The span (or set of all possible linear combinations) of these basis vectors make up the Hilbert space

$$\mathcal{H} = \{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_N\rangle\} \tag{4.421}$$

along with the inner product $\langle \phi_i | \phi_j \rangle = \delta_{ij}$. Writing out these basis vectors, we usually pick an ordering and assign them the unit vectors,

$$|\phi_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ . \\ . \end{pmatrix} \quad , \quad |\phi_2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ . \\ . \end{pmatrix} \quad , \ldots |\phi_N\rangle = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 1 \end{pmatrix} \tag{4.422}$$

Individual quantum systems live in their own individual Hilbert spaces as shown in the Figure 4.4 below.



Figure 4.4: Quantum systems are described by vectors in own Hilbert space

We know that we can write every possible state of the system as some superposition of the basis vectors.

We usually think about the standard projectors associated with the basis vectors

$$\hat{P}_1 = |\phi_1\rangle \langle\phi_1| \tag{4.423}$$

given by the outer products of the different basis vectors with its own dual vectors(or linear functionals) such as,

$$\hat{P}_1 = |\phi_1\rangle \langle\phi_1| = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 & \\ 0 & 0 & \cdot & \cdot & 0 & \\ & \cdot & \cdot & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & \end{pmatrix} \tag{4.424}$$

As we discussed earlier, the essential usefulness of the standard projectors is that their expectation values give us the probability that if we measured a system $|\psi\rangle$ to determine which state it is in, we would get the different basis vectors with probability of the form,

$$Prob(|\phi_1\rangle = \langle\psi| \hat{P}_1 |\psi\rangle = |c_1|^2 \tag{4.425}$$

We also remember that a key property of the projectors is that they sum to the identity

$$\sum_{i=1}^{N} \hat{P}_i = \hat{I} \tag{4.426}$$

Two-Level Systems

To make the discussion less unwieldy, we will work with quantum systems that live in a 2-dimensional Hilbert space, such as a a spin-1/2 particle or photon polarization (both of which will be discussed in detail in later chapters). For now we only need to know that our physical system(particle) has two eigenstates of some observable when it is measured in any direction (in physical space). We will call these states *up* and *down* in the direction of measurement. In particular, we

choose our basis to be given by the (up,down) states measured in the z-direction, which we designate as

$$\mathcal{H} = \{|\uparrow\rangle, |\downarrow\rangle\} \tag{4.427}$$

Notice that we picked an ordering for the basis vectors and we can therefore assign unit vectors to them,

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{4.428}$$

such that any state of the particle can be described as

$$|\psi\rangle = c_\uparrow |\uparrow\rangle + c_\downarrow |\downarrow\rangle \tag{4.429}$$

If we performed a measurement of the observable in the z-direction, we would get two outcomes with probabilities

$$Prob(\uparrow) = \langle\psi| \hat{P}_\uparrow |\psi\rangle = |c_\uparrow|^2 \quad , \quad Prob(\downarrow) = \langle\psi| \hat{P}_\downarrow |\psi\rangle = |c_\downarrow|^2 \tag{4.430}$$

using the corresponding projectors

$$\hat{P}_\uparrow = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad , \quad \hat{P}_\downarrow = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.431}$$

That is basically everything we need to know about a single particle.

Hilbert Space for Composite Systems

Let us begin building up the Hilbert space for two distinct particles (two distinct quantum systems). For example, suppose that I have a particle in my lab and you have a particle in your lab. They have never come in contact with one another and for all intensive purposes, I have no idea what you have done with your particle and vice versa. In this case, it seems to make perfect sense that we could just treat the particles completely independently at the level of their Hilbert spaces. We would have something like that shown in Figure 4.5 below.

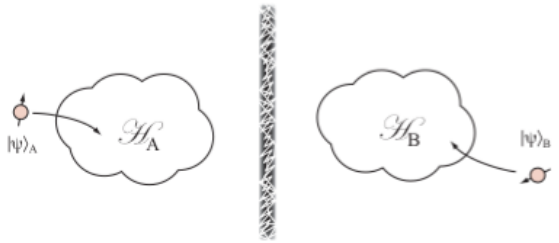


Figure 4.5: Hilbert space for 2 quantum systems independent of one another

We really should be able to think about these systems as entirely disjoint.

Therefore, we can define two different Hilbert spaces,

$$\mathcal{H}_A = \{|\uparrow\rangle_A, |\downarrow\rangle_A\} \tag{4.432}$$

and

$$\mathcal{H}_B = \{|\uparrow\rangle_B, |\downarrow\rangle_B\} \tag{4.433}$$

with operators such as projectors that only act on states in their respective systems

$$\hat{P}^A_\uparrow = |\uparrow\rangle_A \langle\uparrow|_A \ , \ \hat{P}^A_\downarrow = |\downarrow\rangle_A \langle\downarrow|_A \ , \ \hat{I}^A = \hat{P}^A_\uparrow + \hat{P}^A_\downarrow \tag{4.434}$$

and

$$\hat{P}^B_\uparrow = |\uparrow\rangle_B \langle\uparrow|_B \ , \ \hat{P}^B_\downarrow = |\downarrow\rangle_B \langle\downarrow|_B \ , \ \hat{I}^B = \hat{P}^B_\uparrow + \hat{P}^B_\downarrow \tag{4.435}$$

In terms of their matrices, for example,

$$\hat{P}^A_\uparrow = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_A \quad , \quad \hat{P}^B_\uparrow = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_B \tag{4.436}$$

these operators look identical to one another. However, they are not really identical, because you are only allowed to use operators (matrices) with A labels on states with A labels and operators with B labels on states with B labels. These rules reflect the situation that the particles are separate physical systems, possibly in distant locations, that have no idea that the other even exists.

Tensor Product of Hilbert Spaces

Now suppose we bring our individual particles from our independent labs together. In this situation, it is not clear whether we can get away with describing the two systems using two separate Hilbert spaces. For example, what if our particles interact with one another? can we still describe them as independent systems? It is not clear.

In order to be safe, we had better assume that we cannot still treat the systems as living in their own spaces and we must now assemble a suitable composite Hilbert space. The process is shown in Figure 4.6 below.

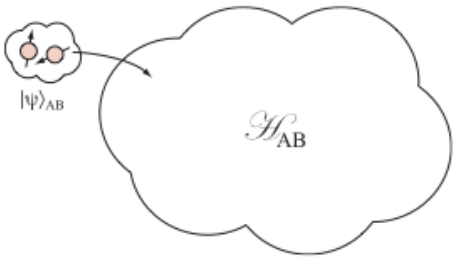


Figure 4.6: Hilbert space for 2 quantum systems independent of one another

In order to do so, let us begin at the level of describing the basis vectors of our new space. System A involves up and down basis vectors and so does system B. So, at the level of the basis vectors, if system A is up, B can be either up or down, corresponding to

$$\{|\uparrow\rangle_A |\uparrow\rangle_B , |\uparrow\rangle_A |\downarrow\rangle_B\} \tag{4.437}$$

or system A can be down and system B could be either up or down

$$\{|\downarrow\rangle_A |\uparrow\rangle_B , |\downarrow\rangle_A |\downarrow\rangle_B\} \tag{4.438}$$

Therefore, we build our composite Hilbert space with the four basis vectors

$$\mathcal{H}_{AB} = \{|\uparrow\rangle_A |\uparrow\rangle_B , |\uparrow\rangle_A |\downarrow\rangle_B , |\downarrow\rangle_A |\uparrow\rangle_B , |\downarrow\rangle_A |\downarrow\rangle_B\} \tag{4.439}$$

What are these funny objects involving some sort of product of basis kets? They cannot be any normal type of matrix-vector multiplication since you cannot multiply a column vector by a column vector. Instead, let us proceed as follows. Given our ordering of the four basis vectors, we can associate them with unit vectors. However, since there are now four basis vectors, our Hilbert space must be 4-dimensional,

$$|\uparrow\rangle_A |\uparrow\rangle_B = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} , \; |\uparrow\rangle_A |\downarrow\rangle_B = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{4.440}$$

$$|\downarrow\rangle_A |\uparrow\rangle_B = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} , \; |\downarrow\rangle_A |\downarrow\rangle_B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

This method of combining two 2-dimensional Hilbert spaces into a single 4-dimensional Hilbert space is known as a *tensor product*. We discussed this earlier and gave it a special symbol $\otimes$,

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B \tag{4.441}$$

to indicate that we are multiplying vector spaces together. Notice that the dimension of the tensor product Hilbert space is the product of the dimensions of the individual spaces.

Of course, since we have four basis vectors, we must now have four standard projectors of the type

$$\hat{P}^{AB}_{\uparrow\uparrow} = |\uparrow\rangle_A |\uparrow\rangle_B {}_A\langle\uparrow| {}_B\langle\uparrow| \tag{4.442}$$

This notation gets really clumsy after a while, so it is convention to shorten it to

$$\begin{aligned} \hat{P}^{AB}_{\uparrow\uparrow} &= |\uparrow\uparrow\rangle_{AB}\langle\uparrow\uparrow| \ , \ \hat{P}^{AB}_{\uparrow\downarrow} = |\uparrow\downarrow\rangle_{AB}\langle\uparrow\downarrow| \\ \hat{P}^{AB}_{\downarrow\uparrow} &= |\downarrow\uparrow\rangle_{AB}\langle\downarrow\uparrow| \ , \ \hat{P}^{AB}_{\downarrow\downarrow} = |\downarrow\downarrow\rangle_{AB}\langle\downarrow\downarrow| \end{aligned} \tag{4.443}$$

Tensor Product of Matrices-Repeating earlier ideas for clarity

We can compute the matrix representations of the projectors for our composite system by multiplying out each basis vector with its dual vector. These are 4-dimensional matrices

$$\begin{aligned} \hat{P}^{AB}_{\uparrow\uparrow} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \ , \ \hat{P}^{AB}_{\uparrow\downarrow} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ \hat{P}^{AB}_{\downarrow\uparrow} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \ , \ \hat{P}^{AB}_{\downarrow\downarrow} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{4.444}$$

Of course, it would be nice to have a systematic method for constructing operators on the tensor product Hilbert space from operators that act only on the individual Hilbert spaces. This is definitely not the standard matrix product, which we can see by looking at the projectors

$$\hat{P}^{AB}_{\uparrow\uparrow} \neq \hat{P}^{A}_{\uparrow}\hat{P}^{B}_{\uparrow} \tag{4.445}$$

since

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{4.446}$$

i.e., clearly their heir dimensions do not match. Instead, we need a tensor product for the projection matrices that reflects the same structure as the composite Hilbert space that we constructed. We symbolize such a product as

$$\hat{P}^{AB}_{\uparrow\uparrow} = \hat{P}^{A}_{\uparrow} \otimes \hat{P}^{B}_{\uparrow} \tag{4.447}$$

How do we perform this tensor product between matrices? Well first, it needs to be an operation that yields a matrix whose dimension is the product of the dimensions of the matrices being multiplied. Second, it has to respect the definition of the ordering that we used to construct the tensor product Hilbert space. Such a product is defined by the following

$$\begin{aligned} X^A \otimes Y^B &= \begin{pmatrix} X_{11}Y & X_{12}Y \\ X_{21}Y & X_{22}Y \end{pmatrix} \\ &= \begin{pmatrix} X_{11}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} & X_{12}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \\ X_{21}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} & X_{22}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \end{pmatrix} \end{aligned} \tag{4.448}$$

In other words, we take each element in the first matrix and replace it with a copy of the second matrix scaled by the element. This process, called the matrix tensor product seems to do the trick. We can check for projectors

$$\begin{aligned} \hat{P}^{AB}_{\uparrow\uparrow} = \hat{P}^A_{\uparrow} \otimes \hat{P}^B_{\uparrow} &= \begin{pmatrix} (1)\hat{P}^B_{\uparrow} & (0)\hat{P}^B_{\uparrow} \\ (0\hat{P}^B_{\uparrow} & (0)\hat{P}^B_{\uparrow} \end{pmatrix} \\ &= \begin{pmatrix} (1)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & (0)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ (0)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & (0)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

and the result checks out. You can verify it for the other projectors on your own.

4.17.2 Quantum Entanglement and the EPR *Paradox*

The operators

$$\hat{P}^A_{\uparrow} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_A \text{ and } \hat{P}^B_{\uparrow} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_B \tag{4.449}$$

look identical to one another. However, they are not really identical because you are only allowed to use operators (matrices) with A labels on states with A labels and operators with B labels in states with B labels. These rules, as we said earlier, reflect the situation that the particles are separate physical systems, possibly in distant locations, that have no idea that the other even exists.

States in the Tensor Product Hilbert Space

These ideas lead to a number of intriguing properties about composite quantum systems that we will now discuss.

Suppose that we took the state

$$|\psi\rangle_A = c_\uparrow^A |\uparrow\rangle_A + c_\downarrow^A |\downarrow\rangle_A \in \mathcal{H}_A \tag{4.450}$$

from Alice's lab, and the state

$$|\psi\rangle_B = c_\uparrow^B |\uparrow\rangle_B + c_\downarrow^B |\downarrow\rangle_B \in \mathcal{H}_B \tag{4.451}$$

from Bob's lab. Alice and Bob both prepared their systems completely independently - with no communication between them to indicate what they were doing in their own labs. When we bring their systems together, we should express them as vectors in the composite Hilbert space

$$|\psi\rangle_A |\psi\rangle_B = (c_\uparrow^A |\uparrow\rangle_A + c_\downarrow^A |\downarrow\rangle_A) \otimes (c_\uparrow^B |\uparrow\rangle_B + c_\downarrow^B |\downarrow\rangle_B) \in \mathcal{H}_{AB} \tag{4.452}$$

After taking the tensor product between these states, we get

$$|\psi\rangle_{AB} = c_\uparrow^A c_\uparrow^B |\uparrow\uparrow\rangle_{AB} + c_\uparrow^A c_\downarrow^B |\uparrow\downarrow\rangle_{AB} + c_\downarrow^A c_\uparrow^B |\downarrow\uparrow\rangle_{AB} + c_\downarrow^A c_\downarrow^B |\downarrow\downarrow\rangle_{AB} \tag{4.453}$$

The important point to notice in this expression is that there is a special relationship between the coefficients: the coefficient for each basis vector is the product of the coefficients for the individual basis vectors in this case.

Quantum Entanglement

In (4.452), we found the expression for two, unrelated, independent quantum systems expressed in the Hilbert space of the composite system. However, we know that, in general, we can describe an arbitrary state in $\mathcal{H}_{AB}$ as an arbitrary superposition of the four basis vectors

$$|\psi\rangle_{AB} = c_{\uparrow\uparrow} |\uparrow\uparrow\rangle_{AB} + c_{\uparrow\downarrow} |\uparrow\downarrow\rangle_{AB} + c_{\downarrow\uparrow} |\downarrow\uparrow\rangle_{AB} + c_{\downarrow\downarrow} |\downarrow\downarrow\rangle_{AB} \tag{4.454}$$

This is the most general possible expression for the composite states since we made sure to use the most general basis possible and we imposed no relationship between the expansion coefficients (other than the state must be normalized).

Here is the big question: can every state of the form (4.453)

$$|\psi\rangle_{AB} = c_{\uparrow\uparrow} |\uparrow\uparrow\rangle_{AB} + c_{\uparrow\downarrow} |\uparrow\downarrow\rangle_{AB} + c_{\downarrow\uparrow} |\downarrow\uparrow\rangle_{AB} + c_{\downarrow\downarrow} |\downarrow\downarrow\rangle_{AB}$$

be written in the product form (4.452)

$$|\psi\rangle_{AB} = c_\uparrow^A c_\uparrow^B |\uparrow\uparrow\rangle_{AB} + c_\uparrow^A c_\downarrow^B |\uparrow\downarrow\rangle_{AB} + c_\downarrow^A c_\uparrow^B |\downarrow\uparrow\rangle_{AB} + c_\downarrow^A c_\downarrow^B |\downarrow\downarrow\rangle_{AB}$$

The answer is a resounding *no*. Consider the simple example

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} |\uparrow\downarrow\rangle_{AB} + \frac{1}{\sqrt{2}} |\downarrow\uparrow\rangle_{AB} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \tag{4.455}$$

Here, if we were to try and find a product state expression for this special composite state vector, we would first infer that the product of the individual basis vectors must satisfy

$$c_{\uparrow}^{A} c_{\downarrow}^{B} = c_{\downarrow}^{A} c_{\uparrow}^{B} = \frac{1}{\sqrt{2}} \tag{4.456}$$

and

$$c_{\uparrow}^{A} c_{\uparrow}^{B} = c_{\downarrow}^{A} c_{\downarrow}^{B} = 0 \tag{4.457}$$

which are not consistent with each other, i.e., there is no solution!

Therefore, we must conclude that the state (4.454)

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} |\uparrow\downarrow\rangle_{AB} + \frac{1}{\sqrt{2}} |\downarrow\uparrow\rangle_{AB} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

cannot be expressed as a product between states in $\mathcal{H}_A$ and $\mathcal{H}_B$. States in a composite Hilbert space, such as this one, that cannot be factorized into a product of states in the constituent Hilbert spaces are referred to as being *entangled*.

Consequences of Entanglement

I strongly believe that entanglement is the most important difference between quantum and classical physics. Let us see why by considering the following situation. Suppose that Alice and Bob get together in a common lab and prepare the joint state (4.454)

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} |\uparrow\downarrow\rangle_{AB} + \frac{1}{\sqrt{2}} |\downarrow\uparrow\rangle_{AB} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

After it is made, Alice takes one of the particles and Bob takes the other (we can assume Alice takes the first one and Bob takes the second without loss of generality). Then both of them return to the individual labs without changing the state of their respective particles. When she gets back home, Alice decides to perform a measurement on her particle to see what state it is in. Such a measurement, performed by Alice on only part of the pair of particles, would be described by operators such as the projectors (remember projectors give us probabilities)

$$\hat{P}_{\uparrow}^{A} \otimes \hat{I}_{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.458}$$

and

$$\hat{P}^A_{\downarrow} \otimes \hat{I}_B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.459}$$

Notice that these projectors have the correct dimension since they are the tensor product of 2-dimensional operators. Also notice that this operator should be interpreted as doing something to Alice's particle (the projector part) and nothing to Bob's particle as indicated by the identity operator acting on Bob's system. The identity operator is the quantum mechanical way of saying that you did not do anything.

Operators of the above form, a tensor product of a projector on one component Hilbert space with the identity on the other, are called *partial projectors*. First, let us compute the probability that when Alice measures her particle she obtains the *up* outcome

$$\begin{aligned} Prob^A(\uparrow) &= {}_{AB}\langle\psi|\, \hat{P}^A_{\uparrow} \otimes \hat{I}_B\, |\psi\rangle_{AB} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2} \end{aligned} \tag{4.460}$$

as well as for the *down* outcome

$$\begin{aligned} Prob^A(\downarrow) &= {}_{AB}\langle\psi|\, \hat{P}^A_{\downarrow} \otimes \hat{I}_B\, |\psi\rangle_{AB} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2} \end{aligned} \tag{4.461}$$

But here is the truly amazing part. What state do we have following Alice's measurement? Suppose that she obtains the *up* outcome, then(by the reduction postulate) we get the measurement eigenstate

$$|\psi'\rangle_{AB} = \frac{\hat{P}^A_{\uparrow} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi|\, \hat{P}^A_{\uparrow} \otimes \hat{I}_B\, |\psi\rangle_{AB}}} |\psi\rangle_{AB} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = |\uparrow\downarrow\rangle_{AB} \tag{4.462}$$

where the denominator is to maintain the normalization.

Similarly, when Alice measures *down* for her particle, the result is the projection.

$$|\psi'\rangle_{AB} = \frac{\hat{P}^A_{\downarrow} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi|\, \hat{P}^A_{\downarrow} \otimes \hat{I}_B\, |\psi\rangle_{AB}}} |\psi\rangle_{AB} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = |\downarrow\uparrow\rangle_{AB} \tag{4.463}$$

The interpretation of this result is that Alice knows what Bob's state is as soon as she performs her measurement. Therefore, a local operation in her lab tells her global information about the composite state.

This prediction of quantum mechanics was considered so bizarre that it prompted Einstein, Podolsky and Rosen to essentially denounce quantum mechanics as either being wrong or at least incomplete(we will define this carefully later in the book). In order to make their point, the three authors proposed the following situation, which today is called the EPR Paradox(after their initials). EPR argued that Alice and Bob could get together and prepare the state just discussed. Then Bob would climb aboard a rocket and fly to a distant planet. At that point, Alice and Bob would both measure their states and instantaneously know the result of the other person's experiment - despite the long distance between them. This seemingly violates special relativity since it would mean that Alice's and Bob's physical systems somehow exchanged information at speeds faster than the speed of light.

4.17.3 Entanglement and Communication

To Communicate Superluminally, or Not

The immediate question we may ask is whether or not this bizarre information gain(what Alice learns about the state of Bob's particle) allows them to communicate faster than the speed of light. The answer is absolutely, definitely, certainly, *NO*! The short answer is that the randomness of the measurement outcomes saves us, but, this question was still of great concern during quantum mechanics' infancy, since it led to a number of apparent paradoxes that took some time to resolve. We will now dispel these rumors of superluminal communication once and for all.

In order to do so, we begin from the state shared by Alice and Bob

$$|\psi_+\rangle_{AB} = \frac{1}{\sqrt{2}}|\uparrow\downarrow\rangle_{AB} + \frac{1}{\sqrt{2}}|\downarrow\uparrow\rangle_{AB} \tag{4.464}$$

and suppose that Alice measures the observable of her state along the z-direction(as we have been doing). Then her possible outcomes correspond to the partial projectors

$$\hat{P}^A_{\uparrow_z} \otimes \hat{I}^B = |\uparrow\rangle_A \langle\uparrow| \otimes \hat{I}^B \text{ and } \hat{P}^A_{\downarrow_z} \otimes \hat{I}^B = |\downarrow\rangle_A \langle\downarrow| \otimes \hat{I}^B \tag{4.465}$$

When Alice measures her up outcome along the z-direction, her state transforms to

$$\uparrow_z : |\psi_+\rangle_{AB} \rightarrow \frac{\hat{P}^A_{\uparrow_z} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi_+|\hat{P}^A_{\uparrow_z} \otimes \hat{I}_B|\psi_+\rangle_{AB}}}|\psi_+\rangle_{AB} = |\uparrow\downarrow\rangle_{AB} \tag{4.466}$$

and when she measures down outcome along the z-direction she gets

$$\downarrow_z\colon |\psi_+\rangle_{AB} \rightarrow \frac{\hat{P}^A_{\downarrow_z} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi_+|\, \hat{P}^A_{\downarrow_z} \otimes \hat{I}_B\, |\psi_+\rangle_{AB}}} |\psi_+\rangle_{AB} = |\downarrow\uparrow\rangle_{AB} \tag{4.467}$$

Now, what if Alice decides to perform an alternative procedure given by measuring whether her particle is up or down along the x-direction. Now the projectors are given by (using corresponding eigenstates for the x-direction which we will derive for spin and polarization in later chapters)

$$\hat{P}^A_{\uparrow_x} \otimes \hat{I}^B = \frac{1}{2}(|\uparrow\rangle + |\downarrow\rangle)\langle\uparrow| + \langle\downarrow|) \otimes \hat{I}^B \tag{4.468}$$

and

$$\hat{P}^A_{\downarrow_x} \otimes \hat{I}^B = \frac{1}{2}(|\uparrow\rangle - |\downarrow\rangle)\langle\uparrow| - \langle\downarrow|) \otimes \hat{I}^B \tag{4.469}$$

i.e., we use

$$|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) \text{ and } |\downarrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle) \tag{4.470}$$

in forming Alice's projectors. If we work though the algebra we find for the resulting state when Alice obtains up outcome in the x-direction

$$\uparrow_x\colon |\psi_+\rangle_{AB} \rightarrow \frac{\hat{P}^A_{\uparrow_x} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi_+|\, \hat{P}^A_{\uparrow_z} \otimes \hat{I}_B\, |\psi_+\rangle_{AB}}} |\psi_+\rangle_{AB} = |\uparrow_x\uparrow_x\rangle_{AB} \tag{4.471}$$

meaning that when Alice measures up in the x-direction, then Bob also measures up in the x-direction. Conversely, when Alice measures down in the x-direction, the resulting state is

$$\downarrow_x\colon |\psi_+\rangle_{AB} \rightarrow \frac{\hat{P}^A_{\downarrow_x} \otimes \hat{I}_B}{\sqrt{{}_{AB}\langle\psi_+|\, \hat{P}^A_{\downarrow_x} \otimes \hat{I}_B\, |\psi_+\rangle_{AB}}} |\psi_+\rangle_{AB} = |\downarrow_x\downarrow_x\rangle_{AB} \tag{4.472}$$

and Bob also measures down in the x-direction.

Using this for a Communication System

We can now see whether it is possible for Alice and Bob to communicate by choosing one basis or another when they want to send different bits of data. For example, suppose that Bob wishes to send Alice binary zero. He might try to do so by selecting to measure in the z-direction (use the z-basis). And when he wishes to send logical one, he would then measure in the x-direction (use the x-basis). The two parties could agree upon the following communication protocol:

1. Bob selects a basis, either z or x, to send logical zero or one, respectively and measures. If he measures up in that basis, the bit is considered good, otherwise he throws away the bit.

2. When Alice chooses to measure in the z-basis (provided Bob also measured in the z-basis), she know that a logical zero has been sent when she measures down (meaning Bob found up since the measurements are anti-correlated). When Alice measures in the x-basis (provided that Bob measured in the x-basis), she knows that Bob sent her a logical one when she measures up (also implying that Bob measured up since the measurements are correlated).

Unfortunately for people like Einstein (who were trying to find quantum mechanical paradoxes), this communication scheme conveys no information from Bob to Alice. This is because Alice must pick her measurement basis randomly. She cannot know the order in which to choose her measurement bases in advance, otherwise she would have just carried the classical information along with her. Instead, she must guess by flipping a coin prior to each measurement. Therefore, she picks the x-basis with probability 1/2 and the z-basis with probability 1/2. Unfortunately, when she picks the wrong basis, her measurement is perfectly uncorrelated with Bob's, so no information can be conveyed.

4.17.4 Nonlocality and Tests of Quantum Entanglement

We have seen that entanglement between different components of a composite quantum system are capable of displaying strong correlations between the outcomes of measurements performed on separate components. In particular, we have been looking at states such as

$$|\psi_+\rangle_{AB} = \frac{1}{\sqrt{2}}|\uparrow\downarrow\rangle_{AB} + \frac{1}{\sqrt{2}}|\downarrow\uparrow\rangle_{AB} = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\1\\0\end{pmatrix} \in \mathcal{H}_{AB} \tag{4.473}$$

in the composite Hilbert space $\mathcal{H}_{AB}$ shared by both Alice and Bob. An essential point to remember, something we learned earlier, is that it is impossible to factorize this state

$$|\psi_+\rangle_{AB} \neq |\psi_A\rangle_A |\psi_B\rangle_B \tag{4.474}$$

for any possible choice of states. In other words, there is no way for Alice and Bob to make this state independently (think of the joint Hilbert space $\mathcal{H}_{AB}$ as being the set of states they can make together and the individual Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ as the sets of states that can be made independently.

Indeed, this prediction of quantum mechanics is somewhat bizarre. It tells us that Alice and Bob working together have more power than they do working separately. The question we want to address now is whether or not it would be possible for Alice or Bob to mimic anything that behaves like quantum entanglement using classical physics. Classically, we would expect to have the following properties hold true

1. *Objective Reality* meaning that even though Alice and Bob do not know the outcome of measurements they might perform on their physical systems, the particle itself *knows* what outcome it will produce when measured.

2. *Local Determinism* meaning that if Alice and Bob have their particles in different space-like separated locations, anything Alice does should have no effect on what Bob measures and vice versa. In other words, Alice's measurement outcome should be determined only by what she has in her lab with her.

Bell Inequalities

In order to test whether these classical assumptions are true, John Bell suggested the following type of experiment back in the 1960's. Suppose that there is a pair source midway between Alice's and Bob's labs (which are distant from one another). That is to say, some device produces arbitrary (possibly random) states in the joint Hilbert space $\mathcal{H}_{AB}$ and then sends one half of the pair to Alice and one half to Bob. Both Alice and Bob know when to expect these particles to arrive, but have no idea what state was made by the source.

Alice and Bob each have a device to measure(up or down) their incoming particle along one of two different possible axes (directions). Therefore, think of the measurement devices each having a switch to select between the two different possible measurement directions and an output meter to indicate whether the measurement outcome is up or down along that direction. We will label the two different possible measurement directions by A,A' and B,B' (for example, A might mean to measure along the z-direction and A' along the y-direction, and so on).

Prior to each particle arrival, Alice and Bob each pick an independent direction to measure by setting their respective switches, wait for the measurement outcome and then record the result. They can repeat this process many times such that their measurement record might look something like:

A	-1	B'	-1
A	-1	B	$+1$
A'	$+1$	B'	$+1$
A	-1	B	-1
A	-1	B'	$+1$
A'	$+1$	B	-1
A'	-1	B'	$+1$
A	$+1$	B	$+1$
A'	-1	B	-1
A	$+1$	B	-1

Table 4.1: Sample Data

where we indicate *up* by $+1$ and *down* by -1.

Given this structure for the experiment, let us return to our concepts of local determinism and objective reality. Well, objective reality suggests that each particle should "know" whether it will be up or down for both possible measurement settings. Local determinism suggests Alice's measurement should be independent of the switch setting on Bob's measurement instrument.

Internal Information

Under the combined assumptions of local determinism and objective reality, each particle must have the following information encoded into their internal states

$$A_n = \pm 1 \quad A'_n = \pm 1 \quad B_n = \pm 1 \quad B'_n = \pm 1 \tag{4.475}$$

Each of the four possible measurement labels are treated as random variables, with the subscript indicating which round of the measurement. So, for example, $A_4 = +1$, would mean that if Alice measures along the direction A in round $n = 4$, she will get *up*.

Now, let us consider the following function of the random variables

$$g_n = A_n B_n + A'_n B_n + A_n B'_n - A'_n B'_n \tag{4.476}$$

Now, were we to tabulate the value of g_n for all possible choices of A_n, A'_n, B_n, B'_n, we would find that $g_n = \pm 2$ always. We can try this out for a set of possible values, for example,

$$A_n = +1, A'_n = -1, B_n = -1, B'_n = -1 \rightarrow g_n = -1 + 1 - 1 - 1 = -2 \tag{4.477}$$

In any case, we may write down the inequality

$$|\frac{1}{N}\sum_{n=1}^{N} g_n| = \frac{1}{N}|\sum_{n=1}^{N} A_n B_n + \sum_{n=1}^{N} A'_n B_n + \sum_{n=1}^{N} A_n B'_n - \sum_{n=1}^{N} A'_n B'_n| \leq 2 \tag{4.478}$$

In other words, since the extrema of g_n are ± 2, then the average over many trials must be no larger than $+2$ and no smaller than -2. Thus, the absolute value of the average must be no greater than 2.

We will redo the discussion in section 4.17.2 and 4.17.3, prove all of these results in detail and greatly expand the discussion of these topics in Chapter 12.

This inequality is one of several that are know as Bell inequalities (this form developed by Clauser, Horne, Shimony and Holt). The true genius of such an inequality is that it provides us with a simple test, based only on probability theory, as to whether or not local determinism and objective reality are valid assumptions.

Violations of Bell's Inequality

As you might have already guessed, quantum mechanics violates this inequality. For a simple example, let us consider the following scenario, which we will prove in detail in Chapter 17. Assume that the pair source produces the following state

$$|\psi_-\rangle_{AB} = \frac{1}{\sqrt{2}}|\uparrow\downarrow\rangle_{AB} - \frac{1}{\sqrt{2}}|\downarrow\uparrow\rangle_{AB} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ +1 \\ -1 \\ 0 \end{pmatrix} \tag{4.479}$$

Also assume that Alice's measurement A corresponds to measuring along the axes

$$A = \hat{z} \ , \ A' = \cos\phi\hat{z} + \sin\phi\hat{x} \tag{4.480}$$

while Bob, on the other hand, measures along the axes

$$B = \hat{z} \ , \ B' = \cos\phi\hat{z} - \sin\phi\hat{x} \tag{4.481}$$

Now, all we have to do is compute the averages for all the terms in the inequality. We will learn how to do this for spin-1/2 particles and photon polarization in Chapters 7 and 9. For now, we just state the results

$$\frac{1}{N}\sum_{n=1}^{N} A_n B_n = -1 \quad , \quad \frac{1}{N}\sum_{n=1}^{N} A'_n B_n = -\cos\phi \tag{4.482}$$

$$\frac{1}{N}\sum_{n=1}^{N} A_n B'_n = -\cos\phi \quad , \quad \frac{1}{N}\sum_{n=1}^{N} A'_n B'_n = \cos 2\phi \tag{4.483}$$

If we combine all of these results we find that

$$|\langle g_n\rangle| = \frac{1}{N}|\sum_{n=1}^{N} g_n| = |-1 - 2\cos\phi + \cos 2\phi| \tag{4.484}$$

We can now plot the expectation of $g_n(\phi)$ and look for violations of the Bell inequality.

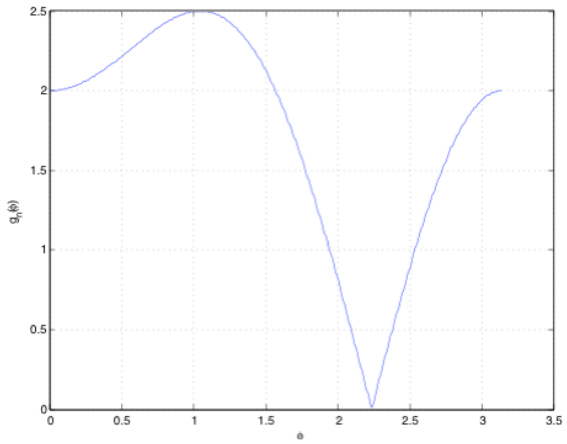


Figure 4.7: Quantum Mechanics versus the Bell Inequality

Clearly, we see violation for a range of values of ϕ.

As we stated earlier, we will have much more to say about this interesting subject in Chapter 12.

Finally, let us summarize once more some of the knowledge we have acquired about density operators.

4.18 Expanding on the Density Operator and the Statistical Description of Quantum Systems

In classical mechanics the state of a system is determined by a point in phase space. If we do not know the exact positions and momenta of all particles in the system, we need to use the probability density function to describe the system statistically

In quantum mechanics the state of a system is characterized by a state vector in Hilbert space that contains all the relevant information. To prepare a vector state for a quantum system at a given time, it suffices to perform a set of measurements on the system corresponding to a complete set of commuting observables. However, in practice such measurements are often impossible. The problem is how to incorporate the incompleteness of our information into the formalism of quantum mechanics in order to describe the state of such a system.

Incompleteness can be formally described in two ways:

(1) One way is to characterize the system as a member of a mixed ensemble

in which we do not know the state vector of every member, but we know only the probability that an arbitrary member can be found in a specified state vector. For example, 70% of the members are characterized by $\left|\psi^{(1)}\right\rangle$ and the remaining by $\left|\psi^{(2)}\right\rangle$ (these states are not necessarily orthogonal). We will investigate this approach later and show that such an ensemble cannot be characterized by an averaged state vector.

(2) The second way in which a description of the system by a state vector is impossible is for systems that interact with others. Such systems are called *open systems* and are of particular interest in quantum optics. In such cases we are interested in only a part of the entire system. It is convenient to separate the system of primary interest from that of secondary interest and to call the former the *system* and the latter the *reservoir*. We can eliminate the reservoir by using the reduced density operator method as we will describe later.

4.18.1 Statistical Description of Quantum Systems and the Nonexistence of an Averaged Quantum State

Consider a mixed ensemble of similar systems such that our information about the state of the members is limited to the probability distribution over some specified state vectors of the system $\{|\alpha\rangle, |\beta\rangle, |\gamma\rangle, ...\}$ that are not necessarily orthogonal. For simplicity we assume that they are the eigenvectors of one of the observables of the system and form a complete basis for the Hilbert space.

Let A be a Hermitian operator of the system with eigenvectors $\{|m\rangle\}$ for $m = 1, 2, ..., N$, where N is the dimension of the Hilbert space of the system. Assume that our information about the system is limited to the set of probabilities $\{P_m\}$ for finding the system in the m^{th} eigenvector. P_m satisfies the conditions

$$\begin{aligned} &0 \leq P_m \leq 1 \quad , \quad (m = 1, ..., N) \\ &\sum_{m=1}^{N} P_m = 1 \end{aligned} \tag{4.485}$$

It is reasonable to ask if the system can be described by a state vector. The state vector of such a system can be any vector such as

$$\left|\psi^{(k)}\right\rangle = \sum_{m=1}^{N} \sqrt{P_m} e^{i\phi_m^{(k)}} |m\rangle \tag{4.486}$$

where the phase $\phi_m^{(k)}$ can take on any real number in the interval $(-\pi, \pi)$. State vectors like Eq. (4.485) are called *accessible states* of the relevant system. The only parameters that discriminate one accessible state from another are the phases $\{\phi_m^{(k)}\}$. Because of their arbitrariness, there is no preference between these states. So the system can be found in each of its accessible states with equal probability. Therefore, the probability distribution over the accessible

states Eq. (4.485) is uniform, although the distribution over the complete basis $\{|m\rangle\}$ is nonuniform distribution like P_m.

If the number of accessible states in this ensemble is Ω, the probability that an arbitrary element of the ensemble is in the state $\left|\psi^{(k)}\right\rangle$ equals $1/\Omega$ (for every $k = 1, ..., \Omega$). So if we want to define an averaged quantum state for this system, we should multiply every accessible state by its corresponding probability and sum over all states:

$$\begin{aligned}\left|\bar{\psi}\right\rangle &= \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left|\psi^{(k)}\right\rangle \\ &= \sum_{k=1}^{\Omega} \frac{1}{\Omega} \sum_{m=1}^{N} \sqrt{P_m} e^{i\phi_m^{(k)}} |m\rangle \\ &= \sum_{m=1}^{N} \sqrt{P_m} \left(\frac{1}{\Omega} \sum_{k=1}^{\Omega} e^{i\phi_m^{(k)}} \right) |m\rangle \end{aligned} \tag{4.487}$$

Because the number of accessible states is very large, we can assume that the $\phi_m^{(k)}$'s vary continuously. Then we can change the sum to an integration, that is

$$\frac{1}{\Omega} \sum_{k=1}^{\Omega} e^{i\phi_m^{(k)}} \to \frac{1}{2\pi} \int_{-\pi}^{\pi} d\phi_m \, e^{i\phi_m} \tag{4.488}$$

Hence, the averaged quantum state is zero,

$$\left|\bar{\psi}\right\rangle = 0 \tag{4.489}$$

Therefore, such a system cannot be described by a state vector and we should look for another way to describe it.

We can assign an operator $\left|\psi^{(k)}\right\rangle \left\langle\psi^{(k)}\right|$ to every accessible state (4.485). Similar to Eq. (4.486), which defines the ensemble average of the accessible state vectors, we can define an ensemble average of these corresponding operators by introducing

$$\rho = \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left|\psi^{(k)}\right\rangle \left\langle\psi^{(k)}\right| \tag{4.490}$$

We will show that in spite of the zero average of the state vector, the average of ρ does not vanish. If we substitute Eq (4.485) into Eq. (4.489), we obtain

$$\begin{aligned}\rho &= \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left(\sum_{m=1}^{N} \sqrt{P_m} e^{i\phi_m^{(k)}} |m\rangle \right) \left(\sum_{n=1}^{N} \sqrt{P_n} e^{-i\phi_n^{(k)}} \langle n| \right) \\ &= \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left(\sum_{n,m=1}^{N} \sqrt{P_m P_n} e^{i\left(\phi_m^{(k)} - \phi_n^{(k)}\right)} |m\rangle \langle n| \right) \end{aligned} \tag{4.491}$$

We separate the double sums in Eq. (4.490) into two separate sums so that the first one contains the terms with the same m and n, and the second one contains

terms with different m and n and obtain

$$\rho = \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left(\sum_{n=1}^{N} P_n |n\rangle \langle n| \right) + \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left(\sum_{n \neq m} \sqrt{P_m P_n} e^{i\left(\phi_m^{(k)} - \phi_n^{(k)}\right)} |m\rangle \langle n| \right) \tag{4.492}$$

In the first sum all the terms are the same. In the second sum we can exchange the order of the sum over k with that over m and n and substitute $\theta_{mn}^{(k)} = \phi_m^{(k)} - \phi_n^{(k)}$:

$$\rho = \sum_{n=1}^{N} P_n |n\rangle \langle n| + \sum_{n \neq m} \sqrt{P_m P_n} \left(\sum_{k=1}^{\Omega} \frac{1}{\Omega} e^{i\theta_{mn}^{(k)}} \right) |m\rangle \langle n| \tag{4.493}$$

The second term in Eq. (4.492) vanishes because the sum over k can be changed by an integration over θ_{mn} just like Eq. (4.487) i.e.,

$$\int_{-\pi}^{\pi} d\theta \, e^{i\theta} = 0$$

Hence Eq. (4.493) reduces to

$$\rho = \sum_{n=1}^{N} P_n |n\rangle \langle n| \tag{4.494}$$

The operator ρ contains both quantum ($\left|\psi^{(k)}\right\rangle$) and classical statistical information (Ω). Hence, all the necessary information for a quantum statistical description of the system is contained in ρ. This idea can be clarified by calculating the ensemble average of an observable O of the system. For this purpose we multiply its expectation value in every accessible state $\left|\psi^{(k)}\right\rangle$ by its corresponding probability, that is, $1/\Omega$, and sum over all states:

$$\langle O \rangle = \sum_{k=1}^{\Omega} \frac{1}{\Omega} \langle O \rangle_k = \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left\langle \left|\psi^{(k)}\right\rangle \middle| O \middle| \psi^{(k)} \right\rangle \tag{4.495}$$

If we substitute the state vector Eq. (4.485) into Eq. (4.494) and follow the same procedure that was done earlier, we find that

$$\begin{aligned} \langle O \rangle &= \sum_{k=1}^{\Omega} \frac{1}{\Omega} \left(\sum_{n=1}^{N} \sqrt{P_n} e^{-i\phi_n^{(k)}} \langle n| \right) O \left(\sum_{m=1}^{N} \sqrt{P_m} e^{i\phi_m^{(k)}} |m\rangle \right) \\ &= \sum_{n=1}^{N} P_n \langle n| O |n\rangle = \sum_{n=1}^{N} P_n O_{nn} \end{aligned} \tag{4.496}$$

Therefore the ensemble average of an arbitrary operator can be evaluated by knowing the probability distribution $\{P_n\}$ and the complete basis $\{|n\rangle\}$, without knowing the state vector.

If we use the completeness of the basis $\{|n\rangle\}$, we can write Eq. (4.495) as

$$\begin{aligned}\langle O\rangle &= \sum_{n=1}^{N} P_n \langle n|\, O\, |n\rangle = \sum_{m,n=1}^{N} P_n \langle n|\, O\, |m\rangle \langle m\,|\, n\rangle \\ &= \sum_{m,n=1}^{N} P_n \langle m\,|\, n\rangle \langle n|\, O\, |m\rangle = \sum_{m=1}^{N} \langle m| \left(\sum_{n=1}^{N} P_n |n\rangle \langle n| \right) O\, |m\rangle \\ &= \sum_{m=1}^{N} \langle m|\, \rho O\, |m\rangle \end{aligned} \tag{4.497}$$

The right-hand side of Eq. (4.496) is the trace of the operator ρO and can be rewritten as

$$\langle O\rangle = Tr(\rho O) \tag{4.498}$$

The density operator, Eq. (4.493), is sufficient to calculate every ensemble average of the operators. Therefore the role of the density operator in the statistical description of quantum systems is similar to the role of the probability distribution function in the statistical description of classical systems.

4.18.2 Open Quantum Systems and the Reduced Density Operator

Most physical systems of interest are not isolated but interact with other systems. For example, an atom cannot be studied as an isolated system even if it is in a vacuum because it interacts with the vacuum state of the surrounding electromagnetic field. This interaction is the source of many interesting phenomena such as spontaneous emission, the natural line width, the Lamb shift, and quantum noise. To study such a system we eliminate the degrees of freedom of the environment. This elimination leads to incompleteness of information about the system of interest so that the description of the system by a state vector is no longer possible.

Suppose we are interested only in making measurements on a system S that interacts with its environment R. The Hilbert space of the composite system $S + R$ is the tensor product

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_R \tag{4.499}$$

where $\mathcal{H}_S$ and $\mathcal{H}_R$ are the Hilbert spaces of the system and reservoir, respectively.

Let H_S and H_R be the corresponding Hamiltonians, and $\{|s\rangle\}$ and $\{|r\rangle\}$ the energy eigenvectors of the system and its reservoir, respectively, in the absence of an interaction. Because the two systems are independent, their operators commute with each other:

$$[H_S, H_R] = 0 \tag{4.500}$$

When they are put into contact, the total Hamiltonian is

$$H = H_S + H_R + V_{RS} \tag{4.501}$$

where V_{RS} is the interaction between the two systems. If we prepare the composite system at the instant t_0 in a product state

$$|\Psi(t_0)\rangle = |\psi_S\rangle \otimes |\psi_r\rangle \tag{4.502}$$

then any measurement on S at this instant depends only on the state $|\psi_S\rangle$ and is independent of the state of the reservoir.

As time evolves, the state of the composite system evolves according to the Schrodinger equation

$$i\hbar \frac{\partial |\Psi(t)\rangle}{\partial t} = H |\Psi(t)\rangle \tag{4.503}$$

where H is the total Hamiltonian as in Eq. (4.500). In general $|\Psi(t)\rangle$ cannot be written in a product form such as Eq. (4.501) because the interaction energy V_{RS} depends on both systems. Such a state is called an *entangled state*.

The question is how we can describe the state of S and make measurements on it. To answer this question, let us consider an observable of the system S such as A_S and calculate its expectation value when the state of the composite system is $|\Psi(t)\rangle$. For this purpose we can use Eq. (4.497):

$$\langle \tilde{A}_S \rangle = Tr_{SR}(\rho \tilde{A}_S) \tag{4.504}$$

where the trace is over the complete basis $\{|s, r\rangle\}$ of the space $\mathcal{H}$, and ρ is the density operator of the composite system and is given by

$$\rho(t) = |\Psi(t)\rangle \langle \Psi(t)| \tag{4.505}$$

$\tilde{A}_S$ is the operator A_S in the composite space $\mathcal{H}$, that is,

$$\tilde{A}_S = A_S \otimes I_R \tag{4.506}$$

where I_R is the identity operator in the space $\mathcal{H}_R$.

A straightforward calculation then gives

$$\langle \tilde{A}_S \rangle = Tr_S(\rho_S \tilde{A}_S) \tag{4.507}$$

where ρ_S is the reduced density operator of the system S in the space $\mathcal{H}_S$ and is defined as

$$\rho_S = Tr_S(\rho) \tag{4.508}$$

As it is seen from Eq. (4.506), we can calculate the expectation values of all observables of the system S in its own state space $\mathcal{H}_S$ by using its reduced density operator.

A necessary and sufficient condition for a system to be in a pure state is that the trace of the square of its density operator equals unity. But in general $Tr\left(\rho_S^2\right)$

is not necessarily equal to unity even if the composite system is in a pure state. Therefore the state of an open system cannot generally be described by a state vector; instead it is described completely by the reduced density operator.

The most complete information about a quantum system is contained in its state vector, but usually our information about the system is not sufficiently complete to determine the state vector of the system. The incompleteness of our information can be incorporated into the formalism of quantum mechanics in two ways. One way is to describe the system as a member of a mixed ensemble in which our information is limited to a probability distribution over some specified state vectors of the system. We showed that the ensemble average of the accessible states of the system is zero and the only way to describe the system is by the density operator.

In the second case, we wish to make measurements on a part of a larger system. This part is considered to be an open system that interacts with the rest of the larger system (the environment or reservoir), which is not of interest to us. Even if we can determine the state vectors of the system and the reservoir at some instant of time, the interaction potential causes the composite system to evolve into another state in which the degrees of freedom of both systems are entangled so that we cannot separate the state of the system from that of the reservoir. In order to focus on the system of interest we eliminate the degrees of freedom of the reservoir. In this way we lose part of the information about the system that had been coded in the state of the composite system so that the description of the system by a state vector is no longer possible, and the system is described by the reduced density operator.

4.19 Problems

4.19.1 Can It Be Written?

Show that a density matrix $\hat{\rho}$ represents a state vector (i.e., it can be written as $|\psi\rangle\langle\psi|$ for some vector $|\psi\rangle$) if, and only if,

$$\hat{\rho}^2 = \hat{\rho}$$

4.19.2 Pure and Nonpure States

Consider an observable σ that can only take on two values +1 or −1. The eigenvectors of the corresponding operator are denoted by $|+\rangle$ and $|-\rangle$. Now consider the following states.

(a) The one-parameter family of pure states that are represented by the vectors

$$|\theta\rangle = \frac{1}{\sqrt{2}}|+\rangle + \frac{e^{i\theta}}{\sqrt{2}}|-\rangle$$

for arbitrary θ.

(b) The nonpure state

$$\rho = \frac{1}{2}\left|+\right\rangle\left\langle+\right| + \frac{1}{2}\left|-\right\rangle\left\langle-\right|$$

Show that $\langle\sigma\rangle = 0$ for both of these states. What, if any, are the physical differences between these various states, and how could they be measured?

4.19.3 Probabilities

Suppose the operator

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

represents an observable. Calculate the probability $Prob(M = 0|\rho)$ for the following state operators:

$$(a)\ \rho = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix},\ (b)\ \rho = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},\ (c)\ \rho = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

4.19.4 Acceptable Density Operators

Which of the following are acceptable as state operators? Find the corresponding state vectors for any of them that represent pure states.

$$\rho_1 = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{3}{4} \end{bmatrix} \quad , \quad \rho_2 = \begin{bmatrix} \frac{9}{25} & \frac{12}{25} \\ \frac{12}{25} & \frac{16}{25} \end{bmatrix}$$

$$\rho_3 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & 0 \end{bmatrix} \quad , \quad \rho_4 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix}$$

$$\rho_5 = \frac{1}{3}\left|u\right\rangle\left\langle u\right| + \frac{2}{3}\left|v\right\rangle\left\langle v\right| + \frac{\sqrt{2}}{3}\left|u\right\rangle\left\langle v\right| + \frac{\sqrt{2}}{3}\left|v\right\rangle\left\langle u\right|$$
$$\langle u \mid u\rangle = \langle v \mid v\rangle = 1 \text{ and } \langle u \mid v\rangle = 0$$

4.19.5 Is it a Density Matrix?

Let $\hat{\rho}_1$ and $\hat{\rho}_2$ be a pair of density matrices. Show that

$$\hat{\rho} = r\hat{\rho}_1 + (1 - r)\hat{\rho}_2$$

is a density matrix for all real numbers r such that $0 \le r \le 1$.

4.19.6 Unitary Operators

An important class of operators are unitary, defined as those that preserve inner products, i.e., if $|\tilde{\psi}\rangle = \hat{U}|\psi\rangle$ and $|\tilde{\varphi}\rangle = \hat{U}|\varphi\rangle$, then $\langle\tilde{\varphi}\,|\,\tilde{\psi}\rangle = \langle\varphi\,|\,\psi\rangle$ and $\langle\tilde{\psi}\,|\,\tilde{\varphi}\rangle = \langle\psi\,|\,\varphi\rangle$.

(a) Show that unitary operators satisfy $\hat{U}\hat{U}^{+} = \hat{U}^{+}\hat{U} = \hat{I}$, i.e., the adjoint is the inverse.

(b) Consider $\hat{U} = e^{i\hat{A}}$, where $\hat{A}$ is a Hermitian operator. Show that $\hat{U}^{+} = e^{-i\hat{A}}$ and thus show that $\hat{U}$ is unitary.

(c) Let $\hat{U}(t) = e^{-i\hat{H}t/\hbar}$ where t is time and $\hat{H}$ is the Hamiltonian. Let $|\psi(0)\rangle$ be the state at time $t = 0$. Show that $|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle = e^{-i\hat{H}t/\hbar}|\psi(0)\rangle$ is a solution of the time-dependent Schrodinger equation, i.e., the state evolves according to a unitary map. Explain why this is required by the conservation of probability in non-relativistic quantum mechanics.

(d) Let $\{|u_n\rangle\}$ be a complete set of energy eigenfunctions, $\hat{H}|u_n\rangle = E_n|u_n\rangle$. Show that $\hat{U}(t) = \sum_n e^{-iE_nt/\hbar}|u_n\rangle\langle u_n|$. Using this result, show that $|\psi(t)\rangle = \sum_n c_n e^{-iE_nt/\hbar}|u_n\rangle$. What is c_n?

4.19.7 More Density Matrices

Suppose we have a system with total angular momentum 1. Pick a basis corresponding to the three eigenvectors of the $z-$component of the angular momentum, J_z, with eigenvalues +1, 0, −1, respectively. We are given an ensemble of such systems described by the density matrix

$$\rho = \frac{1}{4}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

(a) Is ρ a permissible density matrix? Give your reasoning. For the remainder of this problem, assume that it is permissible. Does it describe a pure or mixed state? Give your reasoning.

(b) Given the ensemble described by ρ, what is the average value of J_z?

(c) What is the spread (standard deviation) in the measured values of J_z?

4.19.8 Scale Transformation

Space is invariant under the scale transformation

$$x \to x' = e^{c}x$$

where c is a parameter. The corresponding unitary operator may be written as

$$\hat{U} = e^{-ic\hat{D}}$$

where $\hat{D}$ is the *dilation* generator. Determine the commutators $\left[\hat{D}, \hat{x}\right]$ and $\left[\hat{D}, \hat{p}_x\right]$ between the generators of dilation and space displacements. Determine the operator $\hat{D}$. Not all the laws of physics are invariant under dilation, so the symmetry is less common than displacements or rotations. You will need to use the identity in Problem 6.19.11.

4.19.9 Operator Properties

(a) Prove that if $\hat{H}$ is a Hermitian operator, then $U = e^{iH}$ is a unitary operator.

(b) Show that $\det U = e^{iTrH}$.

4.19.10 An Instantaneous Boost

The unitary operator

$$\hat{U}(\vec{v}) = e^{i\vec{v}\cdot\hat{\vec{G}}}$$

describes the instantaneous ($t = 0$) effect of a transformation to a frame of reference moving at the velocity $\vec{v}$ with respect to the original reference frame. Its effects on the velocity and position operators are:

$$\hat{U}\hat{\vec{V}}\hat{U}^{-1} = \hat{\vec{V}} - \vec{v}\hat{I} \quad , \quad \hat{U}\hat{\vec{Q}}\hat{U}^{-1} = \hat{\vec{Q}}$$

Find an operator $\hat{G}_t$ such that the unitary operator $\hat{U}(\vec{v}, t) = e^{i\vec{v}\cdot\hat{\vec{G}}_t}$ will yield the full Galilean transformation

$$\hat{U}\hat{\vec{V}}\hat{U}^{-1} = \hat{\vec{V}} - \vec{v}\hat{I} \quad , \quad \hat{U}\hat{\vec{Q}}\hat{U}^{-1} = \hat{\vec{Q}} - \vec{v}t\hat{I}$$

Verify that $\hat{G}_t$ satisfies the same commutation relation with $\vec{P}$, $\vec{J}$ and $\hat{H}$ as does $\hat{G}$.

4.19.11 A Very Useful Identity

Prove the following identity, in which $\hat{A}$ and $\hat{B}$ are operators and x is a parameter.

$$e^{x\hat{A}}\hat{B}e^{-x\hat{A}} = \hat{B} + \left[\hat{A}, \hat{B}\right]x + \left[\hat{A}, \left[\hat{A}, \hat{B}\right]\right]\frac{x^2}{2} + \left[\hat{A}, \left[\hat{A}, \left[\hat{A}, \hat{B}\right]\right]\right]\frac{x^3}{6} + ...$$

There is a clever way(see Problem 4.19.12 below if you are having difficulty) to do this problem using ODEs and not just brute-force multiplying everything out.

4.19.12 A Very Useful Identity with some help....

The operator $U(a) = e^{ipa/\hbar}$ is a translation operator in space (here we consider only one dimension). To see this we need to prove the identity

$$e^A B e^{-A} = \sum_0^\infty \frac{1}{n!} \underbrace{[A,[A,...[A,}_{n} B \underbrace{\,]....]]}_{n}$$
$$= B + [A,B] + \frac{1}{2!}[A,[A,B]] + \frac{1}{3!}[A,[A,[A,B]]] + ...$$

(a) Consider $B(t) = e^{tA} B e^{-tA}$, where t is a real parameter. Show that

$$\frac{d}{dt}B(t) = e^{tA}[A,B]e^{-tA}$$

(b) Obviously, $B(0) = B$ and therefore

$$B(1) = B + \int_0^1 dt\, \frac{d}{dt}B(t)$$

Now using the power series $B(t) = \sum_{n=0}^\infty t^n B_n$ and using the above integral expression, show that $B_n = [A, B_{n-1}]/n$.

(c) Show by induction that

$$B_n = \frac{1}{n!} \underbrace{[A,[A,...[A,}_{n} B \underbrace{\,]....]]}_{n}$$

(d) Use $B(1) = e^A B e^{-A}$ and prove the identity.

(e) Now prove $e^{ipa/\hbar} x e^{-ipa/\hbar} = x + a$ showing that $U(a)$ indeed translates space.

4.19.13 Another Very Useful Identity

Prove that

$$e^{\hat{A}+\hat{B}} = e^{\hat{A}} e^{\hat{B}} e^{-\frac{1}{2}[\hat{A},\hat{B}]}$$

provided that the operators $\hat{A}$ and $\hat{B}$ satisfy

$$\left[\hat{A},\left[\hat{A},\hat{B}\right]\right] = \left[\hat{B},\left[\hat{A},\hat{B}\right]\right] = 0$$

A clever solution uses Problem 4.19.11or 64.19.12 result and ODEs.

4.19.14 Pure to Nonpure?

Use the equation of motion for the density operator $\hat{\rho}$ to show that a pure state cannot evolve into a nonpure state and vice versa.

4.19.15 Schur's Lemma

Let G be the space of complex differentiable test functions, $g(x)$, where x is real. It is convenient to extend G slightly to encompass all functions, $\tilde{g}(x)$, such that $\tilde{g}(x) = g(x) + c$, where $g \in G$ and c is any constant. Let us call the extended space $\tilde{G}$. Let $\hat{q}$ and $\hat{p}$ be linear operators on $\tilde{G}$ such that

$$\hat{q}g(x) = xg(x)$$
$$\hat{p}g(x) = -i\frac{dg(x)}{dx} = -ig'(x)$$

Suppose $\hat{M}$ is a linear operator on $\tilde{G}$ that commutes with $\hat{q}$ and $\hat{p}$. Show that

(1) $\hat{q}$ and $\hat{p}$ are hermitian on $\tilde{G}$

(2) $\hat{M}$ is a constant multiple of the identity operator

4.19.16 More About the Density Operator

Let us try to improve our understanding of the density matrix formalism and the connections with *information* or *entropy*.We consider a simple two-state system. Let ρ be any general density matrix operating on the two-dimensional Hilbert space of this system.

(a) Calculate the entropy, $s = -Tr(\rho \ln \rho)$ corresponding to this density matrix. Express the result in terms of a single real parameter. Make a clear interpretation of this parameter and specify its range.

(b) Make a graph of the entropy as a function of the parameter. What is the entropy for a pure state? Interpret your graph in terms of knowledge about a system taken from an ensemble with density matrix ρ.

(c) Consider a system with ensemble ρ a mixture of two ensembles ρ_1 and ρ_2:

$$\rho = \theta\rho_1 + (1-\theta)\rho_2 \quad , \quad 0 \le \theta \le 1$$

As an example, suppose

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \rho_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

in some basis. Prove that

$$s(\rho) \ge \rho = \theta s(\rho_1) + (1-\theta)s(\rho_2)$$

with equality if $\theta = 0$ or $\theta = 1$. This the so-called *von Neumann's mixing theorem.*

4.19.17 Entanglement and the Purity of a Reduced Density Operator

Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Let $|\Psi_{AB}\rangle$ be the state

$$|\Psi_{AB}\rangle = \cos\theta\,|0_A\rangle \otimes |0_B\rangle + \sin\theta\,|1_A\rangle \otimes |1_B\rangle$$

For $0 < \theta < \pi/2$, this is an entangled state. The purity ζ of the reduced density operator $\tilde{\rho}_A = Tr_B[|\Psi_{AB}\rangle\langle\Psi_{AB}|]$ given by

$$\zeta = Tr[\tilde{\rho}_A^2]$$

is a good measure of the *entanglement* of states in $\mathcal{H}_{AB}$. For pure states of the above form, find extrema of ζ with respect to θ $(0 \le \theta \le \pi/2)$. Do entangled states have large ζ or small ζ?

4.19.18 The Controlled-Not Operator

Again let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Consider the *controlled-not* operator on $\mathcal{H}_{AB}$ (very important in quantum computing),

$$U_{AB} = P_0^A \otimes I^B + P_1^A \otimes \sigma_x^B$$

where $P_0^A = |0_A\rangle\langle 0_A|$, $P_1^A = |1_A\rangle\langle 1_A|$ and $\sigma_x^B = |0_B\rangle\langle 1_B| + |1_B\rangle\langle 0_0|$.

Write a matrix representation for U_{AB} with respect to the following (ordered) basis for $\mathcal{H}_{AB}$

$$|0_A\rangle \otimes |0_B\rangle \ , \ |0_A\rangle \otimes |1_B\rangle \ , \ |1_A\rangle \otimes |0_B\rangle \ , \ |1_A\rangle \otimes |1_B\rangle$$

Find the eigenvectors of U_{AB} - you should be able to do this by inspection. Do any of them correspond to entangled states?

4.19.19 Creating Entanglement via Unitary Evolution

Working with the same system as in Problems 6.17 and 6.18, find a factorizable *input* state

$$\left|\Psi_{AB}^{in}\right\rangle = |\Psi_A\rangle \otimes |\Psi_B\rangle$$

such that the *output* state

$$\left|\Psi_{AB}^{out}\right\rangle = U_{AB}\left|\Psi_{AB}^{in}\right\rangle$$

is *maximally entangled*. That is, find any factorizable $\left|\Psi_{AB}^{in}\right\rangle$ such that $Tr[\tilde{\rho}_A^2] = 1/2$, where

$$\tilde{\rho}_A = Tr_B[\left|\Psi_{AB}^{out}\right\rangle\left\langle\Psi_{AB}^{out}\right|]$$

4.19.20 Tensor-Product Bases

Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Consider the following entangled state in the joint Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$,

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}} \left(|0_A 1_B\rangle + |1_A 0_B\rangle \right)$$

where $|0_A 1_B\rangle$ is short-hand notation for $|0_A\rangle \otimes |1_B\rangle$ and so on. Rewrite this state in terms of a new basis $\{|\tilde{0}_A \tilde{0}_B\rangle, |\tilde{0}_A \tilde{1}_B\rangle, |\tilde{1}_A \tilde{0}_B\rangle, |\tilde{1}_A \tilde{1}_B\rangle\}$, where

$$|\tilde{0}_A\rangle = \cos\frac{\phi}{2} |0_A\rangle + \sin\frac{\phi}{2} |1_A\rangle$$
$$|\tilde{1}_A\rangle = -\sin\frac{\phi}{2} |0_A\rangle + \cos\frac{\phi}{2} |1_A\rangle$$

and similarly for $\{|\tilde{0}_B\rangle, |\tilde{1}_B\rangle\}$. Again $|\tilde{0}_A \tilde{0}_B\rangle = |\tilde{0}_A\rangle \otimes |\tilde{0}_B\rangle$, etc. Is our particular choice of $|\Psi_{AB}\rangle$ special in some way?

4.19.21 Matrix Representations

Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Let $|0_A 0_B\rangle = |0_A\rangle \otimes |0_B\rangle$, etc. Let the natural tensor product basis kets for the joint space $\mathcal{H}_{AB}$ be represented by column vectors as follows:

$$|0_A 0_B\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; |0_A 1_B\rangle \leftrightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; |1_A 0_B\rangle \leftrightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; |1_A 1_B\rangle \leftrightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

For parts (a) -(c), let

$$\rho_{AB} = \frac{3}{8} |0_A\rangle \langle 0_A| \otimes \frac{1}{2} \left(|0_B\rangle + |1_B\rangle \right) \left(\langle 0_B| + \langle 1_B| \right) + \frac{5}{8} |1_A\rangle \langle 1_A| \otimes \frac{1}{2} \left(|0_B\rangle - |1_B\rangle \right) \left(\langle 0_B| - \langle 1_B| \right)$$

(a) Find the matrix representation of ρ_{AB} that corresponds to the above vector representation of the basis kets.

(b) Find the matrix representation of the partial projectors $I^A \otimes P_0^B$ and $I^A \otimes P_1^B$ (see problem 4.19.18 for definitions) and then use them to compute the matrix representation of

$$\left(I^A \otimes P_0^B \right) \rho_{AB} \left(I^A \otimes P_0^B \right) + \left(I^A \otimes P_1^B \right) \rho_{AB} \left(I^A \otimes P_1^B \right)$$

(c) Find the matrix representation of $\tilde{\rho}_A = Tr_B[\rho_{AB}]$ by taking the partial trace using Dirac language methods.

4.19.22 Practice with Dirac Language for Joint Systems

Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Let $|0_A 0_B\rangle = |0_A\rangle \otimes |0_B\rangle$, etc. Consider the joint state

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}} \left(|0_A 0_B\rangle + |1_A 1_B\rangle\right)$$

(a) For this particular joint state, find the most general form of an observable O^A acting only on the A subsystem such that

$$\langle \Psi_{AB}| O^A \otimes I^B |\Psi_{AB}\rangle = \langle \Psi_{AB}| \left(I^A \otimes P_0^B\right) O^A \otimes I^B \left(I^A \otimes P_0^B\right) |\Psi_{AB}\rangle$$

where

$$P_0^B = \left|0^B\right\rangle \left\langle 0^B\right|$$

Express your answer in Dirac language.

(b) Consider the specific operator

$$X^A = \left|0^A\right\rangle \left\langle 1^A\right| + \left|1^A\right\rangle \left\langle 0^A\right|$$

which satisfies the general form you should have found in part (a). Find the most general form of the joint state vector $|\Psi'_{AB}\rangle$ such that

$$\langle \Psi'_{AB}| X^A \otimes I^B |\Psi'_{AB}\rangle \neq \langle \Psi_{AB}| \left(I^A \otimes P_0^B\right) X^A \otimes I^B \left(I^A \otimes P_0^B\right) |\Psi_{AB}\rangle$$

(c) Find an example of a reduced density matrix $\tilde{\rho}_A$ for the A subsystem such that no joint state vector $|\Psi'_{AB}\rangle$ of the general form you found in part (b) can satisfy

$$\tilde{\rho}_A = Tr_B \left[|\Psi'_{AB}\rangle \langle \Psi'_{AB}|\right]$$

4.19.23 More Mixed States

Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be a pair of two-dimensional Hilbert spaces with given orthonormal bases $\{|0_A\rangle, |1_A\rangle\}$ and $\{|0_B\rangle, |1_B\rangle\}$. Let $|0_A 0_B\rangle = |0_A\rangle \otimes |0_B\rangle$, etc. Suppose that both the A and B subsystems are initially under your control and you prepare the initial joint state

$$\left|\Psi^0_{AB}\right\rangle = \frac{1}{\sqrt{2}} \left(|0_A 0_B\rangle + |1_A 1_B\rangle\right)$$

(a) Suppose you take the A and B systems prepared in the state $\left|\Psi^0_{AB}\right\rangle$ and give them to your friend, who then performs the following procedure. Your friend flips a biased coin with probability p for heads; if the result of the coin-flip is a head(probability p) the result of the procedure performed by your friend is the state

$$\frac{1}{\sqrt{2}} (|0_a 0_b\rangle - |1_a 1_b\rangle)$$

and if the result is a tail(probability $1-p$) the result of the procedure performed by your friend is the state

$$\frac{1}{\sqrt{2}}(|0_a 0_b\rangle + |1_a 1_b\rangle)$$

i.e., nothing happened. After this procedure what is the density operator you should use to represent your knowledge of the joint state?

(b) Suppose you take the A and B systems prepared in the state $\left|\Psi^0_{AB}\right\rangle$ and give them to your friend, who then performs the alternate procedure. Your friend performs a measurement of the observable

$$O = I^A \otimes U_h$$

but does not tell you the result. After this procedure, what density operator should you use to represent your knowledge of the joint state? Assume that you can use the projection postulate (reduction) for state conditioning (preparation).

4.19.24 Complete Sets of Commuting Observables

Consider a three-dimensional Hilbert space $\mathcal{H}_3$ and the following set of operators:

$$O_\alpha \leftrightarrow \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; O_\beta \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; O_\gamma \leftrightarrow \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Find all possible complete sets of commuting observables(CSCO). That is, determine whether or not each of the sets

$$\{O_\alpha\}, \{O_\beta\}, \{O_\gamma\}, \{O_\alpha, O_\beta\}, \{O_\alpha, O_\gamma\}, \{O_\beta, O_\gamma\}, \{O_\alpha, O_\beta, O_\gamma\}$$

constitutes a valid CSCO.

4.19.25 Conserved Quantum Numbers

Determine which of the CSCO's in problem 4.19.24 (if any) are conserved by the Schrodinger equation with Hamiltonian

$$H = \varepsilon_0 \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \varepsilon_0 \left(O_\alpha\right\} + \{O_\beta\right)$$

Chapter 5

How Does It really Work: Photons, K-Mesons and Stern-Gerlach

5.1 Introduction

Many experiments indicate that electromagnetic waves have a vector property called *polarization*. Suppose that we have an electromagnetic wave (we will just say light from now on) passing through a piece of Polaroid material. The Polaroid material has the property that it lets through only that light whose polarization vector is oriented parallel to a preferred direction in the Polaroid(called the optic axis).

Classically, if an incident beam of light is polarized parallel to the optic axis, then experiment says that *all* of its energy gets through the Polaroid.

If, on the other hand, it is polarized perpendicular to the optic axis, then experiment says that *none* of its energy gets through the Polaroid.

Classically, if it is polarized at an angle α to the optic axis, then experiment says that a *fraction* $\cos^2\alpha$ of its energy gets through the Polaroid.

Many experiments, including several that we will discuss later, indicate that light(and everything else in the universe as it turns out) exhibits both "particle-like" and "wave-like" properties. (We will define both of these terms more carefully later and decide then if this simple statement makes any sense).

The particle associated with light is called a photon.

Can the experimental polarization/Polaroid results be reconciled with this idea of a particle-like photon?

Suppose we assume that a light wave that is polarized in a certain direction is made up of a large number of photons each of which is polarized in that same

direction.

The particle properties of light, as represented by photons, invariably lead to some confusion. It is not possible to eliminate all of this confusion at this elementary discussion level because a satisfactory treatment of photons requires quantum electrodynamics.

We can, however, make many of the more important physical properties clear.

Consider a simple representation of a monochromatic electromagnetic wave (and its associated photons) with angular frequency ω and wavelength λ moving in a direction given by the unit vector $\hat{k}$. Such a monochromatic electromagnetic wave is composed of N (a very large number) photons, each with energy E and momentum $vecp$, such that we have the relationships

$$E = \hbar\omega \quad , \quad \vec{p} = \hbar\vec{k} = \frac{h}{\lambda}\hat{k} \tag{5.1}$$

where $\vec{k}$ is the wave vector, $\hbar = h/2\pi$, $\omega = 2\pi f$, h = Planck's constant, f = frequency = c/λ, and c = speed of light. We note that

$$E = \frac{h}{2\pi}2\pi f = hf = h\frac{c}{\lambda} = pc \tag{5.2}$$

as required by relativity for a particle with zero mass(such as the photon).

The number of photons in the wave is such that the total energy of the N photons, $NE = N\hbar\omega$, is equal to the total energy W in the electromagnetic wave, i.e., $W = NE = N\hbar\omega$.

Here, we are using the fact, derived from many experiments, that the energy of the light wave is quantized and thus can only take on certain discrete values (its value is a multiple of some quantum of energy).

When we specify the *polarization* of light, we are actually giving the direction of the electric field vector $\vec{E}$. Light waves are generally represented by plane electromagnetic waves. This means that the electric field vector $\vec{E}$ and the magnetic field vector $\vec{B}$ are both perpendicular to the direction of propagation specified by $\vec{k}$. If we choose the direction of propagation to be the z-axis, which is specified by the unit vector $\hat{e}_z$, then $\vec{E}$ and $\vec{B}$ lie in the $x-y$ plane. Since we are only considering the polarization property at this point, we can concentrate on the $\vec{E}$ vector alone. Now, any vector in the $x-y$ plane can be specified in terms of a pairs of orthonormal vectors (called the basis) in that plane. The vectors in these orthonormal directions are called the *polarization vectors*.

Two standard sets of orthonormal polarization vectors are often chosen when

one discusses polarization. One of the two sets is

$$\hat{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \hat{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{5.3}$$

which correspond to *plane(or linearly)*-polarized waves. A second set is

$$\hat{e}_R = \hat{e}_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad , \quad \hat{e}_L = \hat{e}_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} \tag{5.4}$$

which correspond to circularly polarized waves.

For classical electromagnetic fields, a light wave propagating in the z-direction is usually described(illustrating with the two orthonormal sets) by electric field vectors of the forms given below.

Plane-polarized basis:

$$\begin{aligned} \vec{E}(\vec{r},t) = \begin{pmatrix} E_x(\vec{r},t) \\ E_x(\vec{r},t) \\ 0 \end{pmatrix} &= E_x(\vec{r},t) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + E_y(\vec{r},t) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ &= E_x(\vec{r},t)\hat{e}_x + E_y(\vec{r},t)\hat{e}_y \end{aligned} \tag{5.5}$$

Circularly-polarized basis:

$$\begin{aligned} \vec{E}(\vec{r},t) &= \begin{pmatrix} E_x(\vec{r},t) \\ E_x(\vec{r},t) \\ 0 \end{pmatrix} \\ &= \frac{E_x(\vec{r},t) + iE_y(\vec{r},t)}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} + \frac{E_x(\vec{r},t) - iE_y(\vec{r},t)}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} \\ &= E_x(\vec{r},t)\hat{e}_R + E_y(\vec{r},t)\hat{e}_L \end{aligned} \tag{5.6}$$

By convention, we represent the field components by

$$E_x(\vec{r},t) = E_x^0 e^{i(kz-\omega t+\alpha_x)} \text{ and } E_y(\vec{r},t) = E_y^0 e^{i(kz-\omega t+\alpha_y)} \tag{5.7}$$

where α_x and α_y are the (real)phases and E_x^0 and E_y^0 are the (real) amplitudes of the electric field components.

Clearly, the polarization state of the light is directly related to the $\vec{E}$ vectors in this formulation.

For example, using the above equations, we have these cases:

1. If $E_y = 0$, then the wave is *plane polarized* in the x-direction

$$\vec{E} = E_x \hat{e}_x = E_x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{5.8}$$

2. If $E_x = 0$, then the wave is *plane polarized* in the y-direction

$$\vec{E} = E_y \hat{e}_y = E_y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{5.9}$$

3. If $E_x = E_y$, then the wave is *plane polarized* at 45°

$$\vec{E} = E_x \hat{e}_x + E_y \hat{e}_y = E_x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{5.10}$$

4. If $E_y = -iE_x = e^{-i\pi/2} E_x$, then the y-component lags the x-component by 90° and the wave is *right circularly* polarized

$$\vec{E} = E_x \hat{e}_R = E_x \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \tag{5.11}$$

5. If $E_y = iE_x = e^{i\pi/2} E_x$, then the y-component leads the x-component by 90° and the wave is *left circularly* polarized

$$\vec{E} = E_x \hat{e}_L = E_x \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} \tag{5.12}$$

For our present discussion, the above properties of polarization are sufficient. We will give a more detailed discussion of the quantum mechanics of photon polarization shortly.

The picture we are proposing assumes that each photon has the same polarization as the light wave, which is, in fact, verified by experiment.

This simple experimental property leads to some fundamental difficulties for classical mechanics.

If the incident beam is polarized parallel or perpendicular to the optic axis of a Polaroid, then classical physics has no problems ..... all the photons (and thus all the energy) either pass through or do not pass(and thus none of the energy) through the Polaroid.

But what about the case where the wave is polarized at 45° to the optic axis of the Polaroid?

For the beam as a whole, the experimental result is that 1/2 ($\cos^2 45^\circ = 1/2$) of the total energy and hence 1/2 of the photons pass through.

But what about any *particular photon*, each of which is polarized at 45° to the optic axis?

Now the answer is not clear at all and the fundamental dilemma of the subatomic world rears its ugly head.

As will become clear during our discussions of quantum mechanics, this question about what will happen to a particular photon under certain conditions is not very precise.

In order for any theory to make clear predictions about experiments, we will have to learn how to ask very precise questions. We must also remember that only questions about the results of experiments have a real significance in physics and it is only such questions that theoretical physics must consider.

All relevant questions and the subsequent experiments devised to answer the questions must be clear and precise, however.

In this case, we can make the question clear by doing the experiment with a beam containing only one photon and observe what happens after it arrives at the Polaroid. In particular, we make a simple observation to see *whether or not* it passes through the Polaroid.

The most important experimental result is that this single photon either passes through the Polaroid or it does not. I will call this type of experiment a *go-nogo* or *yes-no* experiment.

We never observe 1/2 the energy of a single photon. We always observe either zero energy or an energy exactly equal to $\hbar\omega$. One never observes a part of a photon passing through and a part getting absorbed in the Polaroid.

In addition, if a photon gets through, then experiment says that its polarization vector *changes* such that it ends up polarized in a direction parallel to the optic axis of this particular Polaroid (instead of at 45° with respect to that axis as it was polarized beforehand).

In a beam of N photons, each photon will *independently* behave as the single photon did in the description above. No experiment can determine which of the photons will pass through and which will not, even though they are all identical. In each experiment, however, exactly 1/2 of the total energy and 1/2 of the

photons will pass through the 45° Polaroid.

As we shall show later, the only way this result can be interpreted is to say that each photon has a *probability* = 1/2 of passing through the 45° Polaroid.

We are led to this probabilistic point of view because the energy of the electromagnetic wave is quantized(or equivalently, that the electromagnetic wave is made up of photons) and we cannot have fractions of the energy quantum appearing during an experiment.

We have managed to preserve the indivisibility of the photons(the quantization of their energy), but we were able to do this only by abandoning the comforting determinacy of classical physics and introducing probability.

The results in this experiment are not completely determined by the experimental conditions(initial) under control of the experimenter, as they would have been according to classical ideas.

As we shall see, the most that we will be able to predict in any experiment is a set of possible results, with a probability of occurrence for each.

The experiment described above involving a single photon polarized at an angle to the optic axis, allows us to ask only one type of experimental and theoretical question, namely, does the photon go through or is it absorbed - only a *yes-no* experiment? That will turn out to be the only legitimate question we can ask in this case.

It is the first indication of the way we should frame our discussion of theory!

We shall see that questions like....

> What decides whether a particular photon goes through?
>
> When does a particular photon decide whether it will pass through?
>
> How does a particular photon change its polarization direction?

cannot be answered by experiment and, therefore, must be regarded as *outside the domain* of quantum theory and possibly all of physics.

What will our theory of quantum mechanics say about the state of the single photon?

It will be shown that the photon polarized at an angle to the optic axis is in a very special kind of state that we will call a *superposition* of being polarized perpendicular to the optic axis and of being polarized parallel to the optic axis,

i.e., *a superposition of all the possibilities.*

In this state, there will exist an extraordinary kind of relationship between the two kinds(mutually perpendicular directions) of polarization.

The meaning of the word *superposition* will follow clearly from the mathematical formalism and language we have developed in this book. It will, however, require a new physical connection to mathematics.

This, as we shall see later, is suggested by an attempt to express the meaning of superposition in ordinary language(words). If we attempt to explain the behavior of the photon polarized at an angle to the optic axis using ordinary language, then we might be inclined to say something like this:

The photon is

not polarized *parallel* to the optic axis

not polarized *perpendicular* to the optic axis

not simultaneously possessing *both* polarizations

not possessing *neither* polarization

For this experiment with only two possible polarizations, these statements exhaust all the logical possibilities allowed by *ordinary words*!

Superposition is something *completely different* than any of the above and it is *not all* of the above.

Its physical content will, however, be precise and clear in our new mathematical formalism.

When the photon encounters the Polaroid, we are observing it. We are observing whether it is polarized perpendicular or parallel to the optic axis of the Polaroid. The effect of this measurement will be to end up with the photon having one or the other polarizations(the one we measure). In such a measurement, the photon always makes a *jump* from a state of superposition to a state of a *definite* polarization. Which of the two states it jumps to cannot be predicted. We will, as we shall see, be able to predict the probability of each possibility.

If it *jumps* into the parallel state, it has passed through. If it *jumps* into the perpendicular state, it has been absorbed.

We will have a great deal more to say about the two new words, *superposition* and *jump*, as we proceed. Do not attach any classical meaning to the word jump

as used here. It simply my attempt to use words for the moment.

5.1.1 Photon Interference

Another classic experiment involving light *waves* and hence photons is two-slit interference. The ideas of *superposition* and *jump* will reappear in this case and give us another view to enhance our understanding(or maybe confusion) at this point.

We will discuss this type of experiment in great detail and in many different forms in this book. As we shall see, the meaning of this experiment is easily misunderstood.

This experiment looks at the position and momentum properties of photons instead of the polarization properties.

For an approximately monochromatic light wave we have some knowledge of both the position and momentum of the photons. In particular, the photons must be located near the beam of light and their momentum is in the direction of propagation of the beam and has an approximate magnitude

$$p = \frac{E}{c} = \frac{\hbar\omega}{c} = \frac{hf}{c} = \frac{h}{\lambda} \tag{5.13}$$

According to the standard wave interference description, a two-slit interference experiment consists of a device which, in some manner, splits the incoming beam into two beams which are sent over different physical paths and then recombined, in some way, so that they can interfere and produce a distinctive pattern on a screen.

To see the problems directly, we again consider an incident beam consisting of a single photon. What happens as it passes through the apparatus?

The photon has only two *classical* possibilities if it is going to reach the screen and contribute to the interference pattern, i.e., it must follow a path that passes through one of the two slits.

As in the polarization experiment, the correct quantum description will involve describing the single photon as a *superposition* of photons traveling on (at least) two different paths to the same point on a screen.

Once again, as we shall see in detail later on, if we tried to use ordinary language to describe the experimental results in the interference experiment we find ourselves saying that the photon is

not on path #1

not on path #2

not simultaneously on *both* paths

not on *neither* path

What actually happens when we try to determine the energy of a photon in one of the beams?

When we do the measurement, we always find the photon(all of the energy in the single photon system) in one of the beams. When we observe the photon it must somehow *jump* from being in a superposition (which generates the interference pattern) to being entirely in one of the beams (which, as it turns out, does not generate the interference pattern). One possibility is that the *jump* is caused by the measurement of the energy in this case.

Although, we cannot predict on which path the photon will be found, we can, however, predict the probability of either result from the mathematical formalism we will develop for superposition.

Even more striking is the following result. Suppose we have a single beam of light consisting of a large number of photons, which we split up into two beams of equal intensity. On the average we should have about half of the photons in each beam.

Now make the two beams interfere.

If we were to assume that the interference is the result of a photon in one beam interfering with a photon in the other beam, then we would have to allow two photons to annihilate each other some of the time (producing interference minima) and sometimes turn into four photons (producing interference maxima). This process clearly violates conservation of energy.

The new concept of superposition will enable us to deal with this problem. Each photon in the original beam will be in a superposition of being in both beams. The superposition of possibilities will produce the interference effect. Interference between two different photons or of the photon with itself is never required and energy will always be conserved.

We will have a great deal more to say about these and other experiments as we develop the mathematical formalism of quantum mechanics.

Some questions about this approach

Is it really necessary to introduce the new concepts of superposition and jump?

In many simple experiments, the classical model of waves and photons connected in some vague statistical way will be sufficient to explain many experimental results. The new quantum ideas do not contribute anything new in these cases. As we shall see as we progress through this chapter, there are, however, an overwhelming number of experiments where the only correct explanation comes from a quantum theory built around these new concepts and their associated mathematical formalism.

Will this new theory give us a better model of the photon and of single photon processes?

I do not believe that the object of a physical theory is to provide a satisfying model or picture, but instead, to develop a language that enables us to formulate physical laws and make experimental predictions.

Models and pictures are holdovers from classical physics as applied to macroscopic phenomena. In the case of atomic phenomena and beyond we cannot expect to find an appropriate model or picture. All such models and pictures rely on classical concepts and these concepts are totally inadequate in the microscopic arena. Their main goal seems to be to make the reader feel more comfortable with the new and strange phenomena under discussion. Their many misleading implications, however, often cause misunderstandings that lead many a student down dead-end paths as they try to understand quantum mechanics. We will avoid all non-mathematical models.

Unless one has mastered the correct language and adjusted one's mode of thinking, it will not, in general, be possible to understand modern theoretical physics. For me, a self-consistent mathematical formulation of the theory that uses a language appropriate to the physical phenomena involved and is able to make correct predictions for all experiments constitutes a valid physical model.

The reasons for observed physical behavior are contained in the mathematical formalism. There is no deeper physics involved beyond the mathematical formalism. The only *model* is the mathematical formalism and its associated physical connections along with a language that is needed to understand it and express what it means in terms of experiments.

What about determinism?

The problems that the concept of superposition presents to the classical idea of determinism when we create a well-defined system are so devastating that

these old ideas must be abandoned. We will discuss these ideas in great detail throughout the book, but for now a short digression about states, superposition and measurements seems appropriate.

A classical state is specified by giving numerical values for all the coordinates and velocities of the constituents of the system at some instant of time. The subsequent motion is then determined if we know the force laws.

For a *small* system, however, we cannot observe it and determine its properties to the extent required for a complete classical description. Thus, any microscopic system is necessarily specified by fewer or more indefinite data than all the coordinates and velocities at some instant of time.

In the new mathematical formalism, we will carefully define the kind of information we can know about a state, which information we can actually know at a given instant of time and how we will prepare such states. The prepared states will have definite values of a specific set of measurable quantities.

In terms of such states, we will define a superposition such that we have well-defined mathematical relationships between the different states making up the superposition. The mathematical definition of a superposition will not be describable with the classical ideas or pictures or models available, i.e., it will turn out that when we superpose two states, the nature of the relationship that exists between them cannot be explained in terms of classical physical concepts. The system is not partly in one state and partly in the other in any sense.

We will have to deal with a completely new idea here.

During the course of our discussions, we will need to get accustomed to the mathematical formalism behind the concept of superposition and we will need to rely on the formalism and the mathematics, as expressed by the Dirac language, without having any detailed classical models.

The new superposed state will be completely defined by the states involved, the mathematical relationships between them and the physical meaning of those mathematical relationships as defined by the formalism.

As we saw in the polarization and interference experiments, we must have a superposition of two possible states, say A and B. Suppose that if the system were in state A alone, then a particular measurement would, with certainty, produce the result a and if the system were in state B alone, then the same measurement would, with certainty, produce the result b. Then, if the system is in the superposed state(of A and B), it turns out that the same measurement will sometimes produce the result a and sometimes the result b. Given the mathematical definition of the superposed state and the rules that we will specify to relate the physics to the mathematics, we will be able to predict the

probability of getting the result a and the probability of getting the result b.

We will not be able to predict which result we will get in any one measurement, i.e., the measurement process will not be deterministic. Identical measurements on identically prepared states will not yield identical results in any one measurement.

If we repeat the measurements a large number of times, however, we can predict the fraction of time we will obtain the result . That is all we will be able to predict using the formalism of quantum mechanics we will develop. I firmly believe, that nature is such that we will not, under any circumstances, be able to make any more detailed predictions for quantum systems.

So, summarizing the situation before we proceed.....

When physicists try to construct a theory from their experimental observations they are primarily concerned with two things:

how to calculate something that will enable them to
predict the results of further experiments

how to understand what is going on in the experiments

Now, it is not always possible to satisfy both of these concerns at the same time. Sometimes we have a reasonably clear idea of what is going on, but the mathematical details are so complex that calculating is very hard. Sometimes the mathematics is quite clear, but the understanding is difficult.

Quantum mechanics falls into the latter category.

Over the years since quantum mechanics was first proposed, a very pragmatic attitude has formed.

Physicists realized that an understanding of how electrons could behave in such a manner would come once they developed the rules that would enable them to calculate the way they behaved.

So quantum mechanics as a set of rules was developed, as we shall see in this book.

It has been tested in a large number of experiments and never been found to be wrong. It works extremely well.

Yet, we are still not sure how the electrons can behave in such a manner.

In many ways, the situation has even gotten worse over the years. Some people, as we shall see, held out the hope that quantum mechanics was an *incomplete theory* and that as we did more experiments we would discover some loose end or new idea that would allow us to make sense of things.

This has not happened and I believe it never will.

We now carry out the details of a special case that will illustrate how Quantum Mechanics works and also illustrate the mathematical formalism that we have developed in earlier chapters.

5.2 Photon Polarization

As we mentioned earlier, the electric field vector $\vec{E}$ of plane electromagnetic waves lies in a plane perpendicular to the direction of propagation of the wave. If we choose the z-axis as the direction of propagation, we can represent the electric field vector as a 2-dimensional vector in the $x-y$ plane. This means that we will only require two numbers to describe the electric field. Since the polarization state of the light is directly related to the electric field vector, this means that we can also represent the polarization states of the photons by 2-component column vectors or ket vectors of the form

$$|\psi\rangle = \begin{pmatrix} \psi_x \\ \psi_y \end{pmatrix} \text{ where we assume the normalization condition } \langle\psi\,|\,\psi\rangle = 1 \quad (5.14)$$

The components in the *state vectors* depend only on the actual polarization state of the photon. The state vector contains *all* of the information that we can have about the state of polarization of the photon (remember that it consists of just two numbers).

Examples

$$|x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \text{linear polarized photon}$$

$$|y\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \text{linear polarized photon}$$

$$|R\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \rightarrow \text{right circular polarized photon}$$

$$|L\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \rightarrow \text{left circular polarized photon}$$

$$|45\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \text{photon linearly polarized at } 45^\circ \text{ to the x-axis}$$

The bra vector or linear functional corresponding the ket vector $|\psi\rangle$ is given by the *row vector*

$$\langle\psi| = \begin{pmatrix}\psi_x^* & \psi_y^*\end{pmatrix} \tag{5.15}$$

which clearly implies via our inner product rules

$$\langle\psi\,|\,\psi\rangle = \begin{pmatrix}\psi_x^* & \psi_y^*\end{pmatrix}\begin{pmatrix}\psi_x \\ \psi_y\end{pmatrix} = |\psi_x|^2 + |\psi_y|^2 = 1 \tag{5.16}$$

In general, for

$$|\phi\rangle = \begin{pmatrix}\phi_x \\ \phi_y\end{pmatrix} \tag{5.17}$$

the inner product rule says that

$$\langle\phi\,|\,\psi\rangle = \begin{pmatrix}\phi_x^* & \phi_y^*\end{pmatrix}\begin{pmatrix}\psi_x \\ \psi_y\end{pmatrix} = \phi_x^*\psi_x + \phi_y^*\psi_y = \langle\psi\,|\,\phi\rangle^* \tag{5.18}$$

We also have the results

$$\langle x\,|\,x\rangle = 1 = \langle y\,|\,y\rangle \text{ and } \langle x\,|\,y\rangle = 0 = \langle y\,|\,x\rangle \rightarrow \text{ orthonormal set} \tag{5.19}$$

$$\langle R\,|\,R\rangle = 1 = \langle L\,|\,L\rangle \text{ and } \langle R\,|\,L\rangle = 0 = \langle L\,|\,R\rangle \rightarrow \text{ orthonormal set} \tag{5.20}$$

Each of these two sets of state vectors is a basis for the 2-dimensional vector space of polarization states since any other state vector can be written as a linear combination of them, i.e.,

$$|\psi\rangle = \begin{pmatrix}\psi_x \\ \psi_y\end{pmatrix} = \psi_x\begin{pmatrix}1 \\ 0\end{pmatrix} + \psi_y\begin{pmatrix}0 \\ 1\end{pmatrix} = \psi_x\,|x\rangle + \psi_y\,|y\rangle \tag{5.21}$$

or

$$\begin{aligned}|\psi\rangle = \begin{pmatrix}\psi_x \\ \psi_y\end{pmatrix} &= \frac{\psi_x - i\psi_y}{2}\begin{pmatrix}1 \\ i\end{pmatrix} + \frac{\psi_x + i\psi_y}{2}\begin{pmatrix}1 \\ -i\end{pmatrix} \\ &= \frac{\psi_x - i\psi_y}{\sqrt{2}}\,|R\rangle + \frac{\psi_x + i\psi_y}{\sqrt{2}}\,|L\rangle\end{aligned} \tag{5.22}$$

We can find the components along the basis vectors using

$$\langle x\,|\,\psi\rangle = \langle x|\,(\psi_x\,|x\rangle + \psi_y\,|y\rangle) = \psi_x\,\langle x\,|\,x\rangle + \psi_y\,\langle x\,|\,y\rangle = \psi_x \tag{5.23}$$

$$\langle y\,|\,\psi\rangle = \langle y|\,(\psi_x\,|x\rangle + \psi_y\,|y\rangle) = \psi_x\,\langle y\,|\,x\rangle + \psi_y\,\langle y\,|\,y\rangle = \psi_y \tag{5.24}$$

or

$$|\psi\rangle = |x\rangle\,\langle x\,|\,\psi\rangle + |y\rangle\,\langle y\,|\,\psi\rangle \tag{5.25}$$

and similarly

$$|\psi\rangle = |R\rangle\,\langle R\,|\,\psi\rangle + |L\rangle\,\langle L\,|\,\psi\rangle \tag{5.26}$$

Basically, we are illustrating examples of a *superposition principle*, which says that any arbitrary polarization state can be written as a superposition (linear

combination) of x- and y-polarization states or equivalently, as a superposition of right- and left-circularly polarized states.

Our earlier discussions of a beam of light passing through a polaroid can now be *recast* in terms of these polarization states.

Classical physics says that the beam is a superposition of an x-polarized beam and a y-polarized beam and when this beam passes through an x-polaroid, its effect is to remove the y-polarized beam and pass the x-polarized beam through unchanged.

The energy of the classical beam is given by $|\vec{E}|^2$ which for the polarization states is proportional to $|\psi_x|^2 + |\psi_y|^2$. Thus, the beam energy after passing through an x-polaroid is proportional to $|\psi_x|^2$. The fraction of the beam energy or the fraction of the number of photons in the beam that passes through is given by

$$\frac{|\psi_x|^2}{|\psi_x|^2 + |\psi_y|^2} = |\psi_x|^2 = |\langle x \,|\, \psi \rangle|^2 \tag{5.27}$$

for states normalized to 1. Our earlier discussion for the case of a single photon forced us to set this quantity equal to the probability of a single photon in the state $|\psi\rangle$ passing through an x-polaroid or

> probability of a photon in the state $|\psi\rangle$
> passing through an x-polaroid $= |\langle x \,|\, \psi \rangle|^2$

This agrees with the mathematical results we derived in Chapter 4.

This all makes sense in an ensemble interpretation of the state vector, that is, that the state vector represents the *beam* or equivalently, *many copies* of a single photon system.

We then define $\langle x \,|\, \psi \rangle$ as the *probability amplitude* for the individual photon to pass through the x-polaroid.

Another example confirming these results is light passing through a prism. A prism passes right-circularly-polarized(RCP) light and rejects(absorbs) left-circularly-polarized(LCP) light.

Since we can write

$$|\psi\rangle = |R\rangle \langle R \,|\, \psi \rangle + |L\rangle \langle L \,|\, \psi \rangle \tag{5.28}$$

we can generalize the polaroid result to say

> amplitude that a photon in the state $|\psi\rangle$
> passes through a prism $= \langle R \,|\, \psi \rangle$

and

probability that a photon in the state $|\psi\rangle$
passes through a prism = $|\langle R|\psi\rangle|^2$

Polaroids and prisms are examples of *go-nogo* devices. Certain photons are passed through while others are absorbed in these devices.

We note that the probability is independent of the phase of the state, but the amplitude depends on phase. This will be a *crucial point* later on in our discussion.

5.2.1 How Many Basis Sets ?

We have already seen two examples of basis sets for the 2-dimensional vector space of polarization states, namely,

$$\{|x\rangle, |y\rangle\} \text{ and } \{|R\rangle, |L\rangle\} \tag{5.29}$$

In the 2-dimensional vector space there are an infinite number of such orthonormal basis sets related to the $\{|x\rangle, |y\rangle\}$ set. They are all equivalent for describing physical systems (they correspond to different orientations of the polaroid in the experimental measurement). We can obtain the other sets say $\{|x'\rangle, |y'\rangle\}$, by a rotation of the bases (or axes) as shown in Figure 5.1 below.

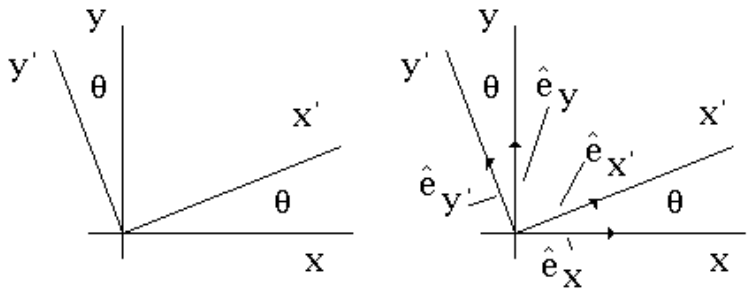


Figure 5.1: Rotation of Axes

We then have in the $x-y$ basis

$$|\psi\rangle = \psi_x |x\rangle + \psi_y |y\rangle = \begin{pmatrix}\psi_x\\ \psi_y\end{pmatrix} = \begin{pmatrix}\langle x|\psi\rangle\\ \langle y|\psi\rangle\end{pmatrix} \tag{5.30}$$

and if we choose to use the equivalent $x'-y'$ basis we have

$$|\psi\rangle = \psi_{x'} |x'\rangle + \psi_{y'} |y'\rangle = \begin{pmatrix}\psi_{x'}\\ \psi_{y'}\end{pmatrix} = \begin{pmatrix}\langle x'|\psi\rangle\\ \langle y'|\psi\rangle\end{pmatrix} \tag{5.31}$$

How are these components related to each other? We have

$$|\psi\rangle = |x\rangle\langle x|\psi\rangle + |y\rangle\langle y|\psi\rangle \tag{5.32}$$

which implies

$$\langle x'|\psi\rangle = \langle x'|x\rangle\langle x|\psi\rangle + \langle x'|y\rangle\langle y|\psi\rangle \tag{5.33}$$

or in matrix notation

$$\langle y'|\psi\rangle = \langle y'|x\rangle\langle x|\psi\rangle + \langle y'|y\rangle\langle y|\psi\rangle \tag{5.34}$$

or in matrix notation

$$\begin{pmatrix}\langle x'|\psi\rangle\\ \langle y'|\psi\rangle\end{pmatrix} = \begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix}\begin{pmatrix}\langle x|\psi\rangle\\ \langle y|\psi\rangle\end{pmatrix} \tag{5.35}$$

So we can transform the basis (transform the components) if we can determine the 2×2 transformation matrix

$$\begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix} \tag{5.36}$$

It turns out that this result is quite general in the sense that it holds for any two bases, not just the linearly polarized bases we used to derive it.

For the linear(plane) polarized case, we can think of an analogy to unit vectors along the axes in ordinary space as shown on the right in Figure 5.1.

Then we have(by analogy)

$$\hat{e}_x\cdot\hat{e}_{x'} = \cos\theta = \langle x'|x\rangle \quad , \quad \hat{e}_{x'}\cdot\hat{e}_y = \sin\theta = \langle x'|y\rangle \tag{5.37}$$

$$\hat{e}_{y'}\cdot\hat{e}_y = \cos\theta = \langle y'|y\rangle \quad , \quad \hat{e}_{y'}\cdot\hat{e}_x = -\sin\theta = \langle y'|x\rangle \tag{5.38}$$

or

$$|x'\rangle = \langle x|x'\rangle|x\rangle + \langle y|x'\rangle|y\rangle = \cos\theta|x\rangle + \sin\theta|y\rangle \tag{5.39}$$

$$|y'\rangle = \langle x|y'\rangle|x\rangle + \langle y|y'\rangle|y\rangle = -\sin\theta|x\rangle + \cos\theta|y\rangle \tag{5.40}$$

and the transformation matrix, $\hat{R}(\theta)$, is

$$\hat{R}(\theta) = \begin{pmatrix}\langle x'|x\rangle & \langle x'|y\rangle\\ \langle y'|x\rangle & \langle y'|y\rangle\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} \tag{5.41}$$

so that

$$\begin{pmatrix}\langle x'|\psi\rangle\\ \langle y'|\psi\rangle\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\langle x|\psi\rangle\\ \langle y|\psi\rangle\end{pmatrix} \tag{5.42}$$

There are two equivalent ways to interpret these results.

First, we could say it tells us the components of $|\psi\rangle$ in the rotated basis (we

keep the vector fixed and rotate the axes). Second, we can rotate the vector and keep the axes fixed(rotate in the opposite direction). In this case, we regard

$$\begin{pmatrix} \langle x' \,|\, \psi \rangle \\ \langle y' \,|\, \psi \rangle \end{pmatrix} \tag{5.43}$$

as a new vector $|\psi'\rangle$ whose components in the fixed $x-y$ basis are the *same* as the components of $|\psi\rangle$ in the $x'-y'$-basis or

$$\langle x' \,|\, \psi \rangle = \langle x \,|\, \psi' \rangle \text{ and } \langle y' \,|\, \psi \rangle = \langle y \,|\, \psi' \rangle \tag{5.44}$$

For *real* ψ_x and ψ_y, $|\psi'\rangle$ is the vector $|\psi\rangle$ rotated clockwise by θ or, regarding $\hat{R}(\theta)$ as a linear operator in the vector space we have

$$|\psi'\rangle = \hat{R}(\theta)\,|\psi\rangle \tag{5.45}$$

It is not a Hermitian operator, however, so it cannot represent an observable. It is a transformation of vectors and according to our earlier discussion, it should be a unitary operator. We can see this as follows:

$$\hat{R}^{-1}(\theta) = \hat{R}(-\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \hat{R}^T(\theta) = \hat{R}^\dagger(\theta) \tag{5.46}$$

Since $\hat{R}(\theta)$ is a unitary transformation operator for rotations, our earlier discussions say that we should be able to express it as an exponential operator involving the angular momentum with respect to the axis of rotation (z-axis) $\hat{J}_z$, of the form

$$\hat{R}(\theta) = e^{i\theta \hat{J}_z/\hbar} \tag{5.47}$$

Now, we can rewrite $\hat{R}(\theta)$ as

$$\begin{aligned} \hat{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} &= \cos\theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + i\sin\theta \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ &= \cos\theta \hat{I} + i\sin\theta \hat{Q} \end{aligned} \tag{5.48}$$

where the physical meaning of the operator

$$\hat{Q} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{5.49}$$

is yet to be determined.

Expanding (5.47) in a power series we have

$$\begin{aligned}\hat{R}(\theta) &= \hat{R}(0) + \frac{1}{1!}\left.\frac{d\hat{R}(\theta)}{d\theta}\right|_{\theta=0}\theta + \frac{1}{2!}\left.\frac{d^2\hat{R}(\theta)}{d\theta^2}\right|_{\theta=0}\theta^2 \\ &\quad + \frac{1}{3!}\left.\frac{d^3\hat{R}(\theta)}{d\theta^3}\right|_{\theta=0}\theta^3 + \frac{1}{4!}\left.\frac{d^4\hat{R}(\theta)}{d\theta^4}\right|_{\theta=0}\theta^4 + \ldots \\ &= \hat{I} + \frac{1}{1!}\left(\frac{i\hat{J}_z}{\hbar}\right)\theta + \frac{1}{2!}\left(\frac{i\hat{J}_z}{\hbar}\right)^2\theta^2 + \frac{1}{3!}\left(\frac{i\hat{J}_z}{\hbar}\right)^3\theta^3 + \frac{1}{4!}\left(\frac{i\hat{J}_z}{\hbar}\right)^4\theta^4 + \ldots \\ &= e^{i\theta\hat{J}_z/\hbar} \end{aligned} \tag{5.50}$$

Now let us assume that $\hbar\hat{Q} = \hat{J}_z$, where $\hat{J}_z$ is defined in (5.47). We then have the following algebraic results.

$$\hat{J}_z^2 = \hbar^2\hat{Q}^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hbar^2\hat{I} \tag{5.51}$$

which implies that

$$\hat{J}_z^3 = \hbar^2\hat{J}_z \quad , \quad \hat{J}_z^4 = \hbar^2\hat{J}_z^2 = \hbar^4\hat{I} \quad , \quad \hat{J}_z^5 = \hbar^4\hat{J}_z \text{ and so on} \tag{5.52}$$

This gives

$$\begin{aligned}\hat{R}(\theta) = e^{i\theta\hat{J}_z/\hbar} &= \hat{I} + (i)\frac{\hat{J}_z}{\hbar}\theta + \frac{(i)^2\hat{I}}{2!}\theta^2 + (i)\frac{\hat{J}_z}{3!\hbar}\theta^3 + \frac{(i)^4\hat{I}}{4!}\theta^4 + \ldots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \ldots\right) \\ &= \cos\theta\hat{I} + i\sin\theta\hbar\hat{J}_z = \cos\theta\hat{I} + i\sin\theta\hat{Q} \end{aligned} \tag{5.53}$$

as in (5.48).

So everything, exponential form of the rotation operator and matrix form of the rotation operator, is consistent if we choose

$$\hat{J}_z = \hbar\hat{Q} = \hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{5.54}$$

This corresponds to the matrix representation of $\hat{J}_z$ in the$\{|x\rangle, |y\rangle\}$ basis.

We now work out the eigenvectors and eigenvalues of $\hat{R}(\theta)$ as given by the equation

$$\hat{R}(\theta)|\psi\rangle = c|\psi\rangle \tag{5.55}$$

where c = the eigenvalue corresponding to the eigenvector $|\psi\rangle$.

Since all vectors are eigenvectors of the identity operator $\hat{I}$, we only need to find

the eigenvectors and eigenvalues of $\hat{J}_z$ in order to solve the problem for $\hat{R}(\theta)$, i.e.,

$$\hat{R}(\theta)\,|\psi\rangle = \cos\theta\hat{I}\,|\psi\rangle + i\sin\theta\hat{Q}\,|\psi\rangle = cos\theta\,|\psi\rangle + \frac{i}{\hbar}\sin\theta\hat{J}_z\,|\psi\rangle = c\,|\psi\rangle \tag{5.56}$$

If we let(since $\hat{R}(\theta)$ and $\hat{J}_z$ commute they have common eigenvectors)

$$\hat{J}_z\,|\psi\rangle = \lambda\,|\psi\rangle \tag{5.57}$$

then we have

$$\hat{R}(\theta)\,|\psi\rangle = \cos\theta\hat{I}\,|\psi\rangle + i\sin\theta\hat{Q}\,|\psi\rangle = \cos\theta\,|\psi\rangle + \frac{i}{\hbar}\sin\theta\hat{J}_z\,|\psi\rangle = c\,|\psi\rangle \tag{5.58}$$

or

$$c = \cos\theta + \frac{i\lambda}{\hbar}\sin\theta \tag{5.59}$$

Now, since $\hat{J}_z^2 = \hbar^2\hat{I}$, we have

$$\hat{J}_z^2\,|\psi\rangle = \lambda^2\,|\psi\rangle = \hbar^2\hat{I}\,|\psi\rangle = \hbar^2\,|\psi\rangle \tag{5.60}$$

which says that

$$\lambda^2 = \hbar^2 \text{ or } \lambda = \pm\hbar = \text{ eigenvalues of } \hat{J}_z \tag{5.61}$$

We can find the corresponding eigenvectors by inserting the eigenvalues into the eigenvalue equation

$$\hat{J}_z\,|J_z = \hbar\rangle = \hbar\,|J_z = \hbar\rangle \tag{5.62}$$

We assume that

$$|J_z = \hbar\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \text{ where } |a|^2 + |b|^2 = 1 \tag{5.63}$$

to get from (5.62)

$$\hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \hbar\begin{pmatrix} -ib \\ ia \end{pmatrix} = \hbar\begin{pmatrix} a \\ b \end{pmatrix} \tag{5.64}$$

This gives the result $ia = b$ which together with the normalization condition says that $a = 1/\sqrt{2}$. We have arbitrarily chosen a to be real since only the relative phase between components is important. This then gives $b = i/\sqrt{2}$. Finally, we have the eigenvector

$$|J_z = \hbar\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} = |R\rangle \tag{5.65}$$

Similarly, we get

$$|J_z = -\hbar\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} = |L\rangle \tag{5.66}$$

So the eigenvectors of $\hat{J}_z$ and hence of $\hat{R}(\theta)$ are the RCP and LCP basis states. We then have

$$\begin{aligned}\hat{R}\theta)\,|R\rangle &= \left(\cos\theta\hat{I} + \frac{i}{\hbar}\sin\theta\hat{J}_z\right)|R\rangle \\ &= (\cos\theta + i\sin\theta)\,|R\rangle \\ &= e^{i\theta}\,|R\rangle \end{aligned} \tag{5.67}$$

Similarly,

$$\hat{R}(\theta)\,|L\rangle = e^{-i\theta}\,|L\rangle \tag{5.68}$$

This agrees with our earlier discussion where we found that the eigenvalues of a unitary operator have complex exponential form.

Physically, this says that $|R\rangle$ and $|L\rangle$ are only changed by an overall phase factor under a rotation of the basis. This allows us to easily specify what happens to an arbitrary vector $|\psi\rangle$ under rotations.

The procedure we will use next is, as we will see over and over again, the standard way to do things in quantum mechanics. First, we expand the arbitrary vector in the $\{|R\rangle, |L\rangle\}$ basis

$$|\psi\rangle = |R\rangle\langle R\,|\,\psi\rangle + |L\rangle\langle L\,|\,\psi\rangle \tag{5.69}$$

We then apply the rotation operator to obtain

$$\begin{aligned}\hat{R}(\theta)\,|\psi\rangle &= \hat{R}(\theta)\,|R\rangle\langle R\,|\,\psi\rangle + \hat{R}(\theta)\,|L\rangle\langle L\,|\,\psi\rangle \\ &= e^{i\theta}\,|R\rangle\langle R\,|\,\psi\rangle + e^{-i\theta}\,|L\rangle\langle L\,|\,\psi\rangle \end{aligned} \tag{5.70}$$

or the RCP component of the vector is multiplied by the phase factor $e^{i\theta}$ and the LCP component of the vector is multiplied by a *different* phase factor $e^{-i\theta}$. Thus, rotations change the *relative* phase of the components, which is a *real* physical change (as opposed to an overall phase change of the state vector).

We interpret (reasons will be clear later) these results to say that the RCP photon is in a state which is an eigenvector of $\hat{J}_z$ with eigenvalue $+\hbar$ or that the photon in that state has z-component of spin $= +\hbar$. Similarly, a LCP photon has z-component of spin $= -\hbar$.

Now, it is an experimental fact that if a photon traveling in the z-direction is absorbed by matter, then the z-component of the angular momentum of the absorber increases by $\hbar$ or decreases by $\hbar$. It never remains the same, nor does it change by any value other than $\pm\hbar$.

One cannot predict, for any single photon, whether the change will be $+\hbar$ or $-\hbar$. We can, however, predict the probability of either value occurring. In particular, according to our probability formalism, we must have

$$|\langle R\,|\,\psi\rangle|^2 = \text{ probability of } +\hbar \quad , \quad |\langle L\,|\,\psi\rangle|^2 = \text{ probability of } -\hbar \tag{5.71}$$

and the average value of the z-component of the angular momentum is

$$\langle \hat{J}_z \rangle = \sum_{all\ possibilities} (\text{eigenvalue}) \times (\text{probability of the eigenvalue}) \tag{5.72}$$

or

$$\langle \hat{J}_z \rangle = +\hbar |\langle R | \psi \rangle|^2 - \hbar |\langle L | \psi \rangle|^2 \tag{5.73}$$

In general, a photon is neither pure RCP nor pure LCP and the angular momentum does not have a definite value.

We can still talk in terms of probabilities, however. The discreteness of the angular momentum spectrum forces a probabilistic interpretation on us.

We can easily see how all of this works using our mathematical formalism for average values as follows:

$$\begin{aligned}
\langle \hat{J}_z \rangle &= \langle \psi | \hat{J}_z | \psi \rangle \\
&= (\langle \psi | R \rangle \langle R | + \langle \psi | L \rangle \langle L |) \hat{J}_z (|R\rangle \langle R | \psi \rangle + |L\rangle \langle L | \psi \rangle) \\
&= (\langle R | \psi \rangle^* \langle R | + \langle L | \psi \rangle^* \langle L |) \hat{J}_z (|R\rangle \langle R | \psi \rangle + |L\rangle \langle L | \psi \rangle) \\
&= \langle R | \hat{J}_z | R \rangle |\langle R | \psi \rangle|^2 + \langle L | \hat{J}_z | L \rangle |\langle L | \psi \rangle|^2 \\
&\qquad + \langle R | \hat{J}_z | L \rangle \langle R | \psi \rangle^* \langle L | \psi \rangle + \langle L | \hat{J}_z | R \rangle \langle L | \psi \rangle^* \langle R | \psi \rangle \\
&= +\hbar |\langle R | \psi \rangle|^2 - \hbar |\langle L | \psi \rangle|^2
\end{aligned} \tag{5.74}$$

as we showed earlier(5.73).

Let us return for a moment to the matrix representation of the $\hat{J}_z$ operator. We have found the following results:

$$\hat{J}_z |R\rangle = +\hbar |R\rangle \text{ and } \hat{J}_z |L\rangle = -\hbar |L\rangle \tag{5.75}$$

In the $\{|R\rangle, |L\rangle\}$ basis, these relations imply the matrix representation

$$\hat{J}_z = \begin{pmatrix} \langle R | \hat{J}_z | R \rangle & \langle R | \hat{J}_z | L \rangle \\ \langle L | \hat{J}_z | R \rangle & \langle L | \hat{J}_z | L \rangle \end{pmatrix} = \hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{5.76}$$

which is the *standard form* of $\hat{J}_z$ in terms of one of the so-called *Pauli matrices*, namely,

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow \hat{J}_z = \hbar \hat{\sigma}_z \tag{5.77}$$

Now

$$|x\rangle = \frac{1}{\sqrt{2}} (|R\rangle + |L\rangle) \text{ and } |y\rangle = \frac{i}{\sqrt{2}} (|R\rangle - |L\rangle) \tag{5.78}$$

and, therefore, in the $\{|x\rangle, |y\rangle\}$ basis we have the matrix representation

$$\hat{J}_z = \begin{pmatrix} \langle x | \hat{J}_z | x \rangle & \langle x | \hat{J}_z | y \rangle \\ \langle y | \hat{J}_z | x \rangle & \langle y | \hat{J}_z | y \rangle \end{pmatrix} = \hbar \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{5.79}$$

which is the form we *guessed and used* earlier.

5.2.2 Projection Operators

Let us now turn our attention to projection operators and density operators in the context of photon polarization.

The general operator $|\psi\rangle\langle\phi|$ can be represented by a 2×2 matrix in the polarization state vector space. It is constructed using the *outer product rule*:

$$\hat{P} = |\psi\rangle\langle\phi| = \begin{pmatrix}\psi_x\\ \psi_y\end{pmatrix}\begin{pmatrix}\phi_x^* & \phi_y^*\end{pmatrix} = \begin{pmatrix}\psi_x\phi_x^* & \psi_x\phi_y^*\\ \psi_y\phi_x^* & \psi_y\phi_y^*\end{pmatrix} \tag{5.80}$$

or *equivalently*, by choosing a basis and finding the matrix representation

$$\begin{aligned}\hat{P} &= \begin{pmatrix}\langle x|\hat{P}|x\rangle & \langle x|\hat{P}|y\rangle\\ \langle y|\hat{P}|x\rangle & \langle y|\hat{P}|y\rangle\end{pmatrix} = \begin{pmatrix}\langle x|\psi\rangle\langle\phi|x\rangle & \langle x|\psi|\langle\phi|y\rangle\\ \langle y|\psi||\phi|x\rangle & \langle y|\psi||\phi|y\rangle\end{pmatrix}\\ &= \begin{pmatrix}\psi_x\phi_x^* & \psi_x\phi_y^*\\ \psi_y\phi_x^* & \psi_y\phi_y^*\end{pmatrix}\end{aligned}$$

In particular, we have for the projection operators

$$\begin{aligned}|x\rangle\langle x| &= \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \quad , \quad |x\rangle\langle y| = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\\ |y\rangle\langle x| &= \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} \quad , \quad |y\rangle\langle y| = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\end{aligned} \tag{5.81}$$

From these results we easily see that

$$|x\rangle\langle x| + |y\rangle\langle y| = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = \hat{I} \tag{5.82}$$

and

$$\begin{aligned}|\psi\rangle &= \hat{I}|\psi\rangle = (|x\rangle\langle x| + |y\rangle\langle y|)|\psi\rangle \qquad (5.83)\\ &= |x\rangle\langle x|\psi\rangle + |y\rangle\langle y|\psi\rangle = \begin{pmatrix}\psi_x\\ \psi_y\end{pmatrix}\end{aligned}$$

as they should. Similarly, we have

$$|R\rangle\langle R| + |L\rangle\langle L| = \hat{I} \tag{5.84}$$

which leads to

$$\begin{aligned}\hat{J}_z &= \hat{J}_z\hat{I} = \hat{J}_z(|R\rangle\langle R| + |L\rangle\langle L|) \qquad (5.85)\\ &= \hbar|R\rangle\langle R| - \hbar|L\rangle\langle L| \qquad (5.86)\end{aligned}$$

which is the expansion of the operator $\hat{J}_z$ in terms of eigenvalues and 1-dimensional subspace projection operators (eigenvectors) that we discussed earlier.

5.2.3 Amplitudes and Probabilities

The *probability interpretation* we have been making follows from the concept of *superposition*. The superposition idea says that we can write any arbitrary photon state as a linear combination of basis states

$$|\psi\rangle = |R\rangle\langle R|\psi\rangle + |L\rangle\langle L|\psi\rangle \tag{5.87}$$

We then interpreted $|\langle R|\psi\rangle|^2$ as the probability that the photon in the state $|\psi\rangle$ will behave as a RCP photon in the state $|R\rangle$.

Generalizing this statement, we say that a system in a state $|\psi\rangle$, in Quantum Mechanics, has a probability $|\langle\phi|\psi\rangle|^2$ of behaving like it was in the state $|\phi\rangle$.

You might now conclude, from the experimental fact that only $\pm\hbar$ is transferred to matter by light, that photons are *always* either in the state $|R\rangle$ with some probability α *or* in the state $|l\rangle$ with probability $1-\alpha$.

FACT: An x-polarized photon never passes through a y-polaroid

PROBLEM: If, the above interpretation of being either $|R\rangle$ or $|L\rangle$ was true, then

1. an x-polarized photon has a probability $|\langle R|x\rangle|^2 = 1/2$ of being RCP and a RCP photon has a probability $|\langle y|R\rangle|^2 = 1/2$ of being a y-polarized photon and thus passing through a y-polaroid.
2. an x-polarized photon has a probability $|\langle L|x\rangle|^2 = 1/2$ of being LCP and a LCP photon has a probability $|\langle y|L\rangle|^2 = 1/2$ of being a y-polarized photon and thus passing through a y-polaroid.

This means that if we assume we can think that the photon is either $|R\rangle$ or $|L\rangle$ but *we do not know which*, i.e., photon properties have an *objective reality*, then the total probability that an x-polarized photon would get through a y-polaroid in this interpretation is

$$\text{total probability } = |\langle R|x\rangle|^2|\langle y|R\rangle|^2 + |\langle L|x\rangle|^2|\langle y|L\rangle|^2 = \frac{1}{2} \tag{5.88}$$

However, as we stated, it *NEVER HAPPENS*. What is wrong?

SOLUTION: When we think of an x-polarized photon as being a RCP photon or a LCP photon with equal probability, we are ruling out the possibility of any *interference* between the RCP and LCP *amplitudes*. We are thinking *classically*!

We give *meaning* to the word *interference* here in this way.

The correct calculation of the probability, which lays the foundation for all of the amplitude mechanics rules in Quantum Mechanics, goes as follows:

1. The probability amplitude of an x-polarized photon passing through a y-polaroid = $\langle y \,|\, x \rangle = 0$, which implies that the probability $|\langle y \,|\, x \rangle|^2 = 0$ also.

2. If we say that the x-polarized photon is in a *superposition* of $|R\rangle$ and $|L\rangle$ (we make no statement about probabilities at this point), this implies that

$$|x\rangle = |R\rangle \langle R \,|\, x\rangle + |L\rangle \langle L \,|\, x\rangle \tag{5.89}$$

which gives

$$\langle y \,|\, x\rangle = \langle y \,|\, R\rangle \langle R \,|\, x\rangle + \langle y \,|\, L\rangle \langle L \,|\, x\rangle \tag{5.90}$$

or the *amplitude* for an x-polarized photon to pass through a y-polaroid is the *sum of two amplitudes*, namely, that it passes through as a RCP photon $\langle y \,|\, R\rangle \langle R \,|\, x\rangle$ and that it passes through as a LCP photon $\langle y \,|\, L\rangle \langle L \,|\, x\rangle$.

3. The *probability* of passing through is then the *absolute square of the total amplitude*

$$\begin{aligned} \textit{probability} &= |\langle y \,|\, R\rangle \langle R \,|\, x\rangle + \langle y \,|\, L\rangle \langle L \,|\, x\rangle|^2 \\ &= (\langle y \,|\, R\rangle^* \langle R \,|\, x\rangle^* + \langle y \,|\, L\rangle^* \langle L \,|\, x\rangle^*)(\langle y \,|\, R\rangle \langle R \,|\, x\rangle + \langle y \,|\, L\rangle \langle L \,|\, x\rangle) \\ &= |\langle R \,|\, x\rangle|^2 |\langle y \,|\, R\rangle|^2 + |\langle L \,|\, x\rangle|^2 |\langle y \,|\, L\rangle|^2 \\ &\quad \langle y \,|\, R\rangle \langle R \,|\, x\rangle \langle y \,|\, L\rangle^* \langle L \,|\, x\rangle^* + \langle y \,|\, R\rangle^* \langle R \,|\, x\rangle^* \langle y \,|\, L\rangle \langle L \,|\, x\rangle \end{aligned}$$

4. The first two terms are the same as the incorrect calculation (6.88) above. The last two terms represent *interference effects* between the two amplitudes (RCP way and LCP way).

A simple calculation shows that the interference terms exactly cancel the first two terms and that the probability equals zero in agreement with experiment!

INTERPRETATION: The way to interpret this result is as follows:

$\langle y \,|\, R\rangle \langle R \,|\, x\rangle$ = probability amplitude for an x-polarized photon to pass through a y-polaroid as a RCP photon

$\langle y \,|\, L\rangle \langle L \,|\, x\rangle$ = probability amplitude for an x-polarized photon to pass through a y-polaroid as a LCP photon

These are *indistinguishable* ways for the process to occur, i.e., no measurement exists that can tell us whether it passes through the system as an RCP photon or as a LCP photon without destroying the interference,i.e., without radically altering the experiment.

To get the correct total probability, we *add all the amplitudes for indistinguishable ways* and then square the resulting total amplitude.

In the incorrect calculation, we found the probability for each indistinguishable way and then added the probabilities.

In one case, we eliminated the interference effects and got the wrong result and, in the other case, we included the interference effects and obtained the correct result.

Summarizing, we have these *rules for amplitude mechanics and probabilities* in Quantum Mechanics:

1. The probability amplitude for two successive events is the *product* of the amplitudes for each event, i.e., the amplitude for the x-polarized photon to pass through the y-polaroid as a RCP polarized photon is the product of the amplitude for an x-polarized photon to be a RCP photon $\langle R \,|\, x \rangle$ and the amplitude for a RCP photon to be a y-polarized photon $\langle y \,|\, R \rangle$

   $$\langle R \,|\, x \rangle \langle y \,|\, R \rangle$$

2. The total amplitude for a process that can take place in several *indistinguishable* ways is the *sum* of the amplitudes for each individual way, i.e.,

   $$\langle y \,|\, x \rangle = \langle R \,|\, x \rangle \langle y \,|\, R \rangle + \langle L \,|\, x \rangle \langle y \,|\, L \rangle$$

   We note here that this is merely a reflection of the property of projection operators that

   $$\hat{I} = |R\rangle \langle R| + |L\rangle \langle L|$$

   which says that

   $$\langle y \,|\, x \rangle = \langle y| \hat{I} |x \rangle = \langle R \,|\, x \rangle \langle y \,|\, R \rangle + \langle L \,|\, x \rangle \langle y \,|\, L \rangle$$

   Thus, the mathematical sum over all projection operators being equal to the identity operator is physically equivalent to the sum over all possible intermediate states and it turns into a sum over all the amplitudes for indistinguishable ways in this interpretation.

3. The total probability for the process to occur is the absolute square of the total amplitude.

So, in classical physics, we

1. find amplitudes and probabilities of each way separately
2. add all probabilities to get total probability

We get *NO* interference effects!!

In Quantum Mechanics, we

1. find the amplitudes for each indistinguishable way the process can occur
2. add all the amplitudes to get a total amplitude
3. square the total amplitude to get the total probability

We get interference effects!!

The important result here is that we must consider *ALL INDISTINGUISHABLE WAYS* in step (2).

An *indistinguishable way* is characterized as follows:

1. If two ways are indistinguishable, then there exists *no measurement* that can decide which of the two ways actually happened without altering the experiment.
2. In particular, if we attempt to find out, then the interference effects will disappear and we will return to the classical result obtained by adding probabilities.

What actually happens is that during any measurement trying distinguish the ways, the relative phase of the components in the superposition becomes completely uncertain and this will wash out the interference. This happens as follows: instead of

$$|x\rangle = |R\rangle\langle R|x\rangle + |L\rangle\langle L|x\rangle \tag{5.91}$$

we would have, if we attempted to add a measurement to determine if the x-polarized photon was RCP or LCP, to consider the state

$$|\tilde{x}\rangle = e^{i\alpha_R}|R\rangle\langle R|x\rangle + e^{i\alpha_L}|L\rangle\langle L|x\rangle \tag{5.92}$$

A probability calculation then gives

$$\begin{aligned} \textit{total probability} &= |\langle y|R\rangle\langle R|x\rangle + \langle y|L\rangle\langle L|x\rangle|^2 \\ &\quad + 2Real\left[\langle y|R\rangle\langle R|x\rangle e^{i(\alpha_R-\alpha_L)}\langle y|L\rangle^*\langle L|x\rangle^*\right] \end{aligned} \tag{5.93}$$

The observed probability, which is the result of many *identical* measurements in the laboratory(in the standard interpretation), is an average over all values of the extra phases(they are random in different experiments).

This involves integrating over the relative phase,i.e.,

$$\frac{1}{2\pi}\int_0^{2\pi} e^{i(\alpha_R-\alpha_L)}\,d(\alpha_R-\alpha_L) = 0 \tag{5.94}$$

It is clear that the interference term averages to zero and we get the classical result! This means we *cannot add* such a measurement and retain any quantum results!

5.2.4 Pure States, Unpure States and Density Operators

If the photon were in the state $|x\rangle$, then we would have, for some linear operator $\hat{A}$

$$\langle \hat{A} \rangle = \langle x | \hat{A} | x \rangle \tag{5.95}$$

From our earlier discussion of density operators, however, we must also have,

$$\langle \hat{A} \rangle = Tr(\hat{W}\hat{A}) \tag{5.96}$$

where $\hat{W}$ is a density operator. Therefore we must have (using the $\{|x\rangle, |y\rangle\}$ basis)

$$\begin{aligned}\langle \hat{A} \rangle = \langle x | \hat{A} | x \rangle = Tr(\hat{W}\hat{A}) &= \langle x | \hat{W}\hat{A} | x \rangle + \langle y | \hat{W}\hat{A} | y \rangle \\ &= \langle x | \hat{W}\hat{I}\hat{A} | x \rangle + \langle y | \hat{W}\hat{I}\hat{A} | y \rangle \\ &= \langle x | \hat{W} | x \rangle \langle x | \hat{A} | x \rangle + \langle x | \hat{W} | y \rangle \langle y | \hat{A} | x \rangle \\ &\quad + \langle y | \hat{W} | x \rangle \langle x | \hat{A} | y \rangle + \langle y | \hat{W} | y \rangle \langle y | \hat{A} | y \rangle\end{aligned}$$

This implies that

$$\langle x | \hat{A} | x \rangle = 1 \text{ and } \langle y | \hat{A} | x \rangle = \langle x | \hat{A} | y \rangle = \langle y | \hat{A} | y \rangle = 0 \tag{5.97}$$

or

$$\hat{W} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |x\rangle \langle x| \tag{5.98}$$

which says that $|x\rangle$ is a *pure state*.

Now suppose that the photon is in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}} |x\rangle + \frac{1}{\sqrt{2}} |y\rangle \tag{5.99}$$

This says that the probability = 1/2 that the photon behaves like $|x\rangle$ and the probability = 1/2 that it behaves like $|y\rangle$. Note that the relative phase between the components is assumed to be *known exactly* in this state. In this case, we have

$$\begin{aligned}\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle &= \frac{1}{2} \left[\langle x | \hat{A} | x \rangle + \langle x | \hat{A} | y \rangle + \langle y | \hat{A} | x \rangle + \langle y | \hat{A} | y \rangle \right] \\ &= Tr(\hat{W}\hat{A}) = \langle x | \hat{W}\hat{A} | x \rangle + \langle y | \hat{W}\hat{A} | y \rangle = \langle x | \hat{W}\hat{I}\hat{A} | x \rangle + \langle y | \hat{W}\hat{I}\hat{A} | y \rangle \\ &= \langle x | \hat{W} | x \rangle \langle x | \hat{A} | x \rangle + \langle x | \hat{W} | y \rangle \langle y | \hat{A} | x \rangle \\ &\quad + \langle y | \hat{W} | y \rangle \langle x | \hat{A} | x \rangle + \langle y | \hat{W} | y \rangle \langle y | \hat{A} | y \rangle\end{aligned}$$

which implies that

$$\langle x | \hat{W} | x \rangle = \frac{1}{2} = \langle y | \hat{W} | y \rangle = \langle y | \hat{W} | x \rangle = \langle x | \hat{W} | y \rangle \tag{5.100}$$

or

$$\hat{W} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix} = |\psi\rangle\langle\psi| \tag{5.101}$$

So, again we have a pure state.

But what happens if we *only know* that the probability = 1/2 that the photon behaves like $|x\rangle$ and the probability = 1/2 that it behaves like $|y\rangle$. This says that we might write the state vector as

$$|\psi\rangle = a|x\rangle + b|y\rangle \tag{5.102}$$

where we only know that $|a|^2 = |b|^2 = 1/2$. Let us choose

$$a = \frac{e^{i\alpha_a}}{\sqrt{2}} \text{ and } b = \frac{e^{i\alpha_b}}{\sqrt{2}} \tag{5.103}$$

We do not have any phase information in this case. In addition, the phases values could be different in each separate experiment. This last fact means that we must average over the relative phase (the only meaningful phase) $\alpha_a - \alpha_b$ when computing the probabilities and this means that all interference effects will vanish.

When we calculate the expectation value we have

$$\begin{aligned}\langle\hat{A}\rangle &= \langle\psi|\hat{A}|\psi\rangle \\ &= \frac{1}{2}\left[\langle x|\hat{A}|x\rangle + \langle x|\hat{A}|y\rangle e^{-i(\alpha_a-\alpha_b)} + \langle y|\hat{A}|x\rangle e^{i(\alpha_a-\alpha_b)} + \langle y|\hat{A}|y\rangle\right]\end{aligned}$$

and when we average over the relative phase we obtain

$$\langle\hat{A}\rangle = \frac{1}{2}\langle x|\hat{A}|x\rangle + \frac{1}{2}\langle y|\hat{A}|y\rangle$$

Again, we must have

$$\begin{aligned}\langle\hat{A}\rangle = Tr(\hat{W}\hat{A}) &= \langle x|\hat{W}\hat{A}|x\rangle + \langle y|\hat{W}\hat{A}|y\rangle = \langle x|\hat{W}\hat{I}\hat{A}|x\rangle + \langle y|\hat{W}\hat{I}\hat{A}|y\rangle \\ &= \langle x|\hat{W}|x\rangle\langle x|\hat{A}|x\rangle + \langle x|\hat{W}|y\rangle\langle y|\hat{A}|x\rangle \\ &\quad + \langle y|\hat{W}|y\rangle\langle x|\hat{A}|x\rangle + \langle y|\hat{W}|y\rangle\langle y|\hat{A}|y\rangle\end{aligned}$$

which implies that

$$\langle x|\hat{W}|x\rangle = \frac{1}{2} = \langle y|\hat{W}|y\rangle \text{ and } \langle y|\hat{W}|x\rangle = \langle x|\hat{W}|y\rangle = 0$$

or

$$\begin{aligned}\hat{W} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} &= \frac{1}{2}|x\rangle\langle x| + \frac{1}{2}|y\rangle\langle y| \\ &= Prob(x)|x\rangle\langle x| + Prob(y)|y\rangle\langle y|\end{aligned}$$

This is a *nonpure* or *mixed state*.

So, *we have a pure state only if the relative phase information is known exactly.*

The way to describe a nonpure state is by the corresponding density matrix, which only requires knowing probabilities and not phases. *It really does not have a state vector.*

5.2.5 Unpolarized Light

Consider the following experiment. We have a beam of monochromatic light that is composed of photons from two sources which output photons in the states $|\psi_1\rangle$ or $|\psi_2\rangle$, respectively. The sources emit the photons randomly and are independent of each other, which implies that we cannot tell which source any particular photon comes from.

We assign these probabilities

$$\begin{aligned} p_1 &= \text{ probability that a photon comes from source \#1} \\ p_2 &= \text{ probability that a photon comes from source \#2} \end{aligned}$$

where $p_1 + p_2 = 1$. Now the probability that a particular observed photon transfers $\hbar$ is

$$p_+ = p_1 |\langle R | \psi_1 \rangle|^2 + p_2 |\langle R | \psi_2 \rangle|^2 \tag{5.104}$$

and the probability that it transfers $-\hbar$ is

$$p_- = p_1 |\langle L | \psi_1 \rangle|^2 + p_2 |\langle L | \psi_2 \rangle|^2 \tag{5.105}$$

This implies that the average value of the angular momentum transfer for the beam of photons is

$$\begin{aligned} \langle \hat{J}_z \rangle &= \hbar p_+ - \hbar p_- \\ &= \hbar p_1 |\langle R | \psi_1 \rangle|^2 + \hbar p_2 |\langle R | \psi_2 \rangle|^2 - \hbar p_1 |\langle L | \psi_1 \rangle|^2 - \hbar p_2 |\langle L | \psi_2 \rangle|^2 \\ &= p_1 \left[\hbar |\langle R | \psi_1 \rangle|^2 - \hbar |\langle L | \psi_1 \rangle|^2 \right] + p_2 \left[\hbar |\langle R | \psi_2 \rangle|^2 - \hbar |\langle L | \psi_2 \rangle|^2 \right] \\ &= p_1 \langle \hat{J}_z \rangle_1 + p_2 \langle \hat{J}_z \rangle_2 \end{aligned}$$

or, the average value of the angular momentum transfer for the beam of photons = sum over the average value in each beam weighted by the probability that photon comes from that beam.

Let me emphasize(once again) at this point that it is important to realize that the statement

> The photon is either in the state $|\psi_1\rangle$ or $|\psi_2\rangle$ but we do not know which

is *NOT* the same statement as

The photon is in a state which is a superposition of $|\psi_1\rangle$ and $|\psi_2\rangle$

In the second case, we are saying the *relative phase* is known as in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}|x\rangle + \frac{1}{\sqrt{2}}|y\rangle \tag{5.106}$$

which we found to be a pure state. *Being in a superposition implies that we know the relative phase of the components.*

In the first case, however, we are saying that the relative phase is unknown and, as we have seen, interference effects will vanish. We can only specify a density matrix in this case.

In pure states, we have superpositions and the probability amplitude rules apply. In nonpure or mixed states, where the system is in one of several states with definite probabilities, we find weighted averages (weighted with the state probabilities) of the value in each state. We use addition of probabilities with no interference effects, which as we have seen, is equivalent to saying the relative phase is unknown.

Unpolarized light has equal probability of being in any polarization state. It is just a special nonpure or mixed state. No relative phase information is known for unpolarized light.

5.2.6 How Does the Polarization State Vector Change?

Up to now we have been considering devices such as polaroids and prisms, which are *go-nogo* devices. Some photons get through and some do not for each of these devices depending on their polarization state.

We now consider devices where all the photons get through no matter what their polarization state is, but, during transit, the device changes the incident polarization state in some way.

In particular, we consider the example of a *birefringent* crystal, such as calcite. A calcite crystal has a *preferred* direction called the *optic* axis. The crystal has a different index of refraction for light polarized parallel to the optic axis than it has for light polarized perpendicular to the optic axis. We assume that the optic axis is in the x-y plane and send a beam of photons in the z-direction. Photons polarized perpendicular to the optic axis are called *ordinary* and are in the state $|o\rangle$ and photons polarized parallel to the optic axis are called *extraordinary* and are in the state $|e\rangle$.

The set $\{|o\rangle, |e\rangle\}$ forms an orthonormal basis and general photon states interacting with a calcite crystal are written as superpositions of these basis states.

This is an example of a general rule in quantum mechanics.

If we are doing an experiment using a particular measuring device that measures the observable $\hat{Q}$, then we should use as the basis for all states, the eigenvectors of $\hat{Q}$. As we shall see, this requirement pushes us to ask the correct experimental questions (those that quantum mechanics can answer). This particular basis is called the *home* space for the experiment.

Now the phase of a light wave with wavelength λ as it propagates through a medium in the z-direction is given by the quantity

$$\phi = e^{ikz} \tag{5.107}$$

with

$$k = \frac{2\pi}{\lambda} = \frac{n\omega}{c} \tag{5.108}$$

where n = index of refraction, $\omega = 2\pi f$, f = frequency and c = speed of light.

Since the phase depends on the index of refraction, the effect of passing through a calcite crystal is to change the *relative phase* of the $|o\rangle$ and $|e\rangle$ components making up the superposition, which is a real physical change that is measurable.

We assume that the photon entering the calcite crystal is in the initial state

$$|\psi_{in}\rangle = |e\rangle \langle e\,|\,\psi_{in}\rangle + |o\rangle \langle o\,|\,\psi_{in}\rangle \tag{5.109}$$

The two components have different indices of refraction n_e and n_o, respectively.

If the beam passes through a length ℓ then the state upon leaving is given by (remember the component phases change differently)

$$|\psi_{out}\rangle = e^{ik_e\ell}\,|e\rangle \langle e\,|\,\psi_{in}\rangle + e^{ik_o\ell}\,|o\rangle \langle o\,|\,\psi_{in}\rangle = \hat{U}_\ell\,|\psi_{in}\rangle \tag{5.110}$$

where

$$\hat{U}_z = e^{ik_e z}\,|e\rangle \langle e| + e^{ik_o z}\,|o\rangle \langle o| \tag{5.111}$$

is a *time development* operator of some sort since ℓ= distance traveled in a time t is proportional to t.

Now we define two new quantities which will be with us throughout our study of Quantum Mechanics. For transitions between two states (*in* and *out* in this case)

$\langle\phi\,|\,\psi_{out}\rangle = \langle\phi|\,\hat{U}_z\,|\psi_{in}\rangle$ = the transition amplitude for a photon to enter the calcite in state $|\psi_{in}\rangle$ and leave in state $|\phi\rangle$

$|\langle\phi\,|\,\psi_{out}|^2\rangle = |\langle\phi|\,\hat{U}_z\,|\psi_{in}\rangle|^2$ = the corresponding transition probability

To proceed any further, we need to find out more about the operator $\hat{U}_z$. Now

$$\begin{aligned} |\psi_z\rangle &= \text{state of photon after traveling a distance z through calcite} \\ &= \hat{U}_z |\psi_{in}\rangle \end{aligned} \tag{5.112}$$

From the form of $\hat{U}_z$ we have

$$\begin{aligned} \hat{U}_{z+\epsilon} &= e^{ik_e(z+\epsilon)} |e\rangle\langle e| + e^{ik_o(z+\epsilon)} |o\rangle\langle o| \\ &= (e^{ik_e\epsilon} |e\rangle\langle e| + e^{ik_o\epsilon} |o\rangle\langle o|)(e^{ik_e z} |e\rangle\langle e| + e^{ik_o z} |o\rangle\langle o|) \end{aligned} \tag{5.113}$$

or

$$\hat{U}_{z+\epsilon} = \hat{U}_\epsilon \hat{U}_z \tag{5.114}$$

This implies that

$$|\psi_{z+\epsilon}\rangle = \hat{U}_{z+\epsilon} |\psi_{in}\rangle = \hat{U}_\epsilon \hat{U}_z |\psi_{in}\rangle = \hat{U}_\epsilon |\psi_z\rangle \tag{5.115}$$

Now let $\epsilon \to 0$ such that $k_o\epsilon \ll 1$ and $k_e\epsilon \ll 1$ and we can then write (to 1^{st}-order)

$$\begin{aligned} \hat{U}_\epsilon &= e^{ik_e\epsilon} |e\rangle\langle e| + e^{ik_o\epsilon} |o\rangle\langle o| \\ &= (1 + ik_e\epsilon) |e\rangle\langle e| + (1 + ik_o\epsilon) |o\rangle\langle o| \\ &= \hat{I} + i\epsilon\hat{K} \end{aligned} \tag{5.116}$$

where

$$\hat{I} = |e\rangle\langle e| + |o\rangle\langle o| \text{ and } \hat{K} = k_e |e\rangle\langle e| + k_o |o\rangle\langle o| \tag{5.117}$$

Now, the relation

$$\hat{K} = k_e |e\rangle\langle e| + k_o |o\rangle\langle o| \tag{5.118}$$

is an expansion of an operator in terms of its eigenvalues and the corresponding projection operators (eigenvectors). It says that the eigenvectors of $\hat{K}$ are $|e\rangle$ and $|o\rangle$ with eigenvalues k_e and k_o, respectively. This illustrates the *awesome power* in these methods we have developed!!

We then have

$$|\psi_{z+\epsilon}\rangle = (\hat{I} + i\epsilon\hat{K}) |\psi_z\rangle \tag{5.119}$$

or

$$|\psi_{z+\epsilon}\rangle - |\psi_z\rangle = i\epsilon\hat{K} |\psi_z\rangle \tag{5.120}$$

or

$$\lim_{\epsilon\to 0} \frac{|\psi_{z+\epsilon}\rangle - |\psi_z\rangle}{\epsilon} = i\hat{K} |\psi_z\rangle \tag{5.121}$$

which gives the differential equation for the *time* development of the state vector

$$\frac{d}{dt} |\psi_z\rangle = i\hat{K} |\psi_z\rangle \tag{5.122}$$

It is clearly similar to the differential equation we obtained earlier(Chapter 4) from the time development operator. If we follow the results from the earlier

case, then we should have

$\hat{K}$ = Hermitian operator and $\hat{U}_z$ = unitary operator

Let us derive these two results. We have, using the $x-y$ basis

$$\begin{aligned}\langle x\,|\,\psi_{z+\epsilon}\rangle - \langle x\,|\,\psi_z\rangle &= i\epsilon\,\langle x|\,\hat{K}\,|\psi_z\rangle \\ &= i\epsilon\,\langle x|\,\hat{K}\hat{I}\,|\psi_z\rangle = i\epsilon\,\langle x|\,\hat{K}\,|x\rangle\,\langle x\,|\,\psi_z\rangle + i\epsilon\,\langle x|\,\hat{K}\,|y\rangle\,\langle y\,|\,\psi_z\rangle\end{aligned} \tag{5.123}$$

or the change in the x-component of ψ_z as we move an infinitesimal amount ϵ has one part proportional to the x-component of ψ_z and a second part proportional to the y-component of ψ_z.

Similarly, we have

$$\begin{aligned}\langle y\,|\,\psi_{z+\epsilon}\rangle - \langle y\,|\,\psi_z\rangle &= i\epsilon\,\langle y|\,\hat{K}\,|\psi_z\rangle \\ &= i\epsilon\,\langle y|\,\hat{K}\hat{I}\,|\psi_z\rangle = i\epsilon\,\langle y|\,\hat{K}\,|x\rangle\,\langle x\,|\,\psi_z\rangle + i\epsilon\,\langle y|\,\hat{K}\,|y\rangle\,\langle y\,|\,\psi_z\rangle\end{aligned} \tag{5.124}$$

Now, since *no photons are lost* as we pass through, we must have

$$\langle\psi_{z+\epsilon}\,|\,\psi_{z+\epsilon}\rangle = \langle\psi_z\,|\,\psi_z\rangle = 1 \tag{5.125}$$

for all z. We then get

$$\begin{aligned}\langle\psi_{z+\epsilon}\,|\,\psi_{z+\epsilon}\rangle = \langle\psi_z\,|\,\psi_z\rangle &+ i\epsilon\left[\langle x|\,\hat{K}\,|x\rangle - \langle x|\,\hat{K}\,|x\rangle^*\right]|\,\langle x\,|\,\psi_z\rangle|^2 \\ &+ i\epsilon\left[\langle y|\,\hat{K}\,|y\rangle - \langle y|\,\hat{K}\,|y\rangle^*\right]|\,\langle y\,|\,\psi_z\rangle|^2 \\ &+ i\epsilon\left[\langle x|\,\hat{K}\,|y\rangle - \langle x|\,\hat{K}\,|y\rangle^*\right]\langle y\,|\,\psi_z\rangle\,\langle x\,|\,\psi_z\rangle^* \\ &+ i\epsilon\left[\langle y|\,\hat{K}\,|x\rangle - \langle y|\,\hat{K}\,|x\rangle^*\right]\langle x\,|\,\psi_z\rangle\,\langle y\,|\,\psi_z\rangle^*\end{aligned}$$

which says that we must have

$$\begin{aligned}\langle x|\,\hat{K}\,|x\rangle &= \langle x|\,\hat{K}\,|x\rangle^* \quad , \quad \langle y|\,\hat{K}\,|y\rangle = \langle y|\,\hat{K}\,|y\rangle^* \\ \langle x|\,\hat{K}\,|y\rangle &= \langle x|\,\hat{K}\,|y\rangle^* \quad , \quad \langle y|\,\hat{K}\,|x\rangle = \langle y|\,\hat{K}\,|x\rangle^*\end{aligned}$$

or that $\hat{K}$ is Hermitian.

Finally, one can show that $\hat{U}_z^\dagger\hat{U}_z = \hat{I}$ so that $\hat{U}_z$ is unitary as expected. From our earlier discussions, we then identify

$\hat{U}_z$ = transformation operator and $\hat{K}$ = generator of transformation

5.2.7 Calculating the Transition Probability

We defined the *transition probability* as

$$T(z) = |\,\langle\phi\,|\,\psi_{z,out}\rangle\,|^2 = |\,\langle\phi|\,\hat{U}_z\,|\psi_{z,in}\rangle\,|^2 \tag{5.126}$$

Using

$$\hat{U}_z = e^{ik_e z}\,|e\rangle\langle e| + e^{ik_o z}\,|o\rangle\langle o|$$

and

$$|\psi_{in}\rangle = a\,|o\rangle + b\,|e\rangle \;\text{ where } |a|^2 + |b|^2 = 1 \tag{5.127}$$

we get

$$\begin{aligned} T(z) &= |\langle\phi|\,(e^{ik_e z}\,|e\rangle\langle e| + e^{ik_o z}\,|o\rangle\langle o|)(a\,|o\rangle + b\,|e\rangle)|^2 \\ &= |\langle\phi|\,(be^{ik_e z}\,|e\rangle + ae^{ik_o z}\,|o\rangle\,|^2 \\ &= |be^{ik_e z}\langle\phi\,|\,e\rangle + ae^{ik_o z}\langle\phi\,|\,o\rangle\,|^2 \end{aligned} \tag{5.128}$$

Now let us ask a specific question.

Suppose $a = 1/\sqrt{2} = -ib$, which means the that photon entering the calcite crystal is an LCP photon.

What is the probability that it will exit as a RCP photon? This means we choose

$$|\phi\rangle = \frac{1}{\sqrt{2}}(|o\rangle + i\,|e\rangle) \tag{5.129}$$

or

$$\langle\phi\,|\,e\rangle = -\frac{i}{\sqrt{2}} \text{ and } \langle\phi\,|\,o\rangle = \frac{1}{\sqrt{2}} \tag{5.130}$$

We then get

$$\begin{aligned} T(z) &= |be^{ik_e z}\langle\phi\,|\,e\rangle + ae^{ik_o z}\langle\phi\,|\,o\rangle\,|^2 \\ &= |\frac{i}{\sqrt{2}}e^{ik_e z}\frac{i}{\sqrt{2}} + \frac{1}{\sqrt{2}}e^{ik_o z}\frac{1}{\sqrt{2}}|^2 = \frac{1}{4}|e^{ik_o z} - -e^{ik_e z}|^2 \\ &= \frac{1}{4}(1 + 1 - e^{i(k_o - k_e)z} - e^{-i(k_o - k_e)z}) \\ &= \frac{1}{2}(1 - \cos(k_o - k_e)z) \end{aligned} \tag{5.131}$$

If we choose $(k_o - k_e)z = \pi$, then $T = 1$ and *all* the LCP photons are turned into RCP photons by a calcite crystal of *just the right length.*

This simple example clearly exhibits the power of these techniques.

5.2.8 Some More Bayesian Thoughts

The essence of quantum theory is its ability to predict *probabilities* for the outcomes of tests based on specified preparations. Quantum mechanics is not a theory about reality; it is a prescription for making the best possible predictions about the future based on certain information (specified) about the past. The quantum theorist can tell you *what the odds* are, if you wish to bet on the occurrence of various events, such as the clicking of this or that detector.

However, a more common activity is the reverse situation where the outcomes of tests are known and it is their preparation(the initial state) that has to be guessed. The formal name for this process is *retrodiction*. Retrodicting is the analog of forecasting an event, but directed oppositely in time, i.e., to the past rather than the future. Just as one might forecast, from a knowledge of physical laws along with specific data about the current position and speed of a comet, where it will be ten years from now, one might retrodict where it was ten years ago.

Suppose we have the experiment described in Figure 5.2 below:

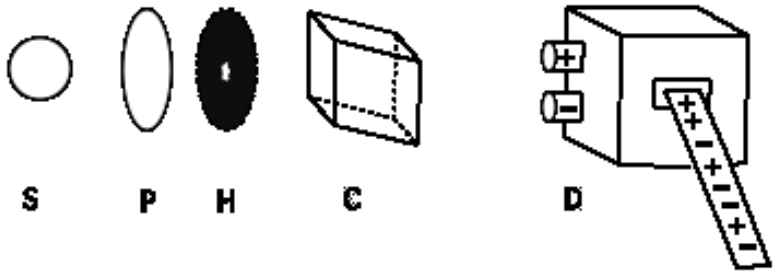


Figure 5.2: Experimental Setup

where S is a thermal source of light, P is a polarizer, H is pinhole, C is a calcite crystal, and D is a detector with separate counters for the two different polarized beams emerging from the calcite crystal. The detector D also makes a permanent record of the measured events. We assume that the light intensity is so weak and the detectors are so fast that individual photons can be registered. The arrivals of photons are recorded by printing + or − on a paper tape, according to whether the upper or lower detector was triggered, respectively. The sequence of + and − marks appears random. As the total number marks, N_+ and N_-, become large, we find that the corresponding probabilities(count ratios), tend to limits

$$\frac{N_+}{N_+ + N_-} \to \cos^2\alpha \quad , \quad \frac{N_-}{N_+ + N_-} \to \sin^2\alpha \tag{5.132}$$

where α is the angle between the polarization axis of the polaroid and the optic axis of the calcite crystal.

Now suppose that we do an experiment and find that the two detectors recorded 4 and 3 events, respectively. What can we *infer* about the orientation of the polarizer?

This is the so-called *inverse probability* problem, which as we have seen from our earlier discussions is an ideal situation to use Bayesian methods.

Consider the following description.

Event B is the outcome of the experiment described above with 4+ detections and 3– detections. This is a *single* experiment and *not* a set of seven experiments.

Event A is the positioning of the polarizer at an angle in the interval θ to $\theta + d\theta$, in that experiment.

Now, in a *statistical ensemble*, that is, an infinite set of conceptual replicas of the same system, the *relative frequencies* of events A and B define the probabilities $Prob(A|I) = Prob(A)$ and $Prob(B|I) = Prob(B)$, where I is all the information about the preparation(conditioning).

In addition, $Prob(A \cap B|I)$ is the joint probability of events A *and* B. This is the relative frequency of the occurrence of *both* events, in the statistical ensemble under consideration. $Prob(A|B \cap I)$ is the *conditional probability* of A, when B is true. As in Chapter 3, we have the relations

$$Prob(A \cap B|I) = Prob(A|B \cap I)Prob(B|I) = Prob(B|A \cap I)Prob(A|I) \quad (5.133)$$

and

$$Prob(A|B \cap I) = \frac{Prob(B|A \cap I)Prob(A|I)}{Prob(B|I)} \quad (5.134)$$

The last equation is Baye's theorem.

In this equation it is assumed that $Prob(B|A \cap I)$ is known from the appropriate physical theory. For example, in the above experiment, the theory tells us that the probabilities for triggering the upper and lower detectors are $\cos^2\theta$ and $\sin^2\theta$. We therefore, have from the Binomial distribution

$$\begin{aligned} Prob(B = \{4,3\}|A \cap I) &= \frac{(n_+ + n_-)!}{n_+!n_-!} Prob(+|A \cap I)^{n_+} Prob(-|A \cap I)^{n_-} \\ &= \frac{7!}{4!3!}(\cos^2\theta)^4(\sin^2\theta)^3 \\ &= 35\cos^8\theta\sin^6\theta \end{aligned} \quad (5.135)$$

In order to determine $Prob(B|A \cap I)$ we still need $Prob(A|I)$ and $Prob(B|I)$. These probabilities cannot be calculated from a theory nor determined empirically. The depend solely on the statistical ensemble that we have mentally constructed.

Let us consider the *complete* set of events of type A and call them $A_1, A_2, \ldots$, etc. For example, A_j represents the positioning of the polarizer at an angle between θ_j and $\theta_j + d\theta_j$. By completeness,

$$\sum_j P(A_j) = 1 \quad (5.136)$$

and therefore

$$P(B) = \sum_j P(B|A_j P(A_j) \tag{5.137}$$

At this point we introduce Baye's postulate (different from Baye's theorem). This postulate which we have used in earlier discussions is also called the *principle of indifference* or the *principle of insufficient reasoning.*

If we have no reason to expect that the person who positioned the polarizer had a preference for some particular orientation, we assume that all orientations are equally likely, so that

$$P(A) = \frac{d\theta}{\pi} \tag{5.138}$$

for every θ (we can always take $0 \leq \theta \leq \pi$ because θ and $\theta + \pi$ are equivalent).

We then have

$$P(B) = \sum_j P(B|A_j P(A_j) = \frac{1}{\pi} \int_0^\pi 35 \cos^8\theta \sin^6\theta \, d\theta = \frac{135}{2^{11}} \tag{5.139}$$

and we then obtain from Baye's theorem that

$$\begin{aligned} Prob(A|B \cap I) &= \frac{Prob(B|A \cap I) Prob(A|I)}{Prob(B|I)} \\ &= \frac{\frac{7!}{4!3!} \cos^8\theta \sin^6\theta \frac{d\theta}{\pi}}{\frac{135}{2^{11}}} = \frac{2^{11}}{5\pi} \cos^8\theta \sin^6\theta d\theta \end{aligned} \tag{5.140}$$

which is the probability that the angle is between θ to $\theta + d\theta$ given that event B is the outcome of a single experiment with 4+ detections and 3− detections.

Suppose, $d\theta = 1^\circ = 0.0175\, rad.$ If we plot

$$Prob(\theta|B = \{4,3\}, d\theta = 0.0175) = \frac{2^{11}}{5\pi} \cos^8\theta \sin^6\theta d\theta = 2.283 \cos^8\theta \sin^6\theta \tag{5.141}$$

versus θ we have

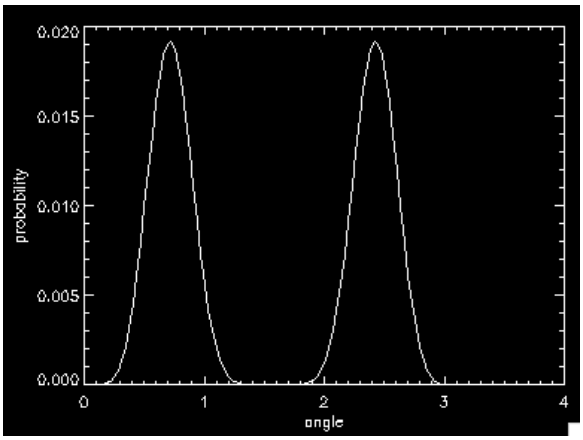


Figure 5.3: Most Likely Angles

This says that, given the single data set, the angle is most likely to be

$$0.72\,rad = 41.3^\circ \text{ or } 2.42\,rad = 138.7^\circ \tag{5.142}$$

Clearly, Bayesian analysis allows us to infer results from one-time experiments on single systems. The key is the use of Baye's postulate of indifference.

5.3 The Strange World of Neutral K-Mesons

We can now use the same formalism we developed for photon polarization to study elementary particles called K-mesons.

K-mesons are produced in high-energy accelerators via the production process

$$\pi^- + p^+ \rightarrow \Lambda^0 + K^0$$

In this reaction, electric charge is conserved. This reaction takes place via the so-called strong interactions. Another physical quantity called strangeness is also conserved in strong interactions. All K^0-mesons have a strangeness equal to +1.

For every particle there always exists an antiparticle. For the $\bar{K}^0$, the antiparticle is called the $\bar{K}^0$. The K^0-mesons have a strangeness equal to −1. A reaction involving the $\bar{K}^0$ is

$$\bar{K}^0 + p^+ \rightarrow \Lambda^0 + \pi^+$$

which is an absorption process.

The K-mesons that exist in the experimental world(the laboratory) are linear superpositions of K^0 and $\bar{K}^0$ states in the same way that RCP and LCP photons

were superpositions of $|x\rangle$ and $|y\rangle$ polarization states. So the world of K-mesons can be represented by a 2-dimensional vector space.

One basis for the vector space is the orthonormal set $\{|K^0\rangle, |\bar{K}^0\rangle\}$ where

$$\langle K^0 | K^0\rangle = 1 = \langle \bar{K}^0 | \bar{K}^0\rangle \text{ and } \langle K^0 | \bar{K}^0\rangle = 0 \tag{5.143}$$

Two linear operators are important for the study of K-mesons.

First, we represent the strangeness operator. We already stated that the states $|K^0\rangle$ and $|\bar{K}^0\rangle$ have definite values of strangeness, which means that they are eigenvectors of the strangeness operator $\hat{S}$ with eigenvalues ±1 (by convention). Using our formalism, this means that

$$\hat{S}|K^0\rangle = |K^0\rangle \text{ and } \hat{S}|\bar{K}^0\rangle = -|\bar{K}^0\rangle \tag{5.144}$$

$$\begin{aligned}\hat{S} &= \begin{pmatrix} \langle K^0|\hat{S}|K^0\rangle & \langle K^0|\hat{S}|\bar{K}^0\rangle \\ \langle \bar{K}^0|\hat{S}|K^0\rangle & \langle \bar{K}^0|\hat{S}|\bar{K}^0\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= |K^0\rangle\langle K^0| - |\bar{K}^0\rangle\langle \bar{K}^0| \end{aligned} \tag{5.145}$$

in the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis.

The second linear operator that is important in the K-meson system is charge conjugation $\hat{C}$. This operator changes particles into antiparticles and vice versa. In the K-meson system using the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis we *define* $\hat{C}$ by the particle-antiparticle changing relations

$$\hat{C}|K^0\rangle = |\bar{K}^0\rangle \text{ and } \hat{C}|\bar{K}^0\rangle = |K^0\rangle \tag{5.146}$$

$$\hat{C} = \begin{pmatrix} \langle K^0|\hat{C}|K^0\rangle & \langle K^0|\hat{C}|\bar{K}^0\rangle \\ \langle \bar{K}^0|\hat{C}|K^0\rangle & \langle \bar{K}^0|\hat{C}|\bar{K}^0\rangle \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.147}$$

We can find the eigenvectors and eigenvalues of the $\hat{C}$ operator as follows

$$\begin{aligned}\hat{C}|\psi\rangle &= \lambda|\psi\rangle \\ \hat{C}^2|\psi\rangle &= \lambda\hat{C}|\psi\rangle = \lambda^2|\psi\rangle = \hat{I}|\psi\rangle = |\psi\rangle\end{aligned} \tag{5.148}$$

where we have used $\hat{C}^2 = \hat{I}$. This result says that $\lambda^2 = 1$ or the eigenvalues of $\hat{C}$ are ±1. If we use the $\{|K^0\rangle, |\bar{K}^0\rangle\}$ basis and assume that

$$|\psi\rangle = a|K^0\rangle + b|\bar{K}^0\rangle \text{ where } |a|^2 + |b|^2 = 1 \tag{5.149}$$

we find for $\lambda = +1$

$$\begin{aligned}\hat{C}|\psi\rangle &= a\hat{C}|K^0\rangle + b\hat{C}|\bar{K}^0\rangle = a|\bar{K}^0\rangle + b|K^0\rangle \\ &= |\psi\rangle = a|K^0\rangle + b|\bar{K}^0\rangle\end{aligned} \tag{5.150}$$

or $a = b = 1/\sqrt{2}$. If we define the +1 eigenvector as $|K_S\rangle$, we then have

$$|K_S\rangle = \frac{1}{\sqrt{2}}\left(\left|K^0\right\rangle + \left|\bar{K}^0\right\rangle\right) \tag{5.151}$$

Similarly, if we define the −1 eigenvector as $|K_L\rangle$, we then have

$$|K_L\rangle = \frac{1}{\sqrt{2}}\left(\left|K^0\right\rangle - \left|\bar{K}^0\right\rangle\right) \tag{5.152}$$

where

$$\hat{C}\left|K^S\right\rangle = \left|K^S\right\rangle \text{ and } \hat{C}\left|K^L\right\rangle = -\left|K^L\right\rangle \tag{5.153}$$

Since the commutator $[\hat{S}, \hat{C}] \neq 0$, these two operators do not have a common set of eigenvectors. This means that both operators cannot have definite values in the same state. In fact, the concept of charge conjugation is meaningless for K-mesons in the $\{\left|K^0\right\rangle, \left|\bar{K}^0\right\rangle\}$ states and the concept of strangeness is meaningless for K-mesons in the $\{\left|K^S\right\rangle, \left|K^L\right\rangle\}$ states.

The $\{\left|K^S\right\rangle, \left|K^L\right\rangle\}$ states form a second orthonormal basis for the vector space (like the RCP and LCP polarization states).

The *standard approach* we will follow when studying physical systems using quantum mechanics will be to

1. define the Hamiltonian for the system
2. find its eigenvalues and eigenvectors
3. investigate the time development operator generated by the Hamiltonian and
4. calculate transition probabilities connected to experiments

Along the way we will define the properties of other operators appropriate to the system under investigation (like $\hat{S}$ and $\hat{C}$ above). It is the job of the theoretical physicist to derive or guess an appropriate Hamiltonian.

Since we are in 2-dimensional vector space, all operators are represented by 2×2 matrices. In the case of the K-meson system, we will assume the most general form constructed from all of the relevant operators. Therefore, we assume

$$\hat{H} = M\hat{I} + A\hat{C} + B\hat{S} = \begin{pmatrix} M + B & A \\ A & M - B \end{pmatrix} \tag{5.154}$$

and investigate the consequences of this assumption. We will assume that the matrix has been written down in the $\{\left|K^0\right\rangle, \left|\bar{K}^0\right\rangle\}$ basis.

Step 1

Investigate the commutators:

$$[\hat{H},\hat{H}]=0 \quad , \quad [\hat{H},\hat{S}]\neq 0 \quad , \quad [\hat{H},\hat{C}]\neq 0 \quad , \quad [\hat{S},\hat{C}]\neq 0 \tag{5.155}$$

Since the Hamiltonian always commutes with itself and it is not explicitly dependent on time, the physical observable connected to the Hamiltonian, namely the energy, is conserved.

Since they do not commute with the assumed form of $\hat{H}$, neither $\hat{S}$ nor $\hat{C}$ is conserved in this model.

When a physical observable is conserved, we say its value corresponds to a good quantum number that can be used to characterize(label) the ket vector representing the physical system.

Step 2

Investigate *special cases* (limits of the most general solution):

Case of A=0:

$$\hat{H}=M\hat{I}+B\hat{S}=\begin{pmatrix} M+B & 0 \\ 0 & M-B \end{pmatrix} \tag{5.156}$$

Now

$$[\hat{H},\hat{S}]=0 \tag{5.157}$$

which means that $\hat{H}$ and $\hat{S}$ share a common set of eigenvectors. We already know the eigenvectors(non-degenerate) for $\hat{S}$ and so the eigenvector/eigenvalue problem for $\hat{H}$ is already solved(*clearly this is a very powerful rule*). We have

$$\hat{H}\left|K^0\right\rangle=(M+B)\left|K^0\right\rangle \text{ and } \hat{H}\left|\bar{K}^0\right\rangle=(M+B)\left|\bar{K}^0\right\rangle \tag{5.158}$$

We could have surmised this from the diagonal form of the matrix representation, since the only way the matrix could be diagonal is for the basis states of the representation to be the eigenvectors.

$\hat{C}$ is not conserved in this case, since $[\hat{H},\hat{C}]\neq 0$.

The energy eigenstates, $\{\left|K^0\right\rangle,\left|\bar{K}^0\right\rangle\}$ in this case, are a basis for the vector space. This is always true for a Hermitian operator. This means that we can write any arbitrary vector as a linear combination of these vectors

$$|\psi\rangle=a\left|K^0\right\rangle+b\left|\bar{K}^0\right\rangle=a\,|E=M+B\rangle+b\,|E=M-B\rangle \tag{5.159}$$

Now, as we derived earlier, energy eigenstates have a simple time dependence. We have

$$\hat{H}\,|E\rangle = E\,|E\rangle \tag{5.160}$$
$$\hat{U}(t)\,|E\rangle = e^{-i\hat{H}t/\hbar}\,|E\rangle = e^{-iEt/\hbar}\,|E\rangle$$

Therefore, in this case, the time dependence of the arbitrary state vector is given by

$$|\psi(t)\rangle = ae^{-i(M+B)t/\hbar}\left|K^0\right\rangle + be^{-i(M-B)t/\hbar}\left|\bar{K}^0\right\rangle \tag{5.161}$$

This will be a general approach we will use, i.e., expand an arbitrary state in energy eigenstates and use the simple time dependence of the energy eigenstates to determine the more complex time dependence of the arbitrary state. Of course, we have to be able to solve the eigenvector/eigenvalue problem for the Hamiltonian(the energy operator) of the system under investigation.

Case of B = 0:

$$\hat{H} = M\hat{I} + A\hat{C} = \begin{pmatrix} M & A \\ A & M \end{pmatrix} \tag{5.162}$$

Now

$$[\hat{H}, \hat{C}] = 0 \tag{5.163}$$

which means that $\hat{H}$ and $\hat{C}$ share a common set of eigenvectors. We already know the eigenvectors(non-degenerate) for $\hat{C}$ and so the eigenvector/eigenvalue problem for $\hat{H}$ is again already solved. We have

$$\hat{H}\,|K_S\rangle = (M + A)\,|K_S\rangle \text{ and } \hat{H}\,|K_L\rangle = (M - A)\,|K_L\rangle \tag{5.164}$$

$\hat{S}$ is not conserved in this case, since $[\hat{H}, \hat{S}] \neq 0$.

The energy eigenstates, $\{|K_S\rangle, |K_L\rangle\}$ in this case, are a basis for the vector space. In this basis

$$\hat{H} = \begin{pmatrix} M + A & 0 \\ 0 & M - A \end{pmatrix} \tag{5.165}$$

as expected.

We could also solve this problem by finding the characteristic equation for the Hamiltonian matrix, i.e., since we have

$$\hat{H}\,|\psi\rangle = \lambda\,|\psi\rangle \rightarrow (\hat{H} - \lambda\hat{I})\,|\psi\rangle = 0 \tag{5.166}$$

the characteristic equation is

$$\begin{aligned} det(\hat{H} - \lambda\hat{I}) &= det\begin{pmatrix} M - \lambda & A \\ A & M = \lambda \end{pmatrix} \\ &= (M - \lambda)^2 - A^2 = 0 \rightarrow \lambda = M \pm A \end{aligned} \tag{5.167}$$

Since we have another basis, we can write any arbitrary vector as a linear combination of these vectors

$$|\psi\rangle = a\,|K_S\rangle + b\,|K_L\rangle = a\,|E = M + A\rangle + b\,|E = M - A\rangle \tag{5.168}$$

Therefore, in this case, we have the time dependence

$$|\psi(t)\rangle = ae^{-i(M+A)t/\hbar}\,|K_S\rangle + be^{-i(M-A)t/\hbar}\,|K_L\rangle \tag{5.169}$$

Step 3

Solve the general Hamiltonian problem (if possible; otherwise we must use approximation methods, which we will discuss later). We have

$$\hat{H} = M\hat{I} + A\hat{C} + B\hat{S} = \begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix} \tag{5.170}$$

We assume that the eigenvectors satisfy $\hat{H}\,|\phi\rangle = E\,|\phi\rangle$ where

$$|\phi\rangle = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \tag{5.171}$$

This gives

$$\begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = E\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \tag{5.172}$$

or

$$\begin{pmatrix} M+B-E & A \\ A & M-B-E \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = 0 \tag{5.173}$$

This is a set of two homogeneous equations in two unknowns. It has a nontrivial solution only if the determinant of the coefficients is zero

$$\begin{vmatrix} M+B-A & A \\ A & M-B-E \end{vmatrix} = (M+B-E)(M-B-E) - A^2 = 0 \tag{5.174}$$

This has solution

$$E_\pm = M \pm \sqrt{A^2 + B^2} \quad \text{(the energy eigenvalues)} \tag{5.175}$$

We solve for the eigenstates by substituting the eigenvalues into the eigenvalue/eigenvector equation

$$\begin{pmatrix} M+B & A \\ A & M-B \end{pmatrix}\begin{pmatrix} \phi_{1\pm} \\ \phi_{2\pm} \end{pmatrix} = E_\pm\begin{pmatrix} \phi_{1\pm} \\ \phi_{2\pm} \end{pmatrix} \tag{5.176}$$

After some algebra we get

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\sqrt{B \pm A^2 + B^2}} = \frac{B \pm \sqrt{A^2 + B^2}}{A} \tag{5.177}$$

We check the validity of this solution by comparing it the limiting cases; that is why we looked at the special cases earlier.

For $B = 0$, we have

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\mp A} = \pm 1 \rightarrow \phi_{1+} = \phi_{2+} \text{ and } \phi_{1-} = -\phi_{2-} \tag{5.178}$$

which says that

$$|\phi_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = |K_S\rangle \text{ and } |\phi_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} = |K_L\rangle \tag{5.179}$$

which agrees with the earlier results for this case.

In the other limiting case, $A = 0$, we have

$$\begin{aligned}\frac{\phi_{1+}}{\phi_{2+}} &= \infty \rightarrow \phi_{1+} = 1 \text{ and } \phi_{2+} = 0\\ \frac{\phi_{1-}}{\phi_{2-}} &= 0 \rightarrow \phi_{1-} = 0 \text{ and } \phi_{2-} = 1\end{aligned} \tag{5.180}$$

which says that

$$|\phi_+\rangle = \begin{pmatrix}1\\0\end{pmatrix} = \left|K^0\right\rangle \text{ and } |\phi_+\rangle = \begin{pmatrix}0\\1\end{pmatrix} = \left|\bar{K}^0\right\rangle \tag{5.181}$$

which again agrees with the earlier results for this case.

If we normalize the general solution

$$\frac{\phi_{1\pm}}{\phi_{2\pm}} = \frac{-A}{\sqrt{B \pm A^2 + B^2}} = \frac{B \pm \sqrt{A^2 + B^2}}{A} \tag{5.182}$$

using

$$|\phi_{1\pm}|^2 + |\phi_{2\pm}|^2 = 1 \tag{5.183}$$

we obtain

$$\begin{aligned}\phi_{1\pm} &= \frac{A}{\sqrt{A^2 + \left(B \mp \sqrt{A^2 + B^2}\right)^2}}\\ \phi_{2\pm} &= \frac{B \mp \sqrt{A^2 + B^2}}{\sqrt{A^2 + \left(B \mp \sqrt{A^2 + B^2}\right)^2}}\end{aligned} \tag{5.184}$$

and

$$|\phi_\pm\rangle = \frac{1}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \begin{pmatrix} A \\ -B \pm \sqrt{A^2+B^2} \end{pmatrix} \tag{5.185}$$

$$= \frac{1}{\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \left[A\left|K^0\right\rangle + \left(-B \pm \sqrt{A^2+B^2}\right)\left|\bar{K}^0\right\rangle \right]$$

$$= \frac{1}{2\sqrt{A^2 + \left(B \mp \sqrt{A^2+B^2}\right)^2}} \times$$

$$\left[\left(A - B \pm \sqrt{A^2+B^2}\right)|K_S\rangle + \left(A + B \pm \sqrt{A^2+B^2}\right)|K_L\rangle \right]$$

Step 4

Look at a realistic physical system that we can relate to experiment.

In the real world of K-mesons, the Hamiltonian is such that $B \ll A$. In this case the states $\{|K_S\rangle, |K_L\rangle\}$ are *almost* energy eigenstates or charge conjugation is *almost* conserved. We expect that instead of being able to write

$$|\phi_+\rangle = |K_S\rangle \text{ and } \phi_- = |K_L\rangle \tag{5.186}$$

which would be true if $B = 0$, we should be able to write

$$|\phi_+\rangle = \cos\frac{\theta}{2}|K_S\rangle + \sin\frac{\theta}{2}|K_L\rangle \tag{5.187}$$

$$|\phi_-\rangle = -\sin\frac{\theta}{2}|K_S\rangle + \cos\frac{\theta}{2}|K_L\rangle$$

where for $\theta \ll 1$ we clearly approximate the $B = 0$ result. Let us see how this works. For $B \ll A$, we choose

$$\frac{\theta}{2} = \frac{B}{2A} \ll 1 \tag{5.188}$$

and get

$$\frac{\theta}{2} \approx \tan\frac{\theta}{2} = \frac{B}{2A} = \frac{\sin\frac{\theta}{2}}{\cos\frac{\theta}{2}} \tag{5.189}$$

To lowest order we can then say

$$\sin\frac{\theta}{2} = \frac{B}{2A} = \delta \ll 1 \text{ and } \cos\frac{\theta}{2} = 1 \tag{5.190}$$

to get

$$|\phi_+\rangle = |K_S\rangle + \delta|K_L\rangle \tag{5.191}$$

$$|\phi_-\rangle = |K_L\rangle - \delta|K_S\rangle$$

This says that if $|\psi_{in}\rangle = |\phi_-\rangle$, then the number

$$\begin{aligned}\frac{|\langle K_S\,|\,\phi_-\rangle|^2}{|\langle K_L\,|\,\phi_-\rangle|^2} &= \frac{\text{probability of observing a } K_S}{\text{probability of observing a } K_L} \\ &= \frac{(\text{Number of } K_S)/(\text{Number of } K_S \text{ and } K_L)}{(\text{Number of } K_L)/(\text{Number of } K_S \text{ and } K_L)} \\ &= \frac{(\text{Number of } K_S)}{(\text{Number of } K_L)}\end{aligned} \tag{5.192}$$

gives the experimental ratio of the number of times we will measure a final state of $|K_S\rangle$ to the number of times we will measure the final state of $|K_L\rangle$. The signature for seeing a final state of $|K_L\rangle$ is to see a decay to 3 π-mesons and that of a final state of $|K_S\rangle$ is to see a decay to 2 π-mesons. The number is

$$\frac{|\langle K_S\,|\,\phi_-\rangle|^2}{|\langle K_L\,|\,\phi_-\rangle|^2} = |\delta|^2 \tag{5.193}$$

Now experiment gives the result $|\delta| = 2\times 10^{-3}$. This number is a measure of how large of an effect strangeness non-conservation has on this system.

If $B = 0$, then charge conjugation is conserved. If $B \neq 0$, then charge conjugation is not absolutely conserved. So δ is a measure of the lack of charge conjugation conservation in the K-meson system. If we identify the energy eigenvalues as the particle *rest energies*

$$M + A = m_S c^2 \quad , \quad M - A = m_L c^2 \tag{5.194}$$

we then have

$$A = \frac{m_S c^2 - m_L c^2}{2} = 10^{-5}\, eV \tag{5.195}$$

and

$$2A\delta = 10^{-17} m_K c^2 \tag{5.196}$$

or

$$\frac{B}{m_K c^2} = 10^{-17} \tag{5.197}$$

Thus, this is a one part in 10^{-17} effect! It is one of the best measurements ever made. It won the Nobel prize for the experimenters in 1963. It is now understood in detail by the standard model of elementary particles.

Now let us look at *Quantum Interference Effects* in this K-meson system.

Suppose that $B = 0$. Then the energy eigenstates are $\{|K_S\rangle, |K_L\rangle\}$ with eigenvalues $M \pm A$. Now let

$$|\psi_{in}\rangle = \left|K^0\right\rangle = \frac{1}{\sqrt{2}}(|K_S\rangle + |K_L\rangle) \tag{5.198}$$

This is not an energy eigenstate so it will evolve in time. Its time evolution is given by

$$\begin{aligned} |\psi(t)\rangle &= e^{-i\hat{H}t/\hbar}|\psi_{in}\rangle = e^{-i\hat{H}t/\hbar}\left|K^0\right\rangle \\ &= \frac{1}{\sqrt{2}}(e^{-i\hat{H}t/\hbar}|K_S\rangle + e^{-i\hat{H}t/\hbar}|K_L\rangle) \\ &= \frac{1}{\sqrt{2}}(e^{-i(M+A)t/\hbar}|K_S\rangle + e^{-i(M-A)t/\hbar}|K_L\rangle) \end{aligned} \tag{5.199}$$

The probability amplitude that the initial meson *changes(oscillates)* into the orthogonal state $\left|\bar{K}^0\right\rangle$ at time t is given by

$$\begin{aligned} \left\langle \bar{K}^0 \middle| \psi(t) \right\rangle &= \frac{1}{\sqrt{2}}(e^{-i(M+A)t/\hbar}\left\langle \bar{K}^0 \middle| K_S \right\rangle + e^{-i(M-A)t/\hbar}\left\langle \bar{K}^0 \middle| K_L \right\rangle) \\ &= \frac{1}{2}(e^{-i(M+A)t/\hbar} - e^{-i(M-A)t/\hbar}) \end{aligned} \tag{5.200}$$

Finally, the probability that the incoming K^0-meson will behave like a $\bar{K}^0$-meson at time t (that it has oscillated into a $\bar{K}^0$) is given by

$$P_{\bar{K}^0}(t) = |\left\langle \bar{K}^0 \middle| \psi(t) \right\rangle|^2 = \frac{1}{2}[1 - \cos\Omega t] \tag{5.201}$$

where

$$\Omega = \frac{1}{\hbar}((M+A) - (M-A)) = \frac{2A}{\hbar} = \frac{m_S c^2 - m_L c^2}{\hbar} \tag{5.202}$$

What is the physics here? If $m_S - m_L = 0$, then $\Omega = 0$ and $P_{\bar{K}^0}(t) = 0$ or the two mesons *do not change* into one another(called *oscillation*) as time passes.

However, if If $m_S - m_L \neq 0$, then $P_{\bar{K}^0}(t) \neq 0$ and the two mesons oscillate back and forth, sometimes being a K^0 and sometimes being a $\bar{K}^0$. This has been observed in the laboratory and is, in fact, the way that the extremely small mass difference is actually measured.

This is also the same mechanism that operates for oscillations between the different flavors of neutrinos. Experiment now indicate that neutrino oscillation is taking place and the explanation is that the neutrinos have different masses, some non-zero!

5.4 Stern-Gerlach Experiments and Measurement

From early in our exploration of quantum mechanics, measurement played a central role. A basic *axiom* of quantum theory is the the *von Neumann projection postulate*:

If an observable $\hat{A}$ is measured, the result is one of

its eigenvalues, a. After the measurement, the system is *projected* into the eigenvector $|a\rangle$. If $|\psi\rangle$ is the state before the measurement, the probability of this occurence is $|\langle a|\psi\rangle|^2$.

Let us now explore the physical consequences of these measurements. The most fundamental example is the measurement of the angular momentum component of a spin-1/2 particle which takes on only two possible values. We will discuss the full theory of spin in later.

The measurement was first carried out by Stern and Gerlach in 1922 to test Bohr's ideas about "space quantization". This was before Uhlenbeck and Goudsmidt's invention of spin angular momentum, so Stern and Gerlach's results were not completely understood. Nonetheless, the results were startling and one of the first real pictures of the strange quantum world.

Consider the force on a magnetic moment which moves through a spatially *inhomogeneous* magnetic field. The potential energy is

$$V = \vec{\mu} \cdot \vec{B}(\vec{r}) \tag{5.203}$$

and thus the force is

$$\vec{F} = -\nabla V = \nabla(\vec{\mu} \cdot \vec{B}(\vec{r})) = (\vec{\mu} \cdot \nabla)\vec{B}(\vec{r}) \tag{5.204}$$

Stern and Gerlach set up a spatially inhomogeneous field with one very large component(call it the z-component).

A schematic diagram of a Stern-Gerlach apparatus is shown in Figure 5.4 below.

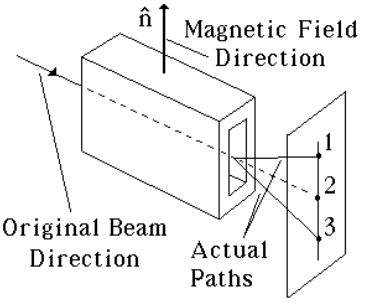


Figure 5.4: Stern-Gerlach Apparatus

We then have

$$\hat{n} = \hat{z} \quad , \quad \vec{B} \propto B(z)\hat{z} + \text{ small x,y components}$$
$$\Rightarrow (\vec{\mu} \cdot \nabla)\vec{B}(\vec{r}) \propto \mu_z \frac{\partial B}{\partial z}\hat{z} \tag{5.205}$$

Now quantum mechanically, $\vec{\mu} = \gamma_J \vec{J}$ which gives

$$\vec{F} = \left(\gamma_J \frac{\partial B}{\partial z} J_z \hat{Z}\right) \tag{5.206}$$

Stern and Gerlach sent Na (sodium) atoms from an oven, collimated into a beam, into the region of ingomogeneous field. Na is an alkali atom with one valence electron with orbital angular momentum $\ell = 0$ (an *s-state*). Thus, all of the angular momentum of the atom is due to the spin angular momentum of the single valence electron (spin 1/2). Thus

$$\vec{\mu} = -\frac{2\mu_B}{\hbar}\vec{S} = -\mu_B \vec{\sigma} \Rightarrow \vec{F} = -\mu_B \frac{\partial B}{\partial z}\sigma_z \hat{z} \tag{5.207}$$

Clearly, the *spin-up* atoms $|\uparrow_x\rangle$ will experience a different interaction than *spin-down* atoms $|\downarrow_x\rangle$ (due to two different eigenvalues of σ_z), thereby splitting the beam into two *spatially separated beams* as shown in Figure 5.4 above. We note that the spins which emerge from the oven have a random orientation.

In general, a Stern-Gerlach apparatus takes an incoming beam of particles with angular momentum and splits the beam into a number of spots on a screen. The number of spots is equal to $2J + 1$, where J = angular momentum value of the incoming beam.

We shall see in later chapters that $2J+1$ = the number of values allowed quantum mechanically for the measurable quantity $\mathbf{J} \cdot \hat{\mathbf{n}}$ where $\hat{\mathbf{n}}$ is unit vector in the direction of the magnetic field inside the apparatus.

In the diagram, spot #2 (undeflected beam) is where the beam would have hit the screen if no magnetic field were present in the apparatus. Spots #1 and #3 are an example the $2J + 1 = 2$ spots we would observe in an experiment when $J = 1/2$ (the original Stern-Gerlach experiment).

The important features of the apparatus for our theoretical discussion are:

1. The breakup into a finite number of discrete beams (we will assume we are working with $J = 1/2$ particles and thus have 2 beams exiting the apparatus)

2. The beams are *separated in 3-dimensional space* and each contains 1/2 of the original particles entering the device

3. The possible values of $\mathbf{B} \cdot \hat{\mathbf{n}}$ for $J = 1/2$ are $\pm\hbar/2$ for any $\hat{\mathbf{n}}$.

4. One exiting beam contains *only* particles with $\mathbf{J}\cdot\hat{\mathbf{n}} = +\hbar/2$ and the other beam contains *only* particles with $\mathbf{J}\cdot\hat{\mathbf{n}} = -\hbar/2$

   We will represent the beam state vectors by the ket vectors

$$\begin{aligned} |+\hat{\mathbf{n}}\rangle &= \text{ beam with } \mathbf{J}\cdot\hat{\mathbf{n}} = +\hbar/2 \\ |-\hat{\mathbf{n}}\rangle &= \text{ beam with } \mathbf{J}\cdot\hat{\mathbf{n}} = -\hbar/2 \end{aligned} \tag{5.208}$$

   and the Stern-Gerlach apparatus with a field in the $\hat{\mathbf{n}}$-direction by SG$\hat{\mathbf{n}}$.

5. The above results occur for all beams no matter what the direction of the unit vector $\hat{\mathbf{n}}$

The S-G experiment has all the essential ingredients of a quantum measurement. The quantum *degree of freedom*, here spin-1/2, is correlated with the final beam direction of the atoms. Once the two spots on the screen are *resolvable*, we can measure the spin state of the atom by determining which spatially separated beam the atom is in.

We now *report* the *results of a series of actual experiments.*

Experiment #1

We send N particles into an SG$\hat{\mathbf{z}}$ device and select out the beam where the particles are in the state $|+\hat{\mathbf{z}}\rangle$ (we block the other beam). It contains $N/2$ particles. We then send this second beam into another SG$\hat{\mathbf{z}}$ device. We find that all $N/2$ exit in the state $|+\hat{\mathbf{z}}\rangle$. There is only one exit beam. Symbolically, this looks like Figure 5.5 below:

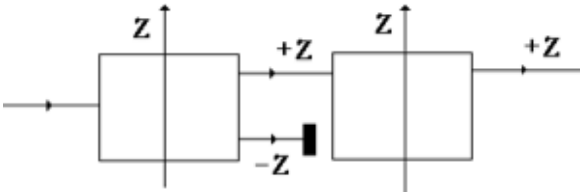


Figure 5.5: Experiment #1

This says that when we make a measurement, say $\mathbf{J}\cdot\hat{\mathbf{z}}$ and then immediately make another measurement of the same quantity, we get the same result as first measurement with probability = 1. Since the only way to a measurement result with certainty is to be in an eigenstate of the observable, the measurement seems to have caused the system to change from a superposition of eigenstates into two beams(separated in physical space) each with a definite state of the observable $\mathbf{J}\cdot\hat{\mathbf{z}}$ (one of its eigenstates).

This experiment is called *state preparation.* We prepare a state by measuring an

observable and finding one of the eigenvalues. Afterwards, by the *von Neumann projection postulate*, the state is the corresponding eigenvector. In the above experiment, by redirecting the *spin-up* beam into the second SG apparatus we are guaranteed to find the eigenstate prepared by the first SG apparatus (with the block). Thus a screen would have only one spot after the second apparatus.

Experiment 2

We send N particles into an SG$\hat{\mathbf{z}}$ device and select out the beam where the particles are in the state $|+\hat{\mathbf{z}}\rangle$. It contains $N/2$ particles. We then send the selected beam into an SG$\hat{\mathbf{x}}$ device We find that $N/4$ exit in the state $|+\hat{\mathbf{x}}\rangle$ and N/4 exit in the state $|-\hat{\mathbf{x}}\rangle$. At the end, there are two exit beams. Symbolically, this looks like Figure 5.6 below:

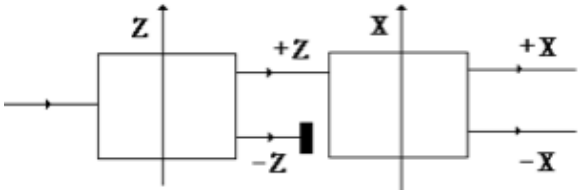


Figure 5.6: Experiment #2

The same thing happens if we stop the $|+\hat{\mathbf{z}}\rangle$ beam and let the $|-\hat{\mathbf{z}}\rangle$ beam into the SG$\hat{\mathbf{x}}$ device. So an SG$\hat{\mathbf{x}}$ device takes a beam with a definite value of $\mathbf{J}\cdot\hat{\mathbf{z}}$ and *randomizes* it, i.e., we once again have two exiting beams with equal numbers of particles. This is saying that for a system in an eigenstate of one observable, the measurement of an *incompatible* observable (observables do not commute) randomizes the value of the original observable. In this case $[\hat{J}_z, \hat{J}_x] \neq 0$.

As we will show in Chapter 7, the $|+\hat{\mathbf{z}}\rangle$ state is not an eigenstate of $\hat{J}_x$. In fact,

$$|+\hat{\mathbf{z}}\rangle = \frac{1}{\sqrt{2}}(|+\hat{\mathbf{x}}\rangle + |-\hat{\mathbf{x}}\rangle) \tag{5.209}$$

i.e., a 50-50 superposition of *spin-up* and *spin-down* in the x-direction. Thus, we see two beams after the SG$\hat{\mathbf{x}}$ apparatus. So even though a *pure* state is entering the SG$\hat{\mathbf{x}}$ apparatus, we do not get a definite value because $[\hat{J}_z, \hat{J}_x] \neq 0]$.

Experiment 3

We now add a third SG device to Experiment 2. It is an $|+\hat{\mathbf{z}}\rangle$ device. We also block the $|+\hat{\mathbf{x}}\rangle$ exiting beam as shown symbolically in Figure 5.7 below.

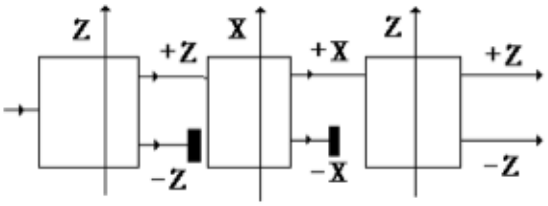


Figure 5.7: Experiment #3

We found above that $N/4$ exited in the state $|+\hat{\mathbf{x}}\rangle$ from the SG$\hat{\mathbf{x}}$ device. After the third device we find that $N/8$ exit in the state $|+\hat{\mathbf{z}}\rangle$ and N$/8$ exit in the state $|-\hat{\mathbf{z}}\rangle$.

What has happened? It seems that making a measurement of $\mathbf{J}\cdot\hat{\mathbf{x}}$ on a beam with *definite* $\mathbf{J}\cdot\hat{\mathbf{z}}$ modifies the system rather dramatically.

We did two successive measurements on these particles. Since we isolated the + beam in each case we might be led to think that the beam entering the last SG$\hat{\mathbf{z}}$ device (because of our selections) has

$$\mathbf{J}\cdot\hat{\mathbf{z}} = +\frac{\hbar}{2} \text{ AND } \mathbf{J}\cdot\hat{\mathbf{x}} = +\frac{\hbar}{2} \tag{5.210}$$

But the experiment says this cannot be so, since 50% of the particles exiting the last device have

$$\mathbf{J}\cdot\hat{\mathbf{z}} = -\frac{\hbar}{2} \tag{5.211}$$

We are forced to say that the SG$\hat{\mathbf{z}}$ device takes a definite value of $\mathbf{J}\cdot\hat{\mathbf{x}}$ and *randomizes* it so that we end up with two exiting beams with equal numbers of particles.

Why? Again, since $[\hat{J}_z, \hat{J}_x] \neq 0$ the two observables cannot share a common set of eigenvectors or they are *incompatible*. Since an observable can *only* have a definite value when the state is one of its eigenvectors, it is not possible for both $\mathbf{J}\cdot\hat{\mathbf{z}}$ and $\mathbf{J}\cdot\hat{\mathbf{x}}$ to simultaneously have definite values.

Our two successive measurements *DO NOT* produce definite values for both observables. Each measurement *only* produces a definite value for the observable it is measuring and *randomizes* the incompatible observable (actually randomizes all other incompatible observables)!

All measurement results depend on the *context* of the measurement.

Another way to think about this is to say the following. An SG$\hat{\mathbf{z}}$ device is a measurement of the angular momentum in the $\hat{\mathbf{z}}$-direction. Any such measurement

of an observable randomizes the next measurement of any other incompatible observable.

In this cascaded measurement, the probability of finding $|\downarrow\rangle$ is the product of conditional probabilities for uncorrelated events

$$P(\downarrow_z \mid \uparrow_x, \uparrow_z) = P(\downarrow_z \mid \uparrow_x)P(\downarrow_z \mid \uparrow_z) = |\langle\downarrow_z \mid \uparrow_x\rangle|^2 |\langle\uparrow_x \mid \uparrow_z\rangle|^2 = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

as is observed.

Now without the intermediate apparatus, *classical theory* would argue as follows. If we don't measure J_x we must sum the probabilities of all possible alternatives or

$$P(\downarrow_z \mid \uparrow_z) = P(\downarrow_z \mid \uparrow_x)P(\uparrow_x \mid \uparrow_z) + P(\downarrow_z \mid \downarrow_x)P(\downarrow_x \mid \uparrow_z) = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$$

which differs from the experimental result(= 0)

As we have discussed, quantum mechanically, if the different possibilities are *indistinguishable* (i.e., there is no information available to distinguish the different alternatives) we must add the probability *amplitudes* (before squaring) so that these different alternatives can interfere.

$$\begin{aligned} P(\downarrow_z \mid \uparrow_z) &= |\langle\downarrow_z \mid \uparrow_z\rangle|^2 = |\langle\downarrow_z| \hat{I} |\uparrow_z\rangle|^2 \\ &= |\langle\downarrow_z| (|\uparrow_x\rangle\langle\uparrow_x| + |\downarrow_x\rangle\langle\downarrow_x|) |\uparrow_z\rangle|^2 \\ &= \underbrace{P(\downarrow_z \mid \uparrow_x)P(\uparrow_x \mid \uparrow_z) + P(\downarrow_z \mid \downarrow_x)P(\downarrow_x \mid \uparrow_z)}_{classical\ terms} + \text{interference terms} \\ &= \frac{1}{2} - \frac{1}{2} = 0 \end{aligned}$$

So, the quantum amplitude interference calculation gives the correct result.

What constitutes a measurement?

If we send atoms with spin-up in the z-direction into an x-oriented SG apparatus, there is a $50-50$ probability of emerging as spin-up or spin-down in the x-direction.

But if we do not detect which port the atom exits from, did we measure the spin in the x-direction?

In a sense, *no*. A measurement is something that *removes coherence*, i.e., the ability for quantum processes to interfere.

In principle, we can recombine these beams as if no measurement had occurred,

i.e., we can still have interference - we do not remove coherence unless we *look*. Again the context of the experiment is the significant feature!

Experiment 4

Now, let us construct a new device to illustrate this point. It is called a modified SG$\hat{\mathbf{x}}$ device. It looks like Figure 5.8 below.

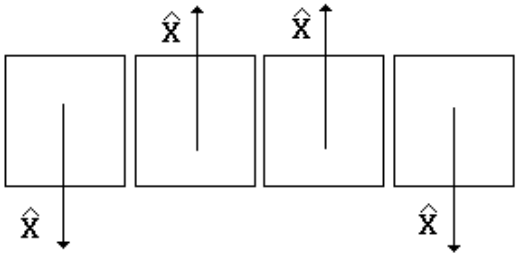


Figure 5.8: Experiment #4

Any beam of particles entering this modified device would experience deflections while traveling though it. However, the device lengths and field strengths have been cleverly chosen so that the *net effect for any beam is no change!* The internal beams are all recombined so that the state of the beam upon exiting the entire device is *identical* to the state of the beam before entering the device. This device might be called a *total-of-nothing* device.

It turns out, however, that we can use this device to make a measurement and select a particular spin state. We can calculate the paths that would be followed in the device by the $|+\hat{\mathbf{x}}\rangle$ and $|-\hat{\mathbf{x}}\rangle$ (these are the relevant beams to consider because these are SG$\hat{\mathbf{x}}$ devices). Using this information, we can block the path that a particle in the state $|-\hat{\mathbf{x}}\rangle$ would follow. Then, all the particles exiting the modified SG$\hat{\mathbf{x}}$ device would be in the state $|+\hat{\mathbf{x}}\rangle$.

Experiment 5

We can confirm this fact by inserting a modified SG$\hat{\mathbf{x}}$ device into Experiment 3 where we replace the SG$\hat{\mathbf{x}}$ device by the modified SG$\hat{\mathbf{x}}$ device. If we block the $|-\hat{\mathbf{x}}\rangle$ beam, the state at the end of the modified SG$\hat{\mathbf{x}}$ device is the same as it was after the original SG$\hat{\mathbf{x}}$ device and we get exactly the same results. In fact, we get the same result whether we block the $|-\hat{\mathbf{x}}\rangle$ beam or the $|+\hat{\mathbf{x}}\rangle$ beam, as should be the case.

Now we set up Experiment 5. It is shown in Figure 5.9 below.

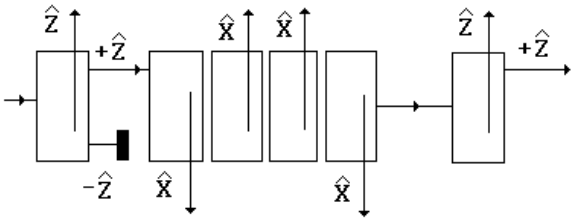


Figure 5.9: Experiment #5

In this experiment a beam enters the SG$\hat{\mathbf{z}}$ device and we block the exiting $|-\hat{\mathbf{z}}\rangle$ beam. The $|+\hat{\mathbf{z}}\rangle$ beam is sent into the modified SG$\hat{\mathbf{x}}$ device and the exit beam is sent into the final SG$\hat{\mathbf{z}}$ device. In this case, however, we *DO NOT* block any of the paths in the modified SG$\hat{\mathbf{x}}$ device. Since the beam entering the modified SG$\hat{\mathbf{x}}$ device is reconstructed before it exits (we already saw that it does this in the Experiment 4), we are *NOT* making a measurement of $\mathbf{J} \cdot \hat{\mathbf{x}}$ using the modified SG$\hat{\mathbf{x}}$ device as we did when we used the original SG$\hat{\mathbf{x}}$ device.

Now we send in N particles and $N/2$ are in the $|+\hat{\mathbf{z}}\rangle$ beam as before. However, now, instead of find $N/8$ particles in the final $|+\hat{\mathbf{z}}\rangle$ beam, we find $N/2$ particles. *ALL* the particles make it through unchanged, even though there are SG$\hat{\mathbf{x}}$ devices in between. It behaves as if the modified SG$\hat{\mathbf{x}}$ device was not there at all(hence its name).

We might have assumed that 50% of the particles in the $|+\hat{\mathbf{z}}\rangle$ beam before the modified SG$\hat{\mathbf{x}}$ device would emerge in the $|+\hat{\mathbf{x}}\rangle$ state and the other 50% would emerge in the $|-\hat{\mathbf{x}}\rangle$ state. Experiment 5 says this cannot be true, since if it were true, then we would have two beams coming out of the final SG$\hat{\mathbf{z}}$ device, each with 50% of the particles. Our results are *incompatible* with the statement that the particles passing through the modified SG$\hat{\mathbf{x}}$ device are *either* in the state $|+\hat{\mathbf{x}}\rangle$ *or* in the state $|-\hat{\mathbf{x}}\rangle$.

In fact, if we carry out this experiment with a very low intensity beam where only one particle at a time is passing through the apparatus, then we observe that *each* particle emerging from the final SG$\hat{\mathbf{z}}$ device is in the state $|+\hat{\mathbf{z}}\rangle$. This eliminates any explanation of the result that would invoke interactions among the particles while they were in the apparatus.

For beams of particles, we have been talking in terms of percentages or fractions of the particles as experimental results. For a single particle, however, it is not possible to predict with certainty the outcome of a measurement in advance. We have seen this earlier in our experiments with photons and polaroids. We are only able to use probability arguments in this case.

In Experiment 2, for instance, before a measurement (passing through the SG$\hat{\mathbf{x}}$ device) of $\mathbf{J}\cdot\hat{\mathbf{x}}$ on a single particle in the $|+\hat{\mathbf{z}}\rangle$ state, all we can say is that there is a 50% probability of obtaining $|+\hat{\mathbf{x}}\rangle$ and a 50% probability of obtaining $|-\hat{\mathbf{x}}\rangle$.

Probabilities alone are not enough, however, to explain Experiment 5. We came to the incorrect conclusion because we made the same mistake as in our earlier discussion of polarization. We added the separate probabilities of the indistinguishable ways to get the total probability, whereas the correct result, as we know, is to add the amplitudes of the indistinguishable ways and then square the total amplitude to get the correct result. We eliminated the interference effects! When we don't actually make a measurement as in the modified SG$\hat{\mathbf{x}}$ device, we must add amplitudes and not probabilities.

We can now introduce a formalism, which should allow us to explain all experiments correctly. We need a 2-dimensional vector space to describe these physical systems. As a basis for this space we can use any of the sets $\{|+\hat{\mathbf{n}}\rangle, |-\hat{\mathbf{n}}\rangle\}$ corresponding to the definite values $\pm\hbar/2$ for $\mathbf{J}\cdot\hat{\mathbf{n}}$. Each of these is an orthonormal basis where

$$\langle+\hat{\mathbf{n}}\,|+\hat{\mathbf{n}}\rangle = 1 = \langle-\hat{\mathbf{n}}\,|-\hat{\mathbf{n}}\rangle \text{ and } \langle+\hat{\mathbf{n}}\,|-\hat{\mathbf{n}}\rangle = 0 \tag{5.212}$$

Any arbitrary state can be written as a superposition of the basis states

$$|\psi\rangle = \langle+\hat{\mathbf{n}}\,|\,\psi\rangle\,|+\hat{\mathbf{n}}\rangle + \langle-\hat{\mathbf{n}}\,|\,\psi\rangle\,|-\hat{\mathbf{n}}\rangle \tag{5.213}$$

and the operators can be written as

$$\mathbf{J}\cdot\hat{\mathbf{n}} = \frac{\hbar}{2}\,|+\hat{\mathbf{n}}\rangle\langle+\hat{\mathbf{n}}| - \frac{\hbar}{2}\,|-\hat{\mathbf{n}}\rangle\langle-\hat{\mathbf{n}}| \tag{5.214}$$

Finally, expectation values are given by

$$\langle\hat{J}_z\rangle = \langle\psi|\,\hat{J}_z\,|\psi\rangle = \frac{\hbar}{2}|\langle+\hat{\mathbf{z}}\,|\,\psi\rangle|^2 - \frac{\hbar}{2}|\langle-\hat{\mathbf{z}}\,|\,\psi\rangle|^2 \tag{5.215}$$

Analysis of Experiment 3

In Experiment 3, the state before entering the first SG$\hat{\mathbf{z}}$ device is

$$|\psi_1\rangle = a\,|+\hat{\mathbf{z}}\rangle + b\,|-\hat{\mathbf{z}}\rangle \text{ where } |a|^2 + |b|^2 = 1 \tag{5.216}$$

Since the SG$\hat{\mathbf{z}}$ device is a measurement of $\hat{J}_z$, after the SG$\hat{\mathbf{z}}$ device the state is $|+\hat{\mathbf{z}}\rangle$ (remember the we blocked the $|-\hat{\mathbf{z}}\rangle$ path).

It is very *important* to realize that we cannot answer a question about a measurement unless we express the state in the basis consisting of the eigenvectors of the operator representing the observable being measured. We call this the *home space* of the observable. That is why we used the $\hat{J}_z$ eigenvectors to discuss the measurement made using the SG$\hat{\mathbf{z}}$ device.

Now, we are going to make a measurement of $\hat{J}_x$ in the SG$\hat{\mathbf{x}}$ device. So we now *switch* to a basis consisting of the $\hat{J}_x$ eigenvectors. We know we can write

$$|+\hat{\mathbf{x}}\rangle = \langle +\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle\,|+\hat{\mathbf{z}}\rangle + \langle -\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle\,|-\hat{\mathbf{z}}\rangle \tag{5.217}$$

One of our experiments tells us that when we send a particle in a $|+\hat{\mathbf{x}}\rangle$ state through an SG$\hat{\mathbf{z}}$ device, the probability = 1/2 that we find $|+\hat{\mathbf{z}}\rangle$ and 1/2 that we find $|-\hat{\mathbf{z}}\rangle$. This means that

$$|\langle +\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle|^2 = |\langle -\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle|^2 = \frac{1}{2} \tag{5.218}$$

or

$$\langle +\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}} \text{ and } \langle -\hat{\mathbf{z}}|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}} \tag{5.219}$$

so that

$$|+\hat{\mathbf{x}}\rangle = \frac{e^{1\alpha_+}}{\sqrt{2}}\,|+\hat{\mathbf{z}}\rangle + \frac{e^{1\alpha_+}}{\sqrt{2}}\,|-\hat{\mathbf{z}}\rangle \tag{5.220}$$

Similarly,

$$|-\hat{\mathbf{x}}\rangle = \frac{e^{1\beta_+}}{\sqrt{2}}\,|+\hat{\mathbf{z}}\rangle - \frac{e^{1\beta_+}}{\sqrt{2}}\,|-\hat{\mathbf{z}}\rangle \tag{5.221}$$

Experiment 6

In this experiment we replace the last SG$\hat{\mathbf{z}}$ device in Experiment 3 with an SG$\hat{\mathbf{y}}$ device. The last part of the setup is shown in Figure 5.10 below.

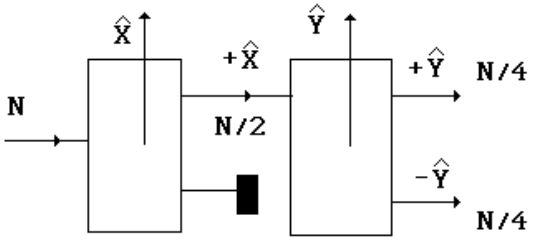


Figure 5.10: Experiment #6

The incoming beam is from the first SG$\hat{\mathbf{z}}$ device, where we blocked the $|-\hat{\mathbf{z}}\rangle$ path, and therefore, we have N particles in the state $|+\hat{\mathbf{z}}\rangle$. They now go through the SG$\hat{\mathbf{x}}$ device, where we block the $|-\hat{\mathbf{x}}\rangle$ path, as shown. The beam entering the SG$\hat{\mathbf{y}}$ device has $N/2$ particles in the $|+\hat{\mathbf{x}}\rangle$ state.

Whether we call the last direction the z-direction or the y-direction cannot affect the results of the experiment, so we get the same result for this experiment as we did for Experiment 3, except that we use the y-label instead of the z-label.

If the SG$\hat{\mathbf{x}}$ device were replaced by an SG$\hat{\mathbf{z}}$ device we would get the same result also, since whether we call the direction the z-direction or the x-direction cannot affect the results of the experiment.

Putting this all together we can say that

$$|+\hat{\mathbf{y}}\rangle = \langle+\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|+\hat{\mathbf{z}}\rangle + \langle-\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|-\hat{\mathbf{z}}\rangle \tag{5.222}$$

with

$$|\langle+\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|^2 = |\langle-\hat{\mathbf{z}}|+\hat{\mathbf{y}}\rangle|^2 = \frac{1}{2} \tag{5.223}$$

and

$$|+\hat{\mathbf{y}}\rangle = \langle+\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|+\hat{\mathbf{x}}\rangle + \langle-\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|-\hat{\mathbf{x}}\rangle \tag{5.224}$$

with

$$|\langle+\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|^2 = |\langle-\hat{\mathbf{x}}|+\hat{\mathbf{y}}\rangle|^2 = \frac{1}{2} \tag{5.225}$$

The conventional choice for the phases factors is such that we have

$$|+\hat{\mathbf{x}}\rangle = \frac{1}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{1}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle \tag{5.226}$$

$$|+\hat{\mathbf{y}}\rangle = \frac{1}{\sqrt{2}}|+\hat{\mathbf{z}}\rangle + \frac{i}{\sqrt{2}}|-\hat{\mathbf{z}}\rangle \tag{5.227}$$

We will derive all of these results more rigorously in Chapter 7.

5.5 Problems

5.5.1 Change the Basis

In examining light polarization in the text, we have been working in the $\{|x\rangle, |y\rangle\}$ basis.

(a) Just to show how easy it is to work in other bases, express $\{|x\rangle, |y\rangle\}$ in the $\{|R\rangle, |L\rangle\}$ and $\{|45°\rangle, |135°\rangle\}$ bases.

(b) If you are working in the $\{|R\rangle, |L\rangle\}$ basis, what would the operator representing a vertical polaroid look like?

5.5.2 Polaroids

Imagine a situation in which a photon in the $|x\rangle$ state strikes a vertically oriented polaroid. Clearly the probability of the photon getting through the vertically oriented polaroid is 0. Now consider the case of two polaroids with the photon in the $|x\rangle$ state striking a polaroid oriented at 45° and then striking a vertically oriented polaroid.

Show that the probability of the photon getting through both polaroids is 1/4.

Consider now the case of three polaroids with the photon in the $|x\rangle$ state striking a polaroid oriented at 30° first, then a polaroid oriented at 60° and finally a vertically oriented polaroid.

Show that the probability of the photon getting through all three polaroids is 27/64.

5.5.3 Calcite Crystal

A photon polarized at an angle θ to the optic axis is sent through a slab of calcite crystal. Assume that the slab is $10^{-2}\,cm$ thick, the direction of photon propagation is the $z-$axis and the optic axis lies in the $x-y$ plane.

Calculate, as a function of θ, he transition probability for the photon to emerge left circularly polarized. Sketch the result. Let the frequency of the light be given by $c/\omega = 5000\,\mathring{\mathrm{A}}$, and let $n_e = 1.50$ and $n_o = 1.65$ for the calcite indices of refraction.

5.5.4 Turpentine

Turpentine is an *optically active* substance. If we send plane polarized light into turpentine then it emerges with its plane of polarization rotated. Specifically, turpentine induces a left-hand rotation of about 5° per cm of turpentine that the light traverses. Write down the transition matrix that relates the incident polarization state to the emergent polarization state. Show that this matrix is unitary. Why is that important? Find its eigenvectors and eigenvalues, as a function of the length of turpentine traversed.

5.5.5 What QM is all about - Two Views

Photons polarized at 30° to the $x-$axis are sent through a $y-$polaroid. An attempt is made to determine how frequently the photons that pass through the polaroid, pass through as *right circularly polarized photons* and how frequently they pass through as *left circularly polarized photons*. This attempt is made as follows:

First, a prism that passes only right circularly polarized light is placed between the source of the 30° polarized photons and the $y-$polaroid, and it is determined how frequently the 30° polarized photons pass through the $y-$polaroid. Then this experiment is repeated with a prism that passes only left circularly polarized photons instead of the one that passes only right.

(a) Show by explicit calculation using standard amplitude mechanics that the sum of the probabilities for passing through the $y-$polaroid measured in these two experiments is different from the probability that one would measure if there were no prism in the path of the photon and only the

$y-$polaroid.

Relate this experiment to the two-slit diffraction experiment.

(b) Repeat the calculation using density matrix methods instead of amplitude mechanics.

5.5.6 Photons and Polarizers

A photon polarization state for a photon propagating in the $z-$direction is given by

$$|\psi\rangle = \sqrt{\frac{2}{3}}\,|x\rangle + \frac{i}{\sqrt{3}}\,|y\rangle$$

(a) What is the probability that a photon in this state will pass through a polaroid with its transmission axis oriented in the $y-$direction?

(b) What is the probability that a photon in this state will pass through a polaroid with its transmission axis y' making an angle φ with the $y-$axis?

(c) A beam carrying N photons per second, each in the state $|\psi\rangle$, is totally absorbed by a black disk with its surface normal in the z-direction. How large is the torque exerted on the disk? In which direction does the disk rotate? **REMINDER**: The photon states $|R\rangle$ and $|L\rangle$ each carry a unit $\hbar$ of angular momentum parallel and antiparallel, respectively, to the direction of propagation of the photon.

5.5.7 Time Evolution

The matrix representation of the Hamiltonian for a photon propagating along the optic axis (taken to be the $z-$axis) of a quartz crystal using the linear polarization states $|x\rangle$ and $|y\rangle$ as a basis is given by

$$\hat{H} = \begin{pmatrix} 0 & -iE_0 \\ iE_0 & 0 \end{pmatrix}$$

(a) What are the eigenstates and eigenvalues of the Hamiltonian?

(b) A photon enters the crystal linearly polarized in the x direction, that is, $|\psi(0)\rangle = |x\rangle$. What is $|\psi(t)\rangle$, the state of the photon at time t? Express your answer in the $\{|x\rangle, |y\rangle\}$ basis.

(c) What is happening to the polarization of the photon as it travels through the crystal?

5.5.8 K-Meson oscillations

An additional effect to worry about when thinking about the time development of K-meson states is that the $|K_L\rangle$ and $|K_S\rangle$ states decay with time. Thus, we expect that these states should have the time dependence

$$|K_L(t)\rangle = e^{-i\omega_L t - t/2\tau_L}\,|K_L\rangle \quad , \quad |K_S(t)\rangle = e^{-i\omega_S t - t/2\tau_S}\,|K_S\rangle$$

where

$$\omega_L = E_L/\hbar \quad , \quad E_L = \left(p^2c^2 + m_L^2c^4\right)^{1/2}$$
$$\omega_S = E_S/\hbar \quad , \quad E_S = \left(p^2c^2 + m_S^2c^4\right)^{1/2}$$

and

$$\tau_S \approx 0.9 \times 10^{-10}\,\text{sec} \quad , \quad \tau_L \approx 560 \times 10^{-10}\,\text{sec}$$

Suppose that a pure K_L beam is sent through a thin absorber whose only effect is to change the relative phase of the K_0 and $\bar{K}_0$ amplitudes by 10°. Calculate the number of K_S decays, relative to the incident number of particles, that will be observed in the first 5 cm after the absorber. Assume the particles have momentum = mc.

5.5.9 What comes out?

A beam of spin 1/2 particles is sent through series of three Stern-Gerlach measuring devices as shown in Figure 7.1 below: The first SGz device transmits

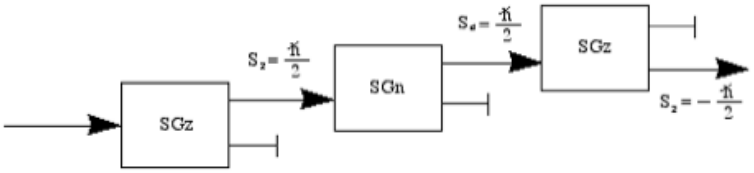


Figure 5.11: Stern-Gerlach Setup

particles with $\hat{S}_z = \hbar/2$ and filters out particles with $\hat{S}_z = -\hbar/2$. The second device, an SGn device transmits particles with $\hat{S}_n = \hbar/2$ and filters out particles with $\hat{S}_n = -\hbar/2$, where the axis $\hat{n}$ makes an angle θ in the $x-z$ plane with respect to the z–axis. Thus the particles passing through this SGn device are in the state

$$|+\hat{n}\rangle = \cos\frac{\theta}{2}\,|+\hat{z}\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|-\hat{z}\rangle$$

with the angle $\varphi = 0$. A last SGz device transmits particles with $\hat{S}_z = -\hbar/2$ and filters out particles with $\hat{S}_z = +\hbar/2$.

(a) What fraction of the particles transmitted through the first SGz device will survive the third measurement?

(b) How must the angle θ of the SGn device be oriented so as to maximize the number of particles the at are transmitted by the final SGz device? What fraction of the particles survive the third measurement for this value of θ?

(c) What fraction of the particles survive the last measurement if the SGz device is simply removed from the experiment?

5.5.10 Orientations

The kets $|h\rangle$ and $|v\rangle$ are states of horizontal and vertical polarization, respectively. Consider the states

$$|\psi_1\rangle = -\frac{1}{2}\left(|h\rangle + \sqrt{3}\,|v\rangle\right) \quad , \quad |\psi_2\rangle = -\frac{1}{2}\left(|h\rangle - \sqrt{3}\,|v\rangle\right) \quad , \quad |\psi_2\rangle = |h\rangle$$

What are the relative orientations of the plane polarization for these three states?

5.5.11 Find the phase angle

If CP is not conserved in the decay of neutral K mesons, then the states of definite energy are no longer the K_L , K_S states, but are slightly different states $|K'_L\rangle$ and $|K'_S\rangle$. One can write, for example,

$$|K'_L\rangle = (1+\varepsilon)\left|K^0\right\rangle - (1-\varepsilon)\left|\bar{K}^0\right\rangle$$

where *varepsilon* is a very small complex number $\left(|\varepsilon| \approx 2\times 10^{-3}\right)$ that is a measure of the lack of CP conservation in the decays. The amplitude for a particle to be in $|K'_L\rangle$ (or $|K'_S\rangle$) varies as $e^{-i\omega_L t - t/2\tau_L}\left(or \quad e^{-i\omega_S t - t/2\tau_S}\right)$ where

$$\hbar\omega_L = \left(p^2c^2 + m_L^2c^4\right)^{1/2}\left(or \quad \hbar\omega_S = \left(p^2c^2 + m_S^2c^4\right)^{1/2}\right)$$

and $\tau_L \gg \tau_S$.

(a) Write out normalized expressions for the states $|K'_L\rangle$ and $|K'_S\rangle$ in terms of $|K_0\rangle$ and $\left|\bar{K}_0\right\rangle$.

(b) Calculate the ratio of the amplitude for a long-lived K to decay to two pions (a $CP = +1$ state) to the amplitude for a short-lived K to decay to two pions. What does a measurement of the ratio of these decay rates tell us about ε?

(c) Suppose that a beam of purely long-lived K mesons is sent through an absorber whose only effect is to change the relative phase of the K_0 and $\bar{K}_0$ components by δ. Derive an expression for the number of two pion events observed as a function of time of travel from the absorber. How well would such a measurement (given δ) enable one to determine the phase of ε?

5.5.12 Quarter-wave plate

A beam of linearly polarized light is incident on a quarter-wave plate (changes relative phase by 90°) with its direction of polarization oriented at 30° to the optic axis. Subsequently, the beam is absorbed by a black disk. Determine the rate angular momentum is transferred to the disk, assuming the beam carries N photons per second.

5.5.13 What is happening?

A system of N ideal linear polarizers is arranged in sequence. The transmission axis of the first polarizer makes an angle φ/N with the $y-$axis. The transmission axis of every other polarizer makes an angle φ/N with respect to the axis of the preceding polarizer. Thus, the transmission axis of the final polarizer makes an angle φ with the $y-$axis. A beam of $y-$polarized photons is incident on the first polarizer.

(a) What is the probability that an incident photon is transmitted by the array?

(b) Evaluate the probability of transmission in the limit of large N.

(c) Consider the special case with the angle 90°. Explain why your result is not in conflict with the fact that $\langle x \,|\, y \rangle = 0$.

5.5.14 Interference

Photons freely propagating through a vacuum have one value for their energy $E = h\nu$. This is therefore a 1−dimensional quantum mechanical system, and since the energy of a freely propagating photon does not change, it must be an eigenstate of the energy operator. So, if the state of the photon at $t = 0$ is denoted as $|\psi(0)\rangle$, then the eigenstate equation can be written $\hat{H}\,|\psi(0)\rangle = E\,|\psi(0)\rangle$. To see what happens to the state of the photon with time, we simply have to apply the time evolution operator

$$\begin{aligned} |\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle = e^{-i\hat{H}t/\hbar}\,|\psi(0)\rangle = e^{-ih\nu t/\hbar}\,|\psi(0)\rangle \\ = e^{-i2\pi\nu t}\,|\psi(0)\rangle = e^{-i2\pi x/\lambda}\,|\psi(0)\rangle \end{aligned}$$

where the last expression uses the fact that $\nu = c/\lambda$ and that the distance it travels is $x = ct$. Notice that the relative probability of finding the photon at various points along the x-axis (the absolute probability depends on the number of photons emerging per unit time) does not change since the modulus-square of the factor in front of $|\psi(0)\rangle$ is 1. Consider the following situation. Two sources of identical photons face each other an emit photons at the same time. Let the distance between the two sources be L.
Notice that we are assuming the photons emerge from each source in state $|\psi(0)\rangle$. In between the two light sources we can detect photons but we do

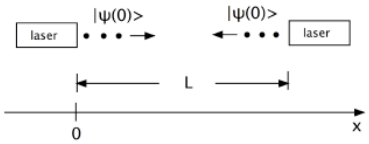


Figure 5.12: Interference Setup

not know from which source they originated. Therefore, we have to treat the photons at a point along the $x-$axis as a superposition of the time-evolved state from the left source and the time-evolved state from the right source.

(a) What is this superposition state $|\psi(t)\rangle$ at a point x between the sources? Assume the photons have wavelength λ.

(b) Find the relative probability of detecting a photon at point x by evaluating $|\langle\psi(t)\,|\,\psi(t)\rangle|^2$ at the point x.

(c) Describe in words what your result is telling you. Does this correspond to anything you have seen when light is described as a wave?

5.5.15 More Interference

Now let us tackle the two slit experiment with photons being shot at the slits one at a time. The situation looks something like the figure below. The distance between the slits, d is quite small (less than a mm) and the distance up the $y-$axis(screen) where the photons arrive is much,much less than L (the distance between the slits and the screen). In the figure, S_1 and S_2 are the lengths of the photon paths from the two slits to a point a distance y up the $y-$axis from the midpoint of the slits. The most important quantity is the difference in length between the two paths. The path length difference or PLD is shown in the figure.

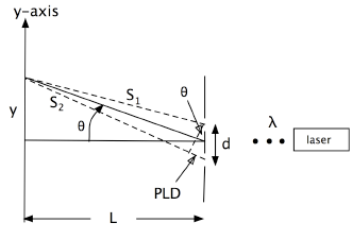


Figure 5.13: Double-Slit Interference Setup

We calculate PLD as follows:

$$PLD = d\sin\theta = d\left[\frac{y}{\left[L^2+y^2\right]^{1/2}}\right] \approx \frac{yd}{L} \quad , \quad y \ll L$$

Show that the relative probability of detecting a photon at various points along the screen is approximately equal to

$$4\cos^2\left(\frac{\pi y d}{\lambda L}\right)$$

5.5.16 The Mach-Zender Interferometer and Quantum Interference

Background information: Consider a single photon incident on a 50-50 beam splitter (that is, a partially transmitting, partially reflecting mirror, with equal coefficients). Whereas classical electromagnetic energy divides equally, the photon is indivisible. That is, if a photon-counting detector is placed at each of the output ports (see figure below), only one of them clicks. Which one clicks is completely random (that is, we have no better guess for one over the other).



Figure 5.14: Beam Splitter

The input-output transformation of the waves incident on 50-50 beam splitters and perfectly reflecting mirrors are shown in the figure below.

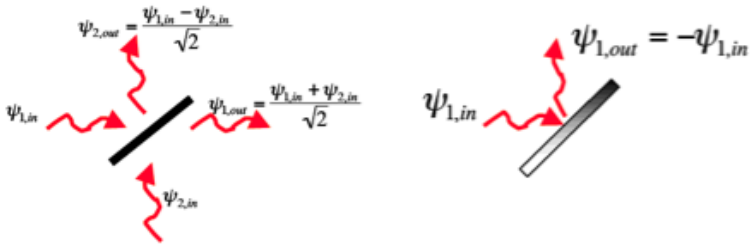


Figure 5.15: Input-Output transformation

(a) Show that with these rules, there is a 50-50 chance of either of the detectors shown in the first figure above to click.

(b) Now we set up a Mach-Zender interferometer(shown below):

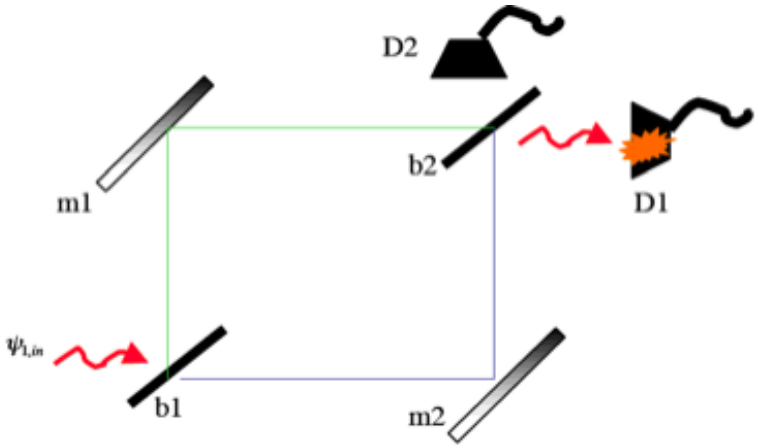


Figure 5.16: Input-Output transformation

The wave is split at beam-splitter b1, where it travels either path b1-m1-b2(call it the green path) or the path b1-m2-b2 (call it the blue path). Mirrors are then used to recombine the beams on a second beam splitter, b2. Detectors D1 and D2 are placed at the two output ports of b2.

Assuming the paths are perfectly balanced (that is equal length), show that the probability for detector D1 to click is 100% - *no randomness*!

(c) Classical logical reasoning would predict a probability for D1 to click given by

$$P_{D1} = P(transmission\ at\ b2|green\ path)P(green\ path) + P(reflection\ at\ b2|blue\ path)P(blue\ path)$$

Calculate this and compare to the quantum result. *Explain.*

(d) How would you set up the interferometer so that detector D2 clicked with 100% probability? How about making them click at random? Leave the *basic geometry the same*, that is, do not change the direction of the beam splitters or the direction of the incident light.

5.5.17 More Mach-Zender

An experimenter sets up two optical devices for single photons. The first, (i) in figure below, is a standard balanced Mach-Zender interferometer with equal path lengths, perfectly reflecting mirrors (M) and 50-50 beam splitters (BS).



Figure 5.17: Mach-Zender Setups

A transparent piece of glass which imparts a phase shift (PS) ϕ is placed in one arm. Photons are detected (D) at one port. The second interferometer, (ii) in figure below, is the same *except* that the final beam splitter is omitted.

Sketch the probability of detecting the photon as a function of ϕ for each device. Explain your answer.

Chapter 6

Schrodinger Wave equation
1-Dimensional Quantum Systems

6.1 The Coordinate Representation

To form a *representation* of an abstract linear vector space we must carry out these steps:

1. Choose a complete, orthonormal set of basis vectors $\{|\alpha_k\rangle\}$.
2. Construct the identity operator $\hat{I}$ as a sum over the one-dimensional subspace projection operators $|\alpha_k\rangle\langle\alpha_k|$
$$\hat{I} = \sum_k |\alpha_k\rangle\langle\alpha_k| \tag{6.1}$$
3. Write an arbitrary vector $|\psi\rangle$ as a linear combination or superposition of basis vectors using the identity operator
$$|\psi\rangle = \hat{I}|\psi\rangle = \left(\sum_k |\alpha_k\rangle\langle\alpha_k|\right)|\psi\rangle = \sum_k \langle\alpha_k|\psi\rangle|\alpha_k\rangle \tag{6.2}$$
It is clear from this last equation, that knowledge about the behavior(say in time) of the expansion coefficients $\langle\alpha_k|\psi\rangle$ will tell us the behavior of the state vector $|\psi\rangle$ and allow us to make predictions. Remember also, that the expansion coefficient is the probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\alpha_k\rangle$.

A particular representation that has become very important in the study of many systems using Quantum Mechanics is formed using the eigenstates of the position operator as a basis. It is called the *coordinate or position representation.*

The eigenstates $\{|\vec{x}\rangle\}$ of the position operator $(\hat{x},\hat{y},\hat{z}) = \hat{\mathbf{Q}}$ satisfy
$$\hat{x}|\vec{x}\rangle = x|\vec{x}\rangle \quad , \quad \hat{y}|\vec{x}\rangle = y|\vec{x}\rangle \quad , \quad \hat{z}|\vec{x}\rangle = z|\vec{x}\rangle \tag{6.3}$$

where the eigenvalues (x, y, z) are *continuous* variables in the range $[-\infty, \infty]$. They form the basis of the coordinate representation.

As we saw earlier, in this case, all summations above become integrals in the continuous spectrum case and we have

$$\hat{I} = \int |\vec{x}\rangle \langle \vec{x}| \, d\vec{x} \qquad , \qquad \text{where } d\vec{x} = dxdydz \tag{6.4}$$

$$|\psi\rangle = \hat{I}\,|\psi\rangle = \int (|\vec{x}\rangle \langle \vec{x}|)\,|\psi\rangle \, d\vec{x} = \int \langle \vec{x}\,|\,\psi\rangle\,|\vec{x}\rangle \, d\vec{x} \tag{6.5}$$

The expansion coefficient in the coordinate representation is given by

$$\psi(\vec{x}) = \langle \vec{x}\,|\,\psi\rangle \tag{6.6}$$

Since the inner product is defined for all states $|\vec{x}\rangle$, this new object is clearly a *function* of the eigenvalues (x, y, z). As we will see, it is the probability amplitude for finding the particle to be in the neighborhood of the point $\vec{x}$ in 3-dimensional space if it is in the state $|\psi\rangle$. It is called the *wave function*.

The bra vector or linear functional corresponding to $|\psi\rangle$ is

$$\langle\psi| = \langle\psi|\,\hat{I} = \int \langle\psi|\,(|\vec{x}\rangle \langle \vec{x}|)\,d\vec{x} = \int \langle\psi\,|\,\vec{x}\rangle \langle \vec{x}| \, d\vec{x} = \int \langle \vec{x}\,|\,\psi\rangle^* \langle \vec{x}| \, d\vec{x} \tag{6.7}$$

The normalization condition takes the form

$$\begin{aligned} \langle\psi\,|\,\psi\rangle = 1 = \langle\psi|\,\hat{I}\,|\psi\rangle &= \int \langle\psi\,|\,\vec{x}\rangle \langle \vec{x}\,|\,\psi\rangle \, d\vec{x} \\ &= \int |\langle \vec{x}\,|\,\psi\rangle|^2 \, d\vec{x} = \int |\psi(\vec{x})|^2 \, d\vec{x} \\ &= \int \psi^*(\vec{x})\psi(\vec{x})\, d\vec{x} \end{aligned} \tag{6.8}$$

The probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\phi\rangle$, where

$$|\phi\rangle = \hat{I}\,|\phi\rangle = \int (|\vec{x}\rangle \langle \vec{x}|)\,|\phi\rangle \, d\vec{x} = \int \langle \vec{x}\,|\,\phi\rangle\,|\vec{x}\rangle \, d\vec{x} \tag{6.9}$$

is given by

$$\begin{aligned} \langle\phi\,|\,\psi\rangle &= \left(\int \langle \vec{x}\,|\,\phi\rangle^* \langle \vec{x}| \, d\vec{x}\right)\left(\int \langle \vec{x}\,'\,|\,\psi\rangle\,|\vec{x}\,'\rangle \, d\vec{x}\,'\right) \\ &= \int d\vec{x} \int d\vec{x}\,' \langle \vec{x}\,|\,\phi\rangle^* \langle \vec{x}\,'\,|\,\psi\rangle \langle \vec{x}\,|\,\vec{x}\,'\rangle \end{aligned} \tag{6.10}$$

Now, in our earlier discussions, we assumed the *normalization* condition

$$\langle \vec{x}\,|\,\vec{x}\,'\rangle = \delta(\vec{x} - \vec{x}\,') \tag{6.11}$$

(position eigenvectors are not normalizable). This normalization condition actually follows (we did not need to assume it) from the expansion in basis states. We have

$$|\psi\rangle = \int \langle \vec{x}\,' \,|\, \psi\rangle \, |\vec{x}\,'\rangle \, d\vec{x}\,'$$
$$\langle \vec{x} \,|\, \psi\rangle = \int \langle \vec{x}\,' \,|\, \psi\rangle \, \langle \vec{x} \,|\, \vec{x}\,'\rangle \, d\vec{x}\,'$$
$$\psi(\vec{x}) = \int \psi(\vec{x}\,') \, \langle \vec{x} \,|\, \vec{x}\,'\rangle \, d\vec{x}\,'$$

which implies the delta-function normalization. Thus, the delta function normalization follows from the completeness property of the projection operators or vice versa.

Using this result in (6.10) we get

$$\begin{aligned} \langle \phi \,|\, \psi\rangle &= \int d\vec{x} \int d\vec{x}\,' \, \langle \vec{x} \,|\, \phi\rangle^* \, \langle \vec{x}\,' \,|\, \psi\rangle \, \delta(\vec{x} - \vec{x}\,') \\ &= \int \langle \vec{x} \,|\, \phi\rangle^* \, \langle \vec{x} \,|\, \psi\rangle \, d\vec{x} = \int \phi^*(\vec{x}) \psi(\vec{x}) \, d\vec{x} \end{aligned} \tag{6.12}$$

We formally write the $\hat{\mathbf{Q}}$ operator using the expansion in eigenvalues and projection operators as

$$\hat{\mathbf{Q}} = \int \vec{x} \, |\vec{x}\rangle \, \langle \vec{x}| \, d\vec{x} \tag{6.13}$$

This represents three integrals, one for each coordinate.

We will also need the properties of the linear momentum operator. The eigenstates $\{|\vec{p}\rangle\}$ of the momentum operator $(\hat{p}_x, \hat{p}_y, \hat{p}_z) = \hat{\mathbf{P}}$ satisfy

$$\hat{p}_x \, |\vec{p}\rangle = p_x \, |\vec{p}\rangle \quad , \quad \hat{p}_y \, |\vec{p}\rangle = p_y \, |\vec{p}\rangle \quad , \quad \hat{p}_z \, |\vec{p}\rangle = p_z \, |\vec{p}\rangle \tag{6.14}$$

where the eigenvalues (p_x, p_y, p_z) are continuous variables in the range $[-\infty, \infty]$. They form the basis of the *momentum representation.*

We then have

$$\hat{I} = \frac{1}{(2\pi\hbar)^3} \int |\vec{p}\rangle \, \langle \vec{p}| \, d\vec{p} \tag{6.15}$$

$$|\psi\rangle = \hat{I} \, |\psi\rangle = \frac{1}{(2\pi\hbar)^3} \int (|\vec{p}\rangle \, \langle \vec{p}|) \, |\psi\rangle \, d\vec{p} = \frac{1}{(2\pi\hbar)^3} \int \langle \vec{p} \,|\, \psi\rangle \, |\vec{p}\rangle \, d\vec{p} \tag{6.16}$$

The expansion coefficient in the momentum representation is

$$\Psi(\vec{p}) = \langle \vec{p} \,|\, \psi\rangle \tag{6.17}$$

It is the probability amplitude for finding the particle with momentum $\vec{p}$ (in the neighborhood of) if it is in the state $|\psi\rangle$.

The bra vector or linear functional corresponding to $|\psi\rangle$ is

$$\langle\psi| = \langle\psi|\hat{I} = \frac{1}{(2\pi\hbar)^3}\int \langle\psi|\,(|\vec{p}\rangle\langle\vec{p}|)\,d\vec{p} = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi\rangle^*\langle\vec{p}|\,d\vec{p} \tag{6.18}$$

The normalization condition takes the form

$$\begin{aligned}\langle\psi\,|\,\psi\rangle = 1 = \langle\psi|\,\hat{I}\,|\psi\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\psi\,|\,\vec{p}\rangle\langle\vec{p}\,|\,\psi\rangle\,d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3}\int |\langle\vec{p}\,|\,\psi\rangle|^2\,d\vec{p}\frac{1}{(2\pi\hbar)^3} = \int |\Psi(\vec{p})|^2\,d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3}\int \Psi^*(\vec{p})\Psi(\vec{p})\,d\vec{p}\end{aligned} \tag{6.19}$$

The probability amplitude for a particle in the state $|\psi\rangle$ to behave like it is in the state $|\phi\rangle$, where

$$|\phi\rangle = \hat{I}\,|\phi\rangle = \frac{1}{(2\pi\hbar)^3}\int (|\vec{p}\rangle\langle\vec{p}|)\,|\phi\rangle\,d\vec{p} = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\phi\rangle\,|\vec{p}\rangle\,d\vec{p} \tag{6.20}$$

is given by

$$\begin{aligned}\langle\phi\,|\,\psi\rangle &= \left(\frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\phi\rangle^*\langle\vec{p}|\,d\vec{p}\right)\left(\frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'\,|\,\psi\rangle\,|\vec{p}\,'\rangle\,d\vec{p}\,'\right) \\ &= \frac{1}{(2\pi\hbar)^6}\int d\vec{p}\int d\vec{p}\,'\,\langle\vec{p}\,|\,\phi\rangle^*\langle\vec{p}\,'\,|\,\psi\rangle\langle\vec{p}\,|\,\vec{p}\,'\rangle\end{aligned} \tag{6.21}$$

The normalization condition follows from

$$\begin{aligned}|\psi\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'\,|\,\psi\rangle\,|\vec{p}\,'\rangle\,d\vec{p}\,' \\ \langle\vec{p}\,|\,\psi\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,'\,|\,\psi\rangle\langle\vec{p}\,|\,\vec{p}\,'\rangle\,d\vec{p}\,' \\ \Psi(\vec{x}) &= \frac{1}{(2\pi\hbar)^3}\int \Psi(\vec{p}\,')\langle\vec{p}\,|\,\vec{p}\,'\rangle\,d\vec{p}\,'\end{aligned}$$

which implies that

$$\frac{1}{(2\pi\hbar)^3}\langle\vec{p}\,|\,\vec{p}\,'\rangle = \delta(\vec{p}-\vec{p}\,') \tag{6.22}$$

Using this result we get

$$\begin{aligned}\langle\phi\,|\,\psi\rangle &= \frac{1}{(2\pi\hbar)^6}\int d\vec{p}\int d\vec{p}\,'\,\langle\vec{p}\,|\,\phi\rangle^*\langle\vec{p}\,'\,|\,\psi\rangle\langle\vec{p}\,|\,\vec{p}\,'\rangle \\ &= \frac{1}{(2\pi\hbar)^3}\int d\vec{p}\int d\vec{p}\,'\,\langle\vec{p}\,|\,\phi\rangle^*\langle\vec{p}\,'\,|\,\psi\rangle\,\delta(\vec{p}-\vec{p}\,') \\ &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\phi\rangle^*\langle\vec{p}\psi\rangle\,d\vec{p} = \frac{1}{(2\pi\hbar)^3}\int \Phi^*(\vec{p})\Psi(\vec{p})\,d\vec{p}\end{aligned} \tag{6.23}$$

We formally write the $\hat{\mathbf{P}}$ operator using the expansion in eigenvalues and projection operators as

$$\hat{\mathbf{P}} = \int \vec{p}\,|\vec{p}\rangle\,\langle\vec{p}|\,d\vec{p} \tag{6.24}$$

We will now derive the *connections* between the two representations.

We have

$$\begin{aligned}\psi(\vec{x}) = \langle\vec{x}\,|\,\psi\rangle = \langle\vec{x}|\,\hat{I}\,|\psi\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}\,|\,\vec{p}\rangle\,\langle\vec{p}\,|\,\psi\rangle\,d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}\,|\,\vec{p}\rangle\,\Psi(\vec{p})\,d\vec{p}\end{aligned} \tag{6.25}$$

and

$$\begin{aligned}\Psi(\vec{p}) = \langle\vec{p}\,|\,\psi\rangle = \langle\vec{p}|\,\hat{I}\,|\psi\rangle &= \int \langle\vec{p}\,|\,\vec{x}\rangle\,\langle\vec{x}\,|\,\psi\rangle\,d\vec{x} \\ &= \int \langle\vec{x}\,|\,\vec{p}\rangle^*\,\psi(\vec{x})\,d\vec{x}\end{aligned} \tag{6.26}$$

Fourier transforms are written

$$g(\vec{p}) = \int e^{-i\vec{p}\cdot\vec{x}/\hbar} f(\vec{x})\,d\vec{x} \quad , \quad f(\vec{x}) = \frac{1}{(2\pi\hbar)^3}\int e^{i\vec{p}\cdot\vec{x}/\hbar} g(\vec{p})\,d\vec{p} \tag{6.27}$$

which gives us the result

$$\langle\vec{x}\,|\,\vec{p}\rangle = e^{i\vec{p}\cdot\vec{x}/\hbar} \tag{6.28}$$

It is not a unique choice, however. It was not unique in our earlier derivation either. It is the choice, however, that allows Quantum mechanics to *make predictions that agree with experiment.* It is Nature's choice!

We might even say that this choice is *another postulate.*

Now, we can use these results to determine the expectation values of operators involving the position and momentum operators.

Since we are interested in the coordinate representation we need only determine the following quantities.

The position operator calculations are straightforward

$$\langle\vec{x}|\,\hat{\mathbf{Q}}\,|\psi\rangle = \vec{x}\,\langle\vec{x}\,|\,\psi\rangle \;\text{ and }\; \langle\vec{x}|\,f(\hat{\mathbf{Q}})\,|\psi\rangle = f(\vec{x})\,\langle\vec{x}\,|\,\psi\rangle \tag{6.25}$$

For the momentum operator we write

$$\begin{aligned}\langle\vec{x}|\,\hat{\mathbf{P}}\,|\psi\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}|\,\hat{\mathbf{P}}\,|\vec{p}\rangle\,\langle\vec{p}\,|\,\psi\rangle\,d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{x}|\,\vec{p}\,|\vec{p}\rangle\,\langle\vec{p}\,|\,\psi\rangle\,d\vec{p} \\ &= \frac{1}{(2\pi\hbar)^3}\int \vec{p}\,\langle\vec{x}\,|\,\vec{p}\rangle\,\langle\vec{p}\,|\,\psi\rangle\,d\vec{p}\end{aligned}$$

Now using $\langle \vec{x} \,|\, \vec{p} \rangle = e^{i\vec{p}\cdot\vec{x}/\hbar}$ we have

$$\vec{p} \,\langle \vec{x} \,|\, \vec{p} \rangle = -i\hbar\nabla \,\langle \vec{x} \,|\, \vec{p} \rangle = \langle \vec{x}| \,\hat{\mathbf{P}}\, |\vec{p}\rangle \tag{6.26}$$

and thus

$$\begin{aligned}
\langle \vec{x}| \,\hat{\mathbf{P}}\, |\psi\rangle &= \frac{1}{(2\pi\hbar)^3} \int \vec{p} \,\langle \vec{x} \,|\, \vec{p} \rangle \,\langle \vec{p} \,|\, \psi \rangle \; d\vec{p} \\
&= (-i\hbar\nabla) \frac{1}{(2\pi\hbar)^3} \int (-i\hbar\nabla) \,\langle \vec{x} \,|\, \vec{p} \rangle \,\langle \vec{p} \,|\, \psi \rangle \; d\vec{p} \\
&= \frac{1}{(2\pi\hbar)^3} \int \langle \vec{x} \,|\, \vec{p} \rangle \,\langle \vec{p} \,|\, \psi \rangle \; d\vec{p} \\
&= -i\hbar\nabla \,\langle \vec{x} \,|\, \psi \rangle
\end{aligned} \tag{6.27}$$

In a similar manner, we can also show that

$$\langle \vec{x}| \,\hat{\mathbf{P}}^2\, |\psi\rangle = (-i\hbar\nabla)^2 = -\hbar^2\nabla^2 \,\langle \vec{x} \,|\, \psi \rangle \tag{6.28}$$

Alternatively, we can show that the gradient is the correct result using the symmetry transformation ideas we developed earlier.

The momentum operator is the generator of displacements in space. We showed earlier that

$$e^{-i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} \,|\vec{x}\rangle = |\vec{x} + \vec{a}\rangle \tag{6.29}$$

Therefore, we have

$$\begin{aligned}
\psi(\vec{x} + \vec{a}) = \langle \vec{x} + \vec{a} \,|\, \psi \rangle &= \langle \vec{x}| \, e^{i\vec{a}\cdot\hat{\mathbf{P}}/\hbar} \,|\psi\rangle \\
&= e^{-\vec{a}\cdot\nabla} \,\langle \vec{x} \,|\, \psi \rangle = e^{-\vec{a}\cdot\nabla} \psi(\vec{x})
\end{aligned} \tag{6.30}$$

Since

$$e^{-\vec{a}\cdot\nabla} = 1 - \vec{a}\cdot\nabla + \frac{(\vec{a}\cdot\nabla)^2}{2!} - \ldots \tag{6.31}$$

we have the standard Taylor series for $\psi(\vec{x} + \vec{a})$. Therefore, the gradient representation of the momentum operator makes sense.

We now use these results to derive the Schrodinger wave equation.

The Schrodinger wave equation is the partial differential equation that corresponds to the eigenvector/eigenvalue equation for the Hamiltonian operator or the energy operator.

The resulting states are the energy eigenstates. We already saw that energy eigenstates are stationary states and thus have simple time dependence. This property will allow us to find the time dependence of amplitudes for very complex systems in a straightforward way.

We have

$$\hat{H}\,|\psi_E\rangle = E\,|\psi_E\rangle \tag{6.32}$$

where E = a number and $\hat{H}$ = the energy operator =(kinetic energy + potential energy) operators, i.e.,

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2m} + V(\hat{\mathbf{Q}}) \tag{6.33}$$

We then have

$$\begin{aligned}
&\langle\vec{x}|\left(\frac{\hat{\mathbf{P}}^2}{2m} + V(\hat{\mathbf{Q}})\right)|\psi_E\rangle = E\,\langle\vec{x}\,|\,\psi_E\rangle \\
&\langle\vec{x}|\,\frac{\hat{\mathbf{P}}^2}{2m}\,|\psi_E\rangle + \langle\vec{x}|\,V(\hat{\mathbf{Q}})\,|\psi_E\rangle = E\,\langle\vec{x}\,|\,\psi_E\rangle \\
&-\frac{\hbar^2}{2m}\nabla^2\,\langle\vec{x}\,|\,\psi_E\rangle + V(\vec{x})\,\langle\vec{x}\,|\,\psi_E\rangle = E\,\langle\vec{x}\,|\,\psi_E\rangle \\
&-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x}) + V(\vec{x})\psi_E(\vec{x}) = E\psi_E(\vec{x})
\end{aligned} \tag{6.34}$$

which is the *time-independent Schrodinger wave equation.* The quantity

$$\psi_E(\vec{x}) = \langle\vec{x}\,|\,\psi_E\rangle \tag{6.35}$$

is the *wave function* or the *energy eigenfunction* in the *position representation* corresponding to energy E.

Now the energy eigenfunctions have a simple time dependence, as we can see from the following. Since

$$\hat{U}(t)\,|\psi_E\rangle = e^{-i\hat{H}t/\hbar}\,|\psi_E\rangle = e^{-iEt/\hbar}\,|\psi_E\rangle \tag{6.36}$$

we have

$$\begin{aligned}
&|\vec{x}\rangle\,\hat{U}(t)\,|\psi_E\rangle = \psi_E(\vec{x},t) = e^{-iEt/\hbar}\,\langle\vec{x}\,|\,\psi_E\rangle \\
&\psi_E(\vec{x},t) = e^{-iEt/\hbar}\psi_E(\vec{x},0)
\end{aligned} \tag{6.37}$$

Therefore,

$$\begin{aligned}
&-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = E\psi_E(\vec{x},t) \\
&-\frac{\hbar^2}{2m}\nabla^2\psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = i\hbar\frac{\partial}{\partial t}\psi_E(\vec{x},t)
\end{aligned} \tag{6.38}$$

which is the *time-dependent Schrodinger wave equation.*

We will have much more to say about these equations and how to use them later on. For now, however, let us look at these ideas in a couple of different ways to try and get a better understanding of what they mean.

6.2 The Free Particle and Wave Packets

Let us assume that the potential energy term in the one-particle Schrodinger equation is equal to zero. The solution, in this case, is called a *free particle*.

We can easily see the form of the solution as a function of space and time as follows: the Hamiltonian of the system is

$$\hat{H} = \frac{\hat{\mathbf{P}}^2}{2m} \text{ with } \hat{H}\,|\psi_E\rangle = E\,|\psi_E\rangle \tag{6.39}$$

This means that $[\hat{H}, \hat{\mathbf{P}}] = 0$ and thus $\hat{H}$ and $\hat{\mathbf{P}}$ share a common set of eigenvectors(eigenfunctions). The eigenvectors(eigenfunctions) corresponding to linear momentum $\hat{\mathbf{P}}$ are easy to find from our earlier derivations. We have

$$\begin{aligned}
\hat{\mathbf{P}}\,|\vec{p}\rangle &= \hat{\mathbf{P}}\,|\psi_E\rangle = \vec{p}\,|\vec{p}\rangle = \vec{p}\,|\psi_E\rangle \\
\langle\vec{x}|\,\hat{\mathbf{P}}\,|\psi_E\rangle &= \vec{p}\,\langle\vec{x}\,|\,\psi_E\rangle = \vec{p}e^{i\vec{p}\cdot\vec{x}/\hbar} \\
&= -i\hbar\nabla e^{i\vec{p}\cdot\vec{x}/\hbar} = -i\hbar\nabla\,\langle\vec{x}\,|\,\psi_E\rangle
\end{aligned} \tag{6.40}$$

which says that the eigenfunctions are

$$\langle\vec{x}\,|\,\vec{p}\rangle = e^{i\vec{p}\cdot\vec{x}/\hbar} = \psi_E(\vec{x}) = \langle\vec{x}\,|\,\psi_E\rangle = \langle\vec{x}\,|\,E\rangle \tag{6.41}$$

We can see that this satisfies the time-independent Schrodinger equation with no potential term. This solution is also called a *plane-wave*.

If we assume that at time $t = 0$, the state of the particle is

$$\psi_E(\vec{x}, 0) = e^{i\vec{p}\cdot\vec{x}/\hbar} \tag{6.42}$$

then the time dependence is given by

$$\begin{aligned}
|\vec{x}, t\rangle &= \hat{U}(t)\,|\vec{x}, 0\rangle = e^{-iEt/\hbar}\,|\vec{x}, 0\rangle \\
\langle\vec{x}, t\,|\,\psi_E\rangle &= e^{-iEt/\hbar}\,\langle\vec{x}, 0\,|\,\psi_E\rangle \\
\psi_E(\vec{x}, t) &= e^{-iEt/\hbar}\psi_E(\vec{x}, 0) = e^{-iEt/\hbar}e^{i\vec{p}\cdot\vec{x}/\hbar} = e^{i(\vec{p}\cdot\vec{x}-Et)/\hbar}
\end{aligned} \tag{6.43}$$

where $E = p^2/2m$.

Now we can write the position state vector the as a linear combination of momentum state vectors or correspondingly we can write position space wave function as a linear combination of momentum eigenfunctions or plane waves(since they are a basis).

$$\begin{aligned}
|\psi_E\rangle &= \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,|\vec{p}\rangle\; d\vec{p} \\
\langle\vec{x}\,|\,\psi_E\rangle &= \psi_E(\vec{x}) = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,\langle\vec{x}\,|\,\vec{p}\rangle\; d\vec{p} \\
\psi_E(\vec{x}, 0) &= \frac{1}{(2\pi\hbar)^3}\int e^{i\vec{p}\cdot\vec{x}/\hbar}\Psi_E(\vec{p}, 0)\, d\vec{p}
\end{aligned} \tag{6.44}$$

Using the same derivation as above (6.48) we can show that

$$|\psi_E\rangle = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,|\vec{p}\rangle\; d\vec{p}$$

$$\langle\vec{x},t\,|\,\psi_E\rangle = \psi_E(\vec{x},t) = \frac{1}{(2\pi\hbar)^3}\int \langle\vec{p}\,|\,\psi_E\rangle\,\langle\vec{x},t\,|\,\vec{p}\rangle\; d\vec{p}$$

$$\psi_E(\vec{x},t) = \frac{1}{(2\pi\hbar)^3}\int e^{i(\vec{p}\cdot\vec{x}-Et)/\hbar}\Psi_E(\vec{p},0)\,d\vec{p} \tag{6.45}$$

We note that in the state $|\vec{p}\rangle$, which is an eigenvector of $\hat{\mathbf{P}}$, the momentum has the value $\vec{p}$ with probability = 1 (a certainty), which is the meaning of an eigenvector! However, the probability of finding the particle at the point $\vec{x}$ is independent of $\vec{x}$, i.e.,

$$|\,\langle\vec{x}\,|\,\vec{p}\rangle\,|^2 = |e^{i\vec{p}\cdot\vec{x}/\hbar}|^2 = 1 \tag{6.46}$$

This means that the particle is equally probable to have any value of momentum. The momentum is completely uncertain in a position eigenstate.

If we consider an intermediate case in which the particle is reasonably well localized in position space, and at the same time, has a fairly well defined momentum, then the wave function for such a state is called a *wave packet*.

Suppose that the function $|\Psi_E(\vec{p},0)|$ is peaked about the value $\vec{p}_0$, with a width $\approx \Delta\vec{p}$. In particular, at $t = 0$, we choose

$$\Psi_E(\vec{p},0) = g(\vec{p})e^{i\alpha(\vec{p})} \tag{6.47}$$

where $g(\vec{p})$, called the weight function, takes the form of a real, nonnegative function such as shown in Figure 6.1 below, and the phase $\alpha(\vec{p})$ is real.

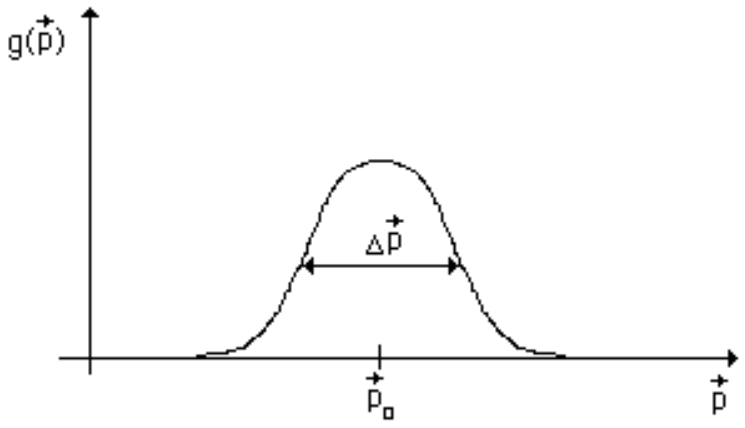


Figure 6.1: Weight Function

We then have

$$\psi_E(\vec{x},t) = \frac{1}{(2\pi\hbar)^3}\int g(\vec{p})e^{i(\vec{p}\cdot\vec{x}-Et+\hbar\alpha(\vec{p}))/\hbar}\,d\vec{p} \tag{6.48}$$

as the wave function of a particle whose momentum is *approximately* $\vec{p}_0$. This means that most of the time, if we measure the momentum of the particle in this state, then we would find a value within $\Delta\vec{p}$ of $\vec{p}_0$.

We now determine how this particle moves in time and why we have called this expression a wave packet.

If for some position $\vec{x}$ (which is a parameter in the integral) the phase of the exponential term is not *slowly varying* for some values of integration variable $\vec{p}$, then the integral will equal zero. This occurs because the exponential term is a very rapidly oscillating function (since $\hbar$ is very small), i.e., if you multiply a very rapidly varying function by a function of the assumed form of $g(\vec{p})$, then *on the average* the integrand is zero. In effect, we are seeing destructive interference between the plane waves.

If, however, $\vec{x}$ is a point such that the phase of the exponential remains *fairly constant* over the *same* range of $\vec{p}$ where the function $g(\vec{p})$ is nonzero, then we will get a nonzero value for the integral and the particle will have some nonzero probability of being at that point.

One way to make this work is as follows:

1. At time t, assume the point $\vec{x}_t$ makes the phase *stationary* near $\vec{p}_0$. By stationary we mean

$$\nabla_{\vec{p}}(\vec{p}\cdot\vec{x}_t - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0} = 0 \tag{6.49}$$

   or we are near an extremum of the phase and hence the phase will be slowly varying.

2. We will then get strong *constructive* interference between the plane waves and $\psi_E(\vec{x},t)$ will be nonzero.

Solving the stationary phase equation for $\vec{x}_t$ we find, using $E = p^2/2m$

$$\begin{aligned}
&(\vec{x}_t - \nabla_{\vec{p}}(E)t + \hbar(\nabla_{\vec{p}}\alpha(\vec{p})))_{\vec{p}=\vec{p}_0} = 0\\
&\vec{x}_t = ((\vec{p}E)t + \hbar(\nabla_{\vec{p}}\alpha(\vec{p})))_{\vec{p}=\vec{p}_0}\\
&\quad = \frac{\vec{p}_0}{m}t - \hbar(\nabla_{\vec{p}}\alpha(\vec{p})) = \frac{\vec{p}_0}{m}t + \vec{x}_0
\end{aligned} \tag{6.50}$$

Therefore, the point $\vec{x}_t$, which is essentially the center of the wave packet, moves in time with a constant velocity $\vec{v} = \vec{p}_0/m$. This is exactly the velocity one expects for a free particle of momentum $\vec{p}_0$ and mass m. It is called the

group velocity of the wave packet. This procedure is called the *stationary phase method*.

How *localized* is this wave packet(*particle*) or what is the spatial extent of the packet?

We will have a nonzero integral and hence a nonzero probability for the particle to be at a particular point as long as we have constructive interference or as long as the exponential undergoes less than one oscillation as $\vec{p}$ varies over the region for which the function $g(\vec{p})$ is also large. The change in the phase of the exponential as we vary the x-component of $\vec{p}$ is approximately

$$\begin{aligned}\Delta\phi &= \Delta(\vec{p}\cdot\vec{x} - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0} = \frac{1}{\hbar}\Delta p_x \frac{\partial}{\partial p_x}(\vec{p}\cdot\vec{x} - Et + \hbar\alpha(\vec{p}))_{\vec{p}=\vec{p}_0} \\ &= \frac{1}{\hbar}\Delta p_x \left(x - \frac{\partial}{\partial p_x}(Et + \hbar\alpha(\vec{p})\right)_{\vec{p}=\vec{p}_0} = \frac{1}{\hbar}\Delta p_x (x - x_t) \end{aligned} \tag{6.51}$$

As long as $\Delta\phi \leq 2\pi$ we will get constructive interference and the integral will be nonzero. This tells us the extent of the wave packet in the x-direction about the point x_t. We get

$$|x - x_t| \geq \frac{2\pi\hbar}{\Delta p_x} \text{ or } |x - x_t|\Delta p_x \geq h \tag{6.52}$$

or

$$\Delta x \Delta p_x \geq h \tag{6.53}$$

Similarly we can show that $\Delta y \Delta p_y \geq h$ and $\Delta z \Delta p_z \geq h$.

These relations between the spatial extent of the wave packet or the uncertainty in its position, and the uncertainty in the particle momentum, are often called the *Heisenberg uncertainty principle*. We will see below that the uncertainty principle involves more than just these simple properties of a wave packet. We will also develop more details about wave packets later on this book.

Note that there is no uncertainty relation between

$$\begin{aligned}&\Delta x \text{ and } \Delta y, \Delta z, \Delta p_y, \Delta p_z \\ &\Delta y \text{ and } \Delta x, \Delta z, \Delta p_x, \Delta p_z \\ &\Delta z \text{ and } \Delta x, \Delta y, \Delta p_x, \Delta p_y\end{aligned}$$

What is the difference between these pairs of variables? The difference is in their commutators. It turns out, in this example of the wave packet, that only those observables that are represented by non-commuting operators have nonzero uncertainty relations.

This result has *nothing to do* with wave packets, Schrodinger equations, wave

functions or Fourier transforms, although all of these quantities can be used to show that uncertainty relations exist.

The existence of uncertainty relations between non-commuting observables was actually built into the theory of Quantum Mechanics when we assumed that we were in an abstract linear vector space. All linear vector spaces have a property called the Schwarz inequality and it always leads to such uncertainty relations for non-commuting operators. So the Heisenberg uncertainty principle, to which much mysticism has been attached, is really an assumption and little more than a lemma that follows from the original assumption of a vector space.

6.3 Derivation of the Uncertainty Relations in General

Given two Hermitian operators $\hat{A}$ and $\hat{B}$ we define the two new operators

$$\hat{D}_A = \hat{A} - \langle \hat{A} \rangle \text{ and } \hat{D}_B = \hat{B} - \langle \hat{B} \rangle \tag{6.54}$$

where $\langle \hat{O} \rangle = \langle \psi | \hat{O} | \psi \rangle$ equals the average or expectation value in the state $|\psi\rangle$. In the statistical analysis of data, we use a quantity called the standard or mean-square deviation as a measure of the uncertainty of an observed quantity. It is defined, for a set of N measurements of the quantity Δq by

$$\begin{aligned}(\Delta q)^2 = (\text{standard deviation})^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i - q_{average})^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(q_i)^2 - \frac{1}{N}\sum_{i=1}^{N}(q_i q_{average}) - \frac{1}{N}\sum_{i=1}^{N}(q_{average} q_i) + \frac{1}{N}\sum_{i=1}^{N}(q_{average})^2 \\ &= (q^2)_{average} - (q_{average})^2 \end{aligned} \tag{6.55}$$

where we have used

$$q_{average} = \frac{1}{N}\sum_{i=1}^{N} q_i \tag{6.56}$$

In analogy, we define the mean-square deviations for $\hat{A}$ and $\hat{B}$ as

$$(\Delta \hat{A})^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 = \langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle = \langle \hat{D}_A^2 \rangle \tag{6.57}$$

$$(\Delta \hat{B})^2 = \langle \hat{B}^2 \rangle - \langle \hat{B} \rangle^2 = \langle (\hat{B} - \langle \hat{B} \rangle)^2 \rangle = \langle \hat{D}_B^2 \rangle \tag{6.58}$$

We then have

$$(\Delta \hat{A})^2 (\Delta \hat{B})^2 = \langle \hat{D}_A^2 \rangle \langle \hat{D}_B^2 \rangle \tag{6.59}$$

Now we assume that

$$[\hat{D}_A, \hat{B}] = [\hat{A}, \hat{B}] = [\hat{D}_A, \hat{D}_B] = i\hat{C} \tag{6.60}$$

where $\hat{C}$ is also a Hermitian operator and we let

$$|\alpha\rangle = \hat{D}_A|\psi\rangle = (\hat{A} - \langle\hat{A}\rangle)|\psi\rangle \text{ and } |\beta\rangle = \hat{D}_B|\psi\rangle = (\hat{B} - \langle\hat{B}\rangle)|\psi\rangle \tag{6.61}$$

Then we have

$$(\Delta\hat{A})^2 = \langle\hat{D}_A^2\rangle = \langle\psi|\hat{D}_A^2|\psi\rangle = (\langle\psi|\hat{D}_A)(\hat{D}_A|\psi\rangle) = \langle\alpha|\alpha\rangle \tag{6.62}$$

$$(\Delta\hat{B})^2 = \langle\hat{D}_B^2\rangle = \langle\psi|\hat{D}_B^2|\psi\rangle = (\langle\psi|\hat{D}_B)(\hat{D}_B|\psi\rangle) = \langle\beta|\beta\rangle \tag{6.63}$$

The Schwarz inequality says that for any two vectors we must have the relation

$$\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2 \tag{6.64}$$

We therefore have

$$\begin{aligned}(\Delta\hat{A})^2(\Delta\hat{B})^2 &= \langle\hat{D}_A^2\rangle\langle\hat{D}_B^2\rangle = \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2 \\ (\Delta\hat{A})^2(\Delta\hat{B})^2 &\geq |\langle\psi|\hat{D}_A\hat{D}_B|\psi\rangle|^2 = |\langle\hat{D}_A\hat{D}_B\rangle|^2\end{aligned} \tag{6.65}$$

Now

$$\begin{aligned}|\langle\hat{D}_A\hat{D}_B\rangle|^2 = |\langle\Delta\hat{A}\Delta\hat{B}\rangle|^2 &= |\langle\frac{1}{2}[\Delta\hat{A},\Delta\hat{B}] + \frac{1}{2}\{\Delta\hat{A},\Delta\hat{B}\}\rangle|^2 \\ &= |\langle\frac{1}{2}[\hat{A},\hat{B}] + \frac{1}{2}\{\hat{A},\hat{B}\}\rangle|^2\end{aligned} \tag{6.66}$$

where

$[\Delta\hat{A},\Delta\hat{B}]^\dagger = -[\Delta\hat{A},\Delta\hat{B}] \rightarrow$ anti-Hermitian $\rightarrow$ expectation value is imaginary

$\{\Delta\hat{A},\Delta\hat{B}\}^\dagger = \{\Delta\hat{A},\Delta\hat{B}\} \rightarrow$ Hermitian $\rightarrow$ expectation value is real

Therefore,

$$\begin{aligned}(\Delta\hat{A})^2(\Delta\hat{B})^2 &\geq |\langle\frac{1}{2}[\hat{A},\hat{B}] + \frac{1}{2}\{\hat{A},\hat{B}\}\rangle|^2 \\ &\geq \frac{1}{4}|\langle i\hat{C}\rangle + a|^2\end{aligned}$$

where a is a real number. Then

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq \frac{1}{4}|a|^2 + \frac{1}{4}|\langle\hat{C}\rangle|^2 \geq \frac{1}{4}|\langle\hat{C}\rangle|^2 \tag{6.67}$$

since $|a|^2/4 \geq 0$. This is the Heisenberg uncertainty principle. *It is simply the Schwarz inequality!!*

The standard form in most texts follows from $[\hat{x},\hat{p}_x] = i\hbar$, which gives

$$\Delta\hat{x}\Delta\hat{p}_x \geq \frac{\hbar}{2} \tag{6.68}$$

We note that if $[\hat{A},\hat{B}] = i\hat{C} = 0$, we then get

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq 0 \tag{6.69}$$

or commuting observables do not have an uncertainty principle!

6.3.1 The Meaning of the Indeterminacy Relations

What is the significance of indeterminacy relations in the world of experimental physics?

Consider the experimental results shown in Figure 6.2 below:

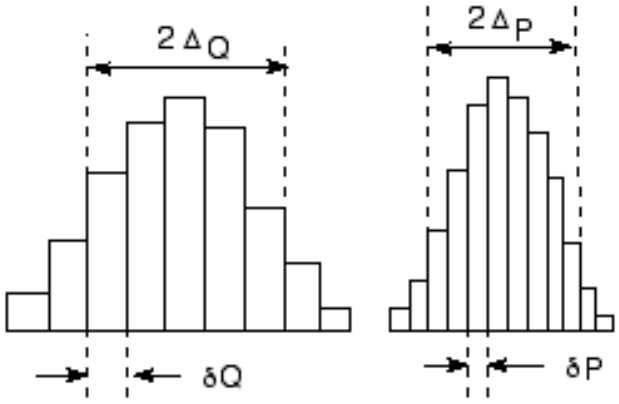


Figure 6.2: Experimental Data

These are frequency distributions for the results of independent measurements of Q and P on an ensemble if similarly prepared systems, i.e., on each of a large number of similarly prepared systems one performs a single measurement (either Q or P). The histograms are the statistical distribution of the results.

The standard deviations (variances) as shown must satisfy (according to the theory) the relation

$$\Delta_Q \Delta_P \geq \frac{\hbar}{2} \tag{6.70}$$

They must be distinguished from the resolution of the individual measurements, δQ and δP.

Let me emphasize these points:

1. The quantities Δ_Q and Δ_P are *not* errors of measurement. The *errors* or preferably the *resolutions* of the Q and P measuring instruments are δ_Q and δ_P. They are logically unrelated to Δ_Q and Δ_P and to the uncertainty relations except for the practical requirement that if

$$\delta_Q > \Delta_Q \ (or\ \delta_P > \Delta_P) \tag{6.71}$$

   then it will not be possible to determine Δ_Q (or Δ_P) in the experiment and the experiment cannot test the uncertainty relation.

2. The experimental test of the indeterminacy relation *does not* involve simultaneous measurements of Q and P, but rather it involves the measurement of one or the other of these dynamical variables on each independently prepared representative of the particular state being studied.

Why am I being so picky here?

The quantities Δ_Q and Δ_P as defined here are often misinterpreted as the errors of individual measurements. This probably arises because Heisenberg's original paper on this subject, published in 1927, was based on an early version of quantum mechanics that predates the systematic formulation and statistical interpretation of quantum mechanics as it exists now. The derivation, as carried out here was not possible in 1927!

6.3.2 Time-Energy Uncertainty Relations

The use of time-energy uncertainty relations in most textbooks is simply incorrect. Let us now derive the most we can say about such relations.

Earlier (6.380), we showed that

$$\frac{d\langle \hat{Q} \rangle_t}{dt} = Tr\left(\hat{W}_0 \frac{d\hat{Q}_H(t)}{dt}\right) = Tr\left[\frac{i}{\hbar}\hat{W}_0[\hat{H}, \hat{Q}_H(t)] + \hat{W}_0\left(\frac{\partial \hat{Q}}{\partial t}\right)_H\right] \tag{6.72}$$

in the Heisenberg picture. So we can write in general that for any operator $\hat{Q}$

$$\frac{d\langle\hat{Q}\rangle}{dt} = \frac{1}{i\hbar}\langle[\hat{Q},\hat{H}]\rangle + \langle\frac{\partial\hat{Q}}{\partial t}\rangle \tag{6.73}$$

Now consider a system whose Hamiltonian $\hat{H}$ does not explicitly depend on time and let $\hat{Q}$ be another observable of this system which does not depend on time explicitly so that

$$\frac{d\langle\hat{Q}\rangle}{dt} = \frac{1}{i\hbar}\langle[\hat{Q},\hat{H}]\rangle \tag{6.74}$$

We consider the dynamical state of the system at a given time t. Let $|\psi\rangle$ be the vector representing that state. Call Δ_Q and Δ_E the root-mean-square deviations of $\hat{Q}$ and $\hat{H}$, respectively. Applying the Schwarz inequality (as earlier) to the vectors $(\hat{Q}-\langle\hat{Q}\rangle)|\psi\rangle$ and $(\hat{H}-\langle\hat{H}\rangle)|\psi\rangle$ and carrying out the same manipulations, we find after some calculations

$$\Delta Q\Delta E \geq \frac{1}{2}|\langle[\hat{Q},\hat{H}]\rangle| \tag{6.75}$$

the equality being realized when $|\psi\rangle$ satisfies the equation

$$(\hat{Q}-\alpha)|\psi\rangle = i\gamma(\hat{H}-\epsilon)|\psi\rangle \tag{6.76}$$

where α, γ, and ϵ are arbitrary real constants. We then have from (6.75)

$$\frac{\Delta Q}{|\frac{d\langle\hat{Q}\rangle}{dt}|}\Delta E \geq \frac{\hbar}{2} \tag{6.77}$$

or

$$\tau_Q\Delta E \geq \frac{\hbar}{2} \tag{6.78}$$

where we have defined

$$\tau_Q = \frac{\Delta Q}{|\frac{d\langle\hat{Q}\rangle}{dt}|} \tag{6.79}$$

τ_Q appears as a time characteristic of the evolution of the expectation value of $\hat{Q}$. It is the time required for the center of the center $\langle\hat{Q}\rangle$ of the statistical distribution of $\hat{Q}$ to be displaced by an amount equal to its width ΔQ. In other words, the time necessary for this statistical distribution to be appreciably modified. In this way we can define a characteristic evolution time for each dynamical variable of the system.

Let τ be the shortest of the times thus defined. τ may be considered as the characteristic time of evolution of the system itself, that is, whatever the measurement carried out on the system at an instant of time t', the statistical distribution of the results is essentially the same as would be obtained at the instant t, as long as the difference $|t-t'|$ is less than τ.

This time τ and the energy spread ΔE satisfy the time-energy uncertainty relation

$$\tau \Delta E \geq \frac{\hbar}{2} \tag{6.80}$$

If, in particular, the system is in a stationary state where $d\langle \hat{Q} \rangle / dt = 0$ no matter what $\hat{Q}$, and consequently τ is infinite, then $\Delta E = 0$ according to (6.80).

Ordinary time t is just a parameter in non-relativistic QM and not an operator! We cannot say that

$$\Delta t \Delta E \geq \frac{\hbar}{2} \tag{6.81}$$

which is an equation that has no meaning!

6.4 The Wave Function and Its Meaning

The Schrodinger equation

$$-\frac{\hbar^2}{2m}\nabla^2 \psi_E(\vec{x},t) + V(\vec{x})\psi_E(\vec{x},t) = i\hbar \frac{\partial}{\partial t}\psi_E(\vec{x},t) \tag{6.82}$$

has the mathematical form similar to that of a type of equation called a *wave equation*.

Since other wave equations imply the existence of real physical fields or waves propagating in real three dimensional space, a possible interpretation of the Schrodinger wave function $\psi_E(\vec{x},t)$ is that of a *wave* in real, three dimensional space. We might also associate the wave field with a particle or a particle could be identified with a wave packet solution to the Schrodinger equation, as we did earlier in our discussion.

These are all misinterpretations!!!

We are being misled here because we are working with a simple system, namely, a single particle in three-dimensional space.

To see that these interpretations are not valid, we must look at the more complicated case of a system of N particles. We now generalize to a coordinate representation for the system of N particles as follows:

1. Choose as a basis the set of vectors that is a common set of eigenvectors for the N position operators $\hat{\mathbf{Q}}^{(\mathbf{1})}, \hat{\mathbf{Q}}^{(\mathbf{2})}, \ldots, \hat{\mathbf{Q}}^{(\mathbf{N})}$ corresponding to the N particles.

2. Assuming that each position operator satisfies an eigenvalue/eigenvector equation of the form

$$\hat{Q}^{(i)} \left| \vec{x}^{(i)} \right\rangle = \vec{x}^{(i)} \left| \vec{x}^{(i)} \right\rangle \quad , \quad i = 1, 2, \ldots, N \tag{6.83}$$

and that each of the sets of single particle eigenvectors forms a basis for the 3-dimensional subspace of the single particle. The basis states for the N particle system are then the *direct product* states among the N sets of single particle eigenvectors. We write them as

$$\left|\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}\right\rangle = \left|\vec{x}^{(1)}\right\rangle \otimes \left|\vec{x}^{(2)}\right\rangle \otimes \ldots \otimes \left|\vec{x}^{(N)}\right\rangle \tag{6.84}$$

The state vector $|\Psi\rangle$ representing an N particle system is then represented in the $3N$-dimensional *configuration space* corresponding to the $3N$ coordinates of the N particles by

$$\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) = \left\langle \vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)} \,\middle|\, \Psi \right\rangle \tag{6.85}$$

The Hamiltonian for the N particle system is given by

$$\hat{H} = \sum_{n=1}^{N} \hat{H}_n + \hat{U}(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) \tag{6.86}$$

where

$$\begin{aligned} \hat{H}_n &= \text{ single particle Hamiltonian for the } n^{th} \text{ particle} \\ &= -\frac{\hbar^2}{2m_n}\nabla_n^2 + V(\vec{x}^{(n)}) \end{aligned} \tag{6.87}$$

and $\hat{U}(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) =$ the interparticle interaction potential energy.

The equation of motion(the N-particle Schrodinger equation) is then

$$\hat{H}\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) = i\hbar\frac{\partial}{\partial t}\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}) \tag{6.88}$$

As we shall now see, this N particle equation *does not allow* any of the interpretations that might seem to be valid for the single particle Schrodinger equation.

If we associated a physical wave in real, three dimensional space with a particle or if a particle were to be identified with a wave packet, then there would have to be N interacting waves in the real, three dimensional space where we actually measure things. It is clear, however, that the N particle Schrodinger equation says that this is NOT the case. There is only one *wave* function in an abstract $3N$-dimensional configuration space.

We are able to make the incorrect interpretation in the one particle case because it just happens, in that case, that the real, 3-dimensional has a one-to-one correspondence to the 3-dimensional configuration space of the Schrodinger equation.

The proper interpretation of Ψ is that it is a *statistical state function.* It is nothing more than a function that enables us to calculate probability distributions

for all observables and their expectation values. *The important physical quantities are the probabilities and the NOT the function used to calculate them.*

$|\Psi(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)})|^2$ = probability density in configuration space for

particle 1 to be at $\vec{x}^{(1)}$

particle 2 to be at $\vec{x}^{(2)}$

...

...

particle N to be at $\vec{x}^{(N)}$

We can demonstrate the necessity for a purely statistical interpretation for the N-particle wave function as follows. Consider the experiment with photons shown in Figure 6.3 below.

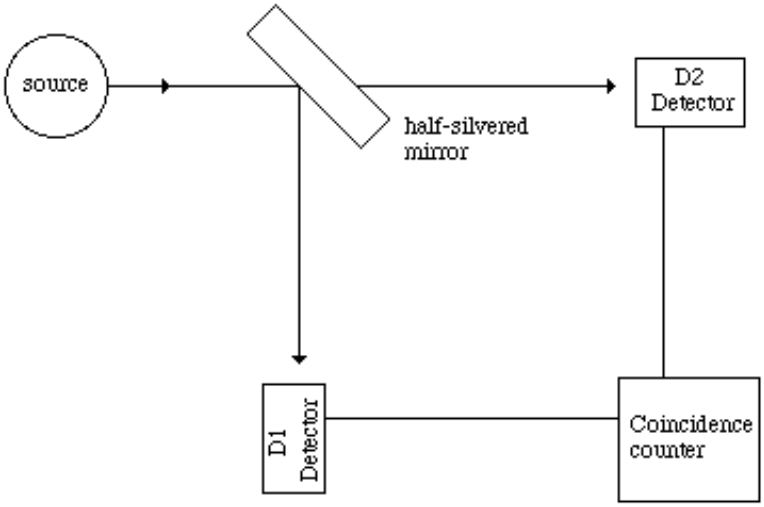


Figure 6.3: Experimental Setup

We have a source which emits light(photons). This light falls on a half-silvered mirror, which allows 1/2 of the light to be transmitted though and 1/2 to be reflected as shown. The transmitted and reflected light is detected with the two photomultipliers $D1$ and $D2$, respectively. Finally signals generated by the detectors when they record a photon hit, are sent to a coincidence counter, which records a count when it receives two signals at the same time (one from each detector) to within some specified accuracy.

Now, suppose that the source can be described as an emitter of identical localized wave packets ($|\Psi|^2$ is nonzero only over a finite region of space). Assume that these packets are emitted at random times, at an average rate R per second and that no more than one packet exists in a small interval Δ. The probability of an emission in any short interval of length Δt is then

$$p = R\Delta t \tag{6.89}$$

In each of these intervals, the detectors either see a photon or they do not. The experimental arrangement guarantees that over long periods of time each detector sees 1/2 of the emitted photons.

We can learn more about the details of what is happening as the detectors record photon hits by monitoring the temporal response of the two detectors, in particular, we can ask whether the detectors have both recorded photon hits during the same interval.

The experimental procedure is to send light into the system and record the number of coincidences relative to number of individual counts of the detectors.

The results are analyzed in terms of an anti-coincidence parameter A given by

$$A = \frac{P_c}{P_1 P_2} \tag{6.90}$$

where

P_1 = experimentally measured probability of detector 1 responding
P_2 = experimentally measured probability of detector 2 responding
P_c = experimentally measured probability of coincidences

If light is composed of single particles (photons) and photons are not divisible and $|\Psi(\vec{x},t)|^2$ = the probability per unit volume at time t that the photon will be located within some small volume about the point $\vec{x}$, then the two detectors should never both record a hit in the same interval (they are mutually exclusive events) and thus $P_c = A = 0$.

As we shall see later on when we analyze the space-time behavior of wave packets, if the photon is a wave packet, then the packet will split into two equal

parts, one on each path. This means that both detectors will always record hits during the same interval and thus $P_c = 1$.

The probability of a detector recording a count is proportional to the amount of the wave amplitude in the region of the detector, in particular, to

$$\int_{\substack{detector \\ volume}} |\Psi|^2 \, d\vec{x} \tag{6.91}$$

For a symmetrical, equal path device, this is exactly $p/2$ for each detector. If the detectors are far apart, the respective triggerings are independent (we assume that a spacelike interval exists between the events). Therefore, the probability of coincidence is given by

$$P_c = \frac{p^2}{4} \tag{6.92}$$

Experiments of this type were carried out by Clauser(1974) and Grangier, Roger and Aspect(1986).

They obtained $A = 0$. This confirmed that light is photons or single particles with a statistical interpretation of the wave function and not particles represented by wave packets or wave fields of some kind.

6.5 One-Dimensional Systems

The time-independent Schrodinger equation in 1-dimension is

$$-\frac{\hbar^2}{2m}\nabla^2\psi_E(x) + V(x)\psi_E(x) = E\psi_E(x) \tag{6.93}$$

The solutions $\psi_E(x)$ are the energy eigenstates (eigenfunctions). As we have seen, their time dependence is given by

$$\psi_E(x,t) = e^{-iEt/\hbar}\psi_E(x,0) \text{ where } \psi_E(x,0) = \langle x \,|\, E \rangle \tag{6.94}$$

and

$$\hat{H}\,|E\rangle = E\,|E\rangle \text{ where } \hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x}) \tag{6.95}$$

We are thus faced with solving an ordinary differential equation with boundary conditions. Since $\psi_E(x)$ is physically related to a probability amplitude and hence to a measurable probability, we *assume* that $\psi_E(x)$ is continuous.

Using this assumption, we can determine the general continuity properties of $d\psi_E(x)/dx$. The continuity property at a particular point, say $x = x_0$, is derived as follows:

$$\begin{aligned} \int_{x_0-\epsilon}^{x_0+\epsilon} \frac{d^2\psi_E(x)}{dx^2}\,dx &= \int_{x_0-\epsilon}^{x_0+\epsilon} d\left(\frac{d\psi_E(x)}{dx}\right) \\ &= -\frac{2m}{\hbar^2}\left[E\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx - \int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx\right] \end{aligned} \tag{6.96}$$

Taking the limit as $\epsilon \to 0$ we have

$$\lim_{\epsilon\to 0}\left(\left.\frac{d\psi_E(x)}{dx}\right|_{x=x_0+\epsilon} - \left.\frac{d\psi_E(x)}{dx}\right|_{x=x_0-\epsilon}\right)$$

$$= -\frac{2m}{\hbar^2}\left[E\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx - \lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}V(x)\psi_E(x)\,dx\right] \tag{6.97}$$

or

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}V(x)\psi_E(x)\,dx \tag{6.98}$$

where we have used the continuity of $\psi_E(x)$ to set

$$\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}\psi_E(x)\,dx = 0 \tag{6.99}$$

This make it clear that whether or not $d\psi_E(x)/dx$ has a discontinuity depends directly on the properties of the potential energy function.

If $V(x)$ is continuous at $x = x_0$ (harmonic oscillator example later), i.e., if

$$\lim_{\epsilon\to 0}[V(x_0+\epsilon) - V(x_0-\epsilon)] = 0 \tag{6.100}$$

then

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}V(x)\psi_E(x)\,dx = 0 \tag{6.101}$$

and $d\psi_E(x)/dx$ is continuous at $x = x_0$.

If $V(x)$ has a finite discontinuity (jump) at $x = x_0$ (finite square well and square barrier examples later), i.e., if

$$\lim_{\epsilon\to 0}[V(x_0+\epsilon) - V(x_0-\epsilon)] = \text{ finite} \tag{6.102}$$

then

$$\Delta\left(\frac{d\psi_E(x)}{dx}\right) = \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon}V(x)\psi_E(x)\,dx = 0 \tag{6.103}$$

and $d\psi_E(x)/dx$ is continuous at $x = x_0$.

Finally, if V(x) has an infinite jump at $x = x_0$ (infinite square well and delta-function examples later), then we have two choices:

1. if the potential is infinite over an extended range of x (the infinite well), then we must force $\psi_E(x) = 0$ in that region and use only the continuity of $\psi_E(x)$ as a boundary condition at the edge of the region

2. if the potential is infinite at a single point, i.e., $V(x) = A\delta(x - x_0)$, then

$$\begin{aligned}\Delta\left(\frac{d\psi_E(x)}{dx}\right) &= \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} V(x)\psi_E(x)\,dx \\ &= \frac{2m}{\hbar^2}\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} A\delta(x-x_0)\psi_E(x)\,dx \\ &= \frac{2mA}{\hbar^2}\lim_{\epsilon\to 0}\psi_E(x_0) = \frac{2mA}{\hbar^2}\psi_E(x_0) \qquad (6.104)\end{aligned}$$

and, thus, $d\psi_E(x)/dx$ is discontinuous at $x = x_0$.

The last thing we must worry about is the validity of our probability interpretation of $\psi_E(x)$, i.e.,

$$\begin{aligned}\psi_E(x) = \langle x\,|\,\psi_E\rangle = &\text{ probability amplitude for the particle} \\ &\text{in the state } |\psi_E\rangle \text{ to be found at } x\end{aligned}$$

which says that we must also have

$$\langle\psi_E\,|\,\psi_E\rangle = \int_{-\infty}^{\infty}|\psi_E(x)|^2\,dx < \infty \qquad (6.105)$$

This means that we must be able to normalize the wave functions and make the total probability that the particle is somewhere on the x-axis equal to one.

A wide range of interesting physical systems can be studied using 1-dimensional potential energy functions. We will consider potentials in the form of square wells and square barriers, delta-functions, linear functions and parabolic functions.

We will now learn how to work with potentials in 1-dimension by doing a variety of different examples. Along the way we will learn some very clever techniques and tricks and expand our understanding of the physics and mathematics of Quantum Mechanics.

We start with a simple system to illustrate the process for solving the Schrodinger equation.

6.5.1 One-Dimensional Barrier

Consider the potential energy function

$$V(x) = \begin{cases} 0 & x \le 0 \\ V_0 & x \ge 0 \end{cases} \qquad (6.106)$$

which looks like Figure 6.4 below.

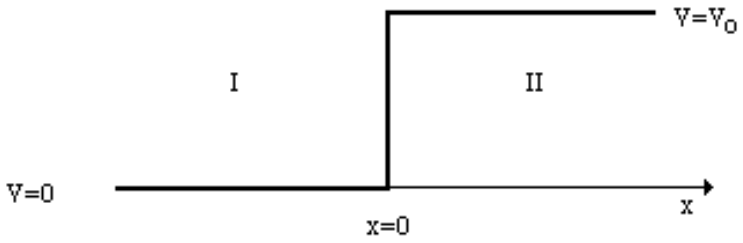


Figure 6.4: Finite Step barrier

Since there are no infinite jumps in the potential energy function, both $\psi(x)$ and $d\psi(x)/dx$ are continuous everywhere. We have two distinct regions to consider, which are labeled I and II in Figure 6.4.

In region I, $V(x) = 0$ and the Schrodinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I(x)}{dx^2} = E\psi_I(x) \tag{6.107}$$

Let us define k by

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \tag{6.108}$$

There are two possible solutions to this equation in region I, namely,

$$\psi_I(x) = e^{\pm ikx} \tag{6.109}$$

each with energy

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \tag{6.110}$$

The most general solution is a linear combination of the two possible solutions

$$\psi_I(x) = Ae^{+ikx} + Be^{-ikx} \tag{6.111}$$

This is a linear combination of two waves. Since this is an energy eigenstate it has a time dependence

$$e^{-iEt/\hbar} \tag{6.112}$$

If we insert the time dependence we have

$$\psi_I(x,t) = Ae^{+i(kx-\omega t)} + Be^{-i(kx+\omega t)} \quad , \quad \omega = \frac{E}{\hbar} \tag{6.113}$$

The first term is a traveling wave moving in the $+x$ direction with phase velocity ω/k and the second term is a traveling wave moving in the $-x$ direction with a phase velocity ω/k.

We can think of this solution in region I as representing an incident wave moving

in the $+x$ direction with amplitude A and a reflected wave moving in the $-x$ direction with an amplitude B. The reflection takes place at the potential discontinuity much the same as a reflection occurs for light on the interface between air and glass. We will show this clearly using wave packets later in this section. The wave packets will also allow us to relate these results to particle motion.

In region II, $V(x) = V_0$ and Schrodinger's equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}(x)}{dx^2} = (E - V_0)\psi_{II}(x) \tag{6.114}$$

If we define

$$\gamma^2 = \frac{2m}{\hbar^2}(E - V_0) \tag{6.115}$$

We then again have two possible solutions

$$\psi_{II}(x) = e^{\pm i\gamma x} \tag{6.116}$$

Two cases arise

$E > V_0 \rightarrow \gamma$ is real $\rightarrow$ traveling wave solutions
$E < V_0 \rightarrow \gamma$ is imaginary$\rightarrow$ real exponential solutions

We first consider $E > V_0$, γ real. We then have the general solution in region II

$$\psi_{II}(x) = Ce^{+i\gamma x} + De^{-i\gamma x} \tag{6.117}$$

Now the actual physical experiment we are considering must always restrict the possibilities (always remember that we are physicists and not just solving an abstract mathematical problem).

If we assume the incident wave is traveling in the $+x$ direction in region I (coming in from $x = -\infty$), then the existence of a wave in region I traveling in the $-x$ direction makes sense. It is a reflected wave (reflection occurs at $x = 0$). In region II, however, we cannot physically have a wave traveling in the $-x$ direction (what is its source?). We can, however, have a wave traveling in the $+x$ direction. It is the transmitted wave (again think of an analogy to an air-glass interface).

So, on physical grounds, we will assume that $D = 0$. We then have the solutions for $E > V_0$, γ real

$$\psi_I(x) = Ae^{+ikx} + Be^{-ikx} \qquad x < 0 \quad , \quad k^2 = \frac{2m}{\hbar^2}E \tag{6.118}$$

$$\psi_{II}(x) = Ce^{+i\gamma x} \qquad x > 0 \quad , \quad \gamma^2 = \frac{2m}{\hbar^2}(E - V_0) \tag{6.119}$$

We now use the continuity conditions at $x = 0$.

$$\psi_I(0) = \psi_{II}(0) \text{ and } \frac{d\psi_I(0)}{dx} = \frac{d\psi_{II}(0)}{dx} \tag{6.120}$$

which give the equations

$$A + B = C \text{ and } ik(A - B) = i\gamma C \tag{6.121}$$

which imply the solutions

$$\frac{C}{A} = \frac{2k}{k+\gamma} \text{ and } \frac{B}{A} = \frac{k-\gamma}{k+\gamma} \tag{6.122}$$

We can see the physical content by looking more deeply at the probability ideas associated with the Schrodinger equation. For a *one-particle state*, the Schrodinger equation formalism says that

$$\int_\Omega \psi^*(x)\psi(x)\, d^3x = \text{ probability particle is located in volume } \Omega \tag{6.123}$$

The time rate of change of this probability is

$$\begin{aligned}\frac{\partial}{\partial t}\int_\Omega \psi^*(x)\psi(x)\, d^3x &= \int_\Omega \left[\psi^* \frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi^*}{\partial t}\right] d^3x \\ &= \frac{i\hbar}{2m}\int_\Omega \left[\psi^* \nabla^2\psi - \psi\nabla^2\psi^*\right] d^3x \\ &= \frac{i\hbar}{2m}\int_\Omega \nabla\cdot\left[\psi^*\nabla\psi - \psi\nabla\psi^*\right] d^3x \end{aligned} \tag{6.124}$$

where we have used the time-dependent Schrodinger equation and its complex conjugate

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V(\vec{x})\psi = i\hbar\frac{\partial\psi}{\partial t} \text{ and } -\frac{\hbar^2}{2m}\nabla^2\psi^* + V(\vec{x})\psi^* = -i\hbar\frac{\partial\psi^*}{\partial t} \tag{6.125}$$

Since the volume Ω is arbitrary, this (8.128) implies a continuity equation of the form

$$\frac{\partial}{\partial t}|\psi(\vec{x},t)|^2 + \nabla\cdot\vec{J}(\vec{x},t) = 0 \tag{6.126}$$

where

$$\vec{J}(\vec{x},t) = -\frac{i\hbar}{2m}\left[\psi^*\nabla\psi - \psi\nabla\psi^*\right] \tag{6.127}$$

is called the *probability flux vector or the probability current density.*

Using the barrier solutions for $E > V_0$, γ real, we have (for x-components)

$$\begin{aligned} J_I(x,t) &= \frac{i\hbar}{2m}(A^*e^{-ikx} + B^*e^{ikx})(ikAe^{ikx} - ikBe^{-ikx}) \\ &\quad - \frac{i\hbar}{2m}(Ae^{ikx} + Be^{-ikx})(ikA^*e^{-ikx} - ikB^*e^{ikx}) \\ &= \frac{\hbar k}{m}\left[|A|^2 - |B|^2\right] = J_{I+} - J_{I-} \end{aligned} \tag{6.128}$$

$$\begin{aligned} J_{II}(x,t) &= \frac{i\hbar}{2m}\left[(C^*e^{-i\gamma x})(i\gamma Ce^{i\gamma x}) - (Ce^{i\gamma x})(-i\gamma C^*e^{-i\gamma x})\right] \\ &= \frac{\hbar\gamma}{m}|C|^2 = J_{II+} \end{aligned} \tag{6.129}$$

We then define the reflection and transmission probabilities in terms of the currents as

$$R = \frac{J_{I-}}{J_{I+}} = \frac{|B|^2}{|A|^2} \text{ and } T = \frac{J_{II+}}{J_{1+}} = \frac{\gamma}{k}\frac{|C|^2}{|A|^2} \tag{6.130}$$

Now, physically, we identify

$|A|^2$ = incident beam intensity
$|B|^2$ = reflected beam intensity
$|C|^2$ = transmitted beam intensity

and therefore

R = probability that an incident wave will be reflected
T = probability that an incident wave will be transmitted

We find

$$R = \frac{(k-\gamma)^2}{(k+\gamma)^2} \text{ and } T = \frac{4k\gamma}{(k+\gamma)^2} \tag{6.131}$$

Notice that

$$R + T = 1 \tag{6.132}$$

which makes physical sense since the total probability for the wave to go somewhere must equal one.

Before proceeding to the case $E < V_0$, let us recast the case $E > V_0$ in terms of the wave packet formalism we derived earlier. This will allow us to make particle interpretations.

Remember we can construct a wavepacket from any traveling wave solution by generating a linear combination of the traveling wave using only a restricted range of momentum values. In particular, the incident wave packet is a linear combination of incident waves

$$\psi_{inc}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p)A(p)e^{i(px-Et)/\hbar} \text{ where } E = \frac{p^2}{2m},\ p = \hbar k \tag{6.133}$$

Similarly, the reflected wave packet is a linear combination of reflected waves

$$\psi_{refl}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p)B(p)e^{-i(px+Et)/\hbar} \text{ where } E = \frac{p^2}{2m},\ p = \hbar k \tag{6.134}$$

We assume, for simplicity, that $f(p)$ is a real function that is nonzero only in a limited range of p-values and has a maximum near $p = p_0$. We also choose $A = 1$. Our solution indicates that $B(p)$ is a real function also and thus does not contribute to the phase of the integrand, where the phase is the part of the integrand of the form

$$e^{i\,phase} \tag{6.135}$$

As we discussed earlier, the integrals are nonzero only when the phase of the integrand is near an extremum at the same place that the rest of the integrand is large(near $p = p_0$ by assumption since we can always choose the form of $f(p)$ so that it dominates). Otherwise, the exponential terms oscillates so rapidly ($\hbar$ is so small) that the integral averages to zero. This is called the stationary phase condition.

For the incident and reflected waves the stationary phase extremum argument gives

$$\frac{\partial}{\partial p}\left(px - \frac{p^2}{2m}t\right)_{p=p_0} = 0 = x - \frac{p_0}{m}t = x - v_0 t \qquad \text{incident wave} \tag{6.136}$$

$$\frac{\partial}{\partial p}\left(px + \frac{p^2}{2m}t\right)_{p=p_0} = 0 = x + \frac{p_0}{m}t = x + v_0 t \qquad \text{reflected wave} \tag{6.137}$$

These equations tell us the location of the maximum of the wave packet (or pulse) given by $|\psi(x,t)|^2$ in space as a function of time. We have

$$x_{inc} = v_0 t \quad , \quad t < 0 \tag{6.138}$$

which says that the incident packet or particle arrives at $x = 0$ at $t = 0$. In this model, a *particle is a localized lump of energy and momentum* where the localization is described by $|\psi(x,t)|^2$.

Note that we obtain the correct kinematic relation of a free particle as we should.

Similarly, for the reflected wave we have

$$x_{refl} = -v_0 t \quad , \quad t > 0 \tag{6.139}$$

which says that the reflected packet or particle leaves $x = 0$ at $t = 0$.

For a finite step barrier of this kind, the reflected packet leaves at the *same time* that the incident packet arrives.

The transmitted packet is given by

$$\psi_{trans}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p) C(p) e^{i(px - Et)/\hbar} \tag{6.140}$$

$$\text{where } E - V_0 = \frac{p^2}{2m} \quad , \quad p = \hbar\gamma$$

Since $C(p)$ is a real function, the stationary phase extremum argument gives

$$\frac{\partial}{\partial p}\left(\gamma x - \frac{p^2}{2m}t - V_0 t\right)_{p=p_0} = 0 = x\frac{\partial\gamma}{\partial p}\bigg|_{p=p_0} - \frac{p_0}{m}t$$
$$= x\,\frac{\partial\sqrt{p^2-2mV_0}}{\partial p}\bigg|_{p=p_0} - \frac{p_0}{m}t$$
$$= x\frac{p_0}{\sqrt{p_0^2-2mV_0}} - \frac{p_0}{m}t$$

or

$$x_{trans} = \frac{\gamma_0}{m}t = \tilde{v}_0 t \qquad \text{transmitted wave} \tag{6.141}$$

This says that the transmitted wave moves in the region $x > 0$, $t > 0$ with a speed

$$\tilde{v}_0 = \frac{\gamma_0}{m} < \frac{p_0}{m} \tag{6.142}$$

as it should.

Summarizing, this wave packet analysis says that the probability amplitude for the particle being at a point in space and time is completely localized in the incident wave for $t < 0$ as the packet travels towards $x = 0$ from the left (it arrives at $x = 0$ at $t = 0$). For $t > 0$ the packet then splits into two packets, namely the reflected and transmitted packets, such that the probability amplitude for the particle is now localized in two regions. The localized packet traveling towards the left represents the probability of the particle being reflected and the localized packet traveling towards the right represents the probability of the particle being transmitted.

Now what happens if $E = p^2/2m < V_0$? We then have the solutions (let $A = 1$)

$$x < 0 \qquad \psi_I = e^{ikx} + Re^{-ikx} \tag{6.143}$$

$$x > 0 \qquad \psi_{II} = Se^{-\beta x} \tag{6.144}$$

where $\hbar k = \sqrt{2mE}$ and $\hbar\beta = \sqrt{2m(V_0 - E)}$. We have excluded the mathematical solution $e^{\beta x}$ for $x > 0$ because it would cause the integral

$$\int_{-\infty}^{\infty} |\psi|^2\, dx \tag{6.145}$$

to diverge, which means it is a *non-normalizable* or *unphysical solution*. This is an example of a solution not being *well-behaved*. At $x = 0$, we have the continuity conditions which give

$$1 + R = S \text{ and } 1 - R = -\frac{\beta}{ik}S \text{ or}$$
$$S = \frac{2ik}{ik-\beta} \text{ and } R = \frac{ik+\beta}{ik-\beta} \tag{6.146}$$

Since the solution for $x > 0$ is not a traveling wave, there is no transmitted packet. The incident wave packet analysis is the same as before. The reflected wave packet analysis is different because $R(p)$ is not a real function. In fact,

$$|R|^2 = 1 \rightarrow R = e^{i\phi} = \frac{ip + \sqrt{2mV_0 - p^2}}{ip - \sqrt{2mV_0 - p^2}}$$
$$= \frac{ip + \sqrt{2mV_0 - p^2}}{ip - \sqrt{2mV_0 - p^2}} \frac{ip + \sqrt{2mV_0 - p^2}}{ip + \sqrt{2mV_0 - p^2}} \tag{6.147}$$

or

$$e^{i\phi} = \frac{p^2 - mV_0 - ip\sqrt{2mV_0 - p^2}}{mV_0} = \cos\phi + i\sin\phi \tag{6.148}$$

with

$$\cos\phi = \frac{p^2 - mV_0}{mV_0} \text{ and } \sin\phi = -\frac{p\sqrt{2mV_0 - p^2}}{mV_0} \tag{6.149}$$

Therefore, we have as wave packet representations

$$\psi_{inc}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p) f(p) e^{i(px - Et)/\hbar} \text{ where } E = \frac{p^2}{2m},\ p = \hbar k \tag{6.150}$$

$$\psi_{refl}(x,t) = \int_0^\infty \frac{dp}{2\pi\hbar} f(p) f(p) e^{-i(px + Et - \hbar\phi(p))/\hbar} \tag{6.151}$$
$$\text{where } E = \frac{p^2}{2m},\ p = \hbar k$$

The stationary phase argument implies

$$x = \frac{p_0}{m} t = v_0 t \text{ for the incident packet} \tag{6.152}$$

which says that the incident packet arrives at $x = 0$ at $t = 0$ and

$$x = -\frac{p_0}{m} t + \hbar \left.\frac{\partial\phi(p)}{\partial p}\right|_{p=p_0} \text{ for the reflected packet} \tag{6.153}$$

This says that the reflected wave packet leaves $x = 0$ at

$$t = t_{delay} = \frac{\hbar m}{p_0} \left.\frac{\partial\phi(p)}{\partial p}\right|_{p=p_0} \tag{6.154}$$

which is NOT the same time that the incident packet arrived!

Continuing we have

$$\frac{\partial \cos\phi(p)}{\partial p} = -\sin\phi(p) \frac{\partial\phi(p)}{\partial p} = \frac{2p}{mV_0} \tag{6.155}$$

$$\frac{\partial\phi(p)}{\partial p} = \frac{2}{\sqrt{2mV_0 - p_0^2}} \tag{6.156}$$

which gives

$$t_{delay} = \frac{m\hbar}{p_0}\frac{2}{\sqrt{2mV_0 - p_0^2}} = \frac{2m}{p_0}\frac{1}{\beta} \tag{6.157}$$

Now, we found that

$$\psi_{II}(x) = Se^{-\beta x} \tag{6.158}$$

which says that the probability amplitude is significantly different from zero up to a distance $d = 1/\beta$ approximately. In other words,

$$\psi_{II}(x = 1/\beta) = \frac{1}{e}\psi_{II}(x = 0) \rightarrow |\psi_{II}(x = 1/\beta)|^2 \approx \frac{1}{7}|\psi_{II}(x = 0)|^2 \tag{6.159}$$

Therefore, to a good approximation, we have the striking result $t_{delay} = 2d/v_0$. What does this mean? It seems as if there is a nonzero probability for finding the *particle* in the classically forbidden region $x > 0$, where K = kinetic energy $= E - V_0 < 0$ since $\psi_{II}(x) = e^{-\beta x} \neq 0$ there. Since K cannot be less than zero, it seems like we must have a violation of conservation of energy if the particle were actually found in the region $x > 0$. This is NOT the case.

If we observe the particle in this forbidden region, then it will no longer be in a state with $E < V_0$. The act of measuring the location of the particle must necessarily introduce an uncertainty in p and hence in E. The particle seems to have an appreciable probability to exist in the region up to a distance $d = 1/\beta$. If we observe the particle on the right, we have then localized it such that $\Delta x \approx 1/\beta$. This says that we have introduced an uncertainty in p

$$\Delta p \geq \frac{\hbar}{\Delta x} \approx \hbar\beta \tag{6.160}$$

and a corresponding energy uncertainty

$$\Delta E \approx \frac{(\Delta p)^2}{2m} \geq \frac{\hbar^2\beta^2}{2m} = V_0 - E \tag{6.161}$$

This implies that E is now uncertain enough that we can no longer claim that energy conservation has been violated! Quantum mechanics has a way of covering its own tracks!

6.5.2 Tunneling

We now change the potential energy function so that the barrier (step) is not infinitely thick (which prevented a traveling wave from existing for $E < V_0$, $x > 0$. The new potential energy function is shown in Figure 6.5 below.

For $E > V_0$ the results are similar to infinitely thick barrier and no new physical ideas appear. For $E < V_0$, however, we get some very interesting new physical

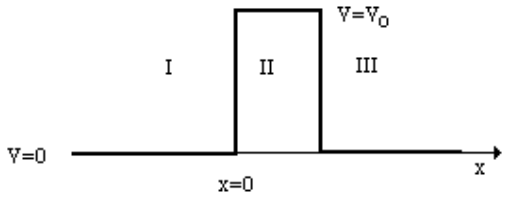


Figure 6.5: Finite Step - Finite Width barrier

results. As we shall see, it turns out that a real traveling wave can appear on the other side of the barrier (even though there are no sources on that side) in this case. This is called quantum tunneling. Let us see how it works.

We have three regions I, II and III to consider as shown in Figure 6.5.

We get the solutions

$$x \le 0 \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I$$
$$\psi_I = A_1 e^{ikx} + B_1 e^{-ikx} \quad , \quad E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \ , \ k \text{ real} \tag{6.162}$$

$$0 \le x \le a \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} + V_0\psi_{II} = E\psi_{II}$$
$$\psi_{II} = Ce^{\gamma x} + De^{-\gamma x} \quad , \quad V_0 - E = \frac{p^2}{2m} = \frac{\hbar^2\gamma^2}{2m} \ , \ \gamma \text{ real} \tag{6.163}$$

$$x \ge a \qquad -\frac{\hbar^2}{2m}\frac{d^2\psi_{III}}{dx^2} = E\psi_{III}$$
$$\psi_{III} = A_2 e^{ikx} + B_2 e^{-ikx} \quad , \quad E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \ , \ k \text{ real} \tag{6.164}$$

The probability current does not vanish at $x = \pm\infty$, which implies that we must assume the existence of distant sources and sinks of probability(or particles).

We have two sets of continuity equations (at $x = 0$ and $x = a$). At $x = 0$ we get

$$\psi_I(0) = \psi_{II}(0) \to A_1 + B_1 = C + D \tag{6.165}$$
$$\frac{d\psi_I(0)}{dx}\frac{d\psi_{II}(0)}{dx} \to ik(A_1 - B_1) = \gamma(C - D) \tag{6.166}$$

and at $x = a$ we get

$$\psi_{II}(a) = \psi_{III}(a) \to Ce^{\gamma a} + De^{-\gamma a} = A_2 e^{ika} + B_2 e^{-ika} \tag{6.167}$$
$$\frac{d\psi_{II}(a)}{dx}\frac{d\psi_{III}(a)}{dx} \to \gamma(Ce^{\gamma a} - De^{-\gamma a}) = ik(A_2 e^{ika} - B_2 e^{-ika}) \tag{6.168}$$

We can restate these equations in matrix form. At $x = 0$ we get

$$\begin{pmatrix} 1 & 1 \\ ik & -ik \end{pmatrix}\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_1\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \gamma & -\gamma \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_2\begin{pmatrix} C \\ D \end{pmatrix} \tag{6.169}$$

and at $x = a$ we get

$$\begin{aligned}\begin{pmatrix} e^{\gamma a} & e^{-\gamma a} \\ \gamma e^{\gamma a} & -\gamma e^{-\gamma a} \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} &= \hat{M}_3\begin{pmatrix} C \\ D \end{pmatrix} \\ &= \begin{pmatrix} e^{ika} & e^{-ika} \\ ike^{ika} & -ike^{-ika} \end{pmatrix}\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} = \hat{M}_4\begin{pmatrix} A_2 \\ B_2 \end{pmatrix}\end{aligned} \tag{6.170}$$

The transmission/reflection properties of the barrier are given by the coefficients A_1, B_1, A_2, B_2, which we assume are related by a transfer matrix $\hat{Y}$, where

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{Y}\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \tag{6.171}$$

We then have

$$\hat{M}_1\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_2\begin{pmatrix} C \\ D \end{pmatrix} \rightarrow \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_1^{-1}\hat{M}_2\begin{pmatrix} C \\ D \end{pmatrix} \tag{6.172}$$

$$\hat{M}_3\begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_4\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \rightarrow \begin{pmatrix} C \\ D \end{pmatrix} = \hat{M}_3^{-1}\hat{M}_4\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \tag{6.173}$$

or

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \hat{M}_1^{-1}\hat{M}_2\hat{M}_3^{-1}\hat{M}_4\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \tag{6.174}$$

and thus

$$\hat{Y} = \hat{M}_1^{-1}\hat{M}_2\hat{M}_3^{-1}\hat{M}_4 \tag{6.175}$$

A great deal of algebra gives

$$Y_{11} = e^{ika}\left[\cosh\gamma a + \frac{1}{2}i\sinh\gamma a\left(\frac{\gamma}{k} - \frac{k}{\gamma}\right)\right] \tag{6.176}$$

$$Y_{12} = \frac{1}{2}e^{-ika}\sinh\gamma a\left(\frac{\gamma}{k} + \frac{k}{\gamma}\right) \tag{6.177}$$

$$Y_{21} = -\frac{1}{2}ie^{ika}\sinh\gamma a\left(\frac{\gamma}{k} + \frac{k}{\gamma}\right) \tag{6.178}$$

$$Y_{22} = e^{-ika}\left[\cosh\gamma a - \frac{1}{2}i\sinh\gamma a\left(\frac{\gamma}{k} - \frac{k}{\gamma}\right)\right] \tag{6.179}$$

If we have sources only on the left of the barrier, then we must choose $B_2 = 0$. The currents are

$$\begin{aligned}J_{LEFT}(x,t) &= \frac{\hbar k}{m}\left[|A_1|^2 - |B_1|^2\right] \\ &= J_{LEFT+}(x,t) - J_{LEFT-}(x,t) \qquad x < 0\end{aligned} \tag{6.180}$$

$$J_{RIGHT}(x,t) = J_{RIGHT+}(x,t) = \frac{\hbar k}{m}|A_2|^2 \qquad x > 0 \tag{6.181}$$

The reflection and transmission probabilities are given by

$$R = \frac{|B_1|^2}{|A_1|^2} = \frac{|Y_{21}|^2}{|Y_{11}|^2} \qquad , \qquad T = \frac{|A_2|^2}{|A_1|^2} = \frac{1}{|Y_{11}|^2} \tag{6.182}$$

Algebra shows $R+T = 1$ as it must in order to conserve probability (or particles). Evaluating the expression for T we get

$$T = \frac{1}{1 + \frac{V_0^2 \sinh^2 \gamma a}{4E(V_0 - e)}} \quad , \quad \gamma^2 = \frac{2m}{\hbar^2}(V_0 - E) \tag{6.183}$$

The fact that $T > 0$ for $E < V_0$ implies the existence of *tunneling*. The probability amplitude leaks through the barrier.

It is important to realize that the fact that $T > 0$, *DOES NOT* say that particles or wave packets passed through the barrier. No measurement can be done on the system that will allow us to observe a particle in the region $0 < x < a$ with $E < V_0$, since this would violate energy conservation.

It is *ONLY* probability that is leaking through. If this causes the probability amplitude and hence the probability to be nonzero on the other side of the barrier, than it *must be possible for us to observe* the particle on the other side, i.e., we can observe the particle on the left side of the barrier with $E < V_0$, but we can never observe it in the region of the barrier with $E < V_0$. *That is what is being said here.*

6.5.3 Bound States

Infinite Square Well

We now consider the potential energy function

$$V(x) = \begin{cases} 0 & -\frac{a}{2} \le x \le \frac{a}{2} \\ \infty & |x| \ge \frac{a}{2} \end{cases} \tag{6.184}$$

This is the so-called *infinite square well* shown in Figure 6.6 below.

This is an example of a potential that is infinite in an extended region. Therefore, we must require that the wave function $\psi(x) = 0$ in these regions or the Schrodinger equation makes no sense mathematically. In this case we have

$$\psi_I(x) = 0 \text{ and } \psi_{III}(x) = 0 \tag{6.185}$$

In region II we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} = E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \tag{6.186}$$

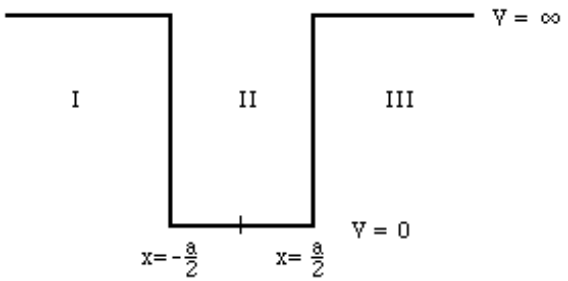


Figure 6.6: Infinite Square Well

and

$$\psi_{II}(x) = Ae^{ikx} + Be^{-ikx} \tag{6.187}$$

The continuity of the wavefunction at $x = \pm a/2$ says that we must have

$$\psi_{II}\left(-\frac{a}{2}\right) = Ae^{-ika/2} + Be^{ika/2} = 0 \tag{6.188}$$

$$\psi_{II}\left(\frac{a}{2}\right) = Ae^{ika/2} + Be^{-ika/2} = 0 \tag{6.189}$$

which imply that

$$\frac{B}{A} = -e^{-ika} = -e^{ika} \tag{6.190}$$

This is an equation for the allowed values(values corresponding to a valid solution) of k. This equation is

$$e^{2ika} = 1 \tag{6.191}$$

The allowed values of k form a discrete spectrum of energy eigenvalues (quantized energies) given by

$$2k_n a = 2n\pi \rightarrow k_n = \frac{n\pi}{a} \rightarrow E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{n^2\pi^2\hbar^2}{2ma^2}\ ,\ n = 1, 2, 3, 4, \ldots \tag{6.192}$$

The corresponding wave functions are

$$\begin{aligned}\psi_{II}^{(n)}(x) &= A_n(e^{ik_n x} - e^{-ik_n a}e^{-ik_n x}) \\ &= A_n e^{-ik_n a/2}(e^{ik_n(x+a/2)} - e^{-ik_n(x+a/2)}) \\ &= \tilde{A}_n \sin k_n(x + a/2)\end{aligned} \tag{6.193}$$

where $\tilde{A}_n$ is determined by the normalization condition

$$\int_{-a/2}^{a/2} |\psi_n(x)|^2\, dx = 1 \tag{6.194}$$

Substituting the value of k_n from (6.196) into (6.197) we get

$$\begin{aligned}\psi_{II}^{(1)}(x) &= \tilde{A}_1 \sin k_1(x + a/2) = \tilde{A}_1 \sin \pi(x + a/2)/a \\ &= \tilde{A}_1 \sin(\pi x/a + \pi/2) = \tilde{A}_1 \cos(\pi x/a)\end{aligned} \tag{6.195}$$

$$\psi_{II}^{(2)}(x) = \tilde{A}_2 \sin k_2(x + a/2) = \tilde{A}_2 \sin 2\pi(x + a/2)/a$$
$$= \tilde{A}_2 \sin(2\pi x/a + \pi) = \tilde{A}_2 \sin(2\pi x/a) \tag{6.196}$$

$$\psi_{II}^{(3)}(x) = \tilde{A}_3 \sin k_3(x + a/2) = \tilde{A}_3 \sin 3\pi(x + a/2)/a$$
$$= \tilde{A}_3 \sin(3\pi x/a + 3\pi/2) = \tilde{A}_3 \cos(3\pi x/a) \tag{6.197}$$

or

$$\psi_{II}(x) = \begin{cases} \sin(n\pi x/a) & \text{n even} \\ \cos(n\pi x/a) & \text{n odd} \end{cases} \quad n = 1, 2, 3, 4, \ldots \tag{6.198}$$

We have mathematically solved the ordinary differential equation problem, now what is the physical meaning of these results?

We find a *discrete spectrum* of allowed energies corresponding to bound states of the Hamiltonian; the energy is quantized. Bound states designate states which are localized in space, i.e., the probability is large only over restricted regions of space and goes to zero far from the potential region.

The *lowest* energy value or lowest energy *level* or *ground state energy* is

$$E_1 = \frac{\pi^2 \hbar^2}{2ma^2} > 0 \tag{6.199}$$

with

$$\psi_1(x) = \begin{cases} A\cos(n\pi x/a) & |x| \le \frac{a}{2} \\ 0 & |x| \ge \frac{a}{2} \end{cases} \quad n = 1, 2, 3, 4, \ldots \tag{6.200}$$

This minimum energy is not zero because of the Heisenberg uncertainty principle. Since the particle has a nonzero amplitude for being in the well, we say that it is localized such that $\Delta x \approx a$ and thus

$$\Delta p \ge \frac{\hbar}{\Delta x} \approx \frac{\hbar}{a} \tag{6.201}$$

This says that the kinetic energy (or energy in this case because the potential energy equals zero in region II) must have a minimum value given approximately by

$$E_{min} = K_{min} \approx \frac{(\Delta p)^2}{2m} \approx \frac{\hbar^2}{2ma^2} \tag{6.202}$$

Let us look more closely at the wave functions.

The integer $n-1$ corresponds to the number of nodes (zeros) of the wave function (other than the well edges).

They also have the property

$$\psi(-x) = \psi(x) \text{ for n odd and } \psi(-x) = -\psi(x) \text{ for n even} \tag{6.203}$$

The above discrete transformation of the wave function corresponds to the *parity* operator $\hat{\mathcal{P}}$ where we have

$$\hat{\mathcal{P}}\psi(x) = \psi(-x) = \psi(x) \text{ means even parity} \tag{6.204}$$

$$\hat{\mathcal{P}}\psi(x) = \psi(-x) = -\psi(x) \text{ means odd parity} \tag{6.205}$$

Let us look more generally at the parity operation. Suppose that the potential energy function obeys the rule $V(\vec{x}) = V(-\vec{x})$ and let $\psi(\vec{x})$ be a solution of the Schrodinger equation with energy E

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x})\right)\psi(\vec{x}) = E\psi(\vec{x}) \tag{6.206}$$

Now let $\vec{x} \to -\vec{x}$ to get the equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(-\vec{x})\right)\psi(-\vec{x}) = E\psi(-\vec{x}) \tag{6.207}$$

or

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x})\right)\psi(-\vec{x}) = E\psi(-\vec{x}) \tag{6.208}$$

This says that, if $\psi(\vec{x})$ is a solution of the Schrodinger equation with energy E, then $\psi(-\vec{x})$ is also a solution of the Schrodinger equation with the same energy E. This says that the combinations

$$\psi(\vec{x}) \pm \psi(-\vec{x}) \tag{6.209}$$

are also solutions of the Schrodinger equation with the same energy E. Now

$$\psi(\vec{x}) + \psi(-\vec{x}) \to \text{ an even parity solution}$$
$$\psi(\vec{x}) - \psi(-\vec{x}) \to \text{ an odd parity solution}$$

This says that if $V(\vec{x}) = V(-\vec{x})$ (a symmetric potential), then we can always choose solutions that have a definite parity (even or odd).

We formally define the *parity operator* by the relation

$$\langle\vec{x}|\,\hat{\mathcal{P}}\,|\psi\rangle = \langle -\vec{x}\,|\,\psi\rangle \tag{6.210}$$

Since

$$\langle\vec{x}|\,\hat{\mathcal{P}}^2\,|\psi\rangle = \langle -\vec{x}|\,\hat{\mathcal{P}}\,|\psi\rangle = \langle\vec{x}\,|\,\psi\rangle \tag{6.211}$$

we must have

$$\hat{\mathcal{P}}^2 = \hat{I} \tag{6.212}$$

which means the eigenvalues of $\hat{\mathcal{P}}$ are ± 1 as we indicated earlier. This also says that $\hat{\mathcal{P}}^{-1} = \hat{\mathcal{P}}$

We can show $[\hat{H}, \hat{\mathcal{P}}] = 0$ for symmetric potentials by

$$\hat{\mathcal{P}}\hat{H}|E\rangle = \hat{\mathcal{P}}E|E\rangle = E\hat{\mathcal{P}}|E\rangle = \pm E|E\rangle$$
$$\hat{H}\hat{\mathcal{P}}|E\rangle = \pm\hat{H}|E\rangle = \pm E|E\rangle$$
$$(\hat{\mathcal{P}}\hat{H} - \hat{H}\hat{\mathcal{P}})|E\rangle = 0$$
$$[\hat{H}, \hat{\mathcal{P}}] = 0$$

since $|E\rangle$ is an arbitrary state. As we saw earlier, this commutator relationship says that

$$\hat{H}\hat{\mathcal{P}} = \hat{\mathcal{P}}\hat{H}$$
$$\hat{\mathcal{P}}\hat{H}\hat{\mathcal{P}} = \hat{\mathcal{P}}^2\hat{H} = \hat{H}$$
$$\hat{\mathcal{P}}^{-1}\hat{H}\hat{\mathcal{P}} = \hat{H} \tag{6.213}$$

which means that $\hat{H}$ is invariant under the $\hat{\mathcal{P}}$ transformation. We have used $\hat{\mathcal{P}}^2 = \hat{I}$ in this derivation. It also says that

$$\hat{H}(\hat{\mathcal{P}}|E\rangle) = \hat{\mathcal{P}}\hat{H}|E\rangle = E(\hat{\mathcal{P}}|E\rangle) \tag{6.214}$$

or $\hat{\mathcal{P}}|E\rangle$ is an eigenstate of $\hat{H}$ with energy E as we stated above. The concept of parity invariance and the fact that $\hat{H}$ and $\hat{\mathcal{P}}$share a common set of eigenfunctions can greatly simplify the solution of the Schrodinger equation in many cases.

The Finite Square Well

We now consider the potential energy function

$$V(x) = \begin{cases} -V_0 & |x| \le \frac{a}{2} \\ 0 & |x| \ge \frac{a}{2} \end{cases} \tag{6.215}$$

This is the so-called *finite square well* (in one dimension) and it is shown in Figure 6.7 below.

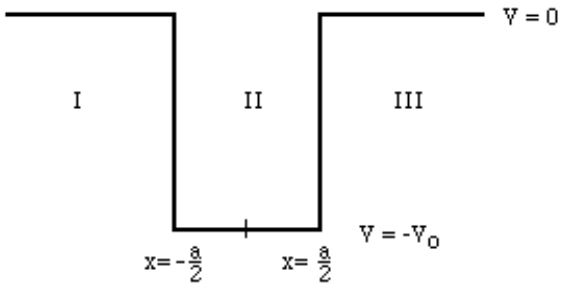


Figure 6.7: Finite Square Well

The solutions are:

Region I : $x < -\frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I \ , 0 \geq E \geq -V_0 \ , \hbar^2 k^2 = 2m|E| \ , E = -|E|$$

with solutions

$$\psi_I(x) = Ae^{-kx} + Be^{kx} \tag{6.216}$$

Since $x = -\infty$ is included in this region, we must exclude the e^{-kx} term by choosing $A = 0$, which gives

$$\psi_I(x) = Be^{kx} \qquad x < -\frac{a}{2} \tag{6.217}$$

Region II: $-\frac{a}{2} \leq x \leq \frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} - V_0\psi_{II} = E\psi_{II} \ , 0 \geq E \geq -V_0 \ , p^2 = 2m(V_0 - |E|) \ , E = -|E|$$

with solutions

$$\psi_{II}(x) = Ce^{ipx/\hbar} + De^{-ipx/\hbar} \tag{6.218}$$

Region III: $x > \frac{a}{2}$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{III}}{dx^2} = E\psi_{III} \ , 0 \geq E \geq -V_0 \ , \hbar^2 k^2 = 2m|E| \ , E = -|E|$$

with solutions

$$\psi_{III}(x) = Fe^{kx} + Ge^{-kx} \tag{6.219}$$

Since $x = +\infty$ is included in this region, we must exclude the e^{kx} term by choosing $F = 0$, which gives

$$\psi_{III}(x) = Ge^{-kx} \qquad x > \frac{a}{2} \tag{6.220}$$

These results represent a general solution to the problem. There seems to be 4 unknown constants, namely, B, C, D, and G. However, since $V(x) = V(-x)$, parity is conserved and we can choose even and odd solutions, or solutions of definite parity.

Even parity implies $\psi(x) = \psi(-x)$ or $G = B$ and $C = D$. This solution is

$$\psi(x) = \begin{cases} C\cos px/\hbar & |x| \leq \frac{a}{2} \\ Be^{-kx} & x > \frac{a}{2} \\ Be^{kx} & x < -\frac{a}{2} \end{cases} \tag{6.221}$$

Odd parity implies $\psi(x) = -\psi(-x)$ or $G = -B$ and $D = -C$. This solution is

$$\psi(x) = \begin{cases} C\sin px/\hbar & |x| \leq \frac{a}{2} \\ Be^{-kx} & x > \frac{a}{2} \\ -Be^{kx} & x < -\frac{a}{2} \end{cases} \tag{6.222}$$

Thus, by using parity we reduce the number of unknowns in the problem to two for each type of solution. We now impose the continuity conditions of the wave function and its derivative only at $x = a/2$ for both solutions. Since these are definite parity solutions the continuity condition at $x = -a/2$ will give no new information and is not needed.

Even Parity Results

$$C\cos pa/2\hbar = Be^{-ka/2} \text{ and } -\frac{p}{\hbar}C\sin pa/2\hbar = -kBe^{-ka/2} \tag{6.223}$$

or

$$\frac{B}{C} = e^{ka/2}\cos pa/2\hbar = \frac{p}{\hbar k}e^{ka/2}\sin pa/2\hbar \tag{6.224}$$

so that

$$p\tan pa/2\hbar = \hbar k \tag{6.225}$$

This last equation is a transcendental equation for E and its solutions determine the allowed E values for the even parity states for this potential energy function. These E values are the even parity energies or energy levels of a particle in the finite square well potential.

Odd Parity Results

$$C\sin pa/2\hbar = Be^{-ka/2} \text{ and } \frac{p}{\hbar}C\cos pa/2\hbar = -kBe^{-ka/2} \tag{6.226}$$

or

$$\frac{B}{C} = e^{ka/2}\sin pa/2\hbar = -\frac{p}{\hbar k}e^{ka/2}\cos pa/2\hbar \tag{6.227}$$

so that

$$p\cot pa/2\hbar = -\hbar k \tag{6.228}$$

Again, this last equation is a transcendental equation for E and its solutions determine the allowed E values for the odd parity states for this potential energy function. These E values are the odd parity energies or energy levels of a particle in the finite square well potential.

In general, at this stage of the solution, we must either devise a clever numerical or graphical trick to find the solutions of the transcendental equations or resort to a computer.

The first thing one should always do is change variables to get rid of as many extraneous constants as possible. In this case we let

$$\beta = ka = \frac{a}{\hbar}\sqrt{2m|E|} \quad , \quad \alpha = \gamma a = \frac{p}{\hbar}a = \frac{a}{\hbar}\sqrt{2m(V_0 - |E|)} \tag{6.229}$$

The first useful equation we can derive is

$$\alpha^2 + \beta^2 = \frac{2mV_0a^2}{\hbar^2} = \text{ constant for a given well} \tag{6.230}$$

This is the equation of a circle of radius

$$\sqrt{\frac{2mV_0a^2}{\hbar^2}} \tag{6.231}$$

With these new variables the two transcendental equations are

$$\beta = \alpha \tan\frac{\alpha}{2} \text{ for even parity and } \beta = -\alpha \cot\frac{\alpha}{2} \text{ for odd parity} \tag{6.232}$$

We can find solutions graphically by plotting as shown in Figure 6.8 below for the case (effectively a choice of the quantity V_0a^2)

$$\text{circle radius } = \sqrt{\frac{2mV_0a^2}{\hbar^2}} = \frac{5\pi}{2} \tag{6.233}$$

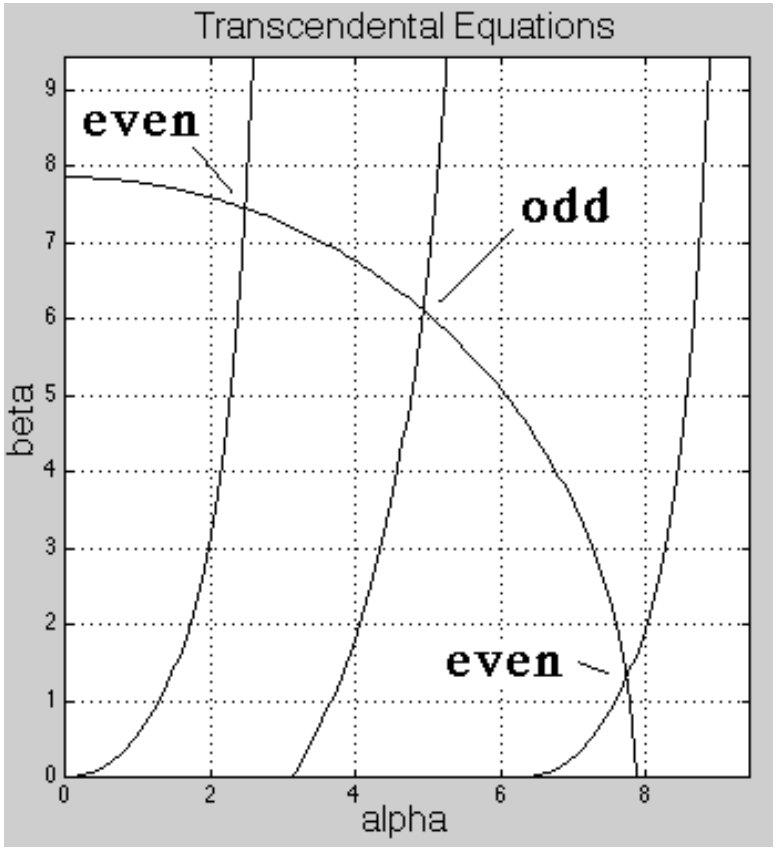


Figure 6.8: Solutions for $\sqrt{\frac{2mV_0a^2}{\hbar^2}} = \frac{5\pi}{2}$

The solutions correspond to the intersections of the circle (fixed for a given

well) and the curves represented by the two transcendental equations. They are shown in Figure 6.8.

For the choice of potential well shown in the figure we have 2 even parity solutions and 1 odd parity solution. These correspond to the allowed energy levels for this particular well and the corresponding wave functions and energies represent bound states of the well.

We can also do a straight numerical solution for even parity by rearranging the equations as follows:

$$\alpha^2 + \beta^2 = \frac{2mV_0a^2}{\hbar^2} \text{ and } \beta = \alpha \tan\frac{\alpha}{2}$$

$$\alpha^2\left(1 + \tan^2\frac{\alpha}{2}\right) = \frac{\alpha^2}{\cos^2\frac{\alpha}{2}} = \frac{2mV_0a^2}{\hbar^2} \tag{6.234}$$

$$\alpha^2 - \frac{25\pi^2}{4}\cos^2\frac{\alpha}{2} = f(\alpha) = 0 \tag{6.235}$$

The numerical solution of this equation can be carried out by any standard technique (Newton-Raphson method, for instance) for finding the zeros of the function $f(\alpha)$. For this case we get

$$\alpha = 2.4950 \text{ and } 7.1416 \tag{6.236}$$

which is clearly in agreement with the graphical result.

Transmission Resonances

As we have seen the spectrum of energies for the square well is made up of a finite number of bound-state levels. The most important feature of a bound-state wave function is that it is localized in space,i.e., it falls off exponentially as we move away from the well.

Another interesting feature of the square well has to do with the continuous part of its energy spectrum. All energy states with $E > 0$ are allowed. Again there are three regions to consider, labeled in the same way as earlier. We will again assume that we have a wave incident from the left (from $x = -\infty$) with unit intensity

$$\psi_I(x) = e^{ikx} + Be^{-ikx} \text{ where } E = \frac{\hbar^2k^2}{2m} \tag{6.237}$$

where B is the amplitude for the reflected wave.

In region II, which is over the well, we have the solution

$$\psi_{II}(x) = Ce^{ik_1x} + De^{-ik_1x} \text{ where } E + V_0 = \frac{\hbar^2k_1^2}{2m} \tag{6.238}$$

In region III, we only have a transmitted wave as the solution

$$\psi_{III}(x) = Fe^{ikx} \text{ where } E = \frac{\hbar^2 k^2}{2m} \tag{6.239}$$

Again, we must match the wave functions and their derivatives at $x = \pm a/2$. This gives us 4 equations for the 4 unknown coefficients. Usually an author now says that *lots of algebra* gives us the transmission coefficient and writes down the answer. Let us actually do it just once.

Continuity of the wave function and its derivative at $x = -a/2$ gives

$$\begin{aligned} e^{-ika/2} + Be^{ika/2} &= Ce^{-ik_1 a/2} + De^{ik_1 a/2} \\ ike^{-ika/2} - ikBe^{ika/2} &= ik_1 Ce^{-ik_1 a/2} - ik_1 De^{ik_1 a/2} \end{aligned}$$

Continuity of the wave function and its derivative at $x = +a/2$ gives

$$\begin{aligned} Fe^{ika/2} &= Ce^{ik_1 a/2} + De^{-ik_1 a/2} \\ ikFe^{ika/2} &= ik_1 Ce^{ik_1 a/2} - ik_1 De^{-ik_1 a/2} \end{aligned}$$

Then

$$\begin{aligned} ik_1 e^{-ika/2} + ik_1 Be^{ika/2} &= ik_1 Ce^{-ik_1 a/2} + ik_1 De^{ik_1 a/2} \\ ike^{-ika/2} - ikBe^{ika/2} &= ik_1 Ce^{-ik_1 a/2} - ik_1 De^{ik_1 a/2} \\ ik_1 Fe^{ika/2} &= ik_1 Ce^{ik_1 a/2} + ik_1 De^{-ik_1 a/2} \\ ikFe^{ika/2} &= ik_1 Ce^{ik_1 a/2} - ik_1 De^{-ik_1 a/2} \end{aligned}$$

Solving for C and D we get

$$\begin{aligned} 2ik_1 Ce^{-ik_1 a/2} &= i(k + k_1)e^{-ika/2} - i(k - k_1)Be^{ika/2} \\ 2ik_1 De^{ik_1 a/2} &= -i(k - k_1)e^{-ika/2} + i(k + k_1)Be^{ika/2} \\ 2ik_1 Ce^{ik_1 a/2} &= i(k + k_1)Fe^{ika/2} \\ 2ik_1 De^{-ik_1 a/2} &= -i(k - k_1)Fe^{ika/2} \end{aligned}$$

Rearranging we have

$$\begin{aligned} C &= \left(\tfrac{k+k_1}{2k_1}\right) Fe^{-ik_1 a/2} e^{ika/2} = e^{ik_1 a/2}\left[\left(\tfrac{k+k_1}{2k_1}\right) e^{-ika/2} - \left(\tfrac{k-k_1}{2k_1}\right) Be^{ika/2}\right] \\ D &= -\left(\tfrac{k-k_1}{2k_1}\right) Fe^{ik_1 a/2} e^{ika/2} = e^{-ik_1 a/2}\left[-\left(\tfrac{k-k_1}{2k_1}\right) e^{-ika/2} + \left(\tfrac{k+k_1}{2k_1}\right) Be^{ika/2}\right] \end{aligned}$$

Therefore

$$\begin{aligned} \left(\tfrac{k+k_1}{2k_1}\right) Fe^{-ik_1 a} e^{ika/2} - \left(\tfrac{k+k_1}{2k_1}\right) e^{-ika/2} &= -\left(\tfrac{k-k_1}{2k_1}\right) Be^{ika/2} \\ -\left(\tfrac{k-k_1}{2k_1}\right) Fe^{ik_1 a} e^{ika/2} + \left(\tfrac{k-k_1}{2k_1}\right) e^{-ika/2} &= \left(\tfrac{k+k_1}{2k_1}\right) Be^{ika/2} \end{aligned}$$

Dividing we get

$$\frac{\left(\tfrac{k+k_1}{2k_1}\right) Fe^{-ik_1 a} e^{ika/2} - \left(\tfrac{k+k_1}{2k_1}\right) e^{-ika/2}}{-\left(\tfrac{k-k_1}{2k_1}\right) Fe^{ik_1 a} e^{ika/2} + \left(\tfrac{k-k_1}{2k_1}\right) e^{-ika/2}} = \frac{-\left(\tfrac{k-k_1}{2k_1}\right)}{\left(\tfrac{k+k_1}{2k_1}\right)}$$

$$(k+k_1)^2\left[Fe^{-ik_1a}e^{ika/2}-e^{-ika/2}\right]=-(k-k_1)^2\left[-Fe^{ik_1a}e^{ika/2}+e^{-ika/2}\right]$$

or

$$\left[(k+k_1)^2e^{-ik_1a}-(k-k_1)^2e^{ik_1a}\right]e^{ika/2}F$$
$$=-(k-k_1)^2e^{-ika/2}+(k+k_1)^2e^{-ika/2}$$

Finally we solve for the transmission coefficient F to get

$$\begin{aligned}F&=\frac{(k+k_1)^2-(k-k_1)^2}{(k+k_1)^2e^{-ik_1a}-(k-k_1)^2e^{ik_1a}}e^{-ika}\\&=\frac{4kk_1}{(k^2+k_1^2)\left(e^{-ik_1a}-e^{ik_1a}\right)+2kk_1\left(e^{-ik_1a}+e^{ik_1a}\right)}e^{-ika}\\&=\frac{e^{-ika}}{\cos k_1a-\frac{i}{2}\left(\frac{k}{k_1}+\frac{k_1}{k}\right)\sin k_1a}\end{aligned}$$

and the transmission probability

$$T(E)=|F(E)|^2=\frac{1}{1+\frac{\sin^2\sqrt{\frac{2m}{\hbar^2}V_0a^2\left(1+\frac{E}{V_0}\right)}}{4\left(\frac{E}{V_0}\right)\left(1+\frac{E}{V_0}\right)}}\tag{6.240}$$

A plot of $T(E)$ versus E is shown below for the case

$$\frac{2mV_0a^2}{\hbar^2}=\frac{25\pi^2}{4}\tag{6.241}$$

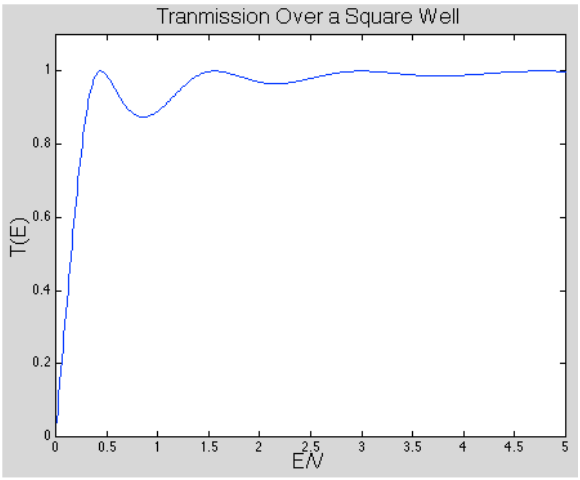


Figure 6.9: Transmission over a Square Well

The peaks are called *resonances*. They occur when $T(E) = 1$ or

$$\sin^2 \sqrt{\frac{2m}{\hbar^2} V_0 a^2 (1 + \frac{E}{V_0})} = 0 \tag{6.242}$$

or when

$$\begin{aligned}
&\sqrt{\tfrac{2m}{\hbar^2} V_0 a^2 (1 + \tfrac{E}{V_0})} = n\pi \\
&\tfrac{E}{V_0} = \tfrac{n^2\pi^2}{\frac{2m}{\hbar^2} V_0 a^2} - 1 = \tfrac{4n^2}{25} - 1 \geq 0 \\
&\tfrac{E}{V_0} = 0.44, 1.56, 3.00, ...
\end{aligned}$$

in agreement with the diagram.

To a particle of these particular energies, the potential looks completely *transparent*; there is no reflected particle

$$R(E) = |B(E)|^2 = 0 \tag{6.243}$$

All of the incident wave is transmitted into region III.

A special feature of these amplitudes, not derivable from non-relativistic quantum mechanics, relates the properties of the transmission probability and the bound state energies of the well.

The transmission amplitude $F(E)$ is a function of the energy E. It was derived for $E > 0$. If we assume that it is an analytic function of E, then it has some interesting properties. In particular, in the regions of $E < 0$ and real, $F(E)$ has poles (goes to infinity) at E values that correspond to the bound state energies of the well.

$F(E)$ being infinite corresponds to having a transmitted wave without having an incident wave. This is exactly the condition to have a bound state, where the transmitted(and reflected) wave does not propagate, but instead fall off exponentially.

The poles of $F(E)$ occur when

$$\begin{aligned}
&\cos k_1 a = \tfrac{i}{2} \left(\tfrac{k}{k_1} + \tfrac{k_1}{k} \right) \sin k_1 a \\
&k_1 \cot \tfrac{k_1 a}{2} = ik \text{ and } k_1 \tan \tfrac{k_1 a}{2} = -ik
\end{aligned}$$

which are the same transcendental equations we had earlier for the bound states of the corresponding finite square well.

This property can be used to find bound state energies after solving the transmission problem for a given potential.

What do wave packets say?

If we construct wave packets representing the incident, reflected and transmitted particles using the formalism we demonstrated earlier, we find the following result.

Away from a resonance energy, the transmitted and reflected packets do not exhibit any strange behavior, i.e., the reflected packet forms and starts to move in the $-x$ direction from $x = -a/2$ at about the same time the transmitted packet forms and starts to move in the $+x$ direction from $x = a/2$. This time is on the order of the classical transit time for a particle moving across the well

$$\Delta t_{classical} = \frac{a}{v} = \frac{ma}{\hbar k_1} \tag{6.244}$$

However, near any resonance, this behavior changes dramatically.

The wave seems to get bottled up in the vicinity of the well, as if some temporary or metastable intermediate state forms. After a time larger than the classical transit time across the well, the reflected and transmitted packets reform and start moving away. An example is given below.

Time Delay at a Square Well

We consider the transmission through a square well as shown below in Figure 6.10.

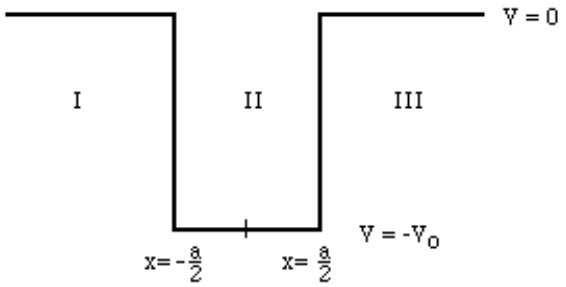


Figure 6.10: Finite Square Well

The transmission amplitude is (from our earlier calculation)

$$F(E) = \frac{e^{-ika}}{\cos k_1 a - \frac{i}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\sin k_1 a} = |F|\, e^{i\theta} \tag{6.245}$$

where we have set

$$e^{-ika} e^{i \tan^{-1}\left(-\frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a\right)} = e^{i\theta} \tag{6.246}$$

We then have

$$\tan(\theta + ka) = \frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a \tag{6.247}$$

Using trigonometric identities we get

$$\begin{aligned} \tfrac{\tan\theta+\tan ka}{1-\tan\theta\tan ka} &= \tfrac{1}{2}\left(\tfrac{k}{k_1} + \tfrac{k_1}{k}\right)\tan k_1 a \\ \tan\theta + \tan ka &= \tfrac{1}{2}\left(\tfrac{k}{k_1} + \tfrac{k_1}{k}\right)\tan k_1 a\,(1 - \tan\theta\tan ka) \end{aligned} \tag{6.248}$$

and

$$\begin{aligned} \tan\theta\left(1 + \frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a\tan ka\right) \\ = \frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a - \tan ka \end{aligned} \tag{6.249}$$

$$\tan\theta = \frac{\frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a - \tan ka}{1 + \frac{1}{2}\left(\frac{k}{k_1} + \frac{k_1}{k}\right)\tan k_1 a\tan ka} \tag{6.250}$$

Now the incident wave packet is given by

$$\psi_{incident}(x,t) = \int f(k)e^{i(kx-\omega(k)t)}\,dk \tag{6.251}$$

and the transmitted wave packet is given by

$$\psi_{transmitted}(x,t) = \int f(k)e^{i(kx-\omega(k)t)}|F(k)|e^{i\theta(k)}\,dk \tag{6.252}$$

where $f(k)$ is nonzero only in a limited region about $k = k_0$ and is peaked at $k = k_0$. The stationary phase argument then gives for the incident wave

$$x = v_0 t \tag{6.253}$$

and for the transmitted wave

$$x = v_0 t - \left.\frac{d\theta}{dk}\right|_{k=k_0} \tag{6.254}$$

Therefore, the incident wave packet arrives at $x = -a/2$ at $t = -a/2v_0$ and the transmitted wave packet leaves $x = +a/2$ at

$$t = \frac{a}{2v_0} + \frac{1}{v_0}\left.\frac{d\theta}{dk}\right|_{k=k_0} \quad , \quad v_0 = \frac{\hbar k_0}{m} \tag{6.255}$$

Therefore, we have a quantum mechanical *time delay* over and above $\Delta t_{classical}$ given by

$$\tau = \frac{1}{v_0}\left.\frac{d\theta}{dk}\right|_{k=k_0} \tag{6.256}$$

Now

$$\frac{d\tan\theta}{dk} = \frac{d\tan\theta}{d\theta}\frac{d\theta}{dk} = \frac{1}{\cos^2\theta}\frac{d\theta}{dk}$$

$$\frac{d\theta}{dk} = \cos^2\theta\frac{d\tan\theta}{dk} = \frac{1}{1+\tan^2\theta}\frac{d\tan\theta}{dk}$$

or

$$\tan\theta = \frac{\frac{1}{2}\left(\frac{k}{k_1}+\frac{k_1}{k}\right)\tan k_1 a - \tan ka}{1+\frac{1}{2}\left(\frac{k}{k_1}+\frac{k_1}{k}\right)\tan k_1 a\tan ka} \tag{6.257}$$

where

$$E = \frac{\hbar^2 k^2}{2m} \text{ and } E + V_0 = \frac{\hbar^2 k_1^2}{2m} \tag{6.258}$$

or

$$k_1^2 = k^2 + \frac{2mV_0}{\hbar^2} \tag{6.259}$$

One can show(Messiah) that these solutions have the following properties:

1. Resonances (when $T = 1$) occur when

   $$k_1 a = n\pi \rightarrow \tan k_1 a = 0$$

2. The time delay due to quantum mechanics near resonance approximately given by

   $$\tau = \frac{V_0}{2k}\tau_{classical} \tag{6.260}$$

   where

   $$\tau_{classical} = \frac{ma}{\hbar k} = \text{ classical transit time} \tag{6.261}$$

   This corresponds physically to the wave (or particle) remaining in region II (the vicinity of the well) during this time delay.

   At that time the wave packet splits up and the reflected and transmitted packets reform and propagate.

3. Off resonance, the transmission probability is very small and the wave practically does not penetrate the region of the well (it is totally reflected at the first potential discontinuity).

6.5.4 Delta-Function Potentials

We now consider the potential energy function

$$V(x) = A\delta(x-a) \tag{6.262}$$

where

$$\begin{aligned} &\delta(x-a) = 0 \quad x \neq a \\ &\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a) \end{aligned} \tag{6.263}$$

and solve the corresponding Schrodinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \tag{6.264}$$

As we discussed earlier the wave function $\psi(x)$ is assumed to be continuous for physical reasons relating to the probability interpretation. The derivative of the wave function, however, is not continuous at $x = a$ for this potential. We can see this as follows. We have

$$-\frac{\hbar^2}{2m}\int_{a-\varepsilon}^{a+\varepsilon}\frac{d^2\psi(x)}{dx^2}dx + A\int_{a-\varepsilon}^{a+\varepsilon}\delta(x-a)V(x)\psi(x)dx = E\int_{a-\varepsilon}^{a+\varepsilon}\psi(x)dx \tag{6.265}$$

In the limit $\varepsilon \to 0$, using the continuity of $\psi(x)$, we get

$$-\frac{\hbar^2}{2m}\left[\left.\frac{d\psi}{dx}\right|_{a+\varepsilon} - \left.\frac{d\psi}{dx}\right|_{a-\varepsilon}\right] = E\psi(a)\int_{a-\varepsilon}^{a+\varepsilon}dx - A\psi(a) \tag{6.266}$$

so that

$$\text{discontinuity } \left(\frac{d\psi}{dx}\right)_{x=a} = \Delta\left(\frac{d\psi}{dx}\right) = \frac{2mA}{\hbar^2}\psi(a) \tag{6.267}$$

For simplicity we choose a = 0. We then have two regions to consider

region I $\quad x < 0$
region II $\quad x > 0$

and the derivative is discontinuous at $x = 0$.

Transmission Problem

We first carry out the calculation of the transmission and reflection probabilities. We assume that $A > 0$(we have a delta function barrier), $E > 0$ and an incident wave of unit intensity coming in from the left.

In region I we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = E\psi_I \to \psi_I(x) = e^{ikx} + Be^{-ikx} \text{ with } E = \frac{\hbar^2k^2}{2m} > 0 \tag{6.268}$$

We have both an incident and a reflected wave.

In region II we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} = E\psi_{II} \to \psi_{II}(x) = Ce^{ikx} \text{ with } E = \frac{\hbar^2k^2}{2m} > 0 \tag{6.269}$$

There is *only* a transmitted wave.

The boundary conditions (at $x = 0$) give

$$\psi_I(0) = \psi_{II}(0) \rightarrow 1 + B = C$$

$$\frac{d\psi_{II}(0)}{dx} - \frac{d\psi_I(0)}{dx} = \frac{2m}{\hbar^2} A\psi_{II}(0) \rightarrow ikC - ik(1 - B) = \frac{2m}{\hbar^2} AC$$

The solutions are

$$C = \frac{ik}{ik - \frac{mA}{\hbar^2}} \text{ and } B = \frac{\frac{mA}{\hbar^2}}{ik - \frac{mA}{\hbar^2}} \tag{6.270}$$

We then have

$$T = \text{ transmission probability } = |C|^2 = \frac{1}{1 + \frac{mA^2}{2\hbar^2 E}} \tag{6.271}$$

$$R = \text{ reflection probability } = |B|^2 = \frac{1}{1 + \frac{2\hbar^2 E}{mA^2}} \tag{6.272}$$

We note that $T + R = 1$ as it must for the probability interpretation to make sense.

From our previous discussion, we suspect that the energy values of the poles of the transmission probability correspond to the bound state energies for the delta function well problem ($A < 0$). For the single delta function potential, T has a single pole at

$$E = -\frac{mA^2}{2\hbar^2} \tag{6.273}$$

Bound-State Problem

We let $A \rightarrow -A\ ,\ A > 0$. In region I we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_I}{dx^2} = -|E|\,\psi_I \rightarrow \psi_I(x) = Be^{\alpha x} \tag{6.274}$$

with

$$E = -|E| = -\frac{\hbar^2\alpha^2}{2m} < 0 \tag{6.275}$$

We have excluded the negative exponential term since it would diverge in region I as $x \rightarrow -\infty$.

In region II we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{II}}{dx^2} = -|E|\,\psi_{II} \rightarrow \psi_{II}(x) = Ce^{-\alpha x} \tag{6.276}$$

with

$$E = -|E| = \frac{\hbar^2\alpha^2}{2m} < 0 \tag{6.277}$$

We have excluded the positive exponential term since it would diverge in region II as $x \to +\infty$.

The boundary conditions give

$$\psi_I(0) = \psi_{II}(0) \to B = C \tag{6.278}$$

$$\frac{d\psi_{II}(0)}{dx} - \frac{d\psi_I(0)}{dx} = -\frac{2m}{\hbar^2} A\psi_I(0) \to -\alpha C - \alpha B = -\frac{2m}{\hbar^2} AB \tag{6.279}$$

The resulting equation for α gives the allowed the bound state energies. We have

$$\alpha = \frac{mA}{\hbar^2} \to \text{ only 1 solution only } \to \text{ 1 bound state} \tag{6.280}$$

$$E = -|E| = -\frac{\hbar^2\alpha^2}{2m} = -\frac{mA^2}{2\hbar^2} \tag{6.281}$$

which is the same value as we obtained from the pole of the transmission probability.

We also note that the solution has definite parity (even) since $\psi(x) = \psi(-x)$. This must occur since $V(x) = V(-x)$ and hence parity commutes with the Hamiltonian. As we also saw in the square well case, if only one solution exists then it is always an even parity solution.

Double Delta Function Potential

We now tackle a more realistic and hence more complex system, namely the *double delta function* potential energy function given by

$$V(x) = u\left[\delta\left(x + \frac{\ell}{2}\right) + \delta\left(x - \frac{\ell}{2}\right)\right] \tag{6.282}$$

We again start by considering the barrier transmission problem for $u > 0$.

Transmission problem

There are three regions to consider

region I	$x < \frac{\ell}{2}$
region II	$-\frac{\ell}{2} < x < \frac{\ell}{2}$
region III	$x > \frac{\ell}{2}$

The solutions in the three regions are

region I	$x < \frac{\ell}{2}$	$\psi_I(x) = Ae^{ikx} + Be^{-ikx}$
region II	$-\frac{\ell}{2} < x < \frac{\ell}{2}$	$\psi_{II}(x) = Ce^{ikx} + De^{-ikx}$
region III	$x > \frac{\ell}{2}$	$\psi_{III}(x) = Fe^{ikx}$

where

$$E = \frac{\hbar^2 k^2}{2m} > 0 \tag{6.283}$$

The boundary conditions (the derivatives are discontinuous at $x = \pm\ell/2$)give the equations

$$x = -\frac{\ell}{2} \qquad Ae^{-ik\frac{\ell}{2}} + Be^{ik\frac{\ell}{2}} = Ce^{-ik\frac{\ell}{2}} + De^{ik\frac{\ell}{2}} \tag{6.284}$$

$$\text{ik}\left(Ce^{-ik\frac{\ell}{2}} - De^{ik\frac{\ell}{2}}\right) - ik\left(Ae^{-ik\frac{\ell}{2}} - Be^{ik\frac{\ell}{2}}\right) = \frac{2mu}{\hbar^2}\left(Ce^{-ik\frac{\ell}{2}} + De^{ik\frac{\ell}{2}}\right) \tag{6.285}$$

$$x = +\frac{\ell}{2} \qquad Ce^{ik\frac{\ell}{2}} + De^{-ik\frac{\ell}{2}} = Fe^{ik\frac{\ell}{2}} \tag{6.286}$$

$$\text{ik}\left(Fe^{ik\frac{\ell}{2}}\right) - ik\left(Ce^{ik\frac{\ell}{2}} - De^{-ik\frac{\ell}{2}}\right) = \frac{2mu}{\hbar^2}\left(Fe^{ik\frac{\ell}{2}}\right) \tag{6.287}$$

We are interested in calculating the transmission probability

$$T(E) = \frac{|F|^2}{|A|^2} \tag{6.288}$$

Much algebra gives these results

$$C = (1 - \frac{mu}{ik\hbar^2})F \text{ and } D = \frac{mu}{ik\hbar^2}e^{ik\ell}F \tag{6.289}$$

$$\frac{F}{A} = \frac{1}{\left[1 - \frac{mu^2}{2\hbar^2 E}(1 - \cos 2k\ell)\right]^2 + i\left[\frac{ku}{E} + \frac{mu^2}{2\hbar^2 E}\sin 2k\ell\right]} = \frac{1}{\left[1 - \frac{mu}{ik\hbar^2}\right]^2 - \left[\frac{mu}{ik\hbar^2}\right]^2 e^{2ik\ell}} \tag{6.290}$$

The poles of the transmission probability occur for the zeroes of the denominator in (6.290). If we let $u = -|u|$ and $E = -|E| < 0$ and $k = i\gamma$, the poles are given by

$$\left[1 - \frac{m|u|}{\gamma\hbar^2}\right]^2 - \left[\frac{m|u|}{\gamma\hbar^2}\right]^2 e^{-2\gamma\ell} = 0 \tag{6.291}$$

Some rearranging and algebra gives

$$e^{\gamma\frac{\ell}{2}} = \frac{m|u|}{\gamma\hbar^2}(e^{\gamma\frac{\ell}{2}} \pm e^{-\gamma\frac{\ell}{2}}) \tag{6.292}$$

as the transcendental equation for the bound state energies of the double delta function well.

We will solve this transcendental equation graphically. Again, it is always best to first clean up the equation as much as possible by changing variables. There are no rules for this part. We let

$$\beta = \gamma\frac{\ell}{2} \text{ and } \varepsilon = \frac{m|u|\ell}{2\hbar^2} \tag{6.293}$$

The transcendental equation then becomes

$$e^{-2\beta} = \pm\left(\frac{\beta}{\varepsilon} - 1\right) \tag{6.294}$$

Bound-State Problem

We now get the same equation in the standard way.

In this case $u < 0$ and we choose

$$E = -|E| = -\frac{\hbar^2\alpha^2}{2m} \tag{6.295}$$

The solutions in the three regions are

region I	$x < \frac{\ell}{2}$	$\psi_I(x) = Ae^{\alpha x}$
region II	$-\frac{\ell}{2} < x < \frac{\ell}{2}$	$\psi_{II}(x) = Be^{\alpha x} + Ce^{-\alpha x}$
region III	$x > \frac{\ell}{2}$	$\psi_{III}(x) = De^{-\alpha x}$

The boundary conditions (the derivatives are discontinuous at $x = \pm\ell/2$)give the equations

$$x = -\frac{\ell}{2} \qquad Ae^{-\alpha\frac{\ell}{2}} + Be^{\alpha\frac{\ell}{2}} = Ce^{-\alpha\frac{\ell}{2}} + De^{\alpha\frac{\ell}{2}} \tag{6.296}$$

$$\alpha\left(Ce^{-\alpha\frac{\ell}{2}} - De^{\alpha\frac{\ell}{2}}\right) - \alpha\left(Ae^{-\alpha\frac{\ell}{2}} - Be^{\alpha\frac{\ell}{2}}\right) = \frac{2mu}{\hbar^2}\left(Ce^{-\alpha\frac{\ell}{2}} + De^{\alpha\frac{\ell}{2}}\right) \tag{6.297}$$

$$x = +\frac{\ell}{2} \qquad Ce^{\alpha\frac{\ell}{2}} + De^{-\alpha\frac{\ell}{2}} = Fe^{\alpha\frac{\ell}{2}} \tag{6.298}$$

$$\alpha\left(Fe^{\alpha\frac{\ell}{2}}\right) - \alpha\left(Ce^{\alpha\frac{\ell}{2}} - De^{-\alpha\frac{\ell}{2}}\right) = \frac{2mu}{\hbar^2}\left(Fe^{\alpha\frac{\ell}{2}}\right) \tag{6.299}$$

We consider two cases:

even parity: $A = D$ and $B = C$
odd parity: $A = -D$ and $B = -C$

Much algebra leads to the transcendental equations

$$\text{even parity} \qquad e^{-\alpha\ell} = -1 + \frac{\hbar^2\alpha}{m|u|} \tag{6.300}$$

$$\text{odd parity} \qquad e^{-\alpha\ell} = +1 - \frac{\hbar^2\alpha}{m\,|u|} \tag{6.301}$$

If we let

$$\beta = \gamma\frac{\ell}{2} \text{ and } \varepsilon = \frac{m\,|u|\,\ell}{2\hbar^2} \tag{6.302}$$

the equations becomes

$$\text{even parity} \qquad e^{-2\beta} = +\left(\frac{\beta}{\epsilon} - 1\right) \tag{6.303}$$

$$\text{odd parity} \qquad e^{-2\beta} = -\left(\frac{\beta}{\epsilon} - 1\right) \tag{6.304}$$

which agree with the result we obtained from the poles of the transmission amplitude.

Graphical Solution

Since

$$\varepsilon = \frac{m\,|u|\,\ell}{2\hbar^2} \tag{6.305}$$

is fixed by the choice of the parameters of the potential, we have transcendental equations for β. The solutions give us the energy eigenvalues. We can simplify the graphical solution with one more change of variables. We let

$$\frac{\beta}{\varepsilon} = \eta \tag{6.306}$$

and get

$$\text{even parity} \qquad e^{-2\epsilon\eta} = \eta - 1 \tag{6.307}$$

$$\text{odd parity} \qquad e^{-2\epsilon\eta} = 1 - \eta \tag{6.308}$$

Procedure:

1. plot $1-\eta$ versus η and $\eta - 1$ versus η
2. plot $e^{-2\epsilon\eta}$ versus η for several values of ϵ
3. the intersections are the even and odd solutions

In Figure 6.11 below, we plot the exponential function for several values of ϵ. As can be seen, the properties of the solutions are:

1. Even solutions always exist
2. Odd solutions do not always exists (see the $\epsilon = 0.25$ line, which only intersects the even curve)
3. As $\epsilon \to 0$ which corresponds to $\ell \to \infty$ we get $E_{odd} = E_{even}$. Physically, the wells are separating so that they no longer influence each other and the system behaves like two separate wells.

We will look at this property again later when we consider a translationally invariant infinite line of delta functions. The interaction between the wells coupled to the translations invariance will lead to energy bands as in real solids.

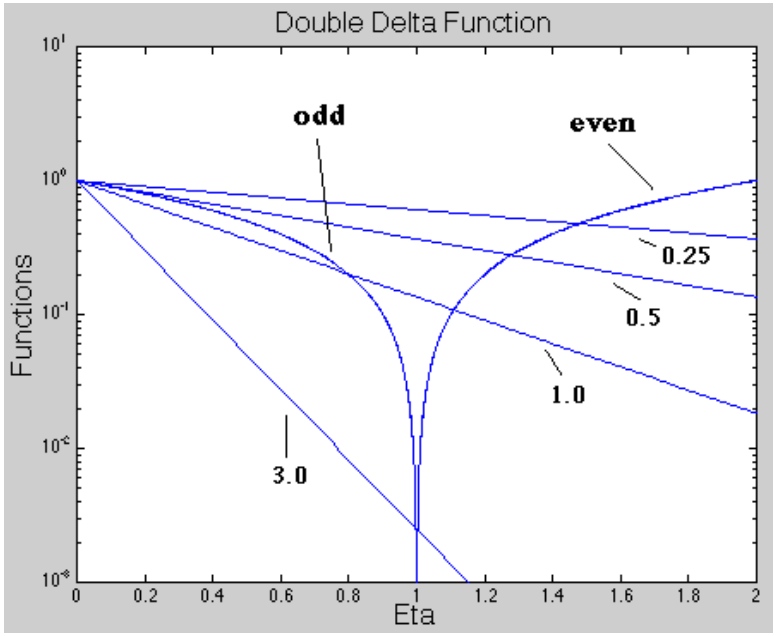


Figure 6.11: Double delta Function Bound States

6.6 Harmonic Oscillators

We now turn our attention to a physical system with a harmonic oscillator potential energy function

$$V(x) = \frac{1}{2}kx^2 = \frac{1}{2}m\omega_0^2 x^2 \tag{6.309}$$

We will first solve this system by going to the position representation and solving the Schrodinger equation using differential equation techniques. We will review all of our methods, introduce Hermite polynomials and define generating functions along the way.

After that we will introduce an algebraic method for solving this system that does not make any reference to the position representation. In the next chapter we will generalize this algebraic method for use in other types of systems.

6.6.1 Differential Equation Method

The Hamiltonian for the system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega_0^2\hat{x}^2 \tag{6.310}$$

The energy eigenvalue equation is

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle \tag{6.311}$$

We put this into the position representation by these steps:

$$\begin{aligned}
\langle x|\,\hat{H}\,|\psi\rangle &= \langle x|\,E\,|\psi\rangle = E\,\langle x\mid\psi\rangle = E\psi(x) \\
\langle x|\left(\frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega_0^2\hat{x}^2\right)|\psi\rangle &= E\psi(x) \\
\langle x|\,\frac{\hat{p}^2}{2m}\,|\psi\rangle + \frac{1}{2}m\omega_0^2\,\langle x|\,\hat{x}^2\,|\psi\rangle &= E\psi(x) \\
\frac{1}{2m}(-i\hbar\frac{d}{dx})^2\,\langle x\mid\psi\rangle + \frac{1}{2}m\omega_0^2x^2\,\langle x\mid\psi\rangle &= E\psi(x) \\
-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}m\omega_0^2x^2\psi(x) &= E\psi(x)
\end{aligned} \tag{6.312}$$

The last line is the standard 1-dimensional time-independent Schrodinger equation for the harmonic oscillator potential.

When solving such equations it is useful to rewrite it in dimensionless form. We do this by making these substitutions (the choice is an art, not a science)

$$\alpha^4 = \frac{m^2\omega_0^2}{\hbar^2}\ ,\ q{=}\alpha x \text{ and } \lambda = \frac{2E}{\hbar\omega_0} \tag{6.313}$$

The equation (6.316) becomes

$$\frac{d^2\psi}{dq^2} + (\lambda - q^2)\psi = 0 \tag{6.314}$$

The solution of such equations is greatly aided by examining and extracting the behavior of ψ in the asymptotic regions $q \to \pm\infty$.

For sufficiently large q the equation becomes

$$\frac{d^2\psi}{dq^2} - q^2\psi = 0 \tag{6.315}$$

which is satisfied by functions of the form

$$q^n e^{\pm\frac{1}{2}q^2} \quad \text{for any finite value of } n \tag{6.316}$$

Since the probability interpretation of the function requires us to be able to normalize it, i.e., we must have

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty \tag{6.317}$$

This rules out the positive exponential. This result suggests that we try a solution of the form

$$\psi(q) = H(q)e^{-\frac{1}{2}q^2} \tag{6.318}$$

where $H(q)$ is a polynomial of finite order in q and we have explicitly extracted the asymptotic behavior.

Substituting into the original differential equation gives a new differential equation for $H(q)$

$$\frac{d^2H}{dq^2} - 2q\frac{dH}{dq} + (\lambda - 1)H = 0 \tag{6.319}$$

We now use the standard Frobenius series substitution method to solve this equation. We assume a solution of the form

$$H(q) = q^s \sum_{j=0}^{\infty} a_j q^j \text{ where } a_0 \neq 0 \text{ and } s \geq 0 \tag{6.320}$$

The nonnegative restriction on s is required in order to guarantee that the solution is well-behaved at $q = 0$ (we need to be able to normalize the wave function). If we substitute this guess into the differential equation for H we get

$$q^s \sum_{j=0}^{\infty} a_j(j+s)(j+s-1)q^{j-2} - 2q^s \sum_{j=0}^{\infty} a_j(j+s)q^j + (\lambda - 1)q^s \sum_{j=0}^{\infty} a_j q^j = 0$$

$$q^s\left[a_0 s(s-1)q^{-2} + a_1 s(s+1)q^{-1} + (a_2(s+2)(s+1) - a_0(2s+1-\lambda))\,q^0\right]$$
$$+ (a_3(s+3)(s+2) - a_1(2s+3-\lambda))\,q^1 + ...+$$
$$(a_{j+2}(s+j+2)(s+j+1) - (2s+2j+1-\lambda)a_j)\,q^j + ... = 0$$

Now for a power series to be identically zero, the coefficient of each power must be zero separately. We thus have

$$s(s-1)a_0 = 0 \tag{6.321}$$
$$s(s+1)a_1 = 0 \tag{6.322}$$
$$(s+2)(s+1)a_2 - (2s+1-\lambda)a_0 = 0 \tag{6.323}$$
$$(s+3)(s+2)a_3 - (2s+3-\lambda)a_1 = 0 \tag{6.324}$$
$$\ldots\ldots$$
$$(s+j+2)(s+j+1)a_{j+2} - (2s+2j+1-\lambda)a_j = 0 \tag{6.325}$$

Now we assumed that $a_0 \neq 0$, $s \geq 0$. Therefore the (6.321) says that we must have $s = 0$ or $s = 1$. (6.322) says that $s = 0$ or $a_1 = 0$, or both. The remaining equations give a_2 in terms of a_0, a_3 in terms of a_1, and, in general, a_{j+2} in terms of a_j. Whether the series will have a finite or an infinite number of terms depends on the choice of s, a_1, and the eigenvalue λ.

If the series does not terminate, then its asymptotic behavior can be inferred from the coefficients of the high power terms. From the last recursion formula (6.325) we have

$$\frac{a_{j+2}}{a_j} \underset{j\to\infty}{\to} \frac{2}{j} \tag{6.326}$$

This ratio is the same as that of the series for $q^n e^{q^2}$ with any finite value of n. This implies that the solution we have found for the wave function

$$\psi(q) = H(q)e^{-\frac{1}{2}q^2} \tag{6.327}$$

will diverge at large q.

Since we cannot have that happen and still retain the probability interpretation, the series must terminate ($H(q)$ must be a polynomial as I mentioned earlier). From the last recursion formula (6.325)we can see that if we choose

$$\lambda = 2s + 2j + 1 \tag{6.328}$$

for some j, then the series will terminate ($a_{j+2} = 0$). Since $a_0 \neq 0$ this cutoff procedure requires that j equals an even integer. This means that the odd series will not terminate unless we choose $a_1 = 0$. The index s can still be either 0 or 1 . Corresponding to these two values, λ is equal to $2j + 1$ or $2j + 3$, where j is an even integer. We can express both cases in terms of an integer n as

$$\lambda_n = 2n + 1 \tag{6.329}$$

This gives us the energy eigenvalues

$$E_n = \frac{\lambda_n}{2}\hbar\omega_0 = \left(n + \frac{1}{2}\right)\hbar\omega_0 \tag{6.330}$$

As in the square well examples, we have a *zero point energy*. It corresponds to the lowest or ground state energy value

$$E_0 = \frac{1}{2}\hbar\omega_0 \tag{6.331}$$

The eigenfunctions have a definite parity, as they should since $V(x) = V(-x)$. From the definition of n, we can see that for a given solution

$$n = \text{ highest value of } s + j \tag{6.332}$$

If we denote the corresponding polynomials by $H_n(q)$ we see that $H_n(q)$ is of degree n in q. Therefore, since the exponential is an even function, we have

n = even → even parity solution
n = odd → odd parity solution

The polynomial of order n, $H_n(q)$, is a solution of the original equation with $\lambda = 2n + 1$

$$\frac{d^2H_n}{dq^2} - 2q\frac{dH_n}{dq} + 2nH_n = 0 \tag{6.333}$$

It is called the *Hermite polynomial*. We could find the exact forms of the polynomials from the recursion relations. However, it more instructive to introduce the idea of a *generating function*. We define

$$S(q,s) = e^{q^2-(s-q)^2} = e^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{s^n}{n!} \tag{6.334}$$

The functions $G_n(q)$ defined in this way are polynomials in q. Now we have

$$\frac{\partial S}{\partial q} = 2se^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{2s^{n+1}}{n!} = \sum_{n=0}^{\infty}\frac{dG_n(q)}{dq}\frac{s^n}{n!}$$

$$\frac{\partial S}{\partial s} = (-2s+2q)e^{-s^2+2sq} = \sum_{n=0}^{\infty} G_n(q)\frac{(-2s+2q)s^n}{n!} = \sum_{n=0}^{\infty} G_n(q)\frac{s^{n-1}}{(n-1)!}$$

Equating equal powers of q in each of these equation we get

$$\frac{dG_n}{dq} = 2nG_{n-1} \text{ and } G_{n+1} = 2qG_n - 2nG_{n-1} \tag{6.335}$$

The lowest order differential equation, which involves only $G_n(q)$, that we can construct out of this pair of equation is

$$\frac{d^2G_n}{dq^2} = 2n\frac{dG_{n-1}}{dq}, \frac{dG_{n+1}}{dq} = 2(n+1)G_n$$

$$\frac{dG_{n+1}}{dq} = 2q\frac{dG_n}{dq} + 2G_n - 2n\frac{dG_{n-1}}{dq}$$

$$\frac{d^2G_n}{dq^2} = 2q\frac{dG_n}{dq} + 2G_n - \frac{dG_{n+1}}{dq} = 2q\frac{dG_n}{dq} + 2G_n - 2(n+1)G_n$$

$$\frac{d^2G_n}{dq^2} - 2q\frac{dG_n}{dq} - 2nG_n = 0 \tag{6.336}$$

which is the Hermite polynomial equation. So $G_n(q) = H_n(q)$ and we have

$$S(q,s) = e^{q^2-(s-q)^2} = e^{-s^2+2sq} = \sum_{n=0}^{\infty} H_n(q)\frac{s^n}{n!} \tag{6.337}$$

We then have

$$\frac{\partial^n S(q,s)}{\partial s^n} = e^{q^2}\frac{\partial^n e^{-(s-q)^2}}{\partial s^n} = (-1)^n e^{q^2}\frac{\partial^n e^{-(s-q)^2}}{\partial q^n}$$

$$= \frac{\partial^n \sum_{m=0}^{\infty} H_m(q)\frac{s^m}{m!}}{\partial s^n} = \sum_{m=n}^{\infty} H_m(q)\frac{s^{m-n}}{(m-n)!} \tag{6.338}$$

where we have used the property

$$\frac{\partial f(s-q)}{\partial s} = -\frac{\partial f(s-q)}{\partial q} \tag{6.339}$$

If we set $s = 0$ in the equation above we get

$$(-1)^n e^{q^2}\frac{\partial^n e^{-q^2}}{\partial q^n} = H_n(q) \tag{6.340}$$

We can then easily generate the polynomials

$$H_0 = 1\,,\ H_1 = 2q\,,\ H_2 = 4q^2 - 2\,,\ \ldots \tag{6.341}$$

The Schrodinger wave functions are then given by

$$\psi_n(x) = A_n H_n(\alpha x)e^{-\frac{1}{2}\alpha^2 x^2} \tag{6.342}$$

We find the normalization constant A_n by

$$\int_{-\infty}^{\infty} |\psi_n(x)|^2\,dx = 1 = \frac{|A_n|^2}{\alpha}\int_{-\infty}^{\infty} H_n^2(q)e^{-q^2}\,dq \tag{6.343}$$

Now we can determine all of the integrals of this type from the generating function as follows. We have

$$\int_{-\infty}^{\infty} e^{-s^2+2sq}e^{-t^2+2tq}e^{-q^2}\,dq = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{s^n t^m}{n!m!}\int_{-\infty}^{\infty} H_n(q)H_m(q)e^{-q^2}\,dq$$

$$= \sqrt{\pi}e^{2st} = \sqrt{\pi}\sum_{n=0}^{\infty}\frac{(2st)^n}{n!} \tag{6.344}$$

which says that

$$\int_{-\infty}^{\infty} H_n(q)H_m(q)e^{-q^2}dq = \sqrt{\pi}2^n n!\delta_{nm} \tag{6.345}$$

The case $n = m$, gives

$$A_n = \left(\frac{\alpha}{\sqrt{\pi}2^n n!}\right)^{1/2} \tag{6.346}$$

The case $n \neq m$ just corresponds to the orthogonality property of eigenfunctions of different eigenvalues.

6.6.2 Algebraic Method

The Hamiltonian for this system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega_0^2\hat{x}^2 \tag{6.347}$$

The energy eigenvalue equation is

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle \tag{6.348}$$

The commutator between the position and momentum operators is

$$[\hat{x}, \hat{p}] = i\hbar \tag{6.349}$$

This information defines a complete solution to the problem when using algebraic methods.

We now define the two new operators (like a change of basis)

$$\hat{a} = \sqrt{\frac{m\omega_0}{2\hbar}}\hat{x} + \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} \text{ and } \hat{a}^+ = \sqrt{\frac{m\omega_0}{2\hbar}}\hat{x} - \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} \tag{6.350}$$

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega_0}}\,(\hat{a} + \hat{a}^+) \text{ and } \hat{p} = -i\sqrt{\frac{m\hbar\omega_0}{2}}\,(\hat{a} - \hat{a}^+) \tag{6.351}$$

These new operators are not Hermitian. They are Hermitian conjugates, however. We then have

$$[\hat{a}, \hat{a}^+] = 1 \text{ and } \hat{H} = \hbar\omega_0(\hat{N}_{op} + \frac{1}{2})\,, \hat{N}_{op} = \hat{a}^+\hat{a} \tag{6.352}$$

Now suppose that the vector $|n\rangle$ is an eigenvector of $\hat{N}_{op} = \hat{a}^\dagger\hat{a}$ with eigenvalue n

$$\hat{N}_{op}\,|n\rangle = n\,|n\rangle \tag{6.353}$$

Since $\hat{N}_{op}$ and $\hat{H}$ differ only by a constant, this implies that $|n\rangle$ is also an eigenvector of $\hat{H}$, i.e.,

$$\hat{H}\,|n\rangle = \hbar\omega_0(\hat{N}_{op} + \frac{1}{2})\,|n\rangle = \hbar\omega_0(n + \frac{1}{2})\,|n\rangle \tag{6.354}$$

Therefore, $|E_n\rangle = |n\rangle$ and the energy eigenvalues are

$$E_n = \hbar\omega_0(n + \frac{1}{2}) \tag{6.355}$$

Our task now is to find the allowed values of n.

Now $\hat{N}_{op}^\dagger = (\hat{a}^\dagger\hat{a})^\dagger = \hat{a}^\dagger\hat{a} = \hat{N}_{op}$, so $\hat{N}_{op}$ is a Hermitian operator. This means that the eigenvalues of $\hat{N}_{op}$ (and $\hat{H}$) are *real* numbers. Now we also have the relation

$$n = \langle n|\,\hat{N}_{op}\,|n\rangle = \langle n|\,\hat{a}^+\hat{a}\,|n\rangle \tag{6.356}$$

If we let $|\phi\rangle = \hat{a}\,|n\rangle$, then the last equation (6.360) can be written

$$n = \langle\phi\,|\,\phi\rangle \geq 0 \tag{6.357}$$

which says that the eigenvalues of $\hat{N}_{op}$ (the values of n) are *non-negative, real*numbers.

Finally, using $[\hat{a}, \hat{a}^\dagger] = 1$, we have

$$\hat{N}_{op}\hat{a} = \hat{a}^+\hat{a}\hat{a} = (\hat{a}\hat{a}^+ - 1)\,\hat{a} = \hat{a}\,(\hat{a}^+\hat{a} - 1) = \hat{a}\left(\hat{N}_{op} - 1\right) \tag{6.358}$$

and similarly,

$$\hat{N}_{op}\hat{a}^+ = \hat{a}^+\left(\hat{N}_{op} + 1\right) \tag{6.359}$$

These operator relations imply that

$$\hat{N}_{op}\hat{a}\,|n\rangle = \hat{a}(\hat{N}_{op} - 1)\,|n\rangle = (n-1)\hat{a}\,|n\rangle \tag{6.360}$$

which says that $\hat{a}\,|n\rangle$ is an eigenstate of $\hat{N}_{op}$ with eigenvalue $n-1$ or that we can write

$$\hat{a}\,|n\rangle = \alpha\,|n-1\rangle \rightarrow \langle n|\,\hat{a}^+ = \alpha^*\,\langle n-1| \tag{6.361}$$

We will assume that the eigenstates are normalized, i.e., $\langle n\,|\,n\rangle = 1$. We then have

$$\begin{aligned}(\langle n|\,\hat{a}^+)\,(\hat{a}\,|n\rangle) = |\alpha|^2\,\langle n-1\,|\,n-1\rangle = |\alpha|^2 \\ = \langle n|\,\hat{a}^+\hat{a}\,|n\rangle = \langle n|\,\hat{N}_{op}\,|n\rangle = n\end{aligned} \tag{6.362}$$

which says that $\alpha = \sqrt{n}$ and we have the relation

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle \tag{6.363}$$

Repeating this procedure we have

$$\hat{a}^2\,|n\rangle = \hat{a}\sqrt{n}\,|n-1\rangle = \sqrt{n}\hat{a}\,|n-1\rangle = \sqrt{n(n-1)}\,|n-2\rangle \tag{6.364}$$

which implies the following:

if the eigenvalue n exists, then
so do the eigenvalues $n-1$, $n-2$, ...

This sequence cannot continue forever since we have already shown that the eigenvalues of $\hat{N}_{op}$ are non-negative.

The only way out of this dilemma is that n is an integer. In this case, we would eventually arrive at an eigenstate and then get

$$\hat{a}\,|1\rangle = |0\rangle \text{ and } \hat{a}\,|0\rangle = 0 \tag{6.365}$$

which says the sequence ends. Thus, the eigenvalues of $\hat{N}_{op}$ are the *non-negative integers*.

We can also go the other way since similar algebra shows that

$$\hat{N}_{op}\hat{a}^{+}\,|n\rangle = \hat{a}^{+}(\hat{N}_{op}+1)\,|n\rangle = (n+1)\hat{a}^{+}\,|n\rangle \tag{6.366}$$

This implies that, in the same way as before,

$$\hat{a}^{+}\,|n\rangle = \sqrt{n+1}\,|n+1\rangle \tag{6.367}$$

Thus, the eigenvalues of $\hat{N}_{op}$ are all the non-negative integers

$$n = 0, 1, 2, 3, 4, \ldots\ldots$$

The operators $\hat{a}$ and $\hat{a}^{\dagger}$ are called *lowering/raising* operators from their property of lowering and raising the eigenvalues by unity. They are also sometimes called *ladder* operators because they generate a ladder of equally-spaced eigenvalues.

Finally, assuming that $\langle 0\,|\,0\rangle = 1$ we have

$$\begin{aligned}
|1\rangle &= \frac{\hat{a}^{+}}{\sqrt{1}}\,|0\rangle = \frac{\hat{a}^{+}}{\sqrt{1!}}\,|0\rangle \\
|2\rangle &= \frac{\hat{a}^{+}}{\sqrt{2}}\,|1\rangle = \frac{\hat{a}^{+}}{\sqrt{2}}\frac{\hat{a}^{+}}{\sqrt{1}}\,|0\rangle = \frac{(\hat{a}^{+})^{2}}{\sqrt{2!}}\,|0\rangle \\
|3\rangle &= \frac{\hat{a}^{+}}{\sqrt{3}}\,|2\rangle = \frac{\hat{a}^{+}}{\sqrt{3}}\frac{(\hat{a}^{+})^{2}}{\sqrt{2!}}\,|0\rangle = \frac{(\hat{a}^{+})^{3}}{\sqrt{3!}}\,|0\rangle \\
&\ldots\ldots \\
|n\rangle &= \frac{(\hat{a}^{+})^{n}}{\sqrt{n!}}\,|0\rangle
\end{aligned} \tag{6.368}$$

and

$$E_n = \hbar\omega_0(n+\frac{1}{2}) \tag{6.369}$$

Notice that we were able to derive all of these results without introducing the wave function (the position representation). If you are desperate to see the

position representation wave functions they can easily be derived from this formalism. We have

$$\begin{aligned} 0 = \langle x| \hat{a} |0\rangle &= \langle x| \sqrt{\frac{m\omega_0}{2\hbar}} \hat{x} + \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} |0\rangle \\ &= \sqrt{\frac{m\omega_0}{2\hbar}} x \langle x \mid 0\rangle + \frac{i(-i\hbar\frac{d}{dx})}{\sqrt{2m\hbar\omega_0}} \langle x \mid 0\rangle \end{aligned} \tag{6.370}$$

This is a differential equation for $\langle x \mid 0\rangle = \psi_0(x)$ which is the ground state wave function. We have

$$\frac{d\psi_o(x)}{dx} + \frac{m\omega_0}{\hbar} x\psi_o(x) = 0 \tag{6.371}$$

which has the solution (when normalized)

$$\langle x \mid 0\rangle = \psi_o(x) = \left(\frac{m\omega_0}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega_0 x^2}{2\hbar}} \tag{6.372}$$

which agrees with our earlier work.

The higher energy wave functions are obtained using

$$\begin{aligned} \langle x \mid 1\rangle = \psi_1(x) &= \langle x| \hat{a}^+ |0\rangle \\ &= \langle x| \sqrt{\frac{m\omega_0}{2\hbar}} \hat{x} - \frac{i\hat{p}}{\sqrt{2m\hbar\omega_0}} |0\rangle \\ &= \sqrt{\frac{2m\omega_0}{\hbar}} x \langle x \mid 0\rangle \end{aligned} \tag{6.373}$$

and so on.

The raising and lowering operator formalism is very powerful and will be very useful beyond non-relativistic quantum mechanics where they will allow the theory to represent the creation and annihilation of particles.

Coherent States

Let us now use this formalism to do an interesting exercise and show the power of the methods.

We consider the following question: what are the eigenstates of the lowering operator $\hat{a}$? That is, we write

$$\hat{a} |\alpha\rangle = \alpha |\alpha\rangle \text{ with } \alpha = |\alpha| e^{i\phi} \tag{6.374}$$

where $|\alpha\rangle$ is the eigenvector of $\hat{a}$ and α is the eigenvalue, which is not necessarily real since $\hat{a}$ is not Hermitian.

Since the vectors $|n\rangle$ are eigenvectors of a Hermitian operator, they form a

orthonormal complete set and can be used as an orthonormal basis for the vector space. We can then write

$$|\alpha\rangle = \sum_{m=0}^{\infty} b_m |m\rangle \tag{6.375}$$

where

$$\langle k \mid \alpha\rangle = \sum_{m=0}^{\infty} b_m \langle k \mid m\rangle = \sum_{m=0}^{\infty} b_m \delta_{km} = b_k \tag{6.376}$$

Now

$$\langle n-1|\,\hat{a}\,|\alpha\rangle = \alpha \langle n-1 \mid \alpha\rangle = \alpha b_{n-1} \tag{6.377}$$

and using

$$\hat{a}^+ |n-1\rangle = \sqrt{n}\,|n\rangle \rightarrow \langle n-1|\,\hat{a} = \sqrt{n}\,\langle n| \tag{6.378}$$

we have

$$\langle n-1|\,\hat{a}\,|\alpha\rangle = \sqrt{n}\,\langle n \mid \alpha\rangle = \sqrt{n} b_n \tag{6.379}$$

or

$$b_n = \frac{\alpha}{\sqrt{n}} b_{n-1} \tag{6.380}$$

This says that

$$b_1 = \frac{\alpha}{\sqrt{1}} b_0 \ , \ b_2 = \frac{\alpha}{\sqrt{2}} b_1 = \frac{\alpha^2}{\sqrt{2!}} b_0 \tag{6.381}$$

or

$$b_n = \frac{\alpha^n}{\sqrt{n!}} b_0 \tag{6.382}$$

We thus get the final result

$$|\alpha\rangle = b_0 \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} |m\rangle \tag{6.383}$$

Let us now normalize this state (choose b_0). We have

$$\begin{aligned}
\langle \alpha \mid \alpha\rangle = 1 &= |b_0|^2 \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\alpha^{*m} \alpha^k}{\sqrt{m!}\sqrt{k!}} \langle k \mid m\rangle \\
&= |b_0|^2 \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\alpha^{*m} \alpha^k}{\sqrt{m!}\sqrt{k!}} \delta_{km} \\
&= |b_0|^2 \sum_{m=0}^{\infty} \frac{|\alpha|^2}{m!} = |b_0|^2 e^{|\alpha|^2}
\end{aligned} \tag{6.384}$$

which says that

$$b_0 = e^{-\frac{1}{2}|\alpha|^2} \tag{6.385}$$

and thus

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} |m\rangle \tag{6.386}$$

Now

$$\langle n \,|\, \alpha \rangle = \text{probability amplitude that the system in the state } |\alpha\rangle \text{ will be found in the state } |n\rangle$$

which then says that

$$P_n = |\langle n \mid \alpha \rangle|^2 = \frac{e^{-|\alpha|^2} |\alpha|^{2n}}{n!} = \frac{e^{-N} N^n}{n!} \tag{6.387}$$

is the probability that the system in the state $|\alpha\rangle$ will be found in the state $|n\rangle$, where we have defined $N = |\alpha|^2$.

We note that

$$\langle \alpha | \, \hat{a}^+ \hat{a} \, | \alpha \rangle = \left|\alpha^2\right| \langle \alpha \mid \alpha \rangle = \left|\alpha^2\right| = N = \langle \alpha | \, \hat{N}_{op} \, | \alpha \rangle \tag{6.388}$$

or N = the average value or expectation value of the $\hat{N}_{op}$ operator in the state $|\alpha\rangle$. This type of probability distribution is called a Poisson distribution, i.e., the state $|\alpha\rangle$ has the number states or energy eigenstates distributed in a *Poisson* manner.

Since the states $|n\rangle$ are energy eigenstates, we know their time dependence, i.e.,

$$|n, t\rangle = e^{-i \frac{E_n}{\hbar} t} \, |n\rangle \tag{6.389}$$

Therefore, we have for the time dependence of the state $|\alpha\rangle$

$$|\alpha, t\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} \, |m, t\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} e^{-i \frac{E_m}{\hbar} t} \, |m\rangle \tag{6.390}$$

This simple operation clearly indicates the fundamental importance of the energy eigenstates when used as a basis set.

If we are able to expand an arbitrary vector representing some physical system in the energy basis, then we immediately know the time dependence of that state vector and hence we know the time dependence of all the probabilities associated with the state vector and the system.

We will use this result to solve for the time dependence of real physical systems in later discussions.

Now let us try to understand the physics contained in the $|\alpha\rangle$ state vector. In a given energy eigenstate the expectation value of the position operator is given

by

$$\begin{aligned}\langle n,t|\,\hat{x}\,|n,t\rangle &= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n,t|\,(\hat{a}+\hat{a}^+)\,|n,t\rangle \\ &= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\,e^{i\frac{E_n}{\hbar}t}(\hat{a}+\hat{a}^+)e^{-i\frac{E_n}{\hbar}t}\,|n\rangle \\ &= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\,(\hat{a}+\hat{a}^+)\,|n\rangle \\ &= \sqrt{\frac{\hbar}{2m\omega_0}}\,\langle n|\left(\sqrt{n}\,|n-1\rangle+\sqrt{n+1}\,|n+1\rangle\right)=0\end{aligned}$$

i.e., it is equal to zero and is independent of time.

On the other hand, in the state $|\alpha\rangle$ we find

$$\langle\alpha,t|\,\hat{x}\,|\alpha,t\rangle = \sqrt{\frac{\hbar}{2m\omega_0}}\sum_m\sum_k b_m^* b_k e^{i\frac{(E_m-E_k)}{\hbar}t}\,\langle m|\,(\hat{a}+\hat{a}^+)\,|k\rangle \tag{6.391}$$

Now

$$\begin{aligned}\langle m|\,(\hat{a}+\hat{a}^+)\,|k\rangle &= \langle m|\left(\sqrt{k}\,|k-1\rangle+\sqrt{k+1}\,|k+1\rangle\right) \\ &= \sqrt{k}\delta_{m,k-1}+\sqrt{k+1}\delta_{m,k+1}\end{aligned} \tag{6.392}$$

Using this result we have

$$\begin{aligned}&\langle\alpha,t|\,\hat{x}\,|\alpha,t\rangle \\ &\quad= \sqrt{\frac{\hbar}{2m\omega_0}}\left(\sum_{k=1}^{\infty} b_{k-1}^* b_k\sqrt{k}e^{i\frac{(E_{k-1}-E_k)}{\hbar}t}+\sum_{k=0}^{\infty} b_{k+1}^* b_k\sqrt{k+1}e^{i\frac{(E_{k+1}-E_k)}{\hbar}t}\right) \\ &\quad= \sqrt{\frac{\hbar}{2m\omega_0}}\left(\sum_{k=1}^{\infty} b_{k-1}^* b_k\sqrt{k}e^{-i\omega_0 t}+\sum_{k=0}^{\infty} b_{k+1}^* b_k\sqrt{k+1}e^{i\omega_0 t}\right) \\ &\quad= \sqrt{\frac{\hbar}{2m\omega_0}}\left(\sum_{k=0}^{\infty} b_k^* b_{k+1}\sqrt{k}e^{-i\omega_0 t}+\sum_{k=0}^{\infty} b_{k+1}^* b_k\sqrt{k+1}e^{i\omega_0 t}\right) \\ &\quad= \sqrt{\frac{\hbar}{2m\omega_0}}b_0^2\left(\sum_{k=0}^{\infty}\frac{\alpha^{*k}\alpha^{k+1}}{\sqrt{(k+1)!k!}}\sqrt{k}e^{-i\omega_0 t}+\sum_{k=0}^{\infty}\frac{\alpha^{*k+1}\alpha^k}{\sqrt{(k+1)!k!}}\sqrt{k+1}e^{i\omega_0 t}\right) \\ &\quad= \sqrt{\frac{\hbar}{2m\omega_0}}b_0^2\sum_k\frac{1}{k!}\,|\alpha|^{2k}\left(\alpha e^{-i\omega_0 t}+\alpha^* e^{i\omega_0 t}\right)\end{aligned} \tag{6.393}$$

Now using $\alpha = |\alpha| e^{i\varphi}$ we get

$$\begin{aligned}\langle \alpha, t | \hat{x} | \alpha, t \rangle &= \sqrt{\frac{\hbar}{2m\omega_0}} b_0^2 2 |\alpha| \sum_k \frac{|\alpha|^{2k}}{k!} \mathrm{Real}(e^{i\phi} e^{-i\omega_0 t}) \\ &= 2x_0 |\alpha| \cos(\omega_0 t - \phi)(b_0^2 \sum_k \frac{|\alpha|^{2k}}{k!}) \quad , \quad x_0 = \sqrt{\frac{\hbar}{2m\omega_0}} \\ &= 2x_0 |\alpha| \cos(\omega_0 t - \phi) \end{aligned} \tag{6.394}$$

The expectation value of the position operator in the state $|\alpha\rangle$ is time-dependent and behaves like a *classical oscillator*.

Now let us look at these states in another way. We consider adding a term to the Hamiltonian of the form

$$V(\hat{x}) = -F_0 \hat{x} \tag{6.395}$$

so that

$$\hat{H} = \hbar\omega_0 (\hat{a}^+ \hat{a} + \frac{1}{2}) - F_0 x_0 (\hat{a} + \hat{a}^+) \tag{6.396}$$

This corresponds to putting a charged oscillator in an electric field and including the effective dipole interaction energy in the Hamiltonian.

In the corresponding classical mechanics problem, we have

$$H = \frac{p^2}{2m} + \frac{1}{2} k x^2 \rightarrow \text{ oscillator with equilibrium point at } x = 0$$

and if we add a linear term in x, we get

$$\begin{aligned} H &= \frac{p^2}{2m} + \frac{1}{2} k x^2 - F_0 x \\ &= \frac{p^2}{2m} + \frac{1}{2} k (x - \frac{F_0}{k})^2 - \frac{F_0^2}{2k} \\ &\rightarrow \text{ oscillator with equilibrium point at } x = \frac{F_0}{k} \end{aligned}$$

This suggests that we look for an oscillator solution to the quantum mechanical problem. In particular, we look for a new set of raising and lowering operators with the same commutator as the original set.

We can accomplish this by letting

$$\hat{A} = \hat{a} + \beta \text{ and } \hat{A}^+ = \hat{a}^+ + \beta \text{ where } \beta = \text{ a real number} \tag{6.397}$$

The commutator $[\hat{a}, \hat{a}^\dagger] = 1$ then implies that $[\hat{A}, \hat{A}^\dagger] = 1$. The Hamiltonian becomes

$$\begin{aligned} \hat{H} &= \hbar\omega_0 ((\hat{A}^+ - \beta)(\hat{A} - \beta) + \frac{1}{2}) - F_0 x_0 ((\hat{A} - \beta) + (\hat{A}^+ - \beta)) \\ &= \hbar\omega_0 (\hat{A}^+ \hat{A} + \frac{1}{2}) - \hbar\omega_0 \beta (\hat{A} + \hat{A}^+) + \hbar\omega_0 \beta^2 - F_0 x_0 (\hat{A} + \hat{A}^+) + 2F_0 x_0 \beta \end{aligned}$$

We are still free to choose β. We choose $\hbar\omega_0\beta = -F_0x_0$, which gives

$$\hat{H} = \hbar\omega_0(\hat{A}^+\hat{A} + \frac{1}{2}) - \frac{F_0^2x_0^2}{\hbar\omega_0} \tag{6.398}$$

which is just a harmonic oscillator with the energies shifted by a constant.

Now choose the new number states $|N\rangle$ where $\hat{A}^\dagger\hat{A}\,|N\rangle = N\,|N\rangle$ as before. We then have

$$\begin{aligned}\hat{H}\,|N\rangle &= (\hbar\omega_0(\hat{A}^+\hat{A} + \frac{1}{2}) - \frac{F_0^2x_0^2}{\hbar\omega_0})\,|N\rangle \\ &= (\hbar\omega_0(N + \frac{1}{2}) - \frac{F_0^2x_0^2}{\hbar\omega_0})\,|N\rangle = E_N\,|N\rangle\end{aligned} \tag{6.399}$$

or we get the same set of energy levels as before, except that they are all displaced by the same constant amount

$$-\frac{F_0^2x_0^2}{\hbar\omega_0}$$

The new ground state is given by

$$\hat{A}\,|N=0\rangle = (\hat{a}+\beta)\,|N=0\rangle = 0 \tag{6.400}$$

$$\hat{a}\,|N=0\rangle = -\beta\,|N=0\rangle = \frac{F_0x_0}{\hbar\omega_0}\,|N=0\rangle \tag{6.401}$$

The new ground state is just an eigenstate of $\hat{a}$ with eigenvalue

$$\alpha = \frac{F_0x_0}{\hbar\omega_0}$$

i.e., the new ground state is a *coherent state*. These states are called *coherent* states because of their properties when they appear in the quantum mechanical theory of the laser. We will discuss these states in many other contexts later.

Now let us get this state another way that illustrates some new powerful tools.

Using the Translation Operator

In general, a displaced state $|\lambda\rangle$ is given in terms of the displacement operator (in one dimension) by

$$|\lambda\rangle = e^{-\frac{i}{\hbar}\hat{p}\lambda}\,|0\rangle \tag{6.402}$$

For the harmonic oscillator system

$$\hat{p} = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a} - \hat{a}^+) \tag{6.403}$$

If we choose $|0\rangle$ to be the ground state of the oscillator, then we have for the corresponding displaced ground-state

$$|\lambda\rangle = e^{\sqrt{\frac{m\omega}{2\hbar}}(\hat{a}^+ - \hat{a})\lambda}\,|0\rangle \tag{6.404}$$

By Glauber's theorem(see last section of this chapter for a derivation)

$$e^{(\hat{A}+\hat{B})} = e^{\hat{A}}e^{\hat{B}}e^{-\frac{1}{2}[\hat{A},\hat{B}]} \tag{6.405}$$

we have

$$\begin{aligned} e^{\sqrt{\frac{m\omega}{2\hbar}}(\hat{a}^+ - \hat{a})\lambda} &= e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda}e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda}e^{\frac{1}{2}\frac{m\omega}{2\hbar}[\hat{a}^+,\hat{a}]\lambda^2} \\ &= e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda}e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda}e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^2} \end{aligned} \tag{6.406}$$

and thus

$$|\lambda\rangle = e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda}e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda}e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^2}\,|0\rangle \tag{6.407}$$

Now

$$\begin{aligned} e^{-\sqrt{\frac{m\omega}{2\hbar}}\hat{a}\lambda}\,|0\rangle &= \left(\hat{I} + \left(-\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}\right) + \frac{1}{2}\left(-\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}\right)^2 + \ldots\right)|0\rangle \\ &= |0\rangle \end{aligned} \tag{6.408}$$

where we used $\hat{a}\,|0\rangle = 0$. Similarly, using $(\hat{a}^\dagger)^n\,|0\rangle = \sqrt{n!}\,|n\rangle$ we have

$$\begin{aligned} e^{\sqrt{\frac{m\omega}{2\hbar}}\hat{a}^+\lambda}\,|0\rangle &= \left(\hat{I} + \left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}^+\right) + \frac{1}{2}\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\hat{a}^+\right)^2 + \ldots\right)|0\rangle \\ &= |0\rangle + \sqrt{\frac{m\omega}{2\hbar}}\lambda\,|1\rangle + \frac{1}{2}\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^2|2\rangle + \ldots \\ &= \sum_{n=0}^{\infty}\frac{\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^n}{\sqrt{n!}}\,|n\rangle \end{aligned} \tag{6.409}$$

or

$$|\lambda\rangle = e^{-\frac{1}{4}\frac{m\omega}{\hbar}\lambda^2}\sum_{n=0}^{\infty}\frac{\left(\sqrt{\frac{m\omega}{2\hbar}}\lambda\right)^n}{\sqrt{n!}}\,|n\rangle \tag{6.410}$$

Thus,

$$|\lambda\rangle = \sum_{n=0}^{\infty} b_n\,|n\rangle \tag{6.411}$$

where

$$b_n = \frac{e^{-\frac{N}{2}}N^{\frac{n}{2}}}{\sqrt{n!}} \quad , \quad \frac{N}{2} = \frac{m\omega}{4\hbar}\lambda^2 \tag{6.412}$$

or

$$\begin{aligned} P_n &= \text{ probability of find the system in the state } |n\rangle \\ &= |b_n|^2 = \frac{e^{-N}N^n}{n!} \end{aligned} \tag{6.413}$$

which is a Poisson distribution. We, thus, obtain the coherent states once again.

6.7 Green's Functions

In this section we will give a simple introduction to the Green's function technique for solving differential equations. We will return to this subject later on in more detail. The general form of the differential equations we have been solving is

$$\hat{L}\psi(x) = Q(x) \text{ where } \hat{L} = E + \frac{\hbar^2}{2m}\frac{d^2}{dx^2} \text{ and } Q(x) = V(x)\psi(x) \tag{6.414}$$

Let us define a new function $G(x, x')$ such that

$$\hat{L}G(x, x') = \delta(x - x') \tag{6.415}$$

Then the solution to the original differential equation is then given by

$$\psi(x) = \psi_0(x) + \int_{-\infty}^{\infty} G(x, x')Q(x')\,dx' \tag{6.416}$$

where $\psi_0(x)$ is any function that satisfies the homogeneous equation

$$\hat{L}\psi_0(x) = 0 \tag{6.417}$$

Proof:

$$\begin{aligned}\hat{L}\psi(x) &= \hat{L}\psi_0(x) + \hat{L}\int G(x, x')Q(x')dx' \\ &= 0 + \int (\hat{L}G(x, x'))Q(x')dx' = \int \delta(x - x')Q(x')dx' = Q(x)\end{aligned}$$

One representation of the delta function is given by

$$\delta(x - x') = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{\frac{i}{\hbar}p(x-x')}dp \tag{6.418}$$

and if we assume that the Fourier transform of $G(x, x')$ exists (label it $\mathfrak{I}$), then we can write

$$G(x, x') = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{\frac{i}{\hbar}p(x-x')}\mathfrak{I}(p)dp \tag{6.419}$$

Substituting into the differential equation for $G(x,x')$ we get

$$\begin{aligned}
\hat{L}G(x,x') &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}(\hat{L}e^{\frac{i}{\hbar}p(x-x')})\mathfrak{I}(p)dp \\
&= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}((E+\frac{\hbar^2}{2m}\frac{d^2}{dx^2})e^{\frac{i}{\hbar}p(x-x')})\mathfrak{I}(p)dp \\
&= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}((E-\frac{p^2}{2m})e^{\frac{i}{\hbar}p(x-x')})\mathfrak{I}(p)dp \\
&= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}e^{\frac{i}{\hbar}p(x-x')}\left(E-\frac{p^2}{2m}\right)\mathfrak{I}(p)dp \\
&= \delta(x-x') = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}e^{\frac{i}{\hbar}p(x-x')}dp
\end{aligned}$$

This says that

$$\mathfrak{I}(p) = \frac{1}{E-\frac{p^2}{2m}} \tag{6.420}$$

and hence

$$G(x,x') = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}e^{\frac{i}{\hbar}p(x-x')}\frac{1}{E-\frac{p^2}{2m}}dp \tag{6.421}$$

and

$$\psi(x) = e^{\frac{i}{\hbar}px} + \int_{-\infty}^{\infty}\left[\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}e^{\frac{i}{\hbar}p(x-x')}\frac{1}{E-\frac{p^2}{2m}}dp\right]Q(x')dx' \tag{6.422}$$

where we have chosen

$$\psi_0(x) = e^{\frac{i}{\hbar}px} \tag{6.423}$$

as the homogeneous solution for $E > 0$.

We can rearrange this as the expression

$$\psi(x) = e^{\frac{i}{\hbar}px} + \int_{-\infty}^{\infty}\frac{dp}{2\pi\hbar}\frac{e^{\frac{i}{\hbar}px}}{E-\frac{p^2}{2m}}\left[\int_{-\infty}^{\infty}e^{-\frac{i}{\hbar}px'}Q(x')dx'\right] \tag{6.424}$$

Example #1 - Single Delta Function

Let us choose $Q(x) = u\delta(x)\psi(x)$, which is a delta function barrier at the origin.

We then have

$$\psi(x) = e^{\frac{i}{\hbar}px} + \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}px}}{E - \frac{p^2}{2m}} \left[\int_{-\infty}^{\infty} e^{-\frac{i}{\hbar}px'} u\delta(x')\psi(x')dx' \right]$$
$$= e^{\frac{i}{\hbar}px} + u\psi(0) \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}px}}{E - \frac{p^2}{2m}} \tag{6.425}$$

How do we evaluate this integral?

There seems to be a mathematical problem since the integrand has poles when

$$p = \pm\sqrt{2mE} \tag{6.426}$$

But being physicists, we know that this problem has a sensible finite solution (we already solved it earlier).

The key here is that we seemingly have a solution to the original differential equation with no unknown constants. However, we have not imposed any boundary conditions yet.

This is a solution to a 2^{nd} order differential equation and any complete solution always requires two boundary conditions.

In this case, the boundary conditions are replaced by a choice of the path of integration or a choice of how we decide to avoid the two poles.

We must treat the integration variable p as a complex variable with valid physics represented along the real axis. Once we accept this idea, we can then complete the definition of the integral by choosing a contour of integration in the complex p-plane.

This choice is equivalent to setting boundary conditions, as we shall see.

This is a very important lesson to learn.

We must always remember that we are physicists dealing with real physical systems. The equations describing these real systems must have solutions since the systems exist. When we run into a mathematical dead-end, the way out of the dilemma is usually to invoke physical arguments. This case of the choice of contours is just the first example of many that we shall see. The creation of the delta function by Dirac was another such case. If this is to make sense, then the possible contour choices should be able to reproduce all of the relevant physics that we obtained earlier using boundary conditions.

Let us see how. The poles are on the real axis as shown in Figure 6.12 below.

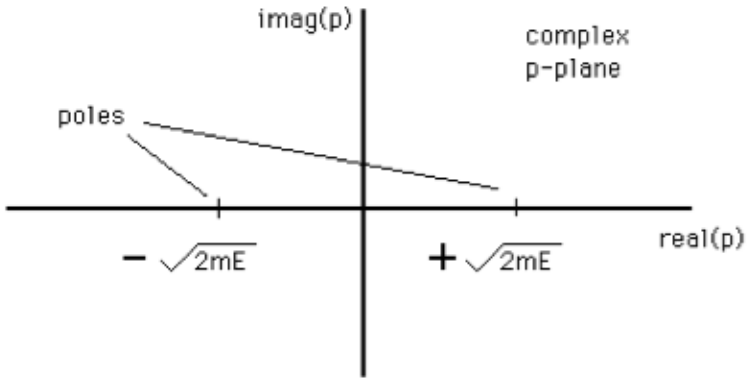


Figure 6.12: Complex p-plane

There are eight possible contour choices which are shown in Figure 6.13 below.

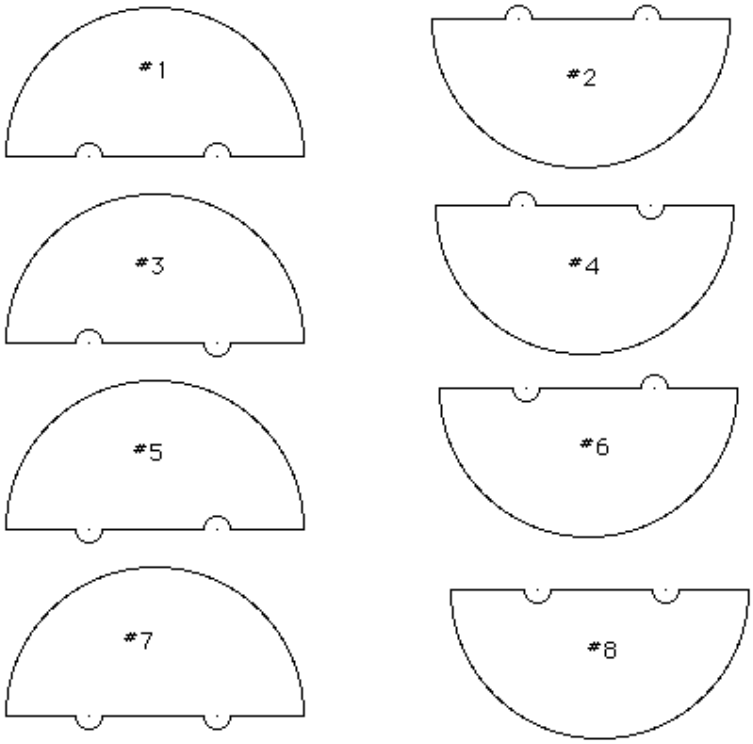


Figure 6.13: Possible Contours

We are looking for the contour that corresponds to the boundary conditions where we end up with only outgoing waves, i.e., where the integral behaves like a wave traveling to the right (the transmitted wave) for $x > 0$ and a wave traveling to the left (the reflected wave) for $x < 0$.

This is accomplished by choosing contour #3 for $x > 0$ and contour #4 for $x < 0$. In each case, the contour encloses only one pole and using the method of residues we can easily evaluate the integral.

We get

$$\int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}px}}{E - \frac{p^2}{2m}} = \frac{m}{i\hbar} \left. \frac{e^{\frac{i}{\hbar}px}}{\sqrt{2mE} + p} \right|_{p=\sqrt{2mE}} = \frac{m}{i\hbar\sqrt{2mE}} e^{\frac{i}{\hbar}\sqrt{2mE}x} \quad , x > 0$$

$$\int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}px}}{E - \frac{p^2}{2m}} = \frac{m}{i\hbar} \left. \frac{e^{\frac{i}{\hbar}px}}{\sqrt{2mE} - p} \right|_{p=-\sqrt{2mE}} = \frac{m}{i\hbar\sqrt{2mE}} e^{-\frac{i}{\hbar}\sqrt{2mE}x} \quad , x < 0$$

The complete solution to the problem is then (using the absolute value function)

$$\psi(x) = \psi_{inc}(x) + \frac{m|u|}{ip\hbar} e^{\frac{i}{\hbar}p|x|}\psi(0) \text{ where } p = \sqrt{2mE} \tag{6.427}$$

Let $x = 0$ in this result and we can solve for $\psi(0)$

$$\psi(0) = \frac{ip\hbar}{ip\hbar - m|u|} \rightarrow \psi(x) = e^{\frac{i}{\hbar}px} + \frac{m|u|}{ip\hbar - m|u|} e^{\frac{i}{\hbar}p|x|} \tag{6.428}$$

We then have

$$\psi_{trans}(x) = \frac{m|u|}{ip\hbar - m|u|} e^{\frac{i}{\hbar}px} \tag{6.429}$$

and

$$S(E) = \text{ transmission amplitude } = \frac{m|u|}{ip\hbar - m|u|} \tag{6.430}$$

and

$$T(E) = \text{ transmission probability } = \frac{E}{E + \frac{mu^2}{2\hbar^2}} \tag{6.431}$$

which is the same result as we found earlier.

Example #2 - Double Delta Function

Let us choose $Q(x) = u(\delta(x+a) + \delta(x-a))\psi(x)$. We then have

$$\begin{aligned}\psi(x) &= e^{\frac{i}{\hbar}px} + \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}px}}{E - \frac{p^2}{2m}} \left[\int_{-\infty}^{\infty} e^{-\frac{i}{\hbar}px'} u(\delta(x'+a) + \delta(x'-a))\psi(x')dx' \right] \\ &= e^{\frac{i}{\hbar}px} + u\psi(a) \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}p(x-a)}}{E - \frac{p^2}{2m}} + u\psi(-a) \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{e^{\frac{i}{\hbar}p(x+a)}}{E - \frac{p^2}{2m}}\end{aligned}$$

Both of these integrals are evaluated using contour #3 for $x > a$(we get outgoing waves moving towards the right only from both terms).

We then get

$$\psi(x) = e^{\frac{i}{\hbar}px} + \frac{mu}{ip\hbar} \left[\psi(a) e^{\frac{i}{\hbar}p(x-a)} + \psi(-a) e^{\frac{i}{\hbar}p(x+a)} \right] \text{ where } p = \sqrt{2mE} \tag{6.432}$$

If we find $\psi(a)$ and $\psi(-a)$ by substituting $\pm a$, we can then evaluate the transmission probability as in the single delta function example. We again get the same result as before.

More about Green's functions later.

6.8 Charged Particles in Electromagnetic Fields

If a charged particle moves in an electromagnetic field, the Lorentz force

$$\vec{F} = q\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right) \tag{6.433}$$

acts on the particle, where q = charge, $\vec{E}$ = electric field, $\vec{v}$ = velocity and $\vec{B}$ = magnetic field.

The electric and magnetic fields can be expressed in terms of the electromagnetic vector and scalar potentials

$$\vec{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} \text{ and } \vec{B} = \nabla \times \vec{A} \tag{6.434}$$

In order to make the transition to a quantum mechanical discussion, we must determine the Lagrangian and Hamiltonian for the charged particle in an electromagnetic field.

The appearance of the velocity vector $\vec{v}$ in the Lorentz force law means that we will need to work with a velocity-dependent potential function. If we insert the expressions for the electric and magnetic fields into the Lorentz force law we get

$$\vec{F} = q\left(-\nabla\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} + \frac{1}{c}\vec{v} \times \left(\nabla \times \vec{A}\right)\right) \tag{6.435}$$

Using the vector identity

$$\vec{B} \times \left(\nabla \times \vec{C}\right) = \nabla\left(\vec{B} \cdot \vec{C}\right) - \left(\vec{B} \cdot \nabla\right)\vec{C} - \left(\vec{C} \cdot \nabla\right)\vec{B} - \vec{C} \times \left(\nabla \times \vec{B}\right) \tag{6.436}$$

we get

$$\vec{v} \times \left(\nabla \times \vec{A}\right) = \nabla\left(\vec{v} \cdot \vec{A}\right) - \left(\vec{v} \cdot \nabla\right)\vec{A} \tag{6.437}$$

where we have used the fact that $\vec{v}$ is not an explicit function of the position vector $\vec{r}$.

Now, the total derivative relationship is given by

$$\begin{aligned} \frac{d\vec{A}}{dt} &= \frac{\partial \vec{A}}{\partial t} + \frac{\partial \vec{A}}{\partial x}\frac{dx}{dt} + \frac{\partial \vec{A}}{\partial y}\frac{dy}{dt} + \frac{\partial \vec{A}}{\partial z}\frac{dz}{dt} \\ &= \frac{\partial \vec{A}}{\partial t} + (\vec{v} \cdot \nabla)\vec{A} \end{aligned} \tag{6.438}$$

Putting all this together we get

$$\vec{F} = q\left(-\nabla\varphi + \frac{1}{c}\nabla(\vec{v} \cdot \vec{A}) - \frac{1}{c}\frac{d\vec{A}}{dt}\right) \tag{6.439}$$

The standard Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \text{ where } L = T - V = \text{ Lagrangian} \tag{6.440}$$

works for a potential V which does not depend on the velocity. However, when the force law takes the following form

$$F_j = -\frac{\partial U}{\partial x_j} + \frac{d}{dt}\left(\frac{\partial U}{\partial \dot{x}_j}\right) \tag{6.441}$$

we can still construct a Lagrange equation of the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \text{ where } L = T - U = \text{ Lagrangian} \tag{6.442}$$

We can get the Lorentz force law into this form by using a mathematical trick

$$\frac{dA_j}{dt} = \frac{d}{dt}\left(\frac{\partial}{\partial v_j}(\vec{v}\cdot\vec{A})\right) = \frac{d}{dt}\left(\frac{\partial}{\partial v_j}(\vec{v}\cdot\vec{A} - q\phi)\right) \tag{6.443}$$

where we have used the fact that the potential ϕ is independent of the velocity, to obtain

$$F_x = -\frac{\partial}{\partial x}\left(q\phi - \frac{q}{c}\vec{v}\cdot\vec{A}\right) + \frac{d}{dt}\left(\frac{\partial}{\partial v_x}\left(q\phi - \frac{q}{c}\vec{v}\cdot\vec{A}\right)\right) \tag{6.444}$$

This says that we need to choose the generalized potential U as

$$U = q\phi - \frac{q}{c}\vec{v}\cdot\vec{A} \tag{6.445}$$

and get the Lagrangian

$$L = T - U = \frac{1}{2}m\vec{v}^2 - q\phi + \frac{q}{c}\vec{v}\cdot\vec{A} \tag{6.446}$$

This gives the canonical momentum

$$\vec{p} = \frac{\partial L}{\partial \vec{v}} = m\vec{v} + \frac{q}{c}\vec{A} \tag{6.447}$$

The canonical momentum is *NOT* the simply the mechanical momentum but now includes an extra term proportional to the vector potential(the momentum of the field!).

The Hamiltonian is derived from the Lagrangian using

$$H = \vec{p}\cdot\vec{v} - L = \frac{1}{2m}\left(\vec{p} - \frac{q}{c}\vec{A}\right)^2 + q\varphi \tag{6.448}$$

This says that the simplest way of coupling the electromagnetic field to the motion of the particle is to replace the momentum by the quantity

$$\vec{p} - \frac{q}{c}\vec{A} \tag{6.449}$$

which includes the momentum associated with the electromagnetic field.

This method of introducing the electromagnetic field into the equations of motion of a particle is called minimal coupling. We will see its relationship to gauge invariance shortly.

The transition to quantum mechanics is now done in the standard manner.

$$\hat{H}\left|\psi(t)\right\rangle = E\left|\psi(t)\right\rangle = ih\frac{\partial}{\partial t}\left|\psi(t)\right\rangle \tag{6.450}$$

$$\left\langle\vec{r}\right|\hat{H}\left|\psi(t)\right\rangle = E\left\langle\vec{r}\mid\psi(t)\right\rangle = ih\frac{\partial}{\partial t}\left\langle\vec{r}\mid\psi(t)\right\rangle \tag{6.451}$$

$$\left\langle\vec{r}\right|\hat{H}\left|\psi(t)\right\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t} \tag{6.452}$$

where

$$\hat{H} = \frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{q}{c}\vec{A}(\hat{\mathbf{r}},t)\right)^2 + q\phi(\hat{\mathbf{r}},t) \tag{6.453}$$

Using

$$\hat{\mathbf{p}} = \frac{\hbar}{i}\nabla \tag{6.454}$$

we then get the Schrodinger equation

$$\left\langle\vec{r}\right|\left[\frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{q}{c}\vec{A}(\hat{\mathbf{r}},t)\right)^2 + q\phi(\hat{\mathbf{r}},t)\right]\left|\psi(t)\right\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t} \tag{6.455}$$

$$\left[\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}(\vec{r},t)\right)^2 + q\phi(\vec{r},t)\right]\left\langle\vec{r}\mid\psi(t)\right\rangle = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t} \tag{6.456}$$

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}(\vec{r},t)\right)^2\psi(\vec{r},t) + q\phi(\vec{r},t)\psi(\vec{r},t) = E\psi(\vec{r},t) = ih\frac{\partial\psi(\vec{r},t)}{\partial t} \tag{6.457}$$

The Schrodinger equation for a charged particle in an electromagnetic field must be invariant under a *gauge transformation* (since Maxwell's equations are so invariant), which is represented by the relations

$$\vec{A} \to \vec{A}' = \vec{A} + \nabla S(\vec{r},t) \tag{6.458}$$

$$\phi \to \phi' = \phi - \frac{1}{c}\frac{\partial S(\vec{r},t)}{\partial t} \tag{6.459}$$

where $S(\vec{r},t)$ is an arbitrary function. In order for the Schrodinger equation to be invariant it must be transformed to

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}'(\vec{r},t)\right)^2\psi'(\vec{r},t) + q\phi'(\vec{r},t)\psi'(\vec{r},t) = ih\frac{\partial\psi'(\vec{r},t)}{\partial t} \tag{6.460}$$

where the wave function is only changed by an overall phase factor, i.e.,

$$\psi'(\vec{r},t) = \psi(\vec{r},t)e^{\frac{iq}{\hbar c}S(\vec{r},t)} \tag{6.461}$$

Proof:

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}'(\vec{r},t)\right)^2 \psi'(\vec{r},t) + q\phi'(\vec{r},t)\psi'(\vec{r},t) = ih\frac{\partial\psi'(\vec{r},t)}{\partial t} \tag{6.462}$$

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A} - \frac{q}{c}\nabla S\right)^2 \psi e^{\frac{iq}{\hbar c}S} + (q\phi - \frac{q}{c}\frac{\partial S}{\partial t})\psi e^{\frac{iq}{\hbar c}S} = ih\frac{\partial\psi e^{\frac{iq}{\hbar c}S}}{\partial t} \tag{6.463}$$

Using the identities

$$\frac{\partial\psi e^{\frac{iq}{\hbar c}S}}{\partial t} = \frac{\partial\psi}{\partial t}e^{\frac{iq}{\hbar c}S} + \frac{iq}{\hbar c}\frac{\partial S}{\partial t}\psi e^{\frac{iq}{\hbar c}S} \tag{6.464}$$

$$\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A} - \frac{q}{c}\nabla S\right)\left(\psi e^{\frac{iq}{\hbar c}S}\right) = e^{\frac{iq}{\hbar c}S}\left(\frac{\hbar}{i}\nabla - \frac{q}{c}\vec{A}\right)\psi \tag{6.465}$$

we get the original Schrodinger equation back.

This result shows that the solutions of the gauge-transformed Schrodinger equation still describe the same physical states.

The wave functions or state vectors differ by a phase factor that depends on space and time and thus, the invariance is *LOCAL* rather than *GLOBAL* (a phase factor independent of space and time).

It is then clear that it is *NOT* the canonical momentum $\hat{\mathbf{p}} = -i\hbar\nabla$ (whose expectation value is *NOT gauge invariant*), but the genuine kinetic momentum

$$\hat{p} - \frac{q}{c}\vec{A}(\hat{r},t) \tag{6.466}$$

(whose expectation value *IS* gauge invariant), that represents a *measurable* quantity.

In any physical system, if the momentum operator $\hat{p}$ appears, then it must always be replaced by

$$\hat{p} - \frac{q}{c}\vec{A}(\hat{r},t)$$

if we turn on an electromagnetic field. This is the only way to guarantee gauge invariance in quantum mechanics.

Quantum mechanics + electromagnetism requires this *minimal coupling* for gauge invariance to be valid. It is important to remember that local gauge transformations of the electromagnetic potentials requires that wave functions or state vectors transform via local phase changes, which means a phase factor that depends on space and time.

Feynman showed that the effect of turning on a magnetic field is to multiply

the wave functions or probability amplitudes by the phase factor

$$\exp\left[\frac{iq}{\hbar c}\int\limits_{path} d\vec{\ell}\cdot\vec{A}(\vec{r},t)\right] \tag{6.467}$$

The Aharonov-Bohm Effect

The extra phase factor has a rather striking and unexpected consequence.

Let $\vec{A}$ be independent of t and consider the possible interference between the motions along two different paths as shown in Figure 6.14 below.

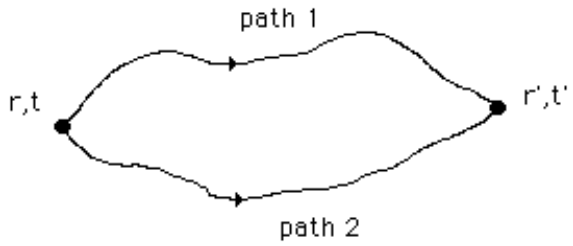


Figure 6.14: 2-Path Experiment

When a magnetic field is turned on, the relative phase for the two paths (1 and 2) is changed by a factor

$$\frac{q}{\hbar c}\left[\int\limits_{path1} d\vec{\ell}\cdot\vec{A} - \int\limits_{path2} d\vec{\ell}\cdot\vec{A}\right] = \frac{q\Phi}{\hbar c} \tag{6.468}$$

where Φ = the total magnetic flux passing through the loop formed by the two paths or the flux enclosed by the loop (this is just Ampere's law).

Now consider the experimental setup shown in Figure 6.15 below.

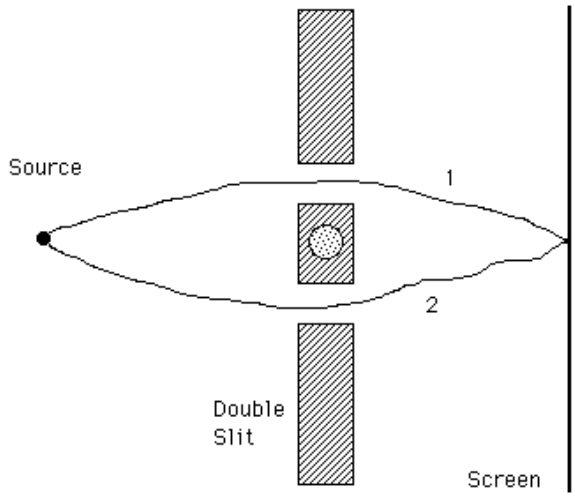


Figure 6.15: Aharonov-Bohm Experiment

The cylindrical solenoid between the two slits generates a magnetic field out of the page and confined to the region of the solenoid. This implies that the particles are traveling in regions *free* of any magnetic fields.

The *enclosed flux* is *NOT* zero, however. As the flux is increased, the relative phase between the two paths changes as described above and the diffraction pattern on the screen changes. This occurs even though there are no magnetic fields present along the paths.

There does exist, however, a nonzero vector potential in these regions and the relative phase is changing because the vector potential is changing.

This illustrates most powerfully that it is the electromagnetic potential, rather than the electromagnetic fields, as one might assume from Maxwell's equations, that are the fundamental physical quantities in quantum mechanics. This is called the *Aharonov-Bohm effect*.

Classically, the electric and magnetic fields, $\vec{E}$ and $\vec{B}$ are the physically relevant quantities, since they determine the Lorentz force. In regions where the electric and magnetic fields are zero, a charged particle feels no force. The vector and scalar potentials $\vec{A}$ and ϕ serve in classical physics only as auxiliary quantities.

In quantum mechanics, the vector potential $\vec{A}(\vec{r})$ is the fundamental physical field. The wave function, however, always has the property that physical variables and physical effects only depend on gauge-invariant quantities.

More about this later.

6.9 Time Evolution of Probabilities

Now that we have solved for the eigenvectors and eigenvalues of several systems we will return to a discussion of the time evolution of probability amplitudes. Along the way we will reinforce the details about the most fundamental procedure in quantum mechanics.

Suppose we have some physical system that is describable by quantum mechanics. We would like to be able to answer the following fundamental question?

> If the system is in the state $|\alpha\rangle$ at time $t = 0$,
> what is the probability that it will be in the
> state $|\beta\rangle$ at time t?

These two states are usually defined as follows:

$|\alpha\rangle$ = initial state = state prepared by experimenter
$|\beta\rangle$ = final state = state defined by measuring device

The formal answer to the question is

$$P_{\beta\alpha}(t) = |Amplitude_{\beta\alpha}(t)|^2 = |A_{\beta\alpha}(t)|^2 = \left|\langle\beta|\,\hat{U}(t)\,|\alpha\rangle\right|^2 \tag{6.469}$$

where $\hat{U}(t)$ is the time evolution operator. Let us assume for simplicity at this stage that the Hamiltonian $\hat{H}$ does not explicitly dependent on time t. From our earlier discussions, we then have

$$\hat{U}(t) = e^{-i\hat{H}t/\hbar} \tag{6.470}$$

so that the probability amplitude becomes

$$A_{\beta\alpha}(t) = \langle\beta|\,\hat{U}(t)\,|\alpha\rangle = \langle\beta|\,e^{-\frac{i}{\hbar}\hat{H}t}\,|\alpha\rangle \tag{6.471}$$

The most important rule in quantum mechanics is the following:

> If you are evaluating an amplitude involving a
> particular operator, $\hat{H}$ in this case, then switch
> to a basis for the space corresponding to the
> eigenvectors of the that operator.

This basis is called the *HOME* space.

This is carried out by inserting *identity* operators as sums of *projection* operators

written in the energy eigenvector basis:

$$\begin{aligned} A_{\beta\alpha}(t) &= \langle\beta| e^{-\frac{i}{\hbar}\hat{H}t} |\alpha\rangle = \langle\beta| \hat{I} e^{-\frac{i}{\hbar}\hat{H}t} \hat{I} |\alpha\rangle \\ &= \sum_E \sum_{E'} \langle\beta \mid E\rangle \langle E| e^{-\frac{i}{\hbar}\hat{H}t} |E'\rangle \langle E' \mid \alpha\rangle \end{aligned} \tag{6.472}$$

where

$$\hat{H}|E\rangle = E|E\rangle \text{ and } \hat{H}|E'\rangle = E'|E'\rangle \tag{6.473}$$

$$\langle E \mid E'\rangle = \delta_{EE'} \tag{6.474}$$

$$\sum_E |E\rangle\langle E| = \hat{I} = \sum_{E'} |E'\rangle\langle E'| \tag{6.475}$$

It is clear that the time evolution calculation is most easily done in this basis.

We then have

$$\begin{aligned} A_{\beta\alpha}(t) &= \sum_E \sum_{E'} \langle\beta \mid E\rangle \langle E| e^{-\frac{i}{\hbar}E't} |E'\rangle \langle E' \mid \alpha\rangle \\ &= \sum_E \sum_{E'} e^{-\frac{i}{\hbar}E't} \langle\beta \mid E\rangle \langle E \mid E'\rangle \langle E' \mid \alpha\rangle \\ &= \sum_E \sum_{E'} e^{-\frac{i}{\hbar}E't} \langle\beta \mid E\rangle \delta_{EE'} \langle E' \mid \alpha\rangle \\ &= \sum_E e^{-\frac{i}{\hbar}E't} \langle\beta \mid E\rangle \langle E \mid \alpha\rangle \end{aligned} \tag{6.476}$$

Let us now describe a typical procedure:

1. Determine the initial state of the system $= |\alpha\rangle$
2. Expand the initial state in terms of the energy basis
$$|\alpha\rangle = \sum_{E''} |E''\rangle\langle E''|\alpha\rangle$$
which means we know the coefficients $\langle E''|\alpha\rangle$
3. Expand the final state in terms of the energy basis
$$|\beta\rangle = \sum_{E''} |E''\rangle\langle E''|\beta\rangle$$
which means we know the coefficients $\langle E''|\beta\rangle$
4. Calculate the amplitude and probability.

Example - Infinite Square Well System

Suppose that a particle in an infinite square well $[-a/2, +a/2]$ is initially in the state $|\alpha\rangle$, such that

$$\langle x|\alpha\rangle = \begin{cases} 0 & x < 0 \\ \frac{2}{\sqrt{a}} & 0 \le x \le \frac{a}{4} \\ 0 & x > \frac{a}{4} \end{cases} \tag{6.477}$$

Now we have found that

$$\langle x \,|\, E \rangle = \begin{cases} \frac{2}{\sqrt{a}} \cos \frac{n\pi x}{a} & n \text{ odd} \\ \frac{2}{\sqrt{a}} \sin \frac{n\pi x}{a} & n \text{ even} \end{cases} \tag{6.478}$$

where

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \qquad n = 1, 2, 3, 4, \ldots \tag{6.479}$$

Therefore, we get

$$\begin{aligned} \langle E_n \mid \alpha \rangle &= \int_{-\frac{a}{2}}^{\frac{a}{2}} \langle E_n \mid x \rangle \langle x \mid \alpha \rangle \, dx = \frac{2}{\sqrt{a}} \int_0^{\frac{a}{4}} \langle E_n \mid x \rangle \, dx \\ &= \frac{2\sqrt{2}}{n\pi} \begin{cases} \sin \frac{n\pi}{4} & n \text{ odd} \\ -\cos \frac{n\pi}{4} & n \text{ even} \end{cases} \end{aligned} \tag{6.480}$$

or

$$\begin{aligned} |\alpha\rangle &= \langle E_1 \mid \alpha \rangle |E_1\rangle + \langle E_2 \mid \alpha \rangle |E_2\rangle + \langle E_3 \mid \alpha \rangle |E_3\rangle + \ldots \\ &= \frac{2}{\pi} |E_1\rangle + \frac{2}{3\pi} |E_3\rangle + \frac{2\sqrt{2}}{4\pi} |E_4\rangle + \frac{2}{5\pi} |E_5\rangle - \frac{2}{7\pi} |E_7\rangle - \frac{2\sqrt{2}}{8\pi} |E_8\rangle + \ldots \end{aligned} \tag{6.481}$$

If we assume that the final state $|\beta\rangle$ is an energy eigenstate, say $|E_n\rangle$, we then have the probability amplitude that the particle started out in state $|\alpha\rangle$ at $t = 0$ is found in the state $|E_n\rangle$ at time t is

$$A_{n\alpha}(t) = \sum_E e^{-\frac{i}{\hbar} Et} \langle E_n \mid E \rangle \langle E \mid \alpha \rangle = e^{-\frac{i}{\hbar} E_n t} \langle E_n \mid \alpha \rangle \tag{6.482}$$

and the corresponding probabilities are

$$P_{n\alpha}(t) = |\langle E_n \mid \alpha \rangle|^2 \tag{6.483}$$

which is independent of time. A plot of these fixed probabilities versus n is shown in Figure 6.16 below.

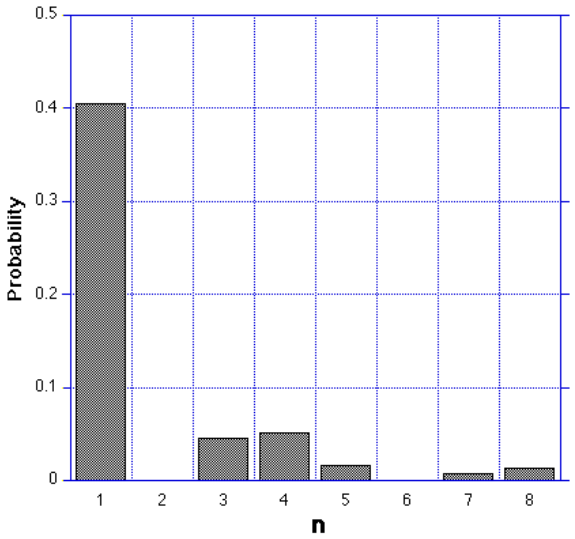


Figure 6.16: Probabilities

If, on the other hand, we assume that the final state is a linear combination of two energy eigenstates, say,

$$|\beta\rangle = \frac{1}{\sqrt{2}}|E_1\rangle + \frac{1}{\sqrt{2}}|E_3\rangle \tag{6.484}$$

then we have the probability amplitude that the particle started out in state $|\alpha\rangle$ at $t = 0$ is found in the state $|\beta\rangle$ at time t is

$$\begin{aligned} A_{\beta\alpha}(t) &= \sum_E e^{-\frac{i}{\hbar}Et}\langle\beta \mid E\rangle\langle E \mid \alpha\rangle \\ &= \frac{1}{\sqrt{2}}\left(e^{-\frac{i}{\hbar}E_1 t}\langle E_1 \mid \alpha\rangle + e^{-\frac{i}{\hbar}E_3 t}\langle E_3 \mid \alpha\rangle\right) \\ &= \frac{1}{\sqrt{2}}|E_1\rangle + \frac{1}{\sqrt{2}}|E_3\rangle \end{aligned} \tag{6.485}$$

and the corresponding probability is

$$\begin{aligned} P_{\beta\alpha}(t) = |\langle\beta \mid \alpha\rangle|^2 &= \frac{1}{2}\left|e^{-\frac{i}{\hbar}E_1 t}\langle E_1 \mid \alpha\rangle + e^{-\frac{i}{\hbar}E_3 t}\langle E_3 \mid \alpha\rangle\right|^2 \\ &= \frac{1}{2}\left|e^{-\frac{i}{\hbar}E_1 t}\frac{2}{\pi} + e^{-\frac{i}{\hbar}E_3 t}\frac{2}{3\pi}\right|^2 = \frac{2}{\pi^2}\left|\frac{1}{3} + e^{\frac{i}{\hbar}(E_3 - E_1)t}\right|^2 \\ &= \frac{2}{\pi^2}\left(\frac{1}{9} + \frac{2}{3}\cos\left[\frac{1}{\hbar}(E_3 - E_1)t\right] + 1\right) \\ &= \frac{20}{9\pi^2}\left(1 + \frac{3}{5}\cos\left[\frac{1}{\hbar}(E_3 - E_1)t\right]\right) \end{aligned} \tag{6.486}$$

which is dependent on time. In this case the probability of observing this final state oscillates between 0.18 and 0.72.

6.10 Numerical Techniques

We now introduce a numerical scheme for solving the 1-dimensional Schrodinger equation. As we have seen the 1-dimensional Schrodinger equations takes the form

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \tag{6.487}$$

We will consider systems with the boundary conditions

$$\psi(x) \text{ and } \frac{d\psi(x)}{dx} \text{ are continuous and finite for all } x \tag{6.488}$$

$$\lim_{x\to\pm\infty} \psi(x) = 0 \tag{6.489}$$

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx < \infty \text{ or the wave function is normalizable} \tag{6.490}$$

Let us illustrate the procedure with the harmonic oscillator potential (so we can check our results against the exact answer). The potential energy function is

$$V(x) = \frac{1}{2}m\omega^2 x^2 \tag{6.491}$$

The Schrodinger equation becomes

$$\frac{d^2\psi(x)}{dx^2} = \frac{2m}{\hbar^2}\left(\frac{1}{2}m\omega^2 x^2 - E\right)\psi(x) \tag{6.492}$$

For real systems the very large or very small numerical values of m, ω, and $\hbar$ can lead to calculational difficulties. For calculational simplicity we will choose

$$\hbar = \sqrt{2} \text{ and } m = \omega = 1 \tag{6.493}$$

Since $\hbar\omega$ has units of energy, when the calculation is finished we can convert our numerical energy values back to real physical units by the operation

$$E_{real} = E_{numerical}\frac{\hbar\omega}{\sqrt{2}} \tag{6.494}$$

We now have the equation

$$\frac{d^2\psi(x)}{dx^2} = \left(0.5x^2 - E\right)\psi(x) \tag{6.495}$$

to solve for the allowed E values.

This potential energy function satisfies $V(x) = V(-x)$, so that the solutions will have definite parity or they will be even ($\psi(x) = \psi(-x)$) or odd ($\psi(x) = -\psi(-x)$).

We arbitrarily choose solutions of the form

even $\psi(0) = 1$ and $\frac{d\psi(0)}{dx} = 0$
odd $\psi(0) = 0$ and $\frac{d\psi(0)}{dx} = 1$

The particular numerical values chosen are arbitrary since the wave functions we find have to be normalized anyway. These choices correspond to choosing wave functions that look like those in Figure 6.17 below

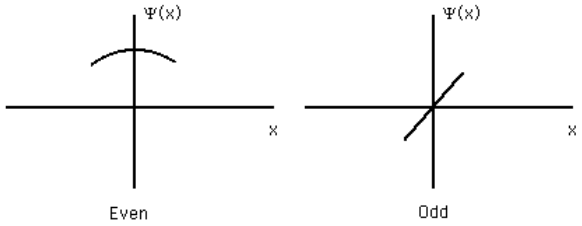


Figure 6.17: Initial Values

in order to satisfy this parity property.

The solution method then follows these steps:

1. choose
$$\psi(0) \text{ and } \frac{d\psi(0)}{dx}$$
(choose an even or odd solution type)

2. pick a starting energy value

3. with the chosen boundary conditions and energy value, break the second-order differential equation into two first-order differential equations and use a Runge-Kutta method to solve for the values of $\psi(x)$ for $x > 0$, i.e., if we let
$$y = \frac{d\psi}{dx} \rightarrow \frac{dy}{dx} = \frac{d^2\psi}{dx^2}$$
then we get the two coupled first-order differential equations
$$\frac{\mathrm{d}y}{\mathrm{dx}} = \left(0.5x^2 - E\right)\psi \text{ and } \frac{\mathrm{d}\psi}{\mathrm{dx}} = \mathrm{y}$$

4. the fourth-order Runge-Kutta equations corresponding to these equations are then
$$x_{k+1} = x_k + h$$
$$\psi_{k+1} = \psi_k + \frac{h}{6}\left(f_1 + 2f_2 + 2f_3 + f_4\right)$$
$$y_{k+1} = y_k + \frac{h}{6}\left(g_1 + 2g_2 + 2g_3 + g_4\right)$$

where

$$f(x,\psi,y,E) = y \text{ and } g(x,\psi,y,E) = (0.5x^2 - E)\psi \tag{6.496}$$

and

$$\begin{aligned}
f_1 &= f(x_k, \psi_k, y_k, E)\\
f_2 &= f(x_k + \frac{h}{2}, \psi_k + \frac{h}{2}f_1, y_k + \frac{h}{2}g_1, E)\\
f_3 &= f(x_k + \frac{h}{2}, \psi_k + \frac{h}{2}f_2, y_k + \frac{h}{2}g_2, E)\\
f_4 &= f(x_k + h, \psi_k + hf_3, y_k + hg_3, E)\\
g_1 &= g(x_k, \psi_k, y_k, E)\\
g_2 &= g(x_k + \frac{h}{2}, \psi_k + \frac{h}{2}f_1, y_k + \frac{h}{2}g_1, E)\\
g_3 &= g(x_k + \frac{h}{2}, \psi_k + \frac{h}{2}f_2, y_k + \frac{h}{2}g_2, E)\\
g_4 &= g(x_k + h, \psi_k + hf_3, y_k + hg_3, E)
\end{aligned}$$

5. an allowed energy value (an eigenvalue of the Hamiltonian) occurs when the numerical solution approaches zero (exponential decrease) for large values of x

6. since it is not possible to find the energy eigenvalues exactly (due to computer roundoff errors) the numerical solution will always diverge (get very large (either positive or negative)) at large values of x

7. the solution method can only home in on the energy eigenvalue

The program code below carries out the following process:

1. choose an energy eigenvalue lower than the desired value

2. choose an energy step value

3. the solution will eventually diverge (say it gets large positively)

4. increase the energy value by the energy step

5. the solution will eventually diverge (say it still gets large positively)

6. increase the energy value by the energy step

7. the solution will eventually diverge negatively

8. an energy eigenvalue is now located between this energy value (where it diverged negatively) and the last energy value (where it diverged positively)

9. change the energy value back to the previous value (where it diverged positively) and decrease the energy step by a factor of 10
10. repeat the process until it diverges negatively again
11. change the energy value back to the previous value (where it diverged positively) and decrease the energy step by a factor of 10
12. continue until the desired accuracy is reached

Sample programs (written in the MATLAB language)

```
function z=sq(x,psi,y,alpha)
if (x < 1)
 z=-15*(1-alpha)*psi;
else
 z=15*alpha*psi;
end
% energy 0 < alpha < 0; search for eigenvalues
% choose E lower than desired eigenvalue; choose estep a reasonable size
% and to move you in direction of an eigenvalue; program will home in
% and stop when estep/E<.0001; uses 4th order Runge-Kutta method
format long
estep=-0.05;alpha=0.99;h=0.05;n=0;Eold=10^25;
while 1
 n=n+1;psi=1;y=0;
 for j=0:1000
  f1=y;
  g1=sq(j*h,psi,y,alpha);f2=y+h*g1/2;
  g2=sq((j+0.5)*h,psi+h*f1/2,y+h*g1,alpha);f3=y+h*g2/2;
  g3=sq((j+0.5)*h,psi+h*f2/2,y+h*g2/2,alpha);f4=y+h*g3;
  g4=sq((j+1)*h,psi+h*f3,y+h*g3,alpha);
  psi=psi+h*(f1+2*f2+2*f3+f4)/6;
  if (abs(psi) > 100)
   if (n == 1)
    check=sign(psi); alpha=alpha+estep;
   else
    if (check ~= sign(psi))
     alpha=alpha-estep; estep = estep/10;
   else
     alpha=alpha+estep;
   end
  end
  break;
 end
 y=y+h*(g1+2*g2+2*g3+g4)/6;
end
 stp=(abs(estep/alpha));
```

```
 [alpha,stp]
 if (stp < 0.0001)
  break;
 end
end
format short
```

A typical run looks like:

$E-Value$	$StepSize$
0.150000	0.333333
0.200000	0.250000
0.250000	0.200000
0.300000	0.166667
0.350000	0.142857
0.400000	0.125000
0.450000	0.111111
0.500000	0.100000
0.550000	0.0909091
0.600000	0.0833333
0.650000	0.0769231
0.700000	0.0714286
0.750000	0.0666667
0.700000	0.0071428
0.705000	0.0070922
0.710000	0.0070422
0.705000	0.0007092
0.705500	0.0007087
0.706000	0.0007082
0.706500	0.0007077
0.707000	0.0007072
0.707500	0.0007067
0.707000	0.0000707

Table 6.1: Run Values

Using the conversion rule the energy eigenvalue is

$$E_{real} = \left(\frac{\sqrt{2}}{2}\right)\frac{\hbar\omega}{\sqrt{2}} = \frac{1}{2}\hbar\omega \tag{6.497}$$

which is exactly correct for the lowest even state eigenvalue (the ground state).

6.11 Translation Invariant Potential - Bands

We now consider a potential that is *periodic in space*, which means that it is *translationally invariant*, i.e.,

$$V(x) = V(x+a) \text{ where } a = \text{ some fixed distance} \tag{6.498}$$

This might represent an electron in a 1-dimensional crystal lattice, where a is the distance between atoms. The change in the allowed energy spectrum caused by the periodicity of the lattice is rather striking.

The Hamiltonian of the system is

$$\hat{H}(x,p) = \frac{p^2}{2m} + V(x) \tag{6.499}$$

This Hamiltonian is invariant if we translate a fixed distance a

$$\hat{H}(x+a,p) = \frac{p^2}{2m} + V(x+a) = \frac{p^2}{2m} + V(x) = \hat{H}(x,p) \tag{6.500}$$

This means that the transformation operator $\hat{T}(a)$ that corresponds to a translation by a distance a *commutes* with the Hamiltonian

$$\left[\hat{H}, \hat{T}(a)\right] = 0 \tag{6.501}$$

and thus they must *share a set of eigenfunctions.*

From our earlier discussion, the translation operator has the defining property that, if the state $|\psi\rangle$ has the wave function $\langle x \,|\, \psi\rangle$ in the position representation, then the state $\hat{T}(a)\,|\psi\rangle$ has the wave function

$$\langle x|\, \hat{T}(a)\, |\psi\rangle = \langle x+a \mid \psi\rangle \tag{6.502}$$

Let us rederive the eigenfunctions of $\hat{T}(a)$. If $|\psi\rangle$ is an eigenstate of the $\hat{T}(a)$ operator, then we must have

$$\hat{T}(a)\,|\psi\rangle = \lambda\,|\psi\rangle \text{ where } \lambda = \text{ the eigenvalue} \tag{6.503}$$

and thus

$$\psi(x+a) = \langle x|\, \hat{T}(a)\, |\psi\rangle = \lambda\, \langle x \mid \psi\rangle = \lambda\psi(x) \tag{6.504}$$

With hindsight, we write λ in the form e^{ika} which defines k. We then have

$$\psi(x+a) = e^{ika}\psi(x) \tag{6.505}$$

If we define

$$u_k(x) = e^{-ikx}\psi(x) \tag{6.506}$$

we then have that

$$\psi(x+a) = e^{ik(x+a)}u_k(x+a) = e^{ika}\psi(x) = e^{ika}e^{ikx}u_k(x) \tag{6.507}$$

$$u_k(x+a) = u_k(x) \tag{6.508}$$

which says that $u_k(x)$ is also periodic in space with period a. The eigenfunctions of $\hat{T}(a)$ are then given by

$$\psi(x) = e^{ikx}u_k(x) \tag{6.509}$$

which is a *plane wave modulated by a function with the periodicity of the lattice* (potential).

This type of wave function is called a *Bloch wave*.

We note that k must be real since if k were not real, then the eigenfunction would diverge as x gets large and thus be unrenormalizable, which is not allowed.

Since $\hat{H}$ and $\hat{T}(a)$ must share a set of eigenfunctions, we can construct a complete set of energy eigenstates in the form of Bloch waves. This means that we can find at least one energy eigenstate for each possible $\lambda = e^{ika}$, since all λ are eigenvalues of $\hat{T}(a)$.

Note also that the values k and $k + 2\pi n/a$ give the same eigenvalue and thus we can restrict our attention to k values in the interval

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a} \tag{6.510}$$

and we will be including all possible eigenvalues of $\hat{T}(a)$.

Example

Let us now consider the specific example of a 1-dimensional periodic array of delta functions

$$V(x) = \sum_{n=-\infty}^{\infty} v_0\delta(x - na) \tag{6.511}$$

which corresponds to the so-called *Kronig-Penney model*.

In the interval $0 < x < a$ the potential vanishes and the solution to the Schrodinger equation is of the form

$$\psi(x) = Ae^{iqx} + Be^{-iqx} \text{ where } E = \frac{\hbar^2 q^2}{2m} \tag{6.512}$$

Therefore, in the interval $0 < x < a$, we must have

$$u_k(x) = Ae^{i(q-k)x} + Be^{-i(q+k)x} \tag{6.513}$$

The coefficients A and B are fully determined by two conditions:

1. $\psi(x)$ and hence $u_k(x)$ must be continuous at the lattices sites

$$\lim_{\eta\to 0}(u_k(\eta) - u_k(-\eta)) \to 0$$

Periodicity then says that $u_k(-\eta) = u_k(a-\eta)$ and hence we must also have

$$\lim_{\eta\to 0}(u_k(\eta) - u_k(a-\eta)) \to 0$$

which gives

$$A + B = Ae^{i(q-k)a} + Be^{-i(q+k)a} \qquad \text{(condition 1)} \tag{6.514}$$

2. $d\psi(x)/dx$ is discontinuous at the lattice points, as we showed earlier, as follows:

$$\lim_{\eta\to 0}\frac{\hbar^2}{2m}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=\eta} - \left.\frac{d\psi(x)}{dx}\right|_{x=-\eta}\right) = v_0\psi(0) \tag{6.515}$$

We have

$$\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=\eta}\right) = iq(A-B)$$

$$\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=-\eta}\right) = e^{-ika}\lim_{\eta\to 0}\left(\left.\frac{d\psi(x)}{dx}\right|_{x=a-\eta}\right)$$

$$= e^{-ika}iq\left(Ae^{iqa} - Be^{-iqa}\right)$$

$$\psi(0) = A + B$$

which gives

$$\frac{\hbar^2}{2m}iq\left(A - B - Ae^{i(q-k)a} + Be^{-i(q+k)a}\right)$$

$$= v_0(A+B) \qquad \text{(condition 2)} \tag{6.516}$$

We only need these two conditions and do not need to worry about all of the rest of the potential because of the periodicity.

We solve for the ratio A/B in each equation and equate the two results to get a transcendental equation for q in terms of k (which then gives us the allowed energy values). We get the equation

$$\cos ka = \cos qa + \frac{mv_0}{q\hbar^2}\sin qa \tag{6.517}$$

Now for any given value of k we must have

$$-1 \le \cos ka \le +1 \tag{6.518}$$

Therefore, if we plot

$$\cos ka = \cos qa + \frac{mv_0 a}{\hbar^2}\frac{\sin qa}{qa} \quad \text{versus} \quad qa \tag{6.519}$$

the only allowed q values are between the two lines representing ± 1 as shown in Figure 6.18 below.

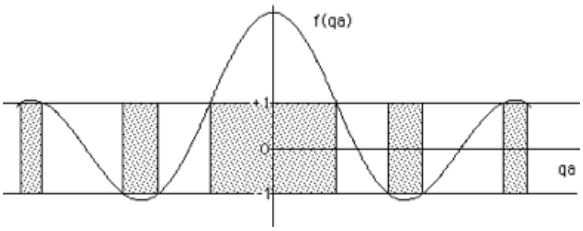


Figure 6.18: Allowed q values

i.e., q must always lie in the unshaded areas. For each value of $\cos ka$ there are an infinite number of solutions. Taking k in the range $-\pi/a \leq k \leq \pi/a$, we can plot

$$E = \frac{\hbar^2 q^2}{2m} \quad \text{versus} \quad ka$$

as shown in Figure 6.19 below.

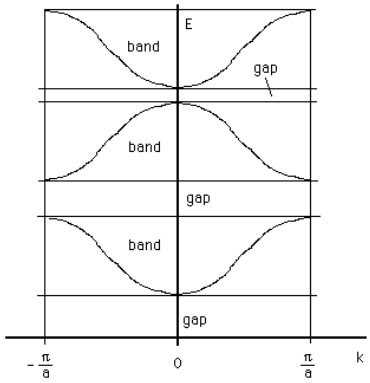


Figure 6.19: Bands and Gaps

The possible energy values lie in bands, with gaps between them. This structure of the energy spectrum with bands and gaps occurs for all periodic potentials.

It is the basis of the explanation of energy bands in solids, which we will discuss in detail later.

6.12 Closed Commutator Algebras and Glauber's Theorem

The earlier example of the solution of the harmonic oscillator potential illustrates a general rule about algebraic solutions.

In that example we had the following commutator algebra

$$[\hat{a}, \hat{a}^+] = \hat{I} \ , \ \left[\hat{a}, \hat{N}\right] = \hat{a} \ , \ \left[\hat{a}^+, \hat{N}\right] = -\hat{a}^+ \tag{6.520}$$

In this algebra, all the commutators only involve a closed set of operators. In this case the set is $\{\hat{a}, \hat{a}^\dagger, \text{ and } \hat{N}\}$.

When the algebra is closed, we call it a *closed commutator algebra.*

If we have a closed commutator algebra, then, in general, we can completely solve the eigenvalue/eigenvector problem for these operators using algebraic methods.

Some examples are the harmonic oscillator, systems that can be factored, and, as we shall show later, angular momentum operators, and the hydrogen atom system.

A very useful algebraic result is *Glauber's Theorem*, which we now prove.

For two operators $\hat{A}$ and $\hat{B}$ assume that

$$\left[\hat{A}, \left[\hat{A}, \hat{B}\right]\right] = \left[\hat{B}, \left[\hat{A}, \hat{B}\right]\right] = 0 \tag{6.521}$$

Now let x be a parameter and consider the operator

$$f(x) = e^{\hat{A}x} e^{\hat{B}x} \tag{6.522}$$

We then get

$$\begin{aligned} \frac{df}{dx} &= \hat{A} e^{\hat{A}x} e^{\hat{B}x} + e^{\hat{A}x} \hat{B} e^{\hat{B}x} = \hat{A} e^{\hat{A}x} e^{\hat{B}x} + e^{\hat{A}x} \hat{B} e^{-\hat{A}x} e^{\hat{A}x} e^{\hat{B}x} \\ &= \left(\hat{A} + e^{\hat{A}x} \hat{B} e^{-\hat{A}x}\right) e^{\hat{A}x} e^{\hat{B}x} = \left(\hat{A} + e^{\hat{A}x} \hat{B} e^{-\hat{A}x}\right) f(x) \end{aligned}$$

Our assumption (6.526) implies that (proof later)

$$\left[\hat{B}, \hat{A}^n\right] = n\hat{A}^{n-1} \left[\hat{B}, \hat{A}\right] \tag{6.523}$$

and

$$\begin{aligned} \left[\hat{B}, e^{-\hat{A}x}\right] &= \sum_n (-1)^n \frac{x^n}{n!} \left[\hat{B}, \hat{A}^n\right] \\ &= \sum_n (-1)^n \frac{x^n}{(n-1)!} \hat{A}^{n-1} \left[\hat{B}, \hat{A}\right] \\ &= -e^{-\hat{A}x} \left[\hat{B}, \hat{A}\right] x \end{aligned} \tag{6.524}$$

which gives

$$\left[\hat{B}, e^{-\hat{A}x}\right] = \hat{B} e^{-\hat{A}x} - e^{-\hat{A}x} \hat{B} = -e^{-\hat{A}x} \left[\hat{B}, \hat{A}\right] x \tag{6.525}$$

$$e^{\hat{A}x} \hat{B} e^{-\hat{A}x} = \hat{B} - \left[\hat{B}, \hat{A}\right] x \tag{6.526}$$

Thus, we have

$$\frac{df}{dx} = \left(\hat{A} + e^{\hat{A}x}\hat{B}e^{-\hat{A}x}\right) f(x) = \left(\hat{A} + \hat{B} + \left[\hat{A}, \hat{B}\right] x\right) f(x) \tag{6.527}$$

Now, our assumption (6.526) says that

$$\left[(\hat{A} + \hat{B}), \left[\hat{A}, \hat{B}\right]\right] = 0 \tag{6.528}$$

This means that we can treat them like numbers to solve the above differential equation, i.e., we do not have to be careful above operator order. The solution is

$$f(x) = e^{\hat{A}x}e^{\hat{B}x} = e^{(\hat{A}+\hat{B})x}e^{\frac{1}{2}[\hat{A},\hat{B}]x^2} \tag{6.529}$$

Letting $x = 1$ we have Baker-Hausdorf identity (Glauber's theorem)

$$e^{(\hat{A}+\hat{B})} = e^{\hat{A}}e^{\hat{B}}e^{-\frac{1}{2}[\hat{A},\hat{B}]} \tag{6.530}$$

This leaves us to prove (6.528). Consider some special cases to see how it works.

$$n = 1 \qquad \left[\hat{B}, \hat{A}\right] = n\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right]$$

$$\begin{aligned} n = 2 \qquad \left[\hat{B}, \hat{A}^2\right] &= \hat{B}\hat{A}^2 - \hat{A}^2\hat{B} = \left[\left[\hat{B}, \hat{A}\right], \hat{A}\right] + 2\hat{A}\hat{B}\hat{A} + 2\hat{A}^2\hat{B} \\ &= 2\hat{A}\hat{B}\hat{A} + 2\hat{A}^2\hat{B} = 2\hat{A}\left[\hat{B}, \hat{A}\right] = n\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] \end{aligned}$$

$$\begin{aligned} n = 3 \qquad \left[\hat{B}, \hat{A}^3\right] &= \hat{B}\hat{A}^3 - \hat{A}^3\hat{B} = \hat{B}\hat{A}^2\hat{A} - \hat{A}\hat{A}^2\hat{B} \\ &= \left[\left[\hat{B}, \hat{A}^2\right] + \hat{A}^2\hat{B}\right]\hat{A} - \hat{A}\left[\hat{B}\hat{A}^2 - \left[\hat{B}, \hat{A}^2\right]\right] \\ &= 2\hat{A}\left[\hat{B}, \hat{A}\right]\hat{A} + \hat{A}^2\hat{B}\hat{A} - \hat{A}\hat{B}\hat{A}^2 + 2\hat{A}^2\left[\hat{B}, \hat{A}\right] \\ &= 4\hat{A}^2\left[\hat{B}, \hat{A}\right] + \hat{A}^2\hat{B}\hat{A} - \hat{A}^3\hat{B} - 2\hat{A}^2\left[\hat{B}, \hat{A}\right] \\ &= 2\hat{A}^2\left[\hat{B}, \hat{A}\right] + \hat{A}^2\left[\hat{B}, \hat{A}\right] = 3\hat{A}^2\left[\hat{B}, \hat{A}\right] = n\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] \end{aligned}$$

This suggests a general proof by induction. Assume that

$$\left[\hat{B}, \hat{A}^{n-1}\right] = (n-1)\hat{A}^{n-2}\left[\hat{B}, \hat{A}\right]$$

is true. Then, we have

$$\begin{aligned} \left[\hat{B}, \hat{A}^n\right] &= \hat{B}\hat{A}^n - \hat{A}^n B = \hat{B}\hat{A}^{n-1}\hat{A} - \hat{A}\hat{A}^{n-1}\hat{B} \\ &= \left[\hat{A}^{n-1}\hat{B} + (n-1)\hat{A}^{n-2}\left[\hat{B}, \hat{A}\right]\right]\hat{A} - \hat{A}\left[\hat{B}\hat{A}^{n-1} - (n-1)\hat{A}^{n-2}\left[\hat{B}, \hat{A}\right]\right] \\ &= 2(n-1)\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] + \hat{A}^{n-1}\hat{B}\hat{A} - \hat{A}\hat{B}\hat{A}^{n-1} \\ &= 2(n-1)\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] + \hat{A}^{n-1}\hat{B}\hat{A} - \hat{A}\left[\hat{A}^{n-1}\hat{B} + (n-1)\hat{A}^{n-2}\left[\hat{B}, \hat{A}\right]\right] \\ &= (n-1)\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] + \hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] = n\hat{A}^{n-1}\left[\hat{B}, \hat{A}\right] \end{aligned}$$

Done!

6.13 Quantum Interference with Slits

In the experiments considered here, we measure the y-component of momentum for a particle passing through a system of slits. The source-slit system is the preparation apparatus that determines the state vector. Recognizing that a system of slits is a position-measuring device allows us to ascertain that the state vector is a position state. Then, writing the state vector in momentum space provides a straightforward calculation for the probability amplitude and its corresponding probability function. Interference effects, if any, are inherent in the probability function. We determine the statistical distribution of scattered particles for four different slit systems. The results are in agreement with the well-known interference patterns obtained in classical wave optics.

6.13.1 Introduction

The double-slit experiment is the archetypical system used to demonstrate quantum mechanical behavior. It is said by Feynman "to contain the only mystery" in all of quantum mechanics. Numerous textbooks and journal articles discuss slit interference, usually in conjunction with wave-particle duality. Most authors emphasize that classical physics cannot describe the double slit experiment with particles. Yet, bolstered by the deBroglie hypothesis, they still ascribe to the classical maxim, "Waves exhibit interference. Particles do not". They then conclude that, "When particles exhibit interference, they are behaving like waves". Then the subsequent analysis is simply wave theory, and any interference effects are made to agree with Young's experiment.

Thus, classical wave optics, rather than quantum mechanics, is used to explain quantum interference. For example, Ohanian states " ... the maxima of this interference pattern are given by a formula familiar from wave optics". Some authors do suggest that a quantum mechanical approach is lacking. Liboff tells us, "The first thing to do is to solve Schrodinger's equation and calculate $|\psi|^2$ at the screen". Ballentine makes a similar statement when discussing diffraction from a periodic array. " ..... solve the Schroedinger equation with boundary conditions corresponding to an incident beam from a certain direction, and hence determine the position probability density $|\Psi(\vec{x})|^2$ at the detectors". But he then says, "an exact solution of this equation would be very difficult to obtain, ... ". The difficulty according to Merzbacher is that, "A careful analysis of the interference experiment would require detailed consideration of the boundary conditions at the slits".

In spite of these misgivings, quantum mechanics does provide a straightforward calculation for the probability distribution of the scattered particles.

Quantum mechanics is a theory about observations and their measurement. Its postulates provide, among other things, a set of instructions for calculating the probability of obtaining a particular result when an observable is measured.

These probability calculations require a state vector $|\psi\rangle$, which is determined by the preparation procedure. Its representation is dictated by the observable being measured

$$|\psi\rangle = \sum_k |a_k\rangle \langle a_k \mid \psi\rangle \tag{6.531}$$

The basis vectors $|a_k\rangle$ are the eigenvectors of the measured observable $\hat{A}$. Having obtained the state vector $|\psi\rangle$, the probability that a measurement of observable $\hat{A}$ yields the value a_k is given by the Born postulate

$$P_k = |\langle a_k \mid \psi\rangle|^2 \tag{6.532}$$

The state vector $|\psi\rangle$ and the probability distribution $|\langle a_k \mid \psi\rangle|^2$ are unique for a given experiment. State preparation and measurement are discussed at length later in the book.

We expect, then, that a quantum mechanical description of a slit experiment will

1. clearly define which observable is being measured,
2. describe the preparation procedure that determines the state vector
3. yield the probability function for the scattered particles.

6.13.2 Probability functions and quantum interference

The experiment considered here consists of the apparatus shown in Figure 6.20 below. For such an experiment, the source-slit system, which determines the possible y-coordinate(s) of the particle at the slits, is the preparation apparatus.

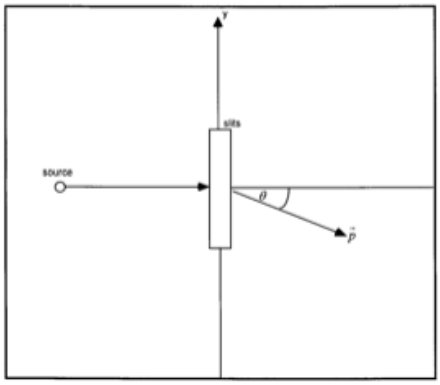


Figure 6.20: Particle Scattering from Slits

A particle originating at the source is scattered at angle θ by the system of slits. Thus, a particle passing through the slits has undergone a unique state preparation that determines the probability of scattering at angle θ.

The state vector is a position state. Because position and momentum are non-commuting observables, a particle passing through slits always has an uncertainty in its y-component of momentum. It can be scattered with any one of the continuum of momentum eigenvalues $p_y = p\sin\theta$, where $-\pi/2 \leq \theta \leq \pi/2$. Measurement of a well-defined scattering angle θ constitutes a measurement of the observable $\hat{p}_y$ and, therefore, the basis vectors in Hilbert space are the momentum eigenvectors $|\hat{p}_y\rangle$.

The probability that a particle leaves the slit apparatus with momentum p_y is, then,

$$P_{p_y} = |\langle p_y \mid \psi\rangle|^2 \tag{6.533}$$

It is this probability function that exhibits quantum interference. Its maxima and minima, if any, correspond to constructive and destructive interference respectively.

In the position representation, the free-particle momentum eigenfunction corresponding to the eigenvalue p_y is

$$\langle y \mid p_y\rangle = \frac{1}{\sqrt{2\pi}} e^{ip_y y/\hbar} \tag{6.534}$$

and the probability amplitude for scattering with momentum p_y is

$$\langle p_y \mid \psi\rangle = \int_{-\infty}^{\infty} \langle p_y \mid y\rangle \langle y \mid \psi\rangle dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ip_y y/\hbar}\psi(y)dy \tag{6.535}$$

An examination of the corresponding probability function

$$P_{p_y} = |\langle p_y \mid \psi\rangle|^2$$

will determine whether or not there is interference.

In the following discussion, we evaluate the integral of amplitude equation by first constructing the position state function

$$\psi(y) = \langle y \mid \psi\rangle \tag{6.536}$$

We do this for four source-slit systems, including the double slit.

6.13.3 Scattering from a narrow slit

A narrow slit of infinitesimal width is an ideal measuring device; it determines the position with infinite resolution. A particle emerging from a slit at $y = y_1$ is

in the position eigenstate $|y_1\rangle$. In the position representation, the eigenfunction of position is the Dirac delta function

$$\psi(y) = \langle y \mid y_1 \rangle = \delta(y - y_1) \tag{6.537}$$

and the probability amplitude for a particle emerging from the slit with momentum p_y is

$$\langle p_y \mid \psi \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ip_y y/\hbar} \delta(y - y_1)\, dy = \frac{1}{\sqrt{2\pi}} e^{-ip_y y_1/\hbar} \tag{6.538}$$

The corresponding probability function is

$$P(p_y) = |\langle p_y \mid \psi \rangle|^2 = \left| \frac{1}{\sqrt{2\pi}} e^{-ip_y y_1/\hbar} \right|^2 = \text{constant} \tag{6.539}$$

It is equally probable that the particle is scattered at any angle. There is no interference.

6.13.4 Scattering from a double slit(narrow)

We again assume that the slits are infinitesimally thin. For such a double slit apparatus, the observable $\hat{y}$ has two eigenvalues, y_1 and y_2. Assuming the source-slit geometry does not favor one slit is over the other, the state vector is the superposition of position eigenvectors

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|y_1\rangle + |y_2\rangle) \tag{6.540}$$

and

$$\psi(y) = \frac{1}{\sqrt{2}} (\delta(y - y_1) + \delta(y - y_2)) \tag{6.541}$$

Here, the amplitude for finding the particle with momentum p_y is

$$\begin{aligned} \langle p_y \mid \psi \rangle &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2}} \left[\int_{-\infty}^{\infty} e^{-ip_y y/\hbar} \delta(y - y_1)\, dy + \int_{-\infty}^{\infty} e^{-ip_y y/\hbar} \delta(y - y_2)\, dy \right] \\ &= \frac{1}{2\sqrt{\pi}} \left(e^{-ip_y y_1/\hbar} + e^{-ip_y y_2/\hbar} \right) \end{aligned} \tag{6.542}$$

From which we get the probability function

$$\begin{aligned} P(p_y) = |\langle p_y \mid \psi \rangle|^2 &= \left| \frac{1}{2\sqrt{\pi}} \left(e^{-ip_y y_1/\hbar} + e^{-ip_y y_2/\hbar} \right) \right|^2 \\ &= \frac{1}{4\pi} \left(2 + e^{ip_y (y_1 - y_2)/\hbar} + e^{-ip_y (y_1 - y_2)/\hbar} \right) \\ &= \frac{1}{2\pi} \left(1 + \cos\left(\frac{p_y d}{\hbar} \right) \right) \end{aligned} \tag{6.543}$$

where $d = y_1 - y_2$ is the distance between the slits. We see that this probability function does have relative maxima and minima and quantum interference does occur. Using $p_y = p\sin\theta$ we obtain the angular distribution of scattered particles

$$P(\theta) = \frac{1}{2\pi}\left[1 + \cos\left(\frac{pd\sin\theta}{\hbar}\right)\right] \tag{6.544}$$

The distance between the slits determines the number of interference fringes. In this example the distance between the slits is $d = 4\lambda$, where λ is the deBroglie wavelength. If we define $\phi = pd\sin\theta/\hbar$ and use the half-angle formula

$$1 + \cos\varphi = 2\cos^2\left(\frac{\varphi}{2}\right) \tag{6.545}$$

the probability function takes the form

$$P(\varphi) = \frac{1}{\pi}\cos^2\left(\frac{\varphi}{2}\right) \tag{6.546}$$

This is the familiar intensity distribution for Fraunhofer diffraction. A plot of (6.550) is shown in Figure 6.21.

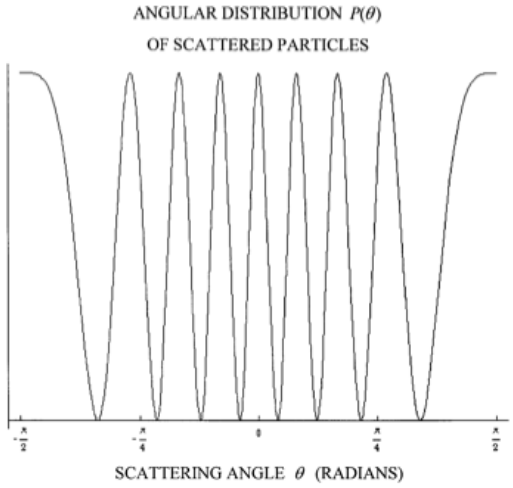


Figure 6.21: Angular Distribution of scattered particles from double narrow slits

6.13.5 Scattering from a slit of finite width

A slit of finite width is an imperfect apparatus for measuring position. It cannot distinguish between different position eigenvalues and a particle emerging from

a slit of width a can have any value of observable $\hat{y}$ in the continuum $-a/2 \le y \le a/2$. Assuming an equal probability of passing through the slit at any point, a particle at the slit is in the superposition state

$$\psi(y) = \langle y \,|\, \psi \rangle = \begin{cases} 1/\sqrt{a} & -a/2 \le y \le a/2 \\ 0 & \text{elsewhere} \end{cases} \tag{6.547}$$

Here, the probability amplitude is

$$\begin{aligned} \langle p_y \mid \psi \rangle &= \frac{1}{\sqrt{2\pi a}} \int_{-a/2}^{a/2} e^{-ip_y y/\hbar} dy \\ &= \frac{i\hbar}{p_y \sqrt{2\pi a}} \left(e^{-ip_y a/2\hbar} - e^{ip_y a/2\hbar} \right) \\ &= \frac{2\hbar}{p_y \sqrt{2\pi a}} \sin(a p_y / 2\hbar) \end{aligned} \tag{6.548}$$

and the corresponding probability function is

$$P(p_y) = |\langle p_y \mid \psi \rangle|^2 = \frac{2\hbar^2}{\pi a p_y^2} \sin^2(a p_y / 2\hbar) \tag{6.549}$$

This result is shown below in Figure 6.22 in terms of the scattering angle θ.

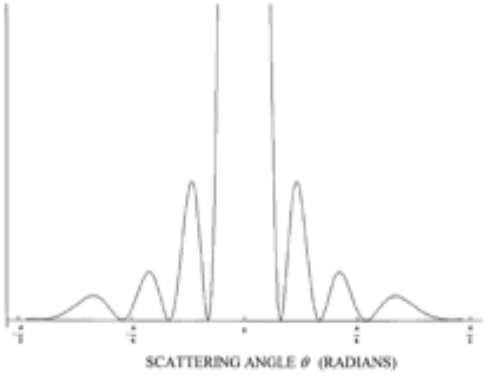


Figure 6.22: Angular Distribution of scattered particles from single slit of finite width

Once again, the distance between the slits determines the number of interference fringes. In this example the distance between the slits is $d = 4\lambda$, where λ is the deBroglie wavelength.

We see that $P(p_y)$ is the well-known diffraction pattern for light if we define $\alpha = ap_y/2\hbar = ap\sin\theta/2\hbar$ and write

$$P(\alpha) = \frac{a}{2\pi}\left(\frac{\sin\alpha}{\alpha}\right)^2 \tag{6.550}$$

6.13.6 Scattering from a double finite-width slit

As a final example, we consider a particle passing through a double-slit apparatus consisting of two slits each of finite width a. This is the most realistic description of the double slit experiment. Here, the state vector is

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|y_1\rangle + |y_2\rangle\right) \tag{6.551}$$

where

$$\langle y\,|\,\psi_1\rangle = \begin{cases} 1/\sqrt{a} & y_1 - a/2 \le y \le y_1 + a/2 \\ 0 & \text{elsewhere} \end{cases} \tag{6.552}$$

and

$$\langle y\,|\,\psi_2\rangle = \begin{cases} 1/\sqrt{a} & y_2 - a/2 \le y \le y_2 + a/2 \\ 0 & \text{elsewhere} \end{cases} \tag{6.553}$$

Again, we calculate the probability amplitude

$$\begin{aligned} \langle p_y\,|\,\psi\rangle &= \frac{1}{\sqrt{2}}\left(\langle p_y\,|\,\psi_1\rangle + \langle p_y\,|\,\psi_2\rangle\right) \\ &= \frac{1}{\sqrt{2\pi a}}\left[\int_{y_1-a/2}^{y_1+a/2} e^{-ip_y y/\hbar}dy + \int_{y_2-a/2}^{y_2+a/2} e^{-ip_y y/\hbar}dy\right] \\ &= \frac{i\hbar}{p_y\sqrt{2\pi a}}\big[e^{-ip_y(y_1+a/2)\hbar} - e^{-ip_y(y_1-a/2)\hbar} \\ &\qquad + e^{-ip_y(y_2+a/2)\hbar} - e^{-ip_y(y_2-a/2)\hbar}\big] \end{aligned} \tag{6.554}$$

With a slight manipulation of terms this amplitude becomes

$$\langle p_y\,|\,\psi\rangle = \frac{2\hbar}{p_y\sqrt{2\pi a}}\left[e^{-ip_y y_1\hbar} - e^{-ip_y y_2\hbar}\right]\sin(ap_y/2\hbar) \tag{6.555}$$

and the corresponding probability distribution is

$$P(p_y) = \frac{4\hbar^2}{\pi a p_y^2}\left(1 + \cos\left(p_y d/\hbar\right)\right)\sin^2(ap_y/2\hbar) \tag{6.556}$$

The angular distribution $P(\theta)$ is shown in Figure 6.23 and Figure 6.24 below for two different slit configurations.

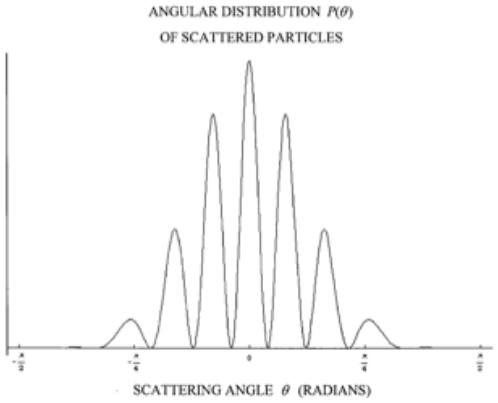


Figure 6.23: Angular Distribution of scattered particles from two slits of finite width The slit width is $\lambda = a$

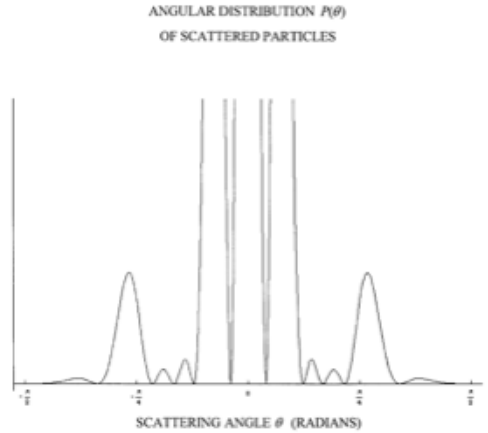


Figure 6.24: Angular Distribution of scattered particles from two slits of finite width The slit width is $\lambda = a/2$

Once again, the distance between the slits determines the number of interference fringes. In this example the distance between the slits is $d = 4\lambda$, where λ is the deBroglie wavelength. Using $\phi = pd\sin\theta/2\hbar$ and $\alpha = ap\sin\theta/2\hbar$, we again get the optical form.

$$P(\varphi) = \frac{2a}{\pi}\cos^2(\varphi/2)\left(\frac{\sin(\alpha)}{\alpha}\right)^2 \tag{6.557}$$

6.13.7 Conclusions

The Born postulate has been used to obtain the interference pattern for particles scattered from a system of slits without referring, a priori, to classical wave theory. Having identified the state vector as a position state and the measured observable the momentum, we obtain explicit expressions for the state vector $|\psi\rangle$ and its corresponding probability function

$$P(p_y) = |\langle p_y \mid \psi\rangle|^2$$

The results are in agreement with wave optics.

Quantum interference can occur only when a large number of identically prepared particles are observed. These particles are detected at different locations, one at a time. A single particle is always detected as a localized entity and no wave properties can be discerned from it.

It is interesting that for particles scattered from a double slit, the probability amplitude that gives rise to the interference is due to a superposition of delta functions.

6.14 Algebraic Methods - Supersymmetric Quantum Mechanics

6.14.1 Generalized Ladder Operators

Earlier in this chapter, we used the ladder or raising/lowering operators and their commutator algebra to carry out an algebraic solution for the harmonic oscillator system.

Can we generalize this procedure to other Hamiltonians?

Let us consider the general Hamiltonian $\hat{H}_0$

$$\hat{H}_0 = -\frac{1}{2}\frac{d^2}{dx^2} + V_0(x) \tag{6.558}$$

where we have set $\hbar = 1$.

For algebraic simplicity, we choose the potential energy term $V_0(x)$ to be

$$V_0(x) = V(x) - E_0 \tag{6.559}$$

where $V(x)$ = actual potential energy and E_0 = ground-state energy, so that the ground state energy is *zero* (this is just a choice of reference level for the potential energy). Suppose that the ground state(zero energy now) wave function is

represented by $\psi_0(x)$. We then have

$$\begin{aligned}\hat{H}_0\psi_0(x) &= -\frac{1}{2}\frac{d^2\psi_0(x)}{dx^2} + V_0(x)\psi_0(x) \\ &= -\frac{1}{2}\frac{d^2\psi_0(x)}{dx^2} + (V(x) - E_0)\psi_0(x) \\ &= \frac{1}{2}\frac{d^2\psi_0(x)}{dx^2} + V(x)\psi_0(x) - E_0\psi_0(x) = 0 \end{aligned} \tag{6.560}$$

which gives

$$\hat{H}_0\psi_0(x) = -\frac{1}{2}\psi_0'' + V_0\psi_0 = 0 \tag{6.561}$$

or

$$V_0 = \frac{1}{2}\frac{\psi_0''}{\psi_0} \tag{6.562}$$

and

$$\hat{H}_0 = \frac{1}{2}\left[-\frac{d^2}{dx^2} + \frac{\psi_0''}{\psi_0}\right] \tag{6.563}$$

Let us now introduce the new operators(a generalization of the raising and lowering operators)

$$\hat{a}^{\pm} = \frac{1}{\sqrt{2}}\left[\mp\frac{d}{dx} - \frac{\psi_0'}{\psi_0}\right] = \frac{1}{\sqrt{2}}\left[\mp\frac{d}{dx} + W(x)\right] \tag{6.564}$$

where

$$W(x) = -\frac{\psi_0'}{\psi_0} \tag{6.565}$$

is called the *superpotential* for the problem.

We then have

$$\begin{aligned}2\hat{a}^{\pm}\hat{a}^{\mp} &= 2\frac{1}{\sqrt{2}}\left[\mp\frac{d}{dx} - \frac{\psi_0'}{\psi_0}\right]\frac{1}{\sqrt{2}}\left[\pm\frac{d}{dx} - \frac{\psi_0'}{\psi_0}\right] \\ &= -\frac{d^2}{dx^2} + \left(\mp\frac{d}{dx}\right)\left(-\frac{\psi_0'}{\psi_0}\right) + \left(-\frac{\psi_0'}{\psi_0}\right)\left(\pm\frac{d}{dx}\right) + \left(-\frac{\psi_0'}{\psi_0}\right)^2 \\ &= -\frac{d^2}{dx^2} + \hat{\alpha} + \left(-\frac{\psi_0'}{\psi_0}\right)^2 \end{aligned} \tag{6.566}$$

To determine a more useful form for the quantity

$$\hat{\alpha} = \pm\frac{d}{dx}\left(\frac{\psi_0'}{\psi_0}\right) \mp \left(\frac{\psi_0'}{\psi_0}\right)\frac{d}{dx} \tag{6.567}$$

we let it operate on an arbitrary function $f(x)$. We get

$$\begin{aligned}\hat{\alpha}f(x) &= \pm\frac{d}{dx}\left(\frac{\psi_0'}{\psi_0}\right)f(x) \mp \left(\frac{\psi_0'}{\psi_0}\right)\frac{d}{dx}f(x) \\ &= \pm\left(\frac{\psi_0'}{\psi_0}\right)\frac{df}{dx} \mp \left(\frac{\psi_0'}{\psi_0}\right)\frac{df}{dx} \pm f(x)\frac{d}{dx}\left(\frac{\psi_0'}{\psi_0}\right) \\ &= \left[\pm\frac{d}{dx}\left(\frac{\psi_0'}{\psi_0}\right)\right]f(x) \end{aligned} \tag{6.568}$$

or

$$\hat{\alpha} = \pm\frac{d}{dx}\left(\frac{\psi_0'}{\psi_0}\right) = \pm\left(\frac{\psi_0''}{\psi_0} - \left(\frac{\psi_0'}{\psi_0}\right)^2\right) \tag{6.569}$$

We then get

$$\begin{aligned} 2\hat{a}^{\pm}\hat{a}^{\mp} &= -\frac{d^2}{dx^2} + \hat{\alpha} + \left(\frac{\psi_0'}{\psi_0}\right)^2 \\ &= -\frac{d^2}{dx^2} \pm \left(\frac{\psi_0''}{\psi_0} - \left(\frac{\psi_0'}{\psi_0}\right)^2\right) + \left(\frac{\psi_0'}{\psi_0}\right)^2 \\ &= -\frac{d^2}{dx^2} \pm \frac{\psi_0''}{\psi_0} + (1 \mp 1)\left(\frac{\psi_0'}{\psi_0}\right)^2 \end{aligned} \tag{6.570}$$

If we define two new quantities by

$$V_1(x) = V_0(x) - \frac{d}{dx}\frac{\psi_0'}{\psi_0} \quad , \quad \hat{H}_1 = -\frac{1}{2}\frac{d^2}{dx^2} + V_1(x) \tag{6.571}$$

then we have

$$\hat{a}^+\hat{a}^- = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}\frac{\psi_0''}{\psi_0} = \frac{1}{2}\left[-\frac{d^2}{dx^2} + \frac{\psi_0''}{\psi_0}\right] = \hat{H}_0 \tag{6.572}$$

and

$$\begin{aligned} \hat{a}^-\hat{a}^+ &= -\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}\frac{\psi_0''}{\psi_0} + \left(\frac{\psi_0'}{\psi_0}\right)^2 \\ &= -\frac{1}{2}\frac{d^2}{dx^2} - V_0 + \frac{\psi_0''}{\psi_0} - \frac{d}{dx}\frac{\psi_0'}{\psi_0} \\ &= -\frac{1}{2}\frac{d^2}{dx^2} + V_0 - \frac{d}{dx}\frac{\psi_0'}{\psi_0} \\ &= -\frac{1}{2}\frac{d^2}{dx^2} + V_1 = \hat{H}_1 \end{aligned} \tag{6.573}$$

$\hat{H}_1$ is called the *supersymmetric (SUSY) partner* of $\hat{H}_0$. $V_1(x)$ and $V_0(x)$ are called the *supersymmetric partner potentials.*

Now consider the operator

$$\hat{\beta} = \frac{d}{dx} \tag{6.574}$$

(remember $-i\hat{\beta} = \hat{p}$ = linear momentum operator) What is $\hat{\beta}^+$? We could figure this out from the momentum operator, but it is instructive to do it from scratch. The formal definition of $\hat{\beta}^+$ is given by the hermiticity condition

$$\int (\hat{\beta}^+ g^*(x)) f(x) dx = \int g^*(x)(\hat{\beta} f(x)) dx \tag{6.575}$$

or

$$\begin{aligned} \int (\hat{\beta}^+ g^*) f dx &= \int g^* \frac{df}{dx} dx = g^* f|_{\text{atlimits}} - \int \frac{dg^*}{dx} f dx \\ &= \int ((-\hat{\beta}) g^*) f dx \end{aligned} \tag{6.576}$$

This says that $\hat{\beta}^+ = -\hat{\beta}$, which we can also see from the Hermitian momentum operator ($\hat{\beta}^+ = (-i\hat{p})^+ = i\hat{p}^+ = i\hat{p} = -\hat{\beta}$). We have assumed that $g, f \to 0$ at the limit points (*this is required for hermiticity*).

Therefore, since

$$a^+ = \frac{1}{\sqrt{2}} \left[-\frac{d}{dx} - \frac{\psi_0'}{\psi_0} \right] = \frac{1}{\sqrt{2}} \left[-\hat{\beta} - \frac{\psi_0'}{\psi_0} \right] \tag{6.577}$$

we get

$$\begin{aligned} (\hat{a}^+)^+ &= \frac{1}{\sqrt{2}} \left[-\hat{\beta}^+ - \frac{\psi_0'}{\psi_0} \right] = \frac{1}{\sqrt{2}} \left[\hat{\beta} - \frac{\psi_0'}{\psi_0} \right] \\ &= \frac{1}{\sqrt{2}} \left[\frac{d}{dx} - \frac{\psi_0'}{\psi_0} \right] = \hat{a}^- \end{aligned} \tag{6.578}$$

Similarly we have $(\hat{a}^-)^+ = \hat{a}^+$ which says that

$$\hat{H}_0 = \hat{a}^+ \hat{a}^- = \hat{a}^+ (\hat{a}^+)^+ = (\hat{a}^-)^+ \hat{a}^- \tag{6.579}$$

which is the same result we had in the harmonic oscillator case, i.e., the Hamiltonian is expressed as the *square* of an operator.

The next thing we need is the commutator

$$[\hat{a}^-, \hat{a}^+] = \hat{a}^- \hat{a}^+ - \hat{a}^+ \hat{a}^- = \hat{H}_1 - \hat{H}_0 \tag{6.580}$$

We find

$$\begin{aligned} [\hat{a}^-, \hat{a}^+] &= -\frac{1}{2} \frac{d^2}{dx^2} + V_1 - \hat{H}_0 \\ &= -\frac{1}{2} \frac{d^2}{dx^2} + V_0 - \frac{d}{dx} \frac{\psi_0'}{\psi_0} + \frac{1}{2} \frac{d^2}{dx^2} - V_0 \\ &= -\frac{d}{dx} \frac{\psi_0'}{\psi_0} \end{aligned} \tag{6.581}$$

In general, this commutator is a function of x for non-harmonic potential (it is equal to a constant for the harmonic oscillator).

In order to work with this system of operators we need to figure out some of their properties.

(1) We have

$$\hat{a}^+\hat{a}^-\psi_0 = \hat{H}_0\psi_0 = 0$$
$$\langle\psi_0|\,\hat{H}_0\,|\psi_0\rangle = 0 = \langle\psi_0|\,\hat{a}^+\hat{a}^-\,|\psi_0\rangle$$

This says that the norm $\|\hat{a}^-\,|\psi_0\rangle\|$ equals zero and hence that

$$\hat{a}^-\psi_0 = 0 \tag{6.582}$$

(2) We have

$$\hat{a}^+\hat{H}_1 - \hat{H}_0\hat{a}^+ = \hat{a}^+\hat{a}^-\hat{a}^+ - \hat{a}^+\hat{a}^-\hat{a}^+ = 0 = \hat{a}^-\hat{H}_0 - \hat{H}_1\hat{a}^- \tag{6.583}$$

(3) Let the state ψ_n^0 be an eigenstate of $\hat{H}_0$ with eigenvalue E_n^0. We then have

$$\hat{H}_0\psi_n^0 = \hat{a}^+\hat{a}^-\psi_n^0 = E_n^0\psi_n^0$$
$$\hat{a}^-\hat{a}^+(\hat{a}^-\psi_n^0) = E_n^0(\hat{a}^-\psi_n^0) = \hat{H}_1(\hat{a}^-\psi_n^0) \tag{6.584}$$

which says that $\hat{a}^-\psi_n^0$ is and eigenstate of $\hat{H}_1$ with eigenvalue E_n^0 (which is also an eigenvalue of $\hat{H}_0$). This is true except for the ground state where $\hat{a}^-\psi_0 = 0$.

(4) Let the state ψ_n^1 be an eigenstate of $\hat{H}_1$ with eigenvalue E_n^1. We then have

$$\hat{H}_1\psi_n^1 = \hat{a}^-\hat{a}^+\psi_n^1 = E_n^1\psi_n^1$$
$$\hat{a}^+\hat{a}^-(\hat{a}^+\psi_n^1) = E_n^1(\hat{a}^+\psi_n^0) = \hat{H}_0(\hat{a}^+\psi_n^0)$$

which says that $\hat{a}^-\psi_n^1$ is and eigenstate of $\hat{H}_0$ with eigenvalue E_n^1 (which is also an eigenvalue of $\hat{H}_0$).

This means that the eigenvalue spectrum of the two Hamiltonians $\hat{H}_0$ and $\hat{H}_1$ can be derived from each other as shown in Figure 6.25 below.

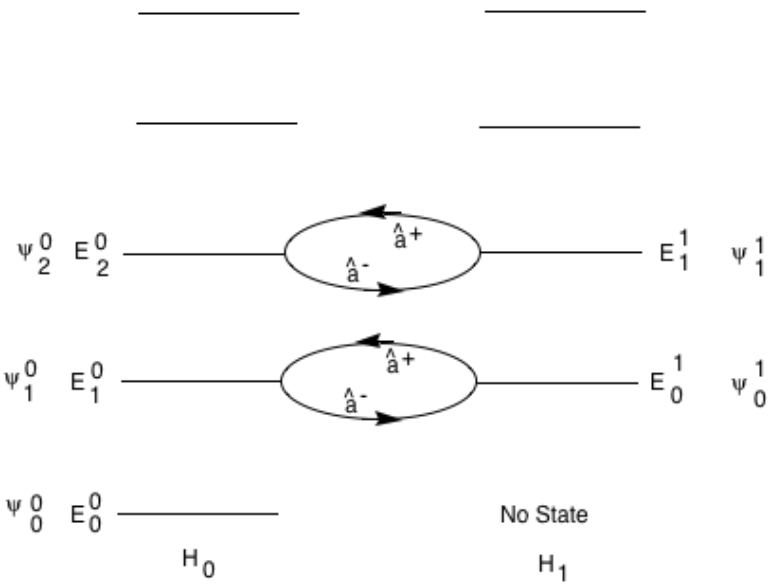


Figure 6.25: Eigenstate/Eigenvalue Relationships

$\hat{H}_0$ and $\hat{H}_1$ have the same energy level spectrum E_n^0,E_n^1 except that the zero energy ground state of $\hat{H}_0$, E_0^0, has no counterpart for $\hat{H}_1$.

The arrows indicate the operator connections between the state vectors.

This is an extremely important result.

If only one of the Hamiltonians is exactly solvable or more easily treated using approximation methods, then the solutions for the other Hamiltonian can be obtained from the solutions of the solvable member of the pair. This is the meaning of *supersymmetric partners*.

(5) Now consider the *normalization* condition. We have

$$\left\langle \psi_n^1 \right| \hat{H}_1 \left| \psi_n^1 \right\rangle = \left\langle \psi_n^1 \right| \hat{a}^- \hat{a}^+ \left| \psi_n^1 \right\rangle = \left\langle \psi_n^1 \right| E_n^1 \left| \psi_n^1 \right\rangle = E_n^1 \left\langle \psi_n^1 \mid \psi_n^1 \right\rangle \qquad (6.585)$$

Therefore, if $\left\langle \psi_n^1 \mid \psi_n^1 \right\rangle = 1$ (normalized to 1), then, if we let $|\alpha\rangle = \hat{a}^+ \left|\psi_n^1\right\rangle$, we must have

$$\langle \alpha \mid \alpha \rangle = E_n^1 \qquad (6.586)$$

The state $|\alpha\rangle = \hat{a}^+ \left|\psi_n^1\right\rangle$ is not normalized. We can insure normalized states in successive steps by choosing

$$\left|\psi_n^0\right\rangle = \frac{1}{\sqrt{E_n^1}} \hat{a}^+ \left|\psi_n^1\right\rangle \quad \text{and} \quad \left|\psi_n^1\right\rangle = \frac{1}{\sqrt{E_n^0}} \hat{a}^- \left|\psi_n^0\right\rangle \qquad (6.587)$$

We now fix one part of the notation by letting $\psi_0^0 = \psi_0$ and define

$$\Phi = -\frac{\psi_0^{0\prime}}{\psi_0^0} \tag{6.588}$$

which implies that

$$\hat{a}^{\pm} = \frac{1}{\sqrt{2}}\left[\mp\frac{d}{dx} + \Phi(x)\right] \tag{6.589}$$

$$V_0 = \frac{1}{2}\frac{\psi_0^{0\prime\prime}}{\psi_0^0} = \frac{1}{2}\left[-\Phi' + \Phi^2\right] \tag{6.590}$$

$$V_1 = \frac{1}{2}\left[\Phi' + \Phi^2\right] \tag{6.591}$$

A compact matrix notation for both Hamiltonians is

$$\begin{pmatrix} \hat{H}_1 & 0 \\ 0 & \hat{H}_0 \end{pmatrix} = \left(\frac{1}{2}\hat{p}^2 + \frac{1}{2}\Phi^2\right)\hat{I} + \frac{1}{2}\Phi'\hat{\sigma}_z \tag{6.592}$$

where

$$\hat{p} = -i\frac{d}{dx} \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{6.593}$$

Now the equation

$$\Phi = -\frac{\psi_0^{0\prime}}{\psi_0^0} \tag{6.594}$$

says that

$$-\Phi(x)dx = \frac{d\psi_0^0}{\psi_0^0} \tag{6.595}$$

$$-\int \Phi(x)dx = \ln(\psi_0^0) - \ln(A) \tag{6.596}$$

$$\psi_0^0(x) = A\exp\left[-\int \Phi(x)dx\right] \tag{6.597}$$

where A is a constant given by the normalization of ψ_0^0.

This result can *inverted*, i.e., we can specify $\Phi(x)$, which then gives ψ_0^0 and a pair of Hamiltonians to study.

6.14.2 Examples

(1) **Harmonic Oscillator** - We choose $\Phi(x) = \omega x$. This gives the two potentials

$$V_0 = \frac{1}{2}\left[-\Phi' + \Phi^2\right] = \frac{1}{2}\left[-\omega + \omega^2 x^2\right] \tag{6.598}$$

$$V_1 = \frac{1}{2}\left[\Phi' + \Phi^2\right] = \frac{1}{2}\left[\omega + \omega^2 x^2\right] \tag{6.599}$$

and the two Hamiltonians

$$\hat{H}_0 = -\frac{1}{2}\frac{d^2}{dx^2} + V_0 = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}\omega^2 x^2 - \frac{1}{2}\omega \tag{6.600}$$

$$\hat{H}_1 = -\frac{1}{2}\frac{d^2}{dx^2} + V_1 = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}\omega^2 x^2 + \frac{1}{2}\omega \tag{6.601}$$

which corresponds to two harmonic oscillators with zero point energies differing by ω. We then find

$$\hat{a}^- = \frac{1}{\sqrt{2}}\left[+\frac{d}{dx} + \Phi(x)\right] = \frac{1}{\sqrt{2}}\left[\frac{d}{dx} + \omega x\right] \tag{6.602}$$

Now we must have

$$\sqrt{2}\hat{a}^-\psi_0^0 = \left[\frac{d}{dx} + \omega x\right]\psi_0^0 = 0 \tag{6.603}$$

which is easy to solve. We get

$$\frac{d\psi_0^0(x)}{\psi_0^0(x)} = -\omega x dx \rightarrow \int \frac{d\psi_0^0(x)}{\psi_0^0(x)} = -\omega \int x dx \tag{6.604}$$

$$\ln\psi_0^0(x) - \ln A = -\frac{1}{2}\omega x^2 \tag{6.605}$$

$$\psi_0^0(x) = A\exp(-\frac{1}{2}\omega x^2) \tag{6.606}$$

as the ground state of $\hat{H}_0$. Since $\hat{H}_1$ differs from $\hat{H}_0$ only by a shift of ω, its lowest eigenstate corresponds to

$$\psi_1^1(x) = A\exp(-\frac{1}{2}\omega x^2) \quad , \quad E_1^1 = \omega \tag{6.607}$$

The first excited state of $\hat{H}_0$ is obtained using $\hat{a}^+$. We have $\hat{a}^+\psi_1^1 = \hat{a}^+\psi_0^0 = \psi_1^0$, so that

$$\begin{aligned}\psi_1^0 &= \frac{1}{\sqrt{2}}\left[-\frac{d}{dx} + \omega x\right]\psi_0^0 \\ &= \frac{1}{\sqrt{2}}\left[-\frac{d}{dx} + \omega x\right]A\exp(-\frac{1}{2}\omega x^2) \\ &= Ax\exp(-\frac{1}{2}\omega x^2)\end{aligned} \tag{6.608}$$

with $E_1^0 = \omega$, and so on. Iterating in this fashion we can generate the entire solution. It clearly agrees with our earlier results for the harmonic oscillator.

(2) **Reflection-Free Potentials** - Let us investigate the case given by

$$\Phi(x) = \tanh(x) \tag{6.609}$$

We then have

$$V_0 = \frac{1}{2}\left[1 - \frac{2}{\cosh^2 x}\right] \quad , \quad V_1 = \frac{1}{2} \tag{6.610}$$

Thus, the potential

$$-\frac{1}{\cosh^2 x} \tag{6.611}$$

has the *constant* potential 1/2, which corresponds to a *free particle*, as its *supersymmetric partner*.

We can find the normalized ground state eigenfunction of $\hat{H}_0$ using

$$\begin{aligned}\psi_0^0 &= A\exp\left[-\int \Phi(x)dx\right] = A\exp\left[-\log(\cosh(x))\right] \\ &= \frac{A}{\cosh(x)}\end{aligned} \tag{6.612}$$

Now, normalization gives

$$\int_{-\infty}^{\infty} \left|\psi_0^0\right|^2 dx = 1 \rightarrow A = \frac{1}{\sqrt{2}} \rightarrow \psi_0^0 = \frac{1}{\sqrt{2}}\frac{1}{\cosh(x)} \tag{6.613}$$

The ground state energy is $E_0^0 = 0$ (by definition). The energy eigenstates for

$$\hat{H}_1 = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2} \tag{6.614}$$

are given by

$$\hat{H}_1\psi_k^1 = -\frac{1}{2}\frac{d^2\psi_k^1}{dx^2} + \frac{1}{2}\psi_k^1 = E_k^1\psi_k^1 \tag{6.615}$$

$$\frac{d^2\psi_k^1}{dx^2} + 2(E_k^1 - \frac{1}{2})\psi_k^1 = 0 \tag{6.616}$$

which has solutions

$$\psi_k^1(x) = e^{ikx} \tag{6.617}$$

where

$$k^2 = 2(E_k^1 - \frac{1}{2}) \text{ or } E_k^1 = \frac{1}{2}(1 + k^2) \tag{6.618}$$

From the formalism, the remaining normalized eigenfunctions of $\hat{H}_0$ are given by

$$\begin{aligned}\psi_k^0(x) &= \frac{1}{\sqrt{E_k^1}}\hat{a}^+\psi_k^1 \\ &= \frac{1}{\sqrt{\frac{1}{2}(1+k^2)}}\frac{1}{\sqrt{2}}\left[-\frac{d}{dx} + \tanh(x)\right]e^{ikx} \\ &= \frac{(-ik + \tanh(x))}{\sqrt{(1+k^2)}}e^{ikx}\end{aligned} \tag{6.619}$$

The corresponding energy eigenvalues are also E_k^1. These continuum states have the property that they do not possess reflected waves, hence the name reflection-free potentials.

6.14.3 Generalizations

While the SUSY method seems to be very powerful, one wonders - how does one find the $\Phi(x)$ functions relevant to a particular Hamiltonian $\hat{H}$ that one wants to solve?

We can see how to do this by re-deriving this result in a different way.

We consider a general Hamiltonian $\hat{H}$ and define a set of operators$\hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3, ...$ and a set of real constants $E_1, E_2, E_3, ...$ such that they satisfy the recursion relations

$$\begin{aligned}
&\hat{\eta}_1^+ \hat{\eta}_1 + E_1 = \hat{H} \\
&\hat{\eta}_2^+ \hat{\eta}_2 + E_2 = \hat{\eta}_1 \hat{\eta}_1^+ + E_1 \\
&\hat{\eta}_3^+ \hat{\eta}_3 + E_3 = \hat{\eta}_2 \hat{\eta}_2^+ + E_2 \\
&..... \\
&.....
\end{aligned}$$

or, in general

$$\hat{\eta}_{j+1}^+ \hat{\eta}_{j+1} + E_{j+1} = \hat{\eta}_j \hat{\eta}_j^+ + E_j \quad , \quad j = 1, 2, 3, ... \tag{6.620}$$

Theorem - If each η_j has an eigenvector $|\xi_j\rangle$ with eigenvalue equal to zero such that

$$\hat{\eta}_j |\xi_j\rangle = 0 \tag{6.621}$$

then

(a) the constant E_j is the j^{th} eigenvalue of $\hat{H}$ (arranged in ascending order)

$$E_1 = ground\,state\,energy,\ E_2 = 1^{st} excited\,state\,energy,\ ..... \tag{6.622}$$

(b) the corresponding eigenvector is (ignoring normalization)

$$|E_j\rangle = \hat{\eta}_1^+ \hat{\eta}_2^+ \hat{\eta}_3^+ .....\hat{\eta}_{j-1}^+ |\xi_j\rangle \tag{6.623}$$

Before proving this theorem, let us talk about the meaning of the theorem.

Statement (a) implies not only that E_j is an eigenvalue, but also that there is no eigenvalue between E_{j-1} and E_j, i.e., if E is an eigenvalue of $\hat{H}$, then E equals one of the E_j or else E is larger than all of the E_j.

If we introduce E_{max} = least upper bound of the sequence $E_1, E_2, E_3, ...$ then

the theorem gives *all* the eigenvalues below E_{max}. If $E_{max} = \infty$ (as in the harmonic oscillator), then this theorem implies all of the eigenvalues. If $E_{max} = 0$ (as in the hydrogen atom), then this theorem implies *all negative* eigenvalues.

The theorem yields all parts of the *discrete spectrum* and says nothing about the continuous part of the spectrum (if it exists).

Proof of the Theorem

We first define new operators

$$\hat{A}_j = \hat{\eta}_j^+ \hat{\eta}_j + E_j \tag{6.624}$$

so that using the recursion relations we get

$$\hat{A}_{j+1} = \hat{\eta}_{j+1}^+ \hat{\eta}_{j+1} + E_{j+1} = \hat{\eta}_j \hat{\eta}_j^+ + E_j \quad \text{for all } j \tag{6.625}$$

We then have

$$\begin{aligned}
\hat{A}_{j+1}\hat{\eta}_j &= \left(\hat{\eta}_j\hat{\eta}_j^+ + E_j\right)\hat{\eta}_j = \hat{\eta}_j\left(\hat{\eta}_j^+\hat{\eta}_j + E_j\right) = \hat{\eta}_j\hat{A}_j \\
\hat{A}_j\hat{\eta}_j^+ &= \left(\hat{\eta}_j^+\hat{\eta}_j + E_j\right)\hat{\eta}_j^+ = \hat{\eta}_j^+\left(\hat{\eta}_j\hat{\eta}_j^+ + E_j\right) = \hat{\eta}_j^+\hat{A}_{j+1} \\
\hat{H} &= \hat{A}_1 = \hat{\eta}_1^+\hat{\eta}_1 + E_1
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{H}\,|E_j\rangle &= \hat{A}_1\hat{\eta}_1^+\hat{\eta}_2^+\hat{\eta}_3^+ \ldots\ldots\ldots\ldots \hat{\eta}_{j-1}^+\,|\xi_j\rangle = \hat{\eta}_1^+\hat{A}_2\hat{\eta}_2^+\hat{\eta}_3^+ \ldots\ldots\ldots\ldots \hat{\eta}_{j-1}^+\,|\xi_j\rangle \\
&= \hat{\eta}_1^+\hat{\eta}_2^+\hat{A}_3\hat{\eta}_3^+ \ldots\ldots\ldots\ldots \hat{\eta}_{j-1}^+\,|\xi_j\rangle = \ldots\ldots \quad \text{and so on until} \\
&= \hat{\eta}_1^+\hat{\eta}_2^+\hat{A}_3\hat{\eta}_3^+ \ldots\ldots\ldots\ldots \hat{\eta}_{j-1}^+\hat{A}_j\,|\xi_j\rangle
\end{aligned}$$

But our assumption $\hat{\eta}_j\,|\xi_j\rangle = 0$ then gives

$$\hat{A}_j\,|\xi_j\rangle = \left(\hat{\eta}_j^+\hat{\eta}_j + E_j\right)|\xi_j\rangle = \hat{\eta}_j^+\left(\hat{\eta}_j\,|\xi_j\rangle\right) + E_j\,|\xi_j\rangle = E_j\,|\xi_j\rangle \tag{6.626}$$

or $|\xi_j\rangle$ is an eigenstate of $\hat{A}_j$ with eigenvalue E_j. This gives

$$\hat{H}\,|E_j\rangle = E_j\,|E_j\rangle \tag{6.627}$$

or $|E_j\rangle$ is an eigenstate of $\hat{H}$ with eigenvalue E_j.

Now consider the quantity $E_{j+1} - E_j$ and assume that $\langle\xi_{j+1}\,|\,\xi_{j+1}\rangle = 1$. We then have

$$\begin{aligned}
E_{j+1} - E_j &= \langle\xi_{j+1}|\left(E_{j+1} - E_j\right)|\xi_{j+1}\rangle = \langle\xi_{j+1}|\left(\hat{\eta}_j\hat{\eta}_j^+ - \hat{\eta}_{j+1}^+\hat{\eta}_{j+1}\right)|\xi_{j+1}\rangle \\
&= \langle\xi_{j+1}|\left(\hat{\eta}_j\hat{\eta}_j^+\,|\xi_{j+1}\rangle \;\text{ since } \hat{\eta}_{j+1}\,|\xi_{j+1}\rangle = 0
\end{aligned} \tag{6.628}$$

Now let $|\beta\rangle = \hat{\eta}_j^+\,|\xi_{j+1}\rangle$. We then have $E_{j+1} - E_j = \langle\beta\,|\,\beta\rangle \geq 0$, which says that

$$E_1 \leq E_2 \leq E_3 \leq E_4 \leq \ldots\ldots\ldots\ldots \tag{6.629}$$

Finally, consider the eigenvector $|E\rangle$ of $\hat{H}$ and define the vector

$$|\rho_n\rangle = \hat{\eta}_n\hat{\eta}_{n-1}\hat{\eta}_{n-2}\hat{\eta}_{n-3}.....\hat{\eta}_2\hat{\eta}_1\,|E\rangle \tag{6.630}$$

We then have

$$\begin{aligned} 0 \le \langle\rho_1\,|\,\rho_1\rangle &= \langle E|\,\hat{\eta_1^+}\hat{\eta}_1\,|E\rangle = \langle E|\,(\hat{A}_1 - E_1)\,|E\rangle \\ &= \langle E|\,(\hat{H} - E_1)KetE = \langle E|\,(E - E_1)\,|E\rangle \end{aligned} \tag{6.631}$$

or

$$E - E_1 \ge 0 \rightarrow E \ge E_1 \tag{6.632}$$

Similarly,

$$\begin{aligned} 0 \le \langle\rho_2\,|\,\rho_2\rangle &= \langle E|\,\hat{\eta_1^+}\hat{\eta_2^+}\hat{\eta}_2\hat{\eta}_1\,|E\rangle = \langle E|\,\hat{\eta_1^+}(\hat{A}_2 - E_2)\hat{\eta}_1\,|E\rangle \\ &= \langle E|\,\hat{\eta_1^+}(\hat{A}_2\hat{\eta}_1 - \hat{\eta}_1E_2)\,|E\rangle = \langle E|\,\hat{\eta_1^+}(\hat{\eta}_1\hat{A}_2 - \hat{\eta}_1E_2)\,|E\rangle \\ &= \langle E|\,\hat{\eta_1^+}\hat{\eta}_1(\hat{H} - E_2)\,|E\rangle = (E - E_2)\,\langle E|\,\hat{\eta_1^+}\hat{\eta}_1\,|E\rangle \\ &= (E - E_2)(E - E_1) \end{aligned} \tag{6.633}$$

But, since $E \ge E_1$, this says that $E \ge E_2$. Generalizing we have

$$0 \le \langle\rho_2\,|\,\rho_2\rangle = (E - E_n)(E - E_{n-1}).....(E - E_2)(E - E_1) \tag{6.634}$$

which implies that $E \ge E_j$ for all j or E = one of the E_j, which concludes the proof.

This theorem does not, however, tell us how to find the operators.

The crucial step in the method is the factorization of the Hamiltonian $\hat{H}$, i.e.,

$$\hat{H} = \hat{\eta}_i^+\hat{\eta}_1 + E_1 \rightarrow \hat{H} - E_1 = \hat{\eta}_i^+\hat{\eta}_1 = \text{ "square" of an operator} \tag{6.635}$$

Once we choose $\hat{\eta}_1$, it is not difficult to construct the others.

During the construction, any adjustable parameters *must* be chosen in such a way that all the are as large as possible, which then guarantees a unique solution to any problem. To see that this requirement is necessary, consider a change in $\hat{\eta}_j$ and E_j with $\hat{\eta}_{j-1}$ and E_{j-1} held fixed. Now

$$\hat{\eta}_j^+\hat{\eta}_j + E_j = \hat{\eta}_{j-1}^+\hat{\eta}_{j-1} + E_{j-1} \tag{6.636}$$

implies that

$$\delta(\hat{\eta}_j^+\hat{\eta}_j) + \delta E_j = \delta(\hat{\eta}_{j-1}\hat{\eta}_{j-1}^+ + E_{j-1}) = 0 \tag{6.637}$$

when $\hat{\eta}_{j-1}$ and E_{j-1} are held fixed. Therefore, we get

$$\begin{aligned} \delta E_j &= \langle\xi_j|\,\delta E_j\,|\xi_j\rangle = \langle\xi_j|\,-\delta(\hat{\eta}_j^+\hat{\eta}_j)\,|\xi_j\rangle = \langle\xi_j|\,-(\delta\hat{\eta}_j^+)\hat{\eta}_j - \hat{\eta}_j^+(\delta\hat{\eta}_j)\,|\xi_j\rangle \\ &= -\langle\xi_j|\,(\delta\hat{\eta}_j^+)\hat{\eta}_j\,|\xi_j\rangle - \langle\xi_j|\,\hat{\eta}_j^+(\delta\hat{\eta}_j)\,|\xi_j\rangle = 0 - \langle\xi_j|\,\hat{\eta}_j^+(\delta\hat{\eta}_j)\,|\xi_j\rangle \\ &= -\langle\xi_j|\,(\delta\hat{\eta}_j^+)^+\hat{\eta}_j\,|\xi_j\rangle^* = 0 \end{aligned}$$

This says that we are at an extremum or physically at a *maximum*.

In the first derivation we simply guessed the function $\Phi(x)$, calculated $V_0(x)$ and then solved the problem corresponding to

$$\hat{H}_0 = -\frac{d^2}{dx^2} + V_0(x) \tag{6.638}$$

t turns out, however, that for the particular form of $\hat{H}$ that we have been considering, we can do more than guess. In particular, for

$$\hat{H} = -\frac{\hat{p}^2}{2m} + V(\hat{x}) \tag{6.639}$$

we can always choose

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}(\hat{p} + i f_j(\hat{x}) \tag{6.640}$$

where $f_j(\hat{x})$ = a *real* function of the operator $\hat{x}$.

6.14.4 Examples

(1) **Harmonic Oscillator revisited** - The Hamiltonian for this system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}x^2 \tag{6.641}$$

We assume

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}(\hat{p} + i f_j(\hat{x})) \tag{6.642}$$

We then have, from earlier discussions,

$$[\hat{x}, \hat{p}] = i\hbar \quad , \quad [f(\hat{x}), \hat{p}] = ih\frac{df(x)}{dx} \tag{6.643}$$

The last commutator follows by operating on an arbitrary function $g(x)$

$$\begin{aligned}
[f(\hat{x}), \hat{p}]\, g(x) &= [f(\hat{x})\hat{p} - \hat{p}f(\hat{x})]g(x) \\
&= -i\hbar[f(x)\frac{d}{dx} - \frac{d}{dx}f(x)]g(x) = -i\hbar[f\frac{dg}{dx} - \frac{d(fg)}{dx}] \\
&= -i\hbar[f\frac{dg}{dx} - f\frac{dg}{dx} - g\frac{df}{dx}] = ih\frac{df}{dx}\mathrm{g(x)}
\end{aligned}$$

which gives the commutation relation.

We then get (using $\hat{p}^+ = \hat{p}$, $\hat{x}^+ = \hat{x}$ and $f^+(\hat{x}) = f^*(x) = f(x)$)

$$\begin{aligned}
\hat{\eta}_j^+\hat{\eta}_j &= \frac{1}{2m}(\hat{p} - i f_j)(\hat{p} + i f_j) \\
&= \frac{1}{2m}\hat{p}^2 + \frac{1}{2m}f_j^2 - \frac{i}{2m}[f_j, \hat{p}] \\
&= \frac{1}{2m}\hat{p}^2 + \frac{1}{2m}f_j^2 + \frac{\hbar}{2m}\frac{df_j}{dx}
\end{aligned}$$

Similarly,

$$\hat{\eta}_j \hat{\eta}_j^+ = \frac{1}{2m}\hat{p}^2 + \frac{1}{2m} f_j^2 - \frac{\hbar}{2m}\frac{df_j}{dx} \tag{6.644}$$

We then have

$$\begin{aligned} \hat{H} &= \hat{\eta}_1^+ \hat{\eta}_1 + E_1 \\ &= \frac{1}{2m}\hat{p}^2 + \frac{1}{2m} f_1^2 + \frac{\hbar}{2m}\frac{df_1}{dx} = \frac{1}{2m}\hat{p}^2 + \frac{1}{2} m\omega^2 x^2 \end{aligned} \tag{6.645}$$

and

$$\frac{1}{2m} f_1^2 + \frac{\hbar}{2m}\frac{df_1}{dx} = \frac{1}{2} m\omega^2 x^2 \tag{6.646}$$

which implies that

$$f_1(x) = \pm m\omega x \quad , \quad E_1 = \mp\frac{1}{2}\hbar\omega \tag{6.647}$$

We need only find a particular solution that guarantees the existence of a vector $|\xi_j\rangle$ such that $\hat{\eta}_j |\xi_j\rangle = 0$.

Which sign does this imply?

The choice of the minus sign guarantees an ascending order of eigenvalues. The choice of the plus sign implies a descending order and no lower limit to the set of eigenvalues, which makes no physical sense.

So we choose

$$f_1(x) = -m\omega x \quad , \quad E_1 = +\frac{1}{2}\hbar\omega \tag{6.648}$$

We now find f_2 using

$$\hat{\eta}_2^+ \hat{\eta}_2 + E_2 = \hat{\eta}_1 \hat{\eta}_1^+ + E_1 \tag{6.649}$$

$$\frac{1}{2m} f_2^2 + \frac{\hbar}{2m}\frac{df_2}{dx} = \frac{1}{2} m\omega^2 x^2 + \hbar\omega \tag{6.650}$$

which gives

$$f_2(x) = -m\omega x \quad , \quad E_2 = +\frac{3}{2}\hbar\omega \tag{6.651}$$

Similarly, the recursion structure gives

$$f_j(x) = -m\omega x \quad , \quad E_j = +(j - 1/2)\hbar\omega \tag{6.652}$$

Thus, all the $\hat{\eta}_j$ are identical

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}(\hat{p} - im\omega\hat{x}) \tag{6.653}$$

and

$$|E_j\rangle \propto \left(\hat{\eta}_j^+\right)^{j-1} |\xi_j\rangle \propto (\hat{p} + im\omega\hat{x})^{j-1} |\xi_j\rangle \tag{6.654}$$

where

$$\hat{\eta}_j \left|\xi_j\right\rangle = \frac{1}{\sqrt{2m}}(\hat{p} - im\omega\hat{x}) \left|\xi_j\right\rangle = 0 \tag{6.655}$$

Finally, we show that the state $|\xi_j\rangle$ exists. We have

$$\hat{\eta}_1 \left|\xi_1\right\rangle = 0 \rightarrow \left|\xi_1\right\rangle = \text{ the ground state} \tag{6.656}$$

This says that

$$(\hat{p} - im\omega\hat{x}) \left|\xi_1\right\rangle = 0 \tag{6.657}$$

$$\left\langle x\right| (\hat{p} - im\omega\hat{x}) \left|\xi_1\right\rangle = \left\langle x\right| \hat{p} \left|\xi_1\right\rangle - im\omega \left\langle x\right| \hat{x} \left|\xi_1\right\rangle = 0 \tag{6.658}$$

$$-i\hbar \frac{d}{dx} \left\langle x \mid \xi_1\right\rangle - im\omega x \left\langle x \mid \xi_1\right\rangle = 0 \tag{6.659}$$

This last differential equation says that

$$\left\langle x \mid \xi_1\right\rangle = A \exp\left[-\frac{m\omega}{2\hbar} x^2\right] \tag{6.660}$$

which is the same ground state wave function as before!

Does this procedure really work for all Hamiltonians $\hat{H}$ of this form?

(2) **One-Dimensional Infinite Square Well** - This is probably the most difficult case to consider because the potential involved is not defined by an explicit, analytic function of x. It only enters the problem implicitly via boundary conditions at the walls. This means that $\hat{H}$ has no explicit dependence on x. We must, however, somehow introduce a dependence on x in the factorization procedure.

Let the well extend from $x = 0$ to $x = L$. We then have (inside the well)

$$\hat{H} = \frac{\hat{p}^2}{2m} \tag{6.661}$$

Again, we assume

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}(\hat{p} + if_j(\hat{x})) \tag{6.662}$$

which gives

$$\frac{1}{2m} f_1^2 + \frac{\hbar}{2m} \frac{df_1}{dx} + E_1 = 0$$

$$\frac{\hbar df_1}{2mE_1 + f_1} = -dx \rightarrow \int \frac{\hbar df_1}{2mE_1 + f_1} = -\int dx$$

$$\hbar \int \frac{df_1}{2mE_1 + f_1} = \hbar \frac{1}{\sqrt{2mE_1}} \tan^{-1} \frac{f_1}{\sqrt{2mE_1}} = -(x - x_0)$$

or

$$f_1(x) = -\sqrt{2mE_1}\tan\left[\frac{\sqrt{2mE_1}}{\hbar}(x - x_0)\right] \tag{6.663}$$

where x_0 is the constant of integration.

The possible choices of E_1 and x_0 are restricted by the behavior of the tangent function. In the position representation, x is a number such that $0 < x < L$. The tangent must remain finite in this interval. The singularities of $\tan y$ occur at

$$y = \frac{\pi}{2}\,,\ \frac{3\pi}{2}\,,\ \frac{5\pi}{2}\,,\ \ldots\ldots\ldots\ldots \tag{6.664}$$

i.e., they are separated by π radians. We attain the largest possible value of E_1 if one of the singularities lies at $x = 0$ and the next at $x = L$ (it is permissible for the tangent to have a singularity at these points (the walls) since the potential and hence $\hat{H}$ are also infinite there. So we assume that

$$\text{at x } = 0 \quad \tan\left[\frac{\sqrt{2mE_1}}{\hbar}x_0\right] = \infty \tag{6.665}$$

$$\text{at x } = L \quad \tan\left[-\frac{\sqrt{2mE_1}}{\hbar}(L - x_0)\right] = \infty \tag{6.666}$$

which imply that

$$\frac{\sqrt{2mE_1}}{\hbar}x_0 = \frac{\pi}{2} \quad , \quad -\frac{\sqrt{2mE_1}}{\hbar}(L - x_0) = \frac{3\pi}{2} = -\frac{\sqrt{2mE_1}}{\hbar}L + \frac{\pi}{2} \tag{6.667}$$

or

$$-\frac{\sqrt{2mE_1}}{\hbar}L = \pi \tag{6.668}$$

This gives

$$E_1 = \frac{\hbar^2\pi^2}{2mL^2} \tag{6.669}$$

and

$$\begin{aligned} f_1 &= -\sqrt{2mE_1}\tan\left[\frac{\sqrt{2mE_1}}{\hbar}(x - \frac{\pi\hbar}{2\sqrt{2mE_1}})\right] \\ &= \frac{\pi\hbar}{L}\tan\left[-\frac{\pi}{L}x - \frac{\pi}{2}\right] = \frac{\pi\hbar}{L}\frac{\sin\left[-\frac{\pi}{L}x - \frac{\pi}{2}\right]}{\cos\left[-\frac{\pi}{L}x - \frac{\pi}{2}\right]} \\ &= \frac{\pi\hbar}{L}\frac{\cos(\frac{\pi x}{L})}{\sin(\frac{\pi x}{L})} = \frac{\pi\hbar}{L}\cot(\frac{\pi x}{L}) \end{aligned} \tag{6.670}$$

This implies that

$$\hat{\eta}_1 = \frac{1}{\sqrt{2m}}\left[\hat{p} + i\frac{\pi\hbar}{L}\cot\frac{\pi x}{L}\right] \tag{6.671}$$

Reflecting on this result we now choose

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}\left[\hat{p} + ic_j \cot(b_j x)\right] \tag{6.672}$$

where c_j and b_j are to be chosen to give the correct recursion relations. Now, in order to guarantee $0 < x < L$ we must have $0 \leq b_j \leq \pi/L$. If we apply the recursion relations we get

$$\hat{\eta}^+_{j+1}\hat{\eta}_{j+1} = \frac{1}{2m}\left[\hat{p}^2 - c_{j+1}b_{j+1}\hbar + c_{j+1}(c_{j+1} - b_{j+1}\hbar)\cot^2 b_{j+1}x\right] \tag{6.673}$$

$$\hat{\eta}_j\hat{\eta}^+_j = \frac{1}{2m}\left[\hat{p}^2 - c_j b_j\hbar + c_j(c_j + b_j\hbar)\cot^2 b_j x\right] \tag{6.674}$$

which implies that

$$\begin{aligned}\frac{1}{2m}&\left[\hat{p}^2 - c_{j+1}b_{j+1}\hbar + c_{j+1}(c_{j+1} - b_{j+1}\hbar)\cot^2 b_{j+1}x\right] + E_{j+1} \\ &= \frac{1}{2m}\left[\hat{p}^2 - c_j b_j\hbar + c_j(c_j + b_j\hbar)\cot^2 b_j x\right] + E_j\end{aligned} \tag{6.675}$$

Equating powers of x gives

$$b_{j+1} = b_j \rightarrow b_j = b_1 = \frac{\pi}{L} \tag{6.676}$$

$$c_{j+1}(c_{j+1} - b_{j+1}\hbar) = c_j(c_j + b_j\hbar) \tag{6.677}$$

$$2mE_{j+1} - c_{j+1}b_{j+1}\hbar = 2mE_j - c_j b_j\hbar \tag{6.678}$$

or

$$2mE_j - (c_j)^2 = 2mE_1 - (c_1)^2 = 0 \tag{6.679}$$

$$E_j = \frac{(c_j)^2}{2m} \tag{6.680}$$

and

$$c_{j+1}(c_{j+1} - \frac{\pi\hbar}{L}) = c_j(c_j + \frac{\pi\hbar}{L}) \tag{6.681}$$

which gives

$$c_{j+1} = -c_j \text{ or } c_{j+1} = c_j + \frac{\pi\hbar}{\mathrm{L}} \tag{6.682}$$

The last choice implies the largest E_1. Therefore,

$$c_j = j\frac{\pi\hbar}{L} \text{ and } E_j = \frac{j^2\pi^2\hbar^2}{2mL^2} \tag{6.683}$$

Finally, we show the existence of $|\xi_j\rangle$ where $\hat{\eta}_j|\xi_j\rangle = 0$. This corresponds to the equation

$$\left[-i\hbar\frac{d}{dx} + i\frac{j\pi\hbar}{L}\cot\left(\frac{\pi x}{L}\right)\right]\langle x \mid \xi_j\rangle = 0 \tag{6.684}$$

or

$$\langle x \mid \xi_j \rangle = \left(\sin \frac{\pi x}{L} \right)^j \tag{6.685}$$

Now

$$|E_j\rangle = \hat{\eta}_1^+ \hat{\eta}_2^+ .....\hat{\eta}_{j-1}^+ |\xi_j\rangle \tag{6.686}$$

implies that

$$\psi_j(x) = \langle x \mid E_j \rangle = \langle x| \hat{\eta}_1^+ \hat{\eta}_2^+ .....\hat{\eta}_{j-1}^+ |\xi_j\rangle \tag{6.687}$$

Therefore,

$$\psi_1(x) = \langle x \mid E_1 \rangle = \langle x \mid \xi_1 \rangle = \sin \frac{\pi x}{L} \tag{6.688}$$

$$\begin{aligned} \psi_2(x) &= \langle x \mid E_2 \rangle = \hat{\eta}_1^+ \langle x \mid \xi_2 \rangle \\ &= \left[-i\hbar \frac{d}{dx} + i \frac{\pi \hbar}{L} \cot \left(\frac{\pi x}{L} \right) \right] \sin^2 \frac{\pi x}{L} \sim \sin \frac{2\pi x}{L} \end{aligned} \tag{6.689}$$

Similarly,

$$\psi_3(x) = \hat{\eta}_1^+ \hat{\eta}_2^+ \langle x \mid \xi_3 \rangle \sim \sin \frac{3\pi x}{L} \tag{6.690}$$

and so on.

We also find

$$V_0(x) = \frac{1}{2} \frac{\psi_0''}{\psi_0} \tag{6.691}$$

where

$$\psi_0(x) = \sqrt{\frac{2}{L}} \sin \frac{\pi x}{l} \tag{6.692}$$

or

$$V_0(x) = \frac{1}{2} \frac{\pi^2}{L^2} \tag{6.693}$$

In addition,

$$W(x) = -\frac{1}{\sqrt{2}} \frac{\psi_0'}{\psi_0} = -\frac{1}{\sqrt{2}} \frac{\pi}{L} \cot \frac{\pi x}{L} \tag{6.694}$$

Finally, the superpartner potential is

$$V_1(x) = V_0(x) - \frac{d}{dx} \frac{\psi_0'}{\psi_0} = \sqrt{\frac{2}{L}} \sin \frac{\pi x}{L} + \frac{1}{\sqrt{2}} \frac{\pi}{L} \frac{d}{dx} \cot \frac{\pi x}{L} \tag{6.695}$$

This completes the solution. The method works even in this case!

(3) **Hydrogen Atom** - As we will see in later discussions, when we write down the 3-dimensional Schrodinger equation for the hydrogen atom and separate the variables, we obtain a 1−dimensional equation in the variable r corresponding to radius. This equation looks like

$$\hat{H}_\ell u_{n\ell}(r) = \left[\frac{1}{2m} \hat{p}_r^2 + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} - \frac{e^2}{r} \right] u_{n\ell}(r) = E_{n\ell} u_{n\ell}(r) \tag{6.696}$$

The factorization method should also work on this equation. We choose

$$\hat{\eta}_1^+ \hat{\eta}_1 = \frac{1}{2m}\hat{p}_r^2 + \frac{\hbar^2 \ell(\ell+1)}{2mr^2} - \frac{e^2}{r} \tag{6.697}$$

$$\hat{\eta}_{j+1}^+ \hat{\eta}_{j+1} + E_{j+1} = \hat{\eta}_j \hat{\eta}_j^+ + E_j \tag{6.698}$$

For convenience, we have suppressed the ℓ dependence ($\hat{\eta}_j \leftrightarrow \hat{\eta}_j^\ell$). Now assume

$$\hat{\eta}_j = \frac{1}{\sqrt{2m}}\left(\hat{p}_r + i\left(b_j + \frac{c_j}{r}\right)\right) \tag{6.699}$$

which gives

$$\hat{\eta}_j^+ \hat{\eta}_j = \frac{1}{2m}\left[\hat{p}_r^2 + b_j^2 + 2b_j\frac{c_j}{r} + \frac{c_j}{r}(c_j - \hbar)\right] \tag{6.700}$$

$$\hat{\eta}_j \hat{\eta}_j^+ = \frac{1}{2m}\left[\hat{p}_r^2 + b_j^2 + 2b_j\frac{c_j}{r} + \frac{c_j}{r}(c_j + \hbar)\right] \tag{6.701}$$

The definition of the Hamiltonian gives

$$\begin{aligned}\hat{\eta}_1^+ \hat{\eta}_1 + E_1 &= \frac{1}{2m}\left[\hat{p}_r^2 + b_1^2 + 2b_1\frac{c_1}{r} + \frac{c_1}{r}(c_1 - \hbar)\right] + E_1 \\ &= \frac{1}{2m}\hat{p}_r^2 + \frac{\hbar^2\ell(\ell+1)}{2mr^2} - \frac{e^2}{r}\end{aligned} \tag{6.702}$$

Equating powers of r gives

$$\frac{1}{2m}c_1(c_1 - \hbar) = \frac{1}{2m}\ell(\ell+1)\hbar^2 \tag{6.703}$$

$$\frac{b_1 c_1}{m} = -e^2 \tag{6.704}$$

$$\frac{b_1^2}{2m} + E_1 = 0 \tag{6.705}$$

which imply

$$c_1 = (\ell+1)\hbar \tag{6.706}$$

$$b_1 = -\frac{me^2}{(\ell+1)\hbar} \tag{6.707}$$

$$E_1 = -\frac{1}{2m}\left(\frac{me^2}{(\ell+1)\hbar}\right)^2 \tag{6.708}$$

The recursion relations then give

$$c_{j+1}(c_{j+1} - \hbar) = c_j(c_j + \hbar) \tag{6.709}$$

$$b_{j+1}c_{j+1} = b_j c_j \tag{6.710}$$

$$\frac{b_{j+1}^2}{2m} + E_{j+1} = \frac{b_j^2}{2m} + E_j \tag{6.711}$$

which imply

$$c_{j+1} = c_j + \hbar \to c_j = (j-1)\hbar + c_1 \to c_j = (\ell + j)\hbar \tag{6.712}$$

$$b_j c_j = b_1 c_1 = -me^2 \tag{6.713}$$

$$E_j = -\frac{b_j^2}{2m} = -\frac{1}{2m}\left(\frac{me^2}{(\ell+j)\hbar}\right)^2 \tag{6.714}$$

If we let $n = \ell + j$ we have

$$E_n = -\frac{1}{2}\alpha^2 \frac{mc^2}{n^2} \tag{6.715}$$

where

$$\alpha = \frac{e^2}{\hbar c} = \text{ the fine structure constant} \tag{6.716}$$

Finally, we get the eigenfunctions. We must have $\hat{\eta}_j^{(\ell)} |\xi_j\rangle = 0$. Using

$$\begin{aligned} \hat{\eta}_j^{(\ell)} &= \frac{1}{\sqrt{2m}}\left[\hat{p}_r + i\left[-\frac{me^2}{(\ell+j)\hbar} + \frac{(\ell+j)\hbar}{r}\right]\right] \\ &= \frac{1}{\sqrt{2m}}\left[\hat{p}_r - \frac{i\hbar}{(\ell+j)a_0} + \frac{i(\ell+j)\hbar}{r}\right] \end{aligned} \tag{6.717}$$

where

$$a_0 = \frac{\hbar^2}{me^2} = \text{ the Bohr radius} \tag{6.718}$$

we get

$$\left[\hat{p}_r - \frac{i\hbar}{(\ell+j)a_0} + \frac{i(\ell+j)\hbar}{r}\right]|\xi_j\rangle = 0 \tag{6.719}$$

where

$$\hat{p}_r = -i\hbar\left[\frac{d}{dr} + \frac{1}{r}\right] \tag{6.720}$$

This implies

$$\left[-i\hbar\left[\frac{d}{dr} + \frac{1}{r}\right] - \frac{i\hbar}{(\ell+j)a_0} + \frac{i(\ell+j)\hbar}{r}\right]\langle n, \ell \mid \xi_j\rangle = 0 \tag{6.721}$$

which has a solution

$$\langle n, \ell \mid \xi_j\rangle = r^{\ell+j-1} e^{-\frac{r}{(\ell+j)a_0}} \sim \psi_j^{(\ell)}(r) \tag{6.722}$$

We thus have

$$\psi_1^{(0)}(r) \sim e^{-\frac{r}{a_0}} \quad , \quad \psi_2^{(0)}(r) \sim \left(1 - \frac{r}{2a_0} e^{-\frac{r}{2a_0}}\right) \tag{6.723}$$

and so on, which are the correct wave functions!

6.14.5 Supersymmetric Quantum Mechanics

This is the study of the relationship between pairs of superpartner potentials. Some of the issues addressed are:

1. To what extent can the observed energy spectrum determine the potential energy function. In other words, can we *invert* the energy levels and find the $V(x)$ that generated them. Is the inversion unique?

2. The number of potentials for which we can analytically solve the Schrodinger equation is small. They are the infinite well, the harmonic oscillator (1–, 2- , and 3–dimensions), the Coulomb potential and a few others.

   All of these potentials can also be solved using the factorization method operator techniques. Thus, they all have supersymmetric analogs. All the superpartner potentials are similar in shape, differing only in the parameters in their definition.

3. Supersymmetric quantum mechanics suggests a connection between fermions and bosons at some fundamental level.

In supersymmetric quantum mechanics we study quantum mechanical systems where the Hamiltonian $\hat{H}$ is constructed from anticommuting charges $\hat{Q}$ which are the square root of $\hat{H}$, i.e.,

$$2\hat{H} = \left\{\hat{Q}, \hat{Q}^+\right\} = \hat{Q}\hat{Q}^+ + \hat{Q}^+\hat{Q} \tag{6.724}$$

$$0 = \left\{\hat{Q}, Q\right\} \rightarrow \hat{Q}^2 = 0 \tag{6.725}$$

These equations imply that

$$\begin{aligned}\left[\hat{H}, \hat{Q}\right] = \left[\hat{Q}, \hat{H}\right] &= \frac{1}{2}\left[\hat{Q}, \hat{Q}\hat{Q}^+ + \hat{Q}^+\hat{Q}\right] \\ &= \frac{1}{2}\left(\hat{Q}^2\hat{Q}^+ + \hat{Q}\hat{Q}^+\hat{Q} - \hat{Q}\hat{Q}^+\hat{Q} - \hat{Q}^+\hat{Q}^2\right) = 0\end{aligned}$$

or that the charge $\hat{Q}$ is a conserved observable.

These Hamiltonians contain coordinates that are quantized by commutators (bosonic coordinates) and anticommutators(fermionic coordinates). These coordinates are mixed by supersymmetry transformations.

For a particle with spin, the position and spin orientation form a pair of such coordinates. An explicit example is given by

$$\hat{Q} = \left(\hat{p} + i\varphi(\hat{x})\right)\hat{\psi}^+ \quad , \quad \hat{Q}^+ = \left(\hat{p} - i\varphi(\hat{x})\right)\hat{\psi} \tag{6.726}$$

where $\hat{x}$ and $\hat{p}$ are bosonic coordinates (degrees of freedom) and $\hat{\psi}$ and $\hat{\psi}^+$ are fermionic coordinates (we have set $\hbar = 1$) satisfying

$$[\hat{x}, \hat{p}] = i \quad , \quad \left\{\hat{\psi}, \hat{\psi}^+\right\} = 1 \quad , \quad \left\{\hat{\psi}, \hat{\psi}\right\} = \left\{\hat{\psi}^+, \hat{\psi}^+\right\} = 0 \tag{6.727}$$

i.e., we have commutators for the bosonic coordinates and anticommutators for the fermionic coordinates.

From these relations we get

$$\left\{\hat{Q},\hat{Q}\right\} = \left\{\hat{Q}^+,\hat{Q}^+\right\} = 0 \tag{6.728}$$

and

$$\hat{H} = \frac{1}{2}\hat{p}^2 + \frac{1}{2}\varphi^2(\hat{x}) - \frac{1}{2}\left[\hat{\psi},\hat{\psi}^+\right]\varphi'(\hat{x}) \tag{6.729}$$

Using a 2×2 representation of the fermionic coordinates , we have

$$\hat{\psi}^+ = \hat{\sigma}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad , \quad \hat{\psi} = \hat{\sigma}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{6.730}$$

$$\left\{\hat{\psi},\hat{\psi}^+\right\} = \hat{\sigma}_+\hat{\sigma}_- + \hat{\sigma}_-\hat{\sigma}_+ = \hat{I} \tag{6.731}$$

$$\left[\hat{\psi},\hat{\psi}^+\right] = -\hat{\sigma}_z = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6.732}$$

so that

$$\begin{aligned} \hat{H} &= \frac{1}{2}\left(\hat{p}^2 + \varphi^2(\hat{x})\right) + \frac{1}{2}\hat{\sigma}_z\varphi'(\hat{x}) \\ &= \quad \hat{H}_0 \quad + \quad \hat{H}_1 \\ &= \quad Bose\,sector \; + Fermi\,sector \end{aligned}$$

The two sectors have the same energy levels. The only exception is the case where the ground state of the Bose sector has zero energy and is thus non-degenerate.

6.14.6 Additional Thoughts

Supersymmetry transformations are represented by the unitary operator

$$\hat{U} = exp(\varepsilon\hat{Q} + \varepsilon^+\hat{Q}^+) \tag{6.733}$$

where ε and ε^+ are anti-commuting c-numbers (called a Grassman algebra).

Supersymmetric one-particle quantum mechanics serves as a model for the investigation of *spontaneous breaking* of supersymmetry, which is supposed to occur in supersymmetric field theories. The ground state $|0\rangle$ is invariant with respect to supersymmetry transformations provided that $\hat{U}|0\rangle = |0\rangle$. This is satisfied if and only if $\hat{Q}|0\rangle = \hat{Q}^+|0\rangle = 0$, that is, if the ground state energy is zero.

If the ground state energy is greater than zero, then supersymmetry is spontaneously broken.

An example of spontaneously broken symmetry is

$$\Phi = g(x^2 - a^2) \tag{6.734}$$

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{g^2}{2}(x^2 - a^2) + gx\hat{\sigma}_z \tag{6.735}$$

The two potentials satisfy $V_0(-x) = -V_1(x)$. There is no normalizable state in this case with $E_{00} = 0$ since

$$\int \Phi(x)\, dx = g\left\{\frac{1}{3}x^3 - a^2 x\right\} \tag{6.736}$$

says that the ground state energy is positive.

The world we actually observe is in one of the degenerate ground states and the supersymmetry is spontaneously broken!

6.15 Problems

6.15.1 Delta function in a well

A particle of mass m moving in one dimension is confined to a space $0 < x < L$ by an infinite well potential. In addition, the particle experiences a delta function potential of strength λ given by $\lambda\delta(x - L/2)$ located at the center of the well as shown in Figure 6.26 below.

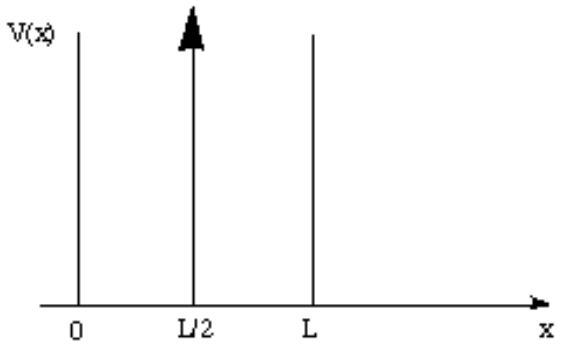


Figure 6.26: Potential Diagram

Find a transcendental equation for the energy eigenvalues E in terms of the mass m, the potential strength λ, and the size of the well L.

6.15.2 Properties of the wave function

A particle of mass m is confined to a one-dimensional region $0 \le x \le a$ (an infinite square well potential). At $t = 0$ its normalized wave function is

$$\psi(x, t = 0) = \sqrt{\frac{8}{5a}}\left(1 + \cos\left(\frac{\pi x}{a}\right)\right)\sin\left(\frac{\pi x}{a}\right)$$

(a) What is the wave function at a later time $t = t_0$?

(b) What is the average energy of the system at $t = 0$ and $t = t_0$?

(c) What is the probability that the particle is found in the left half of the box(i.e., in the region $0 \le x \le a/2$ at $t = t_0$?

6.15.3 Repulsive Potential

A repulsive short-range potential with a strongly attractive core can be approximated by a square barrier with a delta function at its center, namely,

$$V(x) = V_0\Theta(|x| - a) - \frac{\hbar^2 g}{2m}\delta(x)$$

(a) Show that there is a negative energy eigenstate (the ground-state).

(b) If E_0 is the ground-state energy of the delta-function potential in the absence of the positive potential barrier, then the ground-state energy of the present system satisfies the relation $E \ge E_0 + V_0$. What is the particular value of V_0 for which we have the limiting case of a ground-state with zero energy.

6.15.4 Step and Delta Functions

Consider a one-dimensional potential with a step-function component and an attractive delta function component just at the edge of the step, namely,

$$V(x) = V\Theta(x) - \frac{\hbar^2 g}{2m}\delta(x)$$

(a) For $E > V$, compute the reflection coefficient for particle incident from the left. How does this result differ from that of the step barrier alone at high energy?

(b) For $E < 0$ determine the energy eigenvalues and eigenfunctions of any bound-state solutions.

6.15.5 Atomic Model

An approximate model for an atom near a wall is to consider a particle moving under the influence of the one-dimensional potential given by

$$V(x) = \begin{cases} -V_0\delta(x) & x > -d \\ \infty & x < -d \end{cases}$$

as shown in Figure 6.27 below.

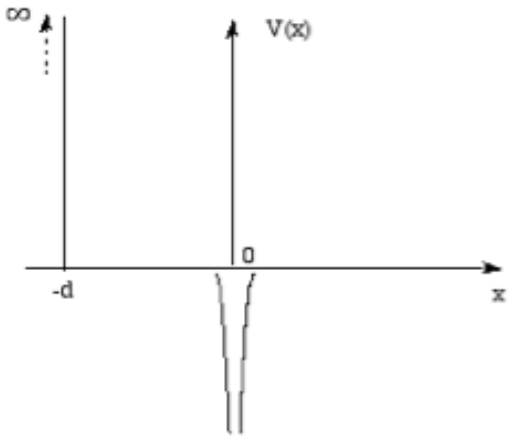


Figure 6.27: Potential Diagram

(a) Find the transcendental equation for the bound state energies.

(b) Find an approximation for the modification of the bound-state energy caused by the wall when it is *far away*. Define carefully what you mean by *far away*.

(c) What is the exact condition on V_0 and d for the existence of at least one bound state?

6.15.6 A confined particle

A particle of mass m is confined to a space $0 < x < a$ in one dimension by infinitely high walls at $x = 0$ and $x = a$. At $t = 0$ the particle is initially in the left half of the well with a wave function given by

$$\psi(x,0) = \begin{cases} \sqrt{2/a} & 0 < x < a/2 \\ 0 & a/2 < x < a \end{cases}$$

(a) Find the time-dependent wave function $\psi(x,t)$.

(b) What is the probability that the particle is in the n^{th} eigenstate of the well at time t?

(c) Derive an expression for average value of particle energy. What is the physical meaning of your result?

6.15.7 $1/x$ potential

An electron moves in one dimension and is confined to the right half-space ($x > 0$) where it has potential energy

$$V(x) = -\frac{e^2}{4x}$$

where e is the charge on an electron.

(a) What is the solution of the Schrodinger equation at large x?

(b) What is the boundary condition at $x = 0$?

(c) Use the results of (a) and (b) to guess the ground state solution of the equation. Remember the ground state wave function has no zeros except at the boundaries.

(d) Find the ground state energy.

(e) Find the expectation value $\langle \hat{x} \rangle$ in the ground state.

6.15.8 Using the commutator

Using the coordinate-momentum commutation relation prove that

$$\sum_n (E_n - E_0) \left| \langle E_n | \hat{x} | E_0 \rangle \right|^2 = \text{constant}$$

where E_0 is the energy corresponding to the eigenstate $|E_0\rangle$. Determine the value of the constant. Assume the Hamiltonian has the general form

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})$$

6.15.9 Matrix Elements for Harmonic Oscillator

Compute the following matrix elements

$$\langle m | \hat{x}^3 | n \rangle \quad , \quad \langle m | \hat{x}\hat{p} | n \rangle$$

6.15.10 A matrix element

Show for the one dimensional simple harmonic oscillator

$$\langle 0 | e^{ik\hat{x}} | 0 \rangle = \exp\left[-k^2 \langle 0 | \hat{x}^2 | 0 \rangle / 2 \right]$$

where $\hat{x}$ is the position operator.

6.15.11 Correlation function

Consider a function, known as the *correlation function*, defined by

$$C(t) = \langle \hat{x}(t)\hat{x}(0) \rangle$$

where $\hat{x}(t)$ is the position operator in the Heisenberg picture. Evaluate the correlation function explicitly for the ground-state of the one dimensional simple harmonic oscillator.

6.15.12 Instantaneous Force

Consider a simple harmonic oscillator in its ground state.

An instantaneous force imparts momentum p_0 to the system such that the new state vector is given by

$$|\psi\rangle = e^{-ip_0\hat{x}/\hbar} |0\rangle$$

where $|0\rangle$ is the ground-state of the original oscillator.

What is the probability that the system will stay in its ground state?

6.15.13 Coherent States

Coherent states are defined to be eigenstates of the annihilation or lowering operator in the harmonic oscillator potential. Each coherent state has a complex label z and is given by $|z\rangle = e^{z\hat{a}^+} |0\rangle$.

(a) Show that $\hat{a} |z\rangle = z |z\rangle$

(b) Show that $\langle z_1 \mid z_2 \rangle = e^{z_1^* z_2}$

(c) Show that the completeness relation takes the form

$$\hat{I} = \sum_n |n\rangle \langle n| = \int \frac{dxdy}{\pi} |z\rangle \langle z| e^{-z^* z}$$

where $|n\rangle$ is a standard harmonic oscillator energy eigenstate, $\hat{I}$ is the identity operator, $z = x + iy$, and the integration is taken over the whole $x - y$ plane(use polar coordinates).

6.15.14 Oscillator with Delta Function

Consider a harmonic oscillator potential with an extra delta function term at the origin, that is,

$$V(x) = \frac{1}{2}m\omega^2 x^2 + \frac{\hbar^2 g}{2m}\delta(x)$$

(a) Using the parity invariance of the Hamiltonian, show that the energy eigenfunctions are even and odd functions and that the simple harmonic oscillator odd-parity energy eigenstates are still eigenstates of the system Hamiltonian, with the same eigenvalues.

(b) Expand the even-parity eigenstates of the new system in terms of the even-parity harmonic oscillator eigenfunctions and determine the expansion coefficients.

(c) Show that the energy eigenvalues that correspond to even eigenstates are solutions of the equation

$$\frac{2}{g} = -\sqrt{\frac{\hbar}{m\pi\omega}} \sum_{k=0}^{\infty} \frac{(2k)!}{2^{2k}(k!)^2} \left(2k + \frac{1}{2} - \frac{E}{\hbar\omega}\right)^{-1}$$

You might need the fact that

$$\psi_{2k}(0) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \frac{\sqrt{(2k)!}}{2^k k!}$$

(d) Consider the following cases:

(1) $g > 0$, $E > 0$

(2) $g < 0$, $E > 0$

(3) $g < 0$, $E < 0$

Show the first and second cases correspond to an infinite number of energy eigenvalues.

Where are they relative to the original energy eigenvalues of the harmonic oscillator?

Show that in the third case, that of an attractive delta function core, there exists a single eigenvalue corresponding to the ground state of the system provided that the coupling is such that

$$\left[\frac{\Gamma(3/4)}{\Gamma(1/4)}\right]^2 < \frac{g^2\hbar}{16m\omega} < 1$$

You might need the series summation:

$$\sum_{k=0}^{\infty} \frac{(2k)!}{4^k(k!)^2} \frac{1}{2k+1-x} = \frac{\sqrt{\pi}}{2} \frac{\Gamma(1/2 - x/2)}{\Gamma(1 - x/2)}$$

You will need to look up other properties of the gamma function to solve this problem.

6.15.15 Measurement on a Particle in a Box

Consider a particle in a box of width a, prepared in the ground state.

(a) What are then possible values one can measure for: (1) energy, (2) position, (3) momentum ?

(b) What are the probabilities for the possible outcomes you found in part (a)?

(c) At some time (call it t = 0) we perform a measurement of position. However, our detector has only finite resolution. We find that the particle is in the middle of the box (call it the origin) with an uncertainty $\Delta x = a/2$, that is, we know the position is, for sure, in the range $-a/4 < x < a/4$, but we are completely uncertain where it is within this range. What is the (normalized) post-measurement state?

(d) Immediately after the position measurement what are the possible values for (1) energy, (2) position, (3) momentum and with what probabilities?

(e) At a later time, what are the possible values for (1) energy, (2) position, (3) momentum and with what probabilities? Comment.

6.15.16 Aharonov-Bohm experiment

Consider an infinitely long solenoid which carries a current I so that there is a constant magnetic field inside the solenoid(see Figure 6.28 below).

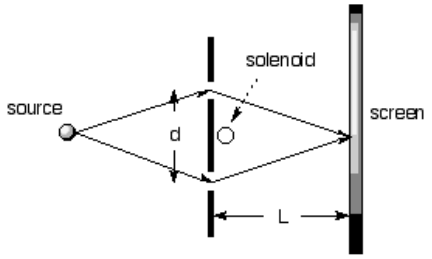


Figure 6.28: Aharonov-Bohm Setup

Suppose that in the region outside the solenoid the motion of a particle with charge e and mass m is described by the Schrodinger equation. Assume that for $I = 0$, the solution of the equation is given by

$$\psi_0(\vec{r}, t) = e^{iE_0 t/\hbar}\psi_0(\vec{r})$$

(a) Write down and solve the Schrodinger equation in the region outside the solenoid in the case $I \neq 0$.

(b) Consider the two-slit diffraction experiment for the particles described above shown in Figure 8.3 above. Assume that the distance d between the two slits is large compared to the diameter of the solenoid.

Compute the shift ΔS of the diffraction pattern on the screen due to the presence of the solenoid with $I \neq 0$. Assume that $L \gg \Delta S$.

6.15.17 A Josephson Junction

A Josephson junction is formed when two superconducting wires are separated by an insulating gap of capacitance C. The quantum states ψ_i , $i = 1, 2$ of the two wires can be characterized by the numbers n_i of Cooper pairs (charge $= -2e$) and phases θ_i, such that $\psi_i = \sqrt{n_i}e^{i\theta_i}$ (Ginzburg-Landau approximation). The (small) amplitude that a pair tunnel across a narrow insulating barrier is $-E_J/n_0$ where $n_0 = n_1 + n_2$ and E_J is the the so-called Josephson energy. The interesting physics is expressed in terms of the differences

$$n = n_2 - n_1 \quad , \quad \varphi = \theta_2 - \theta_1$$

We consider a junction where

$$n_1 \approx n_2 \approx n_0/2$$

When there exists a nonzero difference n between the numbers of pairs of charge $-2e$, where $e > 0$, on the two sides of the junction, there is net charge $-ne$ on side 2 and net charge $+ne$ on side 1. Hence a voltage difference ne/C arises, where the voltage on side 1 is higher than that on side 2 if $n = n_2 - n_1 > 0$. Taking the zero of the voltage to be at the center of the junction, the electrostatic energy of the Cooper pair of charge $-2e$ on side 2 is ne^2/C, and that of a pair on side 1 is $-ne^2/C$. The total electrostatic energy is $C(\Delta V)^2/2 = Q^2/2C = (ne)^2/2C$.

The equations of motion for a pair in the two-state system $(1, 2)$ are

$$i\hbar\frac{d\psi_1}{dt} = U_1\psi_1 - \frac{E_J}{n_0}\psi_2 = -\frac{ne^2}{C}\psi_1 - \frac{E_J}{n_0}\psi_2$$
$$i\hbar\frac{d\psi_2}{dt} = U_2\psi_2 - \frac{E_J}{n_0}\psi_1 = \frac{ne^2}{C}\psi_2 - \frac{E_J}{n_0}\psi_1$$

(a) Discuss the physics of the terms in these equations.

(b) Using $\psi_i = \sqrt{n_i}e^{i\theta_i}$, show that the equations of motion for n and φ are given by

$$\dot{\varphi} = \dot{\theta}_2 - \dot{\theta}_1 \approx -\frac{2ne^2}{\hbar C}$$
$$\dot{n} = \dot{n}_2 - \dot{n}_1 \approx \frac{E_J}{\hbar}\sin\varphi$$

(c) Show that the pair(electric current) from side 1 to side 2 is given by

$$J_S = J_0 \sin\varphi \quad , \quad J_0 = \frac{\pi E_J}{\phi_0}$$

(d) Show that

$$\ddot{\varphi} \approx -\frac{2e^2 E_J}{\hbar^2 C}\sin\varphi$$

For E_J positive, show that this implies there are oscillations about $\varphi = 0$ whose angular frequency (called the Josephson plasma frequency)is given by

$$\omega_J = \sqrt{\frac{2e^2 E_J}{\hbar^2 C}}$$

for small amplitudes.

If E_J is negative, then there are oscillations about $\varphi = \pi$.

(e) If a voltage $V = V_1 - V_2$ is applied across the junction(by a battery), a charge $Q_1 = VC = (-2e)(-n/2) = en$ is held on side 1, and the negative of this on side 2. Show that we then have

$$\dot{\varphi} \approx -\frac{2eV}{\hbar} \equiv -\omega$$

which gives $\varphi = \omega t$.

The battery holds the charge difference across the junction fixed at $VC - en$, but can be a source or sink of charge such that a current can flow in the circuit. Show that in this case, the current is given by

$$J_S = -J_0 \sin\omega t$$

i.e., the DC voltage of the battery generates an AC pair current in circuit of frequency

$$\omega = \frac{2eV}{\hbar}$$

6.15.18 Eigenstates using Coherent States

Obtain eigenstates of the following Hamiltonian

$$\hat{H} = \hbar\omega\hat{a}^+\hat{a} + V\hat{a} + V^*\hat{a}^+$$

for a complex V using coherent states.

6.15.19 Bogoliubov Transformation

Suppose annihilation and creation operators satisfy the standard commutation relations $[\hat{a}, \hat{a}^+] = 1$. Show that the Bogoliubov transformation

$$\hat{b} = \hat{a}\cosh\eta + \hat{a}^+\sinh\eta$$

preserves the commutation relation of the creation and annihilation operators, i.e., $[\hat{b}, \hat{b}^+] = 1$. Use this fact to obtain eigenvalues of the following Hamiltonian

$$\hat{H} = \hbar\omega\hat{(}a)^+\hat{a} + \frac{1}{2}V\left(\hat{a}\hat{a} + \hat{a}^+\hat{a}^+\right)$$

(There is an upper limit on V for which this can be done). Also show that the unitary operator

$$\hat{U} = e^{(\hat{a}\hat{a}+\hat{a}^+\hat{a}^+)\eta/2}$$

can relate the two sets of operators as $\hat{b} = \hat{U}\hat{a}\hat{U}^{-1}$.

6.15.20 Harmonic oscillator

Consider a particle in a 1−dimensional harmonic oscillator potential. Suppose at time $t = 0$, the state vector is

$$|\psi(0)\rangle = e^{-\frac{i\hat{p}a}{\hbar}}|0\rangle$$

where $\hat{p}$ is the momentum operator and a is a real number.

(a) Use the equation of motion in the Heisenberg picture to find the operator $\hat{x}(t)$.

(b) Show that $e^{-\frac{i\hat{p}a}{\hbar}}$ is the translation operator.

(c) In the Heisenberg picture calculate the expectation value $\langle x\rangle$ for $t \geq 0$.

6.15.21 Another oscillator

A 1−dimensional harmonic oscillator is, at time $t = 0$, in the state

$$|\psi(t=0)\rangle = \frac{1}{\sqrt{3}}\left(|0\rangle + |1\rangle + |2\rangle\right)$$

where $|n\rangle$ is the n^{th} energy eigenstate. Find the expectation value of position and energy at time t.

6.15.22 The coherent state

Consider a particle of mass m in a harmonic oscillator potential of frequency ω. Suppose the particle is in the state

$$|\alpha\rangle = \sum_{n=0}^{\infty} c_n |n\rangle$$

where

$$c_n = e^{-|\alpha|^2/2} \frac{\alpha^n}{\sqrt{n!}}$$

and α is a complex number. As we have discussed, this is a *coherent state* or alternatively a *quasi-classical state*.

(a) Show that $|\alpha\rangle$ is an eigenstate of the annihilation operator, i.e., $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$.

(b) Show that in this state $\langle\hat{x}\rangle = x_c \mathrm{Re}(\alpha)$ and $\langle\hat{p}\rangle = p_c \mathrm{Im}(\alpha)$. Determine x_c and p_c.

(c) Show that, in position space, the wave function for this state is $\psi_\alpha(x) = e^{ip_0 x/\hbar} u_0(x - x_0)$ where $u_0(x)$ is the ground state gaussian function and $\langle\hat{x}\rangle = x_0$ and $\langle\hat{p}\rangle = p_0$.

(d) What is the wave function in momentum space? Interpret x_0 and p_0.

(e) Explicitly show that $\psi_\alpha(x)$ is an eigenstate of the annihilation operator using the position-space representation of the annihilation operator.

(f) Show that the coherent state is a minimum uncertainty state (with equal uncertainties in x and p, in characteristic dimensionless units.

(g) If a time $t = 0$ the state is $|\psi(0)\rangle = |\alpha\rangle$, show that at a later time,

$$|\psi(t)\rangle = e^{-i\omega t/2} \left|\alpha e^{-i\omega t}\right\rangle$$

Interpret this result.

(h) Show that, as a function of time, $\langle\hat{x}\rangle$ and $\langle\hat{p}\rangle$ follow the classical trajectory of the harmonic oscillator, hence the name *quasi-classical* state.

(i) Write the wave function as a function of time, $\psi_\alpha(x,t)$. Sketch the time evolving probability density.

(j) Show that in the classical limit

$$\lim_{|\alpha|\to\infty} \frac{\Delta N}{\langle N\rangle} \to 0$$

(k) Show that the probability distribution in n is Poissonian, with appropriate parameters.

(l) Use a *rough* time-energy uncertainty principle($\Delta E \Delta t > \hbar$) , to find an uncertainty principle between the number and phase of a quantum oscillator.

6.15.23 Neutrino Oscillations

It is generally recognized that there are at least three different kinds of neutrinos. They can be distinguished by the reactions in which the neutrinos are created or absorbed. Let us call these three types of neutrino ν_e, ν_μ and ν_τ. It has been speculated that each of these neutrinos has a small but finite rest mass, possibly different for each type. Let us suppose, for this exam question, that there is a small perturbing interaction between these neutrino types, in the absence of which all three types of neutrinos have the same nonzero rest mass M_0. The Hamiltonian of the system can be written as

$$\hat{H} = \hat{H}_0 + \hat{H}_1$$

where

$$\hat{H}_0 = \begin{pmatrix} M_0 & 0 & 0 \\ 0 & M_0 & 0 \\ 0 & 0 & M_0 \end{pmatrix} \rightarrow \text{ no interactions present}$$

and

$$\hat{H}_1 = \begin{pmatrix} 0 & \hbar\omega_1 & \hbar\omega_1 \\ \hbar\omega_1 & 0 & \hbar\omega_1 \\ \hbar\omega_1 & \hbar\omega_1 & 0 \end{pmatrix} \rightarrow \text{ effect of interactions}$$

where we have used the basis

$$|\nu_e\rangle = |1\rangle \quad , \quad |\nu_\mu\rangle = |2\rangle \quad , \quad |\nu_\tau\rangle = |3\rangle$$

(a) First assume that $\omega_1 = 0$, i.e., no interactions. What is the time development operator? Discuss what happens if the neutrino initially was in the state

$$|\psi(0)\rangle = |\nu_e\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ or } |\psi(0)\rangle = |\nu_\mu\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ or } |\psi(0)\rangle = |\nu_\tau\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

What is happening physically in this case?

(b) Now assume that $\omega_1 \neq 0$, i.e., interactions are present. Also assume that at $t = 0$ the neutrino is in the state

$$|\psi(0)\rangle = |\nu_e\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

What is the probability as a function of time, that the neutrino will be in each of the other two states?

(c) An experiment to detect the *neutrino oscillations* is being performed. The flight path of the neutrinos is 2000 meters. Their energy is $100\,GeV$. The sensitivity of the experiment is such that the presence of 1% of neutrinos different from those present at the start of the flight can be measured with confidence. Let $M_0 = 20\,eV$. What is the smallest value of $\hbar\omega_1$ that can be detected? How does this depend on M_0? Don't ignore special relativity.

6.15.24 Generating Function

Use the generating function for Hermite polynomials

$$e^{2xt-t^2} = \sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!}$$

to work out the matrix elements of x in the position representation, that is, compute

$$\langle x\rangle_{nn'} = \int_{-\infty}^{\infty} \psi_n^*(x)x\psi_{n'}(x)dx$$

where

$$\psi_n(x) = N_n H_n(\alpha x)e^{-\frac{1}{2}\alpha^2 x^2}$$

and

$$N_n = \left(\frac{\alpha}{\sqrt{\pi}2^n n!}\right)^{1/2} \quad , \quad \alpha = \left(\frac{m\omega}{\hbar}\right)^{1/2}$$

6.15.25 Given the wave function .....

A particle of mass m moves in one dimension under the influence of a potential $V(x)$. Suppose it is in an energy eigenstate

$$\psi(x) = \left(\frac{\gamma^2}{\pi}\right)^{1/4} \exp\left(-\gamma^2 x^2/2\right)$$

with energy $E = \hbar^2\gamma^2/2m$.

(a) Find the mean position of the particle.

(b) Find the mean momentum of the particle.

(c) Find $V(x)$.

(d) Find the probability $P(p)dp$ that the particle's momentum is between p and $p + dp$.

6.15.26 What is the oscillator doing?

Consider a one dimensional simple harmonic oscillator. Use the number basis to do the following algebraically:

(a) Construct a linear combination of $|0\rangle$ and $|1\rangle$ such that $\langle\hat{x}\rangle$ is as large as possible.

(b) Suppose the oscillator is in the state constructed in (a) at $t = 0$. What is the state vector for $t > 0$? Evaluate the expectation value $\langle\hat{x}\rangle$ as a function of time for $t > 0$ using (i)the Schrodinger picture and (ii) the Heisenberg picture.

(c) Evaluate $\left\langle(\Delta x)^2\right\rangle$ as a function of time using either picture.

6.15.27 Coupled oscillators

Two identical harmonic oscillators in one dimension each have a mass m and frequency ω. Let the two oscillators be coupled by an interaction term Cx_1x_2 where C is a constant and x_1 and x_2 are the coordinates of the two oscillators. Find the exact energy spectrum of eigenvalues for this coupled system.

6.15.28 Interesting operators ....

The operator $\hat{c}$ is defined by the following relations:

$$\hat{c}^2 = 0 \quad , \quad \hat{c}\hat{c}^+ + \hat{c}^+\hat{c} = \{\hat{c}, \hat{c}^+\} = \hat{I}$$

(a) Show that

1. $\hat{N} = \hat{c}^+\hat{c}$ is Hermitian
2. $\hat{N}^2 = \hat{N}$
3. The eigenvalues of $\hat{N}$ are 0 and 1 (eigenstates $|0\rangle$ and $|1\rangle$)
4. $\hat{c}^+|0\rangle = |1\rangle \quad , \quad \hat{c}|0\rangle = 0$

(b) Consider the Hamiltonian

$$\hat{H} = \hbar\omega_0(\hat{c}^+\hat{c} + 1/2)$$

Denoting the eigenstates of $\hat{H}$ by $|n\rangle$, show that the only nonvanishing states are the states $|0\rangle$ and $|1\rangle$ defined in (a).

(c) Can you think of any physical situation that might be described by these new operators?

6.15.29 What is the state?

A particle of mass m in a one dimensional harmonic oscillator potential is in a state for which a measurement of the energy yields the values $\hbar\omega/2$ or $3\hbar\omega/2$, each with a probability of one-half. The average value of the momentum $\langle\hat{p}_x\rangle$ at time $t = 0$ is $(m\omega\hbar/2)^{1/2}$. This information specifies the state of the particle completely. What is this state and what is $\langle\hat{p}_x\rangle$ at time t?

6.15.30 Things about a particle in a box

A particle of mass m moves in a one-dimensional box Infinite well) of length ℓ with the potential

$$V(x) = \begin{cases} \infty & x < 0 \\ 0 & 0 < x < \ell \\ \infty & x > \ell \end{cases}$$

At $t = 0$, the wave function of this particle is known to have the form

$$\psi(x, 0) = \begin{cases} \sqrt{30/\ell^5}\,x(\ell - x) & 0 < x < \ell \\ 0 & \text{otherwise} \end{cases}$$

(a) Write this wave function as a linear combination of energy eigenfunctions

$$\psi_n(x) = \sqrt{\frac{2}{\ell}} \sin\left(\frac{\pi n x}{\ell}\right) \quad , \quad E_n = n^2 \frac{\pi^2 \hbar^2}{2m\ell^2} \quad , \quad n = 1, 2, 3, ...$$

(b) What is the probability of measuring E_n at $t = 0$?

(c) What is $\psi(x, t > 0)$?

6.15.31 Handling arbitrary barriers.....

Electrons in a metal are bound by a potential that may be approximated by a finite square well. Electrons fill up the energy levels of this well up to an energy called the *Fermi energy* as shown in the figure below:

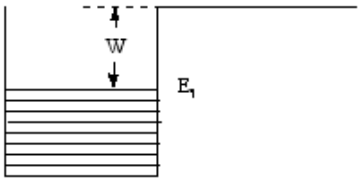


Figure 6.29: Finite Square Well

The difference between the Fermi energy and the top of the well is the *work function* W of the metal. Photons with energies exceeding the work function can eject electrons from the metal - this is the so-called *photoelectric effect*.

Another way to pull out electrons is through application of an external uniform electric field $\vec{\mathcal{E}}$, which alters the potential energy as shown in the figure below:

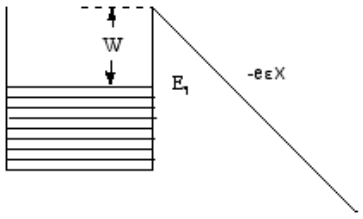


Figure 6.30: Finite Square Well + Electric Field

?By approximating (see notes below) the linear part of the function by a series of square barriers, show that the transmission coefficient for electrons at the Fermi energy is given by

$$T \approx \exp\left(\frac{-4\sqrt{2m}W^{3/2}}{3e\,|\vec{\varepsilon}|\,\hbar}\right)$$

How would you expect this *field- or cold-emission* current to vary with the applied voltage? As part of your problem solution explain the method.

This calculation also plays a role in the derivation of the current-voltage characteristic of a Schottky diode in semiconductor physics.

Approximating an Arbitrary Barrier

For a rectangular barrier of width a and height V_0, we found the transmission coefficient

$$T = \frac{1}{1 + \frac{V_0^2 \sinh^2 \gamma a}{4E(V_0 - E)}}, \gamma^2 = (V_0 - E)\frac{2m}{\hbar^2}, k^2 = \frac{2m}{\hbar^2}E$$

A useful limiting case occurs for $\gamma a \gg 1$. In this case

$$\sinh \gamma a = \frac{e^{\gamma a} - e^{-\gamma a}}{2} \underset{\gamma a >> 1}{\rightarrow} \frac{e^{\gamma a}}{2}$$

so that

$$T = \frac{1}{1 + \left(\frac{\gamma^2 + k^2}{4k\gamma}\right)^2 \sinh^2 \gamma a} \underset{\gamma a >> 1}{\rightarrow} \left(\frac{4k\gamma}{\gamma^2 + k^2}\right)^2 e^{-2\gamma a}$$

Now if we evaluate the natural log of the transmission coefficient we find

$$\ln T \underset{\gamma a >> 1}{\rightarrow} \ln\left(\frac{4k\gamma}{\gamma^2 + k^2}\right)^2 - 2\gamma a \underset{\gamma a \gg 1}{\rightarrow} -2\gamma a$$

where we have dropped the logarithm relative to γa since $\ln(almost\, anything)$ is not very large. This corresponds to only including the exponential term.

We can now use this result to calculate the probability of transmission through a non-square barrier, such as that shown in the figure below:

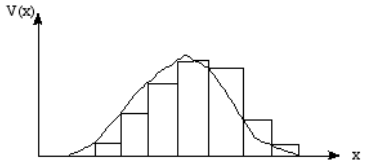


Figure 6.31: Arbitrary Barrier Potential

When we only include the exponential term, the probability of transmission through an arbitrary barrier, as above, is just the product of the individual transmission coefficients of a succession of rectangular barrier as shown above. Thus, if the barrier is sufficiently smooth so that we can approximate it by a series of rectangular barriers (each of width Δx) that are not too thin for the condition $\gamma a \gg 1$ to hold, then for the barrier as a whole

$$\ln T \approx \ln \prod_i T_i = \sum_i \ln T_i = -2 \sum_i \gamma_i \Delta x$$

If we now assume that we can approximate this last term by an integral, we find

$$T \approx \exp\left(-2\sum_i \gamma_i \Delta x\right) \approx \exp\left(-2\int \sqrt{\frac{2m}{\hbar^2}}\sqrt{V(x)-E}dx\right)$$

where the integration is over the region for which the square root is real.

You may have a somewhat uneasy feeling about this crude derivation. Clearly, the approximations made break down near the turning points, where $E = V(x)$. Nevertheless, a more detailed treatment shows that it works amazingly well.

6.15.32 Deuteron model

Consider the motion of a particle of mass $m = 0.8 \times 10^{-24}\, gm$ in the well shown in the figure below:

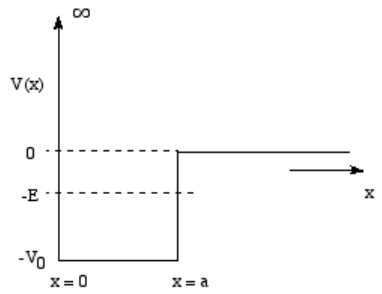


Figure 6.32: Deuteron Model

The size of the well (range of the potential) is $a = 1.4 \times 10^{-13}\, cm$. If the binding energy of the system is $2.2\, MeV$, find the depth of the potential V_0 in MeV. This is a model of the deuteron in one dimension.

6.15.33 Use Matrix Methods

A one-dimensional potential barrier is shown in the figure below.

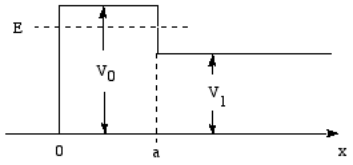


Figure 6.33: A Potential Barrier

Define and calculate the transmission probability for a particle of mass m and energy E $(V_1 < E < V_0)$ incident on the barrier from the left. If you let $V_1 \to 0$

and $a \to 2a$, then you can compare your answer to other textbook results. Develop matrix methods (as in the text) to solve the boundary condition equations.

6.15.34 Finite Square Well Encore

Consider the symmetric finite square well of depth V_0 and width a.

(a) Let $k_0 = \sqrt{2mV_0/\hbar^2}$. Sketch the bound states for the following choices of $k_0 a/2$.

(i) $\frac{k_0 a}{2} = 1$, $(ii)\frac{k_0 a}{2} = 1.6$, $(iii)\frac{k_0 a}{2} = 5$

(b) Show that no matter how shallow the well, there is at least one bound state of this potential. Describe it.

(c) Let us re-derive the bound state energy for the delta function well directly from the limit of the the finite potential well. Use the graphical solution discussed in the text. Take the limit as $a \to 0$, $V_0 \to \infty$, but $aV_0 \to U_0$(constant) and show that the binding energy is $E_b = mU_0^2/2\hbar^2$.

(d) Consider now the half-infinite well, half-finite potential well as shown below.

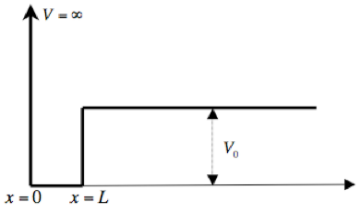


Figure 6.34: Half-Infinite, Half-Finite Well

Without doing any calculation, show that there are no bound states unless $k_0 L > \pi/2$. HINT: think about erecting an infinite wall down the center of a symmetric finite well of width $a = 2L$. Also, think about parity.

(e) Show that in general, the binding energy eigenvalues satisfy the eigenvalue equation

$$\kappa = -k \cot kL$$

where

$$\kappa = \sqrt{\frac{2mE_b}{\hbar^2}} \quad and \quad k^2 + \kappa^2 = k_0^2$$

6.15.35 Half-Infinite Half-Finite Square Well Encore

Consider the unbound case ($E > V_0$) eigenstate of the potential below.

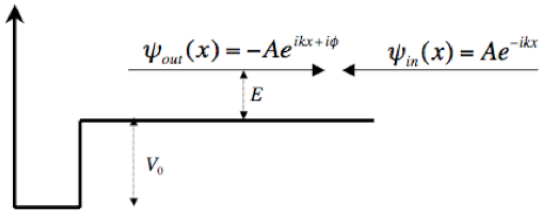


Figure 6.35: Half-Infinite, Half-Finite Well Again

Unlike the potentials with finite wall, the scattering in this case has only one output channel - reflection. If we send in a plane wave towards the potential, $\psi_{in}(x) = Ae^{-ikx}$, where the particle has energy $E = (\hbar k)^2/2m$, the reflected wave will emerge from the potential with a phase shift, $\psi_{out}(x) = Ae^{ikx+\phi}$,

(a) Show that the reflected wave is phase shifted by

$$\phi = 2\tan^{-1}\left(\frac{k}{q}\tan qL\right) - 2kL$$

where

$$q^2 = k^2 + k_0^2 \quad , \quad \frac{\hbar^2 k_0^2}{2m} = V_0$$

(b) Plot the function of ϕ as a function of $k_0 L$ for fixed energy. Comment on your plot.

(c) The phase shifted reflected wave is equivalent to that which would arise from a hard wall, but moved a distance L' from the origin.

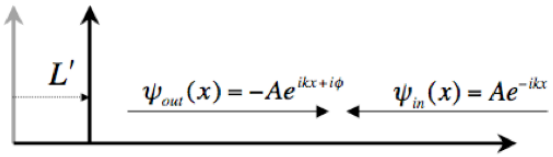


Figure 6.36: Shifted Wall

What is the effective L' as a function of the phase shift ϕ induced by our semi-finite well? What is the maximum value of L'? Can L' be negative? From your plot in (b), sketch L' as a function of $k_0 L$, for fixed energy. Comment on your plot.

6.15.36 Nuclear α Decay

Nuclear *alpha*–decays $(A, Z) \to (A-2, Z-2) + \alpha$ have lifetimes ranging from nanoseconds (or shorter) to millions of years (or longer). This enormous range

was understood by George Gamov by the exponential sensitivity to underlying parameters in tunneling phenomena. Consider $\alpha = {}^4He$ as a point particle in the potential given schematically in the figure below.

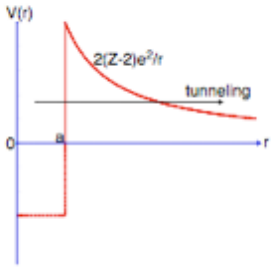


Figure 6.37: Nuclear Potential Model

The potential barrier is due to the Coulomb potential $2(Z-2)e^2/r$. The probability of tunneling is proportional to the so-called Gamov's transmission coefficients obtained in Problem 8.31

$$T = exp\left[-\frac{2}{\hbar}\int_a^b \sqrt{2m(V(x)-E)}\,dx\right]$$

where a and b are the classical turning points (where $E = V(x)$) Work out numerically T for the following parameters: $Z = 92$ (Uranium), size of nucleus $a = 5\,fm$ and the kinetic energy of the α particle $1\,MeV$, $3\,MeV$, $10\,MeV$, $30\,MeV$.

6.15.37 One Particle, Two Boxes

Consider two boxes in 1-dimension of width a, with infinitely high walls, separated by a distance $L = 2a$. We define the box by the potential energy function sketched below.

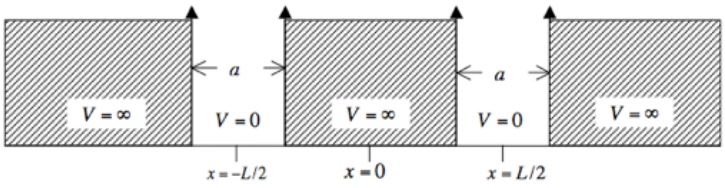


Figure 6.38: Two Boxes

A particle experiences this potential and its state is described by a wave function. The energy eigenfunctions are *doubly* degenerate, $\left\{\phi_n^{(+)}, \phi_n^{(-)} \,|\, n = 1, 2, 3, 4, ...\right\}$ so

that

$$E_n^{(+)} = E_n^{(-)} = n^2 \frac{\pi^2 \hbar^2}{2ma^2}$$

where $\phi_n^{(\pm)} = u_n(x \pm L/2)$ with

$$u_n(x) = \begin{cases} \sqrt{2/a}\cos\left(\frac{n\pi x}{a}\right), \ n = 1, 3, 5, ... & -a/2 < x < a/2 \\ \sqrt{2/a}\sin\left(\frac{n\pi x}{a}\right), \ n = 2, 4, 6, ... & -a/2 < x < a/2 \\ 0 & |x| > a/2 \end{cases}$$

Suppose at time $t = 0$ the wave function is

$$\psi(x) = \frac{1}{2}\phi_1^{(-)}(x) + \frac{1}{2}\phi_2^{(-)}(x) + \frac{1}{\sqrt{2}}\phi_1^{(+)}(x)$$

At this time, answer parts (a) - (d)

(a) What is the probability of finding the particle in the state $\phi_1^{(+)}(x)$?

(b) What is the probability of finding the particle with energy $\pi^2\hbar^2/2ma^2$?

(c) CLAIM: At $t = 0$ there is a 50-50 chance for finding the particle in either box. Justify this claim.

(d) What is the state at a later time assuming no measurements are done?

Now let us generalize. Suppose we have an arbitrary wave function at $t = 0$, $\psi(x, 0)$, that satisfies all the boundary conditions.

(e) Show that, in general, the probability to find the particle in the left box does not change with time. Explain why this makes sense physically.

Switch gears again .....

(f) Show that the state $\Phi_n(x) = c_1\phi_n^{(+)}(x) + c_2\phi_n^{(-)}(x)$ (where c_1 and c_2 are arbitrary complex numbers) is a stationary state.

Consider then the state described by the wave function $\psi(x) = (\phi_1^{(+)}(x) + c_2\phi_1^{(-)}(x))/\sqrt{2}$.

(g) Sketch the probability density in x. What is the mean value $\langle x \rangle$? How does this change with time?

(h) Show that the momentum space wave function is

$$\tilde{\psi}(p) = \sqrt{2}\cos(pL/2\hbar)\tilde{u}_1(p)$$

where

$$\tilde{u}_1(p) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} u_1(x)e^{-ipx/\hbar}$$

is the momentum-space wave function of $u_1(x)$.

(i) Without calculation, what is the mean value $\langle p \rangle$? How does this change with time?

(j) Suppose the potential energy was somehow *turned off* (don't ask me how, just imagine it was done) so the particle is now free.

Without doing any calculation, sketch how you expect the position-space wave function to evolve at later times, showing all important features. Please explain your sketch.

6.15.38 A half-infinite/half-leaky box

Consider a one dimensional potential

$$V(x) = \begin{cases} \infty & x < 0 \\ U_0 \delta(x-a) & x > 0 \end{cases}$$

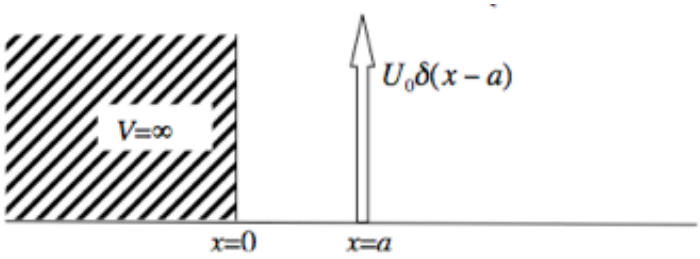


Figure 6.39: Infinite Wall + Delta Function

(a) Show that the stationary states with energy E can be written

$$u(x) = \begin{cases} 0 & x < 0 \\ A \frac{\sin(ka+\phi(k))}{\sin(ka)} \sin(kx) & 0 < x < a \\ A \sin(kx + \phi(k)) & x > a \end{cases}$$

where

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad , \quad \phi(k) = \tan^{-1}\left[\frac{k \tan(ka)}{k - \gamma_0 \tan(ka)}\right] \quad , \quad \gamma_0 = \frac{2mU_0}{\hbar^2}$$

What is the nature of these states - bound or unbound?

(b) Show that the limits $\gamma_0 \to 0$ and $\gamma_0 \to \infty$ give reasonable solutions.

(c) Sketch the energy eigenfunction when $ka = \pi$. Explain this solution.

(d) Sketch the energy eigenfunction when $ka = \pi/2$. How does the probability to find the particle in the region $0 < x < a$ compare with that found in part (c)? Comment.

(e) In a scattering scenario, we imagine sending in an incident plane wave which is reflected with unit probability, but phase shifted according to the conventions shown in the figure below:

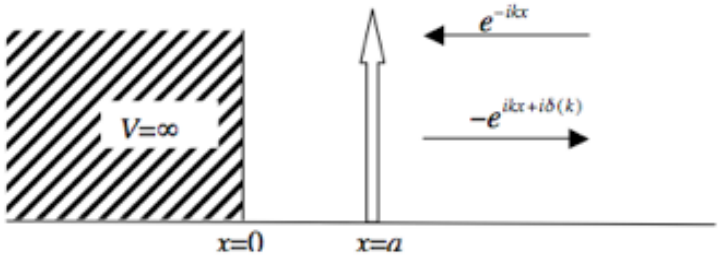


Figure 6.40: Scattering Scenario

Show that the phase shift of the scattered wave is $\delta(k) = 2\phi(k)$.

There exist mathematical conditions such that the so-called *S-matrix element* $e^{i\delta(k)}$ blows up. For these solutions is k real, imaginary, or complex? Comment.

6.15.39 Neutrino Oscillations Redux

Read the article T. Araki *et al*, "Measurement of Neutrino Oscillations with Kam LAND: Evidence of Spectral Distortion," *Phys. Rev. Lett.* **94**, 081801 (2005), which shows the neutrino oscillation, a quantum phenomenon demonstrated at the largest distance scale yet (about $180\,km$).

(a) The Hamiltonian for an ultrarelativistic particle is approximated by

$$H = \sqrt{p^2c^2 + m^2c^4} \approx pc + \frac{m^2c^3}{2p}$$

for p=$|\vec{p}|$. Suppose in a basis of two states, m^2 is given as a 2×2 matrix

$$m^2 = m_0^2 I + \frac{\Delta m^2}{2}\begin{pmatrix} -\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix}$$

Write down the eigenstates of m^2.

(b) Calculate the probability for the state

$$|\psi\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

to be still found in the same state after time interval t for definite momentum p.

(c) Using the data shown in Fig. 3 of the article, estimate approximately values of Δm^2 and $\sin^2 2\theta$.

6.15.40 Is it in the ground state?

An infinitely deep one-dimensional potential well runs fro $x = 0$ to $x = a$. The normalized energy eigenstates are

$$u_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \ , \quad n = 1, 2, 3, ...$$

A particle is placed in the left-hand half of the well so that its wavefunction is ψ = constant for $x < a/2$. If the energy of the particle is now measured, what is the probability of finding it in the ground state?

6.15.41 Some Thoughts on T-Violation

Any Hamiltonian can be recast to the form

$$H = U \begin{pmatrix} E_1 & 0 & \dots & 0 \\ 0 & E_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & E_n \end{pmatrix} U^+$$

where U is a general $n \times n$ unitary matrix.

(a) Show that the time evolution operator is given by

$$e^{-iHt/\hbar} = U \begin{pmatrix} e^{-iE_1 t/\hbar} & 0 & \dots & 0 \\ 0 & e^{-iE_2 t/\hbar} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{-iE_n t/\hbar} \end{pmatrix} U^+$$

(b) For a two-state problem, the most general unitary matrix is

$$U = e^{i\theta} \begin{pmatrix} \cos\theta e^{i\phi} & -\sin\theta e^{i\eta} \\ \sin\theta e^{-i\eta} & \cos\theta e^{-i\phi} \end{pmatrix}$$

Work out the probabilities $P(1 \to 2)$ and $P(2 \to 1)$ over time interval t and verify that they are the same despite the the apparent T-violation due to complex phases. NOTE: This is the same problem as the neutrino oscillation (problem 8.39) if you set $E_i = \sqrt{p^2c^2 + m^2c^4} \approx pc + \frac{m^2c^3}{2p}$ and set all phases to zero.

(c) For a three-state problem, however, the time-reversal invariance can be broken. Calculate the difference $P(1 \to 2) - P(2 \to 1)$ for the following form of the unitary matrix

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where five unimportant phases have been dropped. The notation is $s_{12} = \sin\theta_{12}, c_{23} = \cos\theta_{23}$, etc.

(d) For CP-conjugate states (e.g.., anti-neutrinos($\bar{\nu}$) vs neutrinos(ν), the Hamiltonian is given by substituting U^* in place of U. Show that the probabilities $P(1 \to 2)$ and $P(\bar{1} \to \bar{2})$ can differ (CP violation) yet CPT is respected, ie., $P(1 \to 2) = P(\bar{2} \to \bar{1})$.

6.15.42 Kronig-Penney Model

Consider a periodic repulsive potential of the form

$$V = \sum_{n=-\infty}^{\infty} \lambda\delta(x - na)$$

with $\lambda > 0$. The general solution for $-a < x < 0$ is given by

$$\psi(x) = Ae^{i\kappa x} + Be^{-i\kappa x}$$

with $\kappa = \sqrt{2mE}/\hbar$. Using Bloch's theorem, the wave function for the next period $0 < x < a$ is given by

$$\psi(x) = e^{ika}\left(Ae^{i\kappa(x-a)} + Be^{-i\kappa(x-a)}\right)$$

for $|k| \le \pi/a$. Answer the following questions.

(a) Write down the continuity condition for the wave function and the required discontinuity for its derivative at $x = 0$. Show that the phase e^{ika} under the discrete translation $x \to x + a$ is given by κ as

$$e^{ika} = \cos\kappa a + \frac{1}{\kappa d}\sin\kappa a \pm i\sqrt{1 - \left(\cos\kappa a + \frac{1}{\kappa d}\sin\kappa a\right)^2}$$

Here and below, $d = \hbar^2/m\lambda$.

(b) Take the limit of zero potential $d \to \infty$ and show that there are no gaps between the bands as expected for a free particle.

(c) When the potential is weak but finite (largte d) show analytically that there appear gaps between the bands at $k = \pm\pi/a$.

(d) Plot the relationship between κ and k for a weak potential ($d = 3a$) and a strong potential ($d = a/3$) (both solutions together).

(e) You always find two values of k at the same energy (or κ). What discrete symmetry guarantees this degeneracy?

6.15.43 Operator Moments and Uncertainty

Consider an observable O_A for a finite-dimensional quantum system with spectral decomposition

$$O_A = \sum_i \lambda_i P_i$$

(a) Show that the exponential operator $E_A = exp(O_A)$ has spectral decomposition

$$E_A = \sum_i e^{\lambda_i P_i}$$

Do this by inserting the spectral decomposition of O_A into the power series expansion of the exponential.

(b) Prove that for any state $|\Psi_A\rangle$ such that $\Delta O_A = 0$, we automatically have $\Delta E_A = 0$.

6.15.44 Uncertainty and Dynamics

Consider the observable

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

and the initial state

$$|\Psi_A(0)\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(a) Compute the uncertainty $\Delta O_X = 0$ with respect to the initial state $|\Psi_A(0)\rangle$.

(b) Now let the state evolve according to the Schrodinger equation, with Hamiltonian operator

$$H = \hbar \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

Compute the uncertainty $\Delta O_X = 0$ as a function of t.

(c) Repeat part (b) but replace O_X with the observable

$$O_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

That is, compute the uncertainty ΔO_Z as a function of t assuming evolution according to the Schrodinger equation with the Hamiltonian above.

(d) Show that your answers to parts (b) and (c) always respect the Heisenberg Uncertainty Relation

$$\Delta O_X \Delta O_Z \geq \frac{1}{2} |\langle [O_X, O_Z] \rangle|$$

Are there any times t at which the Heisenberg Uncertainty Relation is satisfied with equality?

Chapter 7

Angular Momentum; 2- and 3-Dimensions

7.1 Angular Momentum Eigenvalues and Eigenvectors

Earlier, we derived the commutation relations that define angular momentum operators

$$\left[\hat{J}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{J}_k \equiv i\hbar\sum_k \varepsilon_{ijk}\hat{J}_k \tag{7.1}$$

where

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } ijk = \text{even permutation of } 123 \\ -1 & \text{if } ijk = \text{odd permutation of } 123 \\ 0 & \text{if any two indices are identical} \end{cases} \tag{7.2}$$

and the *Einstein summation convention* over *repeated indices* is understood if the summation sign is left out (unless an explicit override is given).

In addition, these are all Hermitian operators, i.e., $(\hat{J}_i)^\dagger = \hat{J}_i^\dagger = \hat{J}_i$.

Since these three operators form a *closed commutator algebra*, we can solve for the eigenvectors and eigenvalues using only the commutators.

The three operators $\hat{J}_1$, $\hat{J}_2$ and $\hat{J}_3$ do not commute with each other and hence do not share a common set of eigenvectors (called a *representation*).

However, there exists another operator that commutes with each of the angular momentum components separately. If we define

$$\hat{J}^2 = \sum_{i=1}^{3} \hat{J}_i^2 \tag{7.3}$$

which is the square of the total angular momentum vector, we have

$$\left[\hat{J}^2, \hat{J}_i\right] = 0 \text{ for } i = 1, 2, 3 \tag{7.4}$$

In addition, we have $(\hat{J}^2)^\dagger = \hat{J}^2$, so that $\hat{J}^2$ is Hermitian also.

The commutation relations and the Hermitian property say that $\hat{J}^2$ and *any one* of the components *share a complete set of common* eigenvectors. By convention, we choose to use $\hat{J}^2$ and $\hat{J}_3$ as the two operators, whose eigenvalues(good quantum numbers) will characterize the set of common eigenvectors.

As we shall see, $\hat{J}^2$ is a so-called *Casimir invariant operator* that characterizes the representations(set of eigenvectors). In particular, the eigenvalue of $\hat{J}^2$ *characterizes the representation* and the eigenvalues of one of the components of the angular momentum (usually $\hat{J}_3$) will *characterize the eigenvectors within a representation.*

We define the eigenvector/eigenvalue relations by the equations

$$\hat{J}^2 |\lambda m\rangle = \lambda \hbar^2 |\lambda m\rangle \tag{7.5}$$

$$\hat{J}_3 |\lambda m\rangle = m\hbar |\lambda m\rangle \tag{7.6}$$

where the appropriate factors of $\hbar$ that have been explicitly put into the relations make m and λ dimensionless numbers.

We now define some other operators and their associated commutators so that we can use them in our derivations.

$$\hat{J}_\pm = \hat{J}_1 \pm i\hat{J}_2 \tag{7.7}$$

$$\hat{J}_- = \left(\hat{J}_+\right)^+ \rightarrow \text{ they are not Hermitian operators} \tag{7.8}$$

We then have

$$\left[\hat{J}^2, \hat{J}_\pm\right] = \left[\hat{J}^2, \hat{J}_1\right] \pm i\left[\hat{J}^2, \hat{J}_2\right] = 0 \tag{7.9}$$

$$\begin{aligned}\left[\hat{J}_3, \hat{J}_\pm\right] &= \left[\hat{J}_3, \hat{J}_1\right] \pm i\left[\hat{J}_3, \hat{J}_2\right] \\ &= i\hbar\hat{J}_2 \mp i\left(i\hbar\hat{J}_1\right) = \pm\hbar\hat{J}_\pm\end{aligned} \tag{7.10}$$

and

$$\begin{aligned}\left[\hat{J}_+, \hat{J}_-\right] &= \left[\hat{J}_1, \hat{J}_1\right] + i\left[\hat{J}_2, \hat{J}_1\right] - i\left[\hat{J}_1, \hat{J}_2\right] - \left[\hat{J}_2, \hat{J}_2\right] \\ &= -2i\left[\hat{J}_1, \hat{J}_2\right] = 2\hbar\hat{J}_3\end{aligned} \tag{7.11}$$

and

$$\begin{aligned}\hat{J}_+\hat{J}_- &= \left(\hat{J}_1 + i\hat{J}_2\right)\left(\hat{J}_1 - i\hat{J}_2\right) \\ &= \hat{J}_1^2 + \hat{J}_2^2 - i\left[\hat{J}_1, \hat{J}_2\right] = \hat{J}^2 - \hat{J}_3^2 + \hbar\hat{J}_3\end{aligned} \tag{7.12}$$

and

$$\hat{J}_-\hat{J}_+ = \hat{J}^2 - \hat{J}_3^2 - \hbar\hat{J}_3 \tag{7.13}$$

Finally, we have

$$\hat{J}^2 = \frac{\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+}{2} + \hat{J}_3^2 \tag{7.14}$$

7.1.1 Derivation of Eigenvalues

Now the definitions (7.5) and (7.6) tell us that

$$\langle \lambda m | \hat{J}^2 | \lambda m \rangle = \lambda \hbar^2 \langle \lambda m \mid \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i^2 | \lambda m \rangle \tag{7.15}$$

$$\lambda \hbar^2 \langle \lambda m \mid \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i \hat{J}_i | \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i^+ \hat{J}_i | \lambda m \rangle \tag{7.16}$$

Let us define the new vector $|\alpha_i\rangle = \hat{J}_i |\lambda m\rangle$. Remember that the norm of any vector is non-negative, i.e., $\langle a \,|\, a \rangle \geq 0$. Therefore

$$\langle \lambda m \mid \lambda m \rangle \geq 0 \text{ and } \langle \alpha_i \mid \alpha_i \rangle \geq 0 \tag{7.17}$$

Now since $\langle \alpha_i | = \langle \lambda m | \hat{J}_i^\dagger$ we have

$$\lambda \hbar^2 \langle \lambda m \mid \lambda m \rangle = \sum_i \langle \lambda m | \hat{J}_i^+ \hat{J}_i | \lambda m \rangle = \sum_i \langle \alpha_i \mid \alpha_i \rangle \geq 0 \tag{7.18}$$

or we have

$$\lambda \geq 0 \text{ or the eigenvalues of } \hat{J}^2 \text{ are greater than } 0 \tag{7.19}$$

In fact, we can even say more than this using these equations. We have

$$\begin{aligned} \lambda \hbar^2 \langle \lambda m \mid \lambda m \rangle &= \sum_i \langle \alpha_i \mid \alpha_i \rangle \\ &= \langle \alpha_1 \mid \alpha_1 \rangle + \langle \alpha_2 \mid \alpha_2 \rangle + \langle \alpha_3 \mid \alpha_3 \rangle \\ &= \langle \alpha_1 \mid \alpha_1 \rangle + \langle \alpha_2 \mid \alpha_2 \rangle + \langle \lambda m | \hat{J}_3^2 | \lambda m \rangle \\ &= \langle \alpha_1 \mid \alpha_1 \rangle + \langle \alpha_2 \mid \alpha_2 \rangle + m^2 \hbar^2 \langle \lambda m \mid \lambda m \rangle \geq 0 \end{aligned}$$

which says that

$$\lambda \hbar^2 \langle \lambda m \mid \lambda m \rangle \geq m^2 \hbar^2 \langle \lambda m \mid \lambda m \rangle \Rightarrow \lambda \geq m^2 \tag{7.20}$$

This says that for a *fixed value* of λ (the eigenvalue of $\hat{J}^2$), which characterizes the representation, there must be *maximum and minimum* values of m (the eigenvalue of $\hat{J}_3$), which characterizes the eigenvectors within a representation.

Now we have

$$\begin{aligned} \hat{J}_3 \hat{J}_+ |\lambda m\rangle &= \hat{J}_3 (\hat{J}_+ |\lambda m\rangle) \\ &= (\hat{J}_+ \hat{J}_3 + \left[\hat{J}_3, \hat{J}_+\right]) |\lambda m\rangle = (\hat{J}_+ \hat{J}_3 + \hbar \hat{J}_+) |\lambda m\rangle \\ &= \hbar(m+1) \hat{J}_+ |\lambda m\rangle = \hbar(m+1)(\hat{J}_+ |\lambda m\rangle) \end{aligned} \tag{7.21}$$

which says that $\hat{J}_+ |\lambda m\rangle$ is an eigenvector of $\hat{J}_3$ with the *raised* eigenvalue $\hbar(m+1)$, i.e., $\hat{J}_+ |\lambda m\rangle \propto |\lambda, m+1\rangle$ (remember the harmonic oscillator discussion).

Since we already showed that for fixed λ, there must be a maximum value of m, say m_{max}, then it must be the case that for that particular m-value we have

$$\hat{J}_+ |\lambda m_{\max}\rangle = 0 \tag{7.22}$$

If this were not true, then we would have

$$\hat{J}_+ |\lambda m_{\max}\rangle \propto |\lambda, m_{\max}+1\rangle \tag{7.23}$$

but this violates the statement that m_{max} was was the maximum m-value.

Using this result we find

$$\begin{aligned}
&\hat{J}_-\hat{J}_+ |\lambda m_{\max}\rangle = 0 = \left(\hat{J}^2 - \hat{J}_3^2 - \hbar\hat{J}_3\right)|\lambda m_{\max}\rangle \\
&\hbar^2\left(\lambda - m_{\max}^2 - m_{\max}\right)|\lambda m_{\max}\rangle = 0 \\
&\lambda - m_{\max}^2 - m_{\max} = 0 \rightarrow \lambda = m_{\max}^2 + m_{\max} \\
&\lambda = m_{\max}(m_{\max}+1)
\end{aligned} \tag{7.24}$$

It is convention to define

$$m_{\max} = j \text{ and hence } \lambda = j(j+1) \tag{7.25}$$

In the same way we can show

$$\begin{aligned}
\hat{J}_3\hat{J}_- |\lambda m\rangle &= \hat{J}_3(\hat{J}_- |\lambda m\rangle) \\
&= (\hat{J}_-\hat{J}_3 + \left[\hat{J}_3, \hat{J}_-\right])|\lambda m\rangle = (\hat{J}_-\hat{J}_3 - \hbar\hat{J}_+)|\lambda m\rangle \\
&= \hbar(m-1)\hat{J}_- |\lambda m\rangle = \hbar(m-1)(\hat{J}_- |\lambda m\rangle)
\end{aligned} \tag{7.26}$$

which says that $\hat{J}_- |\lambda m\rangle$ is an eigenvector of $\hat{J}_3$ with the *lowered* eigenvalue $\hbar(m-1)$, i.e., $\hat{J}_- |\lambda m\rangle \propto |\lambda, m-1\rangle$.

If we let the minimum value of m be m_{min}, then as before we must have

$$\hat{J}_- |\lambda m_{\min}\rangle = 0 \tag{7.27}$$

or m_{min} is not the minimum value of m. This says that

$$\begin{aligned}
&\hat{J}_+\hat{J}_- |\lambda m_{\min}\rangle = 0 = \left(\hat{J}^2 - \hat{J}_3^2 + \hbar\hat{J}_3\right)|\lambda m_{\min}\rangle \\
&\hbar^2\left(\lambda - m_{\min}^2 + m_{\min}\right)|\lambda m_{\min}\rangle = 0 \\
&\lambda - m_{\min}^2 + m_{\min} = 0 \rightarrow \lambda = m_{\min}^2 - m_{\min} \\
&\lambda = m_{\min}(m_{\min}-1) = j(j+1)
\end{aligned} \tag{7.28}$$

which says that

$$m_{\min} = -j \tag{7.29}$$

We have thus shown that the pair of operators $\hat{J}^2$ and $\hat{J}_3$ have a common set of eigenvectors $|jm\rangle$ (we now use the labels j and m), where we have found that

$$-j \le m \le j \tag{7.30}$$

and the allowed m-values change by steps of one, i.e., for a given j-value, the allowed m-values are

$$-j, -j+1, -j+2, \ldots\ldots, j-2, j-1, j \tag{7.31}$$

which implies that

$$2j = \text{ integer} \tag{7.32}$$

or

$$j = \frac{\text{integer}}{2} \geq 0 \quad \text{the allowed values} \tag{7.33}$$

Thus, we have the *allowed sets* or *representations of the angular momentum commutation relations* given by

$$\begin{aligned} j &= 0 \quad , \quad m = 0 \\ j &= \frac{1}{2} \quad , \quad m = \frac{1}{2}, -\frac{1}{2} \\ j &= 1 \quad , \quad m = 1, 0, -1 \\ j &= \frac{3}{2} \quad , \quad m = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2} \end{aligned} \tag{7.34}$$

$$\tag{7.35}$$

and so on.

For each value of j, there are $2j + 1$ allowed m-values in the eigenvalue spectrum(representation) and

$$\hat{J}^2 |jm\rangle = \hbar^2 j(j+1) |jm\rangle \tag{7.36}$$

$$\hat{J}_3 |jm\rangle = m\hbar |jm\rangle \tag{7.37}$$

Before proceeding, we need a few more relations. We found earlier that

$$\hat{J}_+ |jm\rangle = C_+ |j, m+1\rangle = |\alpha_+\rangle \tag{7.38}$$

$$\hat{J}_- |jm\rangle = C_- |j, m-1\rangle = |\alpha_-\rangle \tag{7.39}$$

and from these we have

$$\langle\alpha_+| = C_+^* \langle j, m+1| \tag{7.40}$$

$$\langle\alpha_-| = C_-^* \langle j, m-1| \tag{7.41}$$

We can then say that

$$\begin{aligned} \langle\alpha_+ \mid \alpha_+\rangle &= |C_+|^2 \langle j, m+1 \mid j, m+1\rangle = |C_+|^2 \\ &= (\langle jm| (\hat{J}_+)^+)(\hat{J}_+ |jm\rangle) \\ &= \langle jm| \hat{J}_- \hat{J}_+ |jm\rangle \\ &= \langle jm| \left(\hat{J}^2 - \hat{J}_3^2 - \hbar\hat{J}_3\right) |jm\rangle \\ &= \langle jm| \hbar^2 (j(j+1) - m^2 - m) |jm\rangle \\ &= \hbar^2 (j(j+1) - m^2 - m) \langle jm \mid jm\rangle \\ &= \hbar^2 (j(j+1) - m^2 - m) \end{aligned} \tag{7.42}$$

or

$$C_+ = \hbar\sqrt{j(j+1) - m(m+1)} \tag{7.43}$$

and similarly

$$\begin{aligned}
\langle \alpha_- \mid \alpha_- \rangle &= |C_-|^2 \langle j, m-1 \mid j, m-1 \rangle = |C_-|^2 \\
&= (\langle jm| (\hat{J}_-)^+)(\hat{J}_- |jm\rangle) \\
&= \langle jm| \hat{J}_+ \hat{J}_- |jm\rangle \\
&= \langle jm| \left(\hat{J}^2 - \hat{J}_3^2 + \hbar\hat{J}_3\right) |jm\rangle \\
&= \langle jm| \hbar^2(j(j+1) - m^2 + m) |jm\rangle \\
&= \hbar^2(j(j+1) - m^2 + m) \langle jm \mid jm\rangle \\
&= \hbar^2(j(j+1) - m^2 + m)
\end{aligned} \tag{7.44}$$

or

$$C_- = \hbar\sqrt{j(j+1) - m(m-1)} \tag{7.45}$$

Therefore, we have the very important relations for the raising/lowering or ladder operators

$$\begin{aligned}
\hat{J}_\pm |jm\rangle &= \hbar\sqrt{j(j+1) - m(m\pm 1)}\, |j, m\pm 1\rangle \\
&= \hbar\sqrt{(j \pm m + 1)(j \mp m)}\, |j, m\pm 1\rangle
\end{aligned} \tag{7.46}$$

7.2 Transformations and Generators; Spherical Harmonics

There are many vector operators and vector component operators in the following discussions. To avoid confusing notation, we will adopt the following conventions:

$\vec{A}_{op}$ = a vector operator

$\hat{\mathrm{A}}_j$ = a vector component operator

$\hat{B}$ = any non - vector operator or it might be a unit vector (context will decide)

$\vec{q}$ = an ordinary vector

As we showed earlier, the angular momentum operators are the generators of rotations. The unitary transformation operator for a rotation through an angle θ about an axis along the direction specified by the unit vector $\hat{n}$ is given by

$$\hat{U}_{\hat{n}}(\theta) = e^{-\frac{i}{\hbar}\theta\hat{n}\cdot\vec{J}_{op}} \tag{7.47}$$

where $\vec{J}_{op}$ is the angular momentum operator.

What does the $\hat{J}_3$ operator look like in the position representation?

We will quickly get an idea of the answer and then step back and do things in more detail.

Suppose that we have a position representation wave function $\psi(\vec{r})$. For a rotation about the 3-axis, we showed earlier that for an infinitesimal angle ϵ

$$\hat{U}_3(\varepsilon)\psi(x_1,x_2,x_3) = \psi(x_1\cos\varepsilon + x_2\sin\varepsilon, -x_1\sin\varepsilon + x_2\cos\varepsilon, x_3) \tag{7.48}$$

In general, for infinitesimal shifts(a in this case) in the coordinates, we have, to lowest order in the infinitesimals, for a function of one variable

$$f(x+a) = f(x) + a\frac{\partial f}{\partial x} \tag{7.49}$$

Extending this to two variables, we have, to first order in infinitesimals a and b,

$$f(x+a,y+b) = f(x,y) + a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y} \tag{7.50}$$

Therefore, for an infinitesimal angle of rotation ϵ,

$$\begin{aligned}\hat{U}_3(\varepsilon)\psi(x_1,x_2,x_3) &= \psi(x_1\cos\varepsilon + x_2\sin\varepsilon, -x_1\sin\varepsilon + x_2\cos\varepsilon, x_3)\\ &= \psi(x_1,x_2,x_3) + \varepsilon x_2\frac{\partial\psi}{\partial x_1} - \varepsilon x_1\frac{\partial\psi}{\partial x_2}\\ &= \left(1 + \varepsilon\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\right)\right)\psi(x_1,x_2,x_3)\end{aligned} \tag{7.51}$$

But we also have (to first order)

$$\hat{U}_3(\varepsilon)\psi(x_1,x_2,x_3) = (1 - \frac{i}{\hbar}\varepsilon\hat{J}_3)\psi(x_1,x_2,x_3) \tag{7.52}$$

Putting these two equations together we have

$$\begin{aligned}\hat{J}_3 &= -i\hbar\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\right) = (\vec{r}_{op}\times(-i\hbar\nabla))_3\\ &= (\vec{r}_{op}\times\vec{p}_{op})_3 = \left(\vec{L}_{op}\right)_3 = \hat{L}_3\end{aligned} \tag{7.53}$$

where

$$\begin{aligned}&\vec{L}_{op} = \text{orbital angular momentum operator}\\ &\vec{r}_{op} = (\hat{x},\hat{y},\hat{z}) = (\hat{x}_1,\hat{x}_2,\hat{x}_3)\\ &\vec{p}_{op} = (\hat{p}_x,\hat{p}_y,\hat{p}_z) = (\hat{p}_1,\hat{p}_2,\hat{p}_3)\end{aligned}$$

Since $\vec{L}_{op}$ is an angular momentum, it must have the commutation relations

$$\left[\hat{L}_i,\hat{L}_j\right] = i\hbar\varepsilon_{ijk}\hat{L}_k \tag{7.54}$$

where as we indicated above

$$\hat{L}_1 = \hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2 \tag{7.55}$$

$$\hat{L}_2 = \hat{x}_3\hat{p}_1 - \hat{x}_1\hat{p}_3 \tag{7.56}$$

$$\hat{L}_3 = \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1 \tag{7.57}$$

and

$$\vec{L}^{+}_{op} = \vec{L}_{op} \tag{7.58}$$

Using

$$[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij} \tag{7.59}$$

we get

$$\left[\hat{L}_i, x_j\right] = i\hbar\varepsilon_{ijk}x_k \tag{7.60}$$

and for $\hat{n}$ = a unit vector

$$\sum_i \hat{n}_i \left[\hat{L}_i, x_j\right] = i\hbar \sum_i \hat{n}_i \varepsilon_{ijk} x_k \tag{7.61}$$

$$\left[\hat{n}\cdot\vec{L}_{op}, x_j\right] = i\hbar\left(\vec{r}_{op}\times\hat{n}\right)_j \tag{7.62}$$

$$\begin{aligned}\left[\hat{n}\cdot\vec{L}_{op}, \vec{r}_{op}\right] &= \sum_j \left[\hat{n}\cdot\vec{L}_{op}, x_j\right]\hat{e}_j \\ &= i\hbar\sum_j \left(\vec{r}_{op}\times\hat{n}\right)_j \hat{e}_j = i\hbar\left(\vec{r}_{op}\times\hat{n}\right)\end{aligned} \tag{7.63}$$

where we have used

$$\vec{A}\times\vec{B} = \sum_{ijk}\varepsilon_{ijk}A_jB_k\hat{e}_i \tag{7.64}$$

Similarly, we get

$$\left[\hat{n}\cdot\vec{L}_{op}, \vec{p}_{op}\right] = i\hbar\left(\vec{p}_{op}\times\hat{n}\right) \tag{7.65}$$

Now let us step back and consider rotations in 3-dimensional space and try to get a better physical understanding of what is happening.

Consider the operator

$$\vec{r}\,'_{op} = \vec{r}_{op} + \vec{\alpha}\times\vec{r}_{op} \tag{7.66}$$

where $\vec{\alpha}$ = ordinary *infinitesimal* vector.

Now let $|\vec{r}_0\rangle$ = an eigenstate of $\vec{r}_{op}$ with eigenvalue $\vec{r}_0$, i.e.,

$$\vec{r}_{op}\,|\vec{r}_0\rangle = \vec{r}_0\,|\vec{r}_0\rangle \tag{7.67}$$

We then have

$$\vec{r}\,'_{op}\,|\vec{r}_0\rangle = \vec{r}\,'_0\,|\vec{r}_0\rangle = \left(\vec{r}_0 + \vec{\alpha}\times\vec{r}_0\right)|\vec{r}_0\rangle \tag{7.68}$$

which says that $|\vec{r}_0\rangle$ is also an eigenstate of $\vec{r}\,'_{op}$ (different eigenvalue).

Now, for simplicity, let $\vec{\alpha} = \alpha\hat{e}_3$, which says that

$$\vec{\alpha} \times \vec{r}_0 = \alpha\hat{e}_3 \times (x_{01}\hat{e}_1 + x_{02}\hat{e}_2 + x_{03}\hat{e}_3) = \alpha x_{01}\hat{e}_2 - \alpha x_{02}\hat{e}_1 \tag{7.69}$$

$$\vec{r}_0 + \vec{\alpha} \times \vec{r}_0 = (x_{01} - \alpha x_{02})\hat{e}_1 + (x_{02} + \alpha x_{01})\hat{e}_2 + x_{03}\hat{e}_3 \tag{7.70}$$

This last expression is the vector we get if we rotate $\vec{r}_0$ about $\hat{\alpha} = \vec{\alpha}/|\vec{\alpha}|$ ($\hat{e}_3$ in this case) by an infinitesimal angle $\hat{\alpha} = |\vec{\alpha}|$. This result generalizes for $\vec{\alpha}$ in any direction.

Since the eigenvalues of $\vec{r}\,'_{op}$ are those of $\vec{r}_{op}$ rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$, we conclude that $\vec{r}\,'_{op}$ is the position operator rotated by rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$.

Alternatively, $\vec{r}\,'_{op}$ is the position operator in a coordinate frame rotated by $|\vec{\alpha}|$ about $\vec{\alpha}$ (remember our earlier discussions about the active/passive views).

To connect this to generators, unitary transformations, and angular momentum, we proceed as follows. We can rewrite $\vec{r}\,'_{op}$ as

$$\vec{r}\,'_{op} = \vec{r}_{op} + \frac{i}{\hbar}\left[\vec{\alpha} \cdot \vec{L}_{op}, \vec{r}_{op}\right] \tag{7.71}$$

which is equivalent (to first order in α) to

$$\vec{r}\,'_{op} = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \tag{7.72}$$

i.e.,

$$\begin{aligned}\vec{r}\,'_{op} &= e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \\ &= (\hat{I} + \frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op})\vec{r}_{op}(\hat{I} - \frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}) + O(\alpha^2) \\ &= \vec{r}_{op} + \frac{i}{\hbar}\left[\vec{\alpha}\cdot\vec{L}_{op}, \vec{r}_{op}\right] + O(\alpha^2)\end{aligned} \tag{7.73}$$

Earlier, however, we showed that, if the state vector transforms as

$$|\psi'\rangle = \hat{U}\,|\psi\rangle \tag{7.74}$$

then the operators transform as

$$\hat{U}^{-1}\hat{O}\hat{U} \tag{7.75}$$

which then implies that the rotation operator is

$$\hat{U}(\vec{\alpha}) = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \tag{7.76}$$

as we expect.

This result derived for infinitesimal rotation angles holds for finite rotation angles also.

Let us look at the effect on state vectors

$$\vec{r}\,'_{op}\,|\vec{r}_0\rangle = \vec{r}\,'_0\,|\vec{r}_0\rangle = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle \tag{7.77}$$

$$\vec{r}_{op}e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle = \vec{r}\,'_0e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle \tag{7.78}$$

or

$$e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle \tag{7.79}$$

is an eigenstate of $\vec{r}_{op}$ with eigenvalue $\vec{r}\,'_0$ or

$$|\vec{r}\,'_0\rangle = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\vec{r}_0\rangle \text{ and } \langle\vec{r}\,'_0| = \langle\vec{r}_0|\,e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} \tag{7.80}$$

and thus $\vec{L}_{op}$ is the *generator of ordinary rotations in 3-dimensional space.*

For wave functions we have $\psi(\vec{r}_0) = \langle\vec{r}_0\,|\,\psi\rangle$. Using this result, the wave function transforms as

$$\begin{aligned}\psi(\vec{r}\,'_0) &= \langle\vec{r}\,'_0\mid\psi\rangle = \langle\vec{r}_0|\,e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}}\,|\psi\rangle \\ &= \langle\vec{r}_0\mid\psi\,'\rangle = \psi\,'(\vec{r}_0) = \text{ wave function at rotated point } \vec{r}\,'_0\end{aligned} \tag{7.81}$$

Now, we have seen that

$$\vec{L}_{op} = \vec{r}_{op}\times\vec{p}_{op} = -i\hbar\vec{r}_{op}\times\nabla \tag{7.82}$$

This can easily be evaluated in different coordinate systems.

Cartesian Coordinates

$$\vec{L}_{op} = -i\hbar\sum_{ijk}\varepsilon_{ijk}x_j\frac{\partial}{\partial x_k}\hat{e}_i \tag{7.83}$$

$$\hat{L}_1 = -i\hbar\left(x_2\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_2}\right) \tag{7.84}$$

$$\hat{L}_2 = -i\hbar\left(x_3\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_3}\right) \quad , \quad \hat{L}_3 = -i\hbar\left(x_1\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_1}\right) \tag{7.85}$$

as we saw at the beginning of this discussion.

Spherical-Polar Coordinates

We have

$$\vec{r} = r\hat{e}_r \text{ and } \nabla = \hat{e}_r\frac{\partial}{\partial r} + \hat{e}_\theta\frac{1}{r}\frac{\partial}{\partial\theta} + \hat{e}_\varphi\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi} \tag{7.86}$$

where

$$\hat{e}_r = \sin\theta\cos\varphi\hat{e}_1 + \sin\theta\sin\varphi\hat{e}_2 + \cos\theta\hat{e}_3 \tag{7.87}$$

$$\hat{e}_\theta = \cos\theta\cos\varphi\hat{e}_1 + \cos\theta\sin\varphi\hat{e}_2 + \sin s\theta\hat{e}_3 \tag{7.88}$$

$$\hat{e}_\varphi = -\sin\theta\hat{e}_1 + \cos\theta\hat{e}_2 \tag{7.89}$$

which gives

$$\vec{L}_{op} = -i\hbar\left[\hat{e}_\varphi \frac{\partial}{\partial\theta} - \hat{e}_\theta \frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right] \tag{7.90}$$

and

$$\hat{L}_3 = \hat{L}_z = \hat{e}_z \cdot \vec{L}_{op} = -i\hbar\frac{\partial}{\partial\varphi} \tag{7.91}$$

$$\vec{L}^2_{op} = \vec{L}_{op} \cdot \vec{L}_{op} = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right] \tag{7.92}$$

Similarly, we have

$$\hat{L}_1 = \hat{L}_x = i\hbar\left[\sin\varphi\frac{\partial}{\partial\theta} + \cot\theta\cos\varphi\frac{\partial}{\partial\varphi}\right] \tag{7.93}$$

$$\hat{L}_2 = \hat{L}_y = -i\hbar\left[\cos\varphi\frac{\partial}{\partial\theta} - \cot\theta\sin\varphi\frac{\partial}{\partial\varphi}\right] \tag{7.94}$$

and

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}r - \frac{\vec{L}^2_{op}}{\hbar^2 r^2} \tag{7.95}$$

7.2.1 Eigenfunctions; Eigenvalues; Position Representation

Since

$$\left[\hat{L}_i, \hat{L}_j\right] = i\hbar\varepsilon_{ijk}\hat{L}_k \text{and} \left[\hat{L}^2_{op}, \hat{L}_j\right] = 0 \tag{7.96}$$

the derivation of the eigenvalues and eigenvectors follows our earlier work, i.e., the equations

$$\hat{L}^2_{op}|\ell m\rangle = \hbar^2\ell(\ell+1)|\ell m\rangle \text{ and } \hat{L}_3|\ell m\rangle = \hbar m|\ell m\rangle \tag{7.97}$$

$$\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y \tag{7.98}$$

imply

$$\ell = \frac{\text{integer}}{2} \geq 0 \tag{7.99}$$

and for a given value of ℓ, m takes on the $2\ell+1$ values

$$m = -\ell, -\ell+1, -\ell+2, ..., \ell-2, \ell-1, \ell \tag{7.100}$$

If we move back into 3-dimensional space and define

$$Y_{\ell m}(\theta,\varphi) = \langle\theta\varphi \mid \ell m\rangle = \text{ spherical harmonic} \tag{7.101}$$

then we have the defining equations for the $Y_{\ell m}(\theta,\varphi)$ given by

$$\begin{aligned}\langle\theta\varphi|\vec{L}^2_{op}|\ell m\rangle &= \vec{L}^2_{op}\langle\theta\varphi \mid \ell m\rangle = \vec{L}^2_{op}Y_{\ell m}(\theta,\varphi)\\ &= \hbar^2\ell(\ell+1)\langle\theta\varphi \mid \ell m\rangle = \hbar^2\ell(\ell+1)Y_{\ell m}(\theta,\varphi)\end{aligned} \tag{7.102}$$

$$\begin{aligned}\langle\theta\varphi|\hat{L}_3|\ell m\rangle &= \hat{L}_3\langle\theta\varphi \mid \ell m\rangle = \hat{L}_3 Y_{\ell m}(\theta,\varphi)\\ &= \hbar m\langle\theta\varphi \mid \ell m\rangle = \hbar m Y_{\ell m}(\theta,\varphi)\end{aligned} \tag{7.103}$$

Before determining the functional form of the $Y_{\ell m}{}'s$, we must step back and see if there are any restrictions that need to be imposed on the possible eigenvalues ℓ and m due to the fact that we are in a real 3-dimensional space.

In general, eigenvalue restrictions come about only from the imposition of physical *boundary conditions*. A standard boundary condition that is usually imposed is the following:

In real 3-dimensional space, if we rotate the system (or the axes by 2π, then we get back the *same* world. This means that the $Y_{\ell m}(\theta, \varphi)$ should be *single-valued* under such rotations or that

$$Y_{\ell m}(\theta, \varphi + 2\pi) = Y_{\ell m}(\theta, \varphi) \tag{7.104}$$

Now from the general rules we have developed, this gives

$$\langle \theta, \varphi + 2\pi \mid \ell m \rangle = \langle \theta\varphi | \, e^{2\pi i \hat{L}_3/\hbar} \, | \ell m \rangle = e^{2\pi m i} \langle \theta, \varphi \mid \ell m \rangle \tag{7.105}$$

or that single-valuedness of the wave function requires that

$$\langle \theta, \varphi + 2\pi \mid \ell m \rangle = \langle \theta\varphi | \, e^{2\pi i \hat{L}_3/\hbar} \, | \ell m \rangle = e^{2\pi m i} \langle \theta, \varphi \mid \ell m \rangle \tag{7.106}$$

This says, that for orbital angular momentum in real 3-dimensional space, no 1/2-integer values are allowed for m and hence that ℓ must be an integer. The allowed sets are:

$$\begin{aligned} \ell &= 0 \quad , \quad m = 0 \\ \ell &= 1 \quad , \quad m = -1, 0, 1 \\ \ell &= 2 \quad , \quad m = -2, -1, 0, 1, 2 \end{aligned}$$

and so on.

It is important to note that we are imposing a much stronger condition than is necessary. In general, as we have stated several times, it is not the state vectors, operators or wave functions that have any physical meaning in quantum mechanics. The only quantities that have physical meaning are those directly related to measurable quantities, namely, the probabilities, and the expectation values. This single-valued condition on m is not needed for the single-valuedness of the expectation values, since the extra phase factors involving m will cancel out during the calculation of the expectation value. Experiment, however, says that ℓ is an integer only, so it seems that the strong condition is valid. We *cannot prove* that this is so, however.

Let us now figure out the $Y_{\ell m}(\theta, \varphi)$. We have

$$\hat{L}_3 Y_{\ell m}(\theta, \varphi) = \frac{\hbar}{i} \frac{\partial}{\partial \varphi} Y_{\ell m}(\theta, \varphi) = \hbar m Y_{\ell m}(\theta, \varphi) \tag{7.107}$$

which tells us the φ-dependence.

$$Y_{\ell m}(\theta,\varphi) = e^{im\varphi} P_{\ell m}(\theta) \tag{7.108}$$

Now we also must have, since $\ell =$ maximum value of m

$$\langle\theta\varphi|\,\hat{L}_+\,|\ell\ell\rangle = 0 = \hat{L}_+\,\langle\theta\varphi\mid\ell\ell\rangle = \hat{L}_+ Y_{\ell\ell}(\theta,\varphi) \tag{7.109}$$

Using the expressions for $\hat{L}_x, \hat{L}_y, \hat{L}_+$ and $\hat{L}_-$ we have

$$\frac{\hbar}{i} e^{i\varphi}\left[i\frac{\partial}{\partial\theta} - \cot\theta\frac{\partial}{\partial\varphi}\right] Y_{\ell\ell}(\theta,\varphi) = 0 \tag{7.110}$$

Using

$$\frac{\hbar}{i}\frac{\partial}{\partial\varphi} Y_{\ell\ell}(\theta,\varphi) = \ell\hbar Y_{\ell\ell}(\theta,\varphi) \tag{7.111}$$

we get

$$\left[\frac{\partial}{\partial\theta} - \ell\cot\theta\right] P_{\ell\ell}(\theta) = 0 \tag{7.112}$$

or

$$P_{\ell\ell}(\theta) = (\sin\theta)^\ell \tag{7.113}$$

Therefore, the final expression is

$$Y_{\ell\ell}(\theta,\varphi) = A_{\ell m} e^{i\ell\varphi}(\sin\theta)^\ell \tag{7.114}$$

Now that we have generated the topmost spherical harmonic. We can generate all the others for a given ℓ using the lowering operators, i.e.,

$$\hat{L}_- Y_{\ell m}(\theta,\varphi) = \hbar\sqrt{\ell(\ell+1) - m(m-1)}\, Y_{\ell m-1}(\theta,\varphi) \tag{7.115}$$

where

$$\hat{L}_- = \frac{\hbar}{i} e^{-i\varphi}\left[i\frac{\partial}{\partial\theta} - \cot\theta\frac{\partial}{\partial\varphi}\right] \tag{7.116}$$

In general, we choose the $A_{\ell m}$ so that the $Y_{\ell m}(\theta,\varphi)$ are normalized, i.e.,

$$1 = \int d\Omega\,|Y_{\ell m}(\theta,\varphi)|^2 = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta\,|Y_{\ell m}(\theta,\varphi)|^2 \tag{7.117}$$

Since the $Y_{\ell m}(\theta,\varphi)$ are eigenfunctions of a Hermitian operators, they form a compete set which we can always make orthogonal so that we also assume that we have

$$\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta Y^*_{\ell' m'}(\theta,\varphi) Y_{\ell m}(\theta,\varphi) = \delta_{\ell'\ell}\delta_{m' m} \tag{7.118}$$

The algebra is complicated. The general result is

$$Y_{\ell m}(\theta,\varphi) - \frac{(-1)\ell}{2^\ell \ell!}\sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell+m)!}{(\ell-m)!}}\frac{e^{im\varphi}}{(\sin\theta)^m}\left(\frac{d}{d\cos\theta}\right)^{\ell-m}(\sin\theta)^{2\ell} \tag{7.119}$$

Some *examples* are:

$$Y_{00} = \frac{1}{\sqrt{4\pi}} \quad , \quad Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta \tag{7.120}$$

$$Y_{1,\pm1} = \mp\sqrt{\frac{3}{8\pi}}e^{\pm i\varphi}\sin\theta \quad , \quad Y_{20} = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1) \tag{7.121}$$

$$Y_{2,\pm1} = \mp\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{\pm i\varphi} \quad , \quad Y_{2,\pm2} = \sqrt{\frac{15}{32\pi}}\sin^2\theta e^{\pm 2i\varphi} \tag{7.122}$$

Some Properties

$$Y_{\ell,-m}(\theta,\varphi) = (-1)^m Y^*_{\ell,m}(\theta,\varphi) \tag{7.123}$$

Under the parity operation

$$\vec{r} \to -\vec{r} \text{ or } r \to r, \theta \to \pi - \theta, \varphi \to \varphi + \pi \tag{7.124}$$

This says that

$$\begin{aligned} e^{im\varphi} &\to e^{im\varphi}e^{im\pi} = (-1)^m e^{im\varphi} \\ \sin\theta &\to \sin(\pi - \theta) \to \sin\theta \\ \cos\theta &\to \cos(\pi - \theta) \to -\cos\theta \end{aligned}$$

which imply that

$$Y_{\ell,m}(\theta,\varphi) \to (-1)^\ell Y_{\ell,m}(\theta,\varphi) \tag{7.125}$$

Therefore,

if ℓ is even, then we have an even parity state

if ℓ is odd, then we have an odd parity state

Since they form a complete set, any function of (θ,φ) can be expanded in terms of the $Y_{\ell m}(\theta,\varphi)$ (the $Y_{\ell m}(\theta,\varphi)$ are a basis), i.e., we can write

$$f(\theta,\varphi) = \sum_{\ell,m} f_{\ell m} Y_{\ell,m}(\theta,\varphi) \tag{7.126}$$

where

$$f_{\ell m} = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta Y^*_{\ell' m'}(\theta,\varphi) f(\theta,\varphi) \tag{7.127}$$

7.3 Spin

As we discussed earlier, a second kind of angular momentum exists in quantum mechanics. It is related to *internal degrees of freedom* of particles and is not related in any way to ordinary 3-dimensional space properties.

We designate this new angular momentum by *spin* and represent it by the

operator $\vec{S}_{op}$, where since it is an angular momentum, its components must satisfy the commutators

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k \tag{7.128}$$

where we have used the Einstein summation convention over repeated indices.

The analysis for the eigenvalues and eigenvectors follows the same path as for earlier discussions. We have

$$\vec{S}^2_{op}\,|s, m_s\rangle = \hbar^2 s(s+1)\,|s, m_s\rangle \tag{7.129}$$

$$\hat{S}_3\,|s, m_s\rangle = \hbar m_s\,|s, m_s\rangle \tag{7.130}$$

which, together with the commutators, gives the following results.

For a given value of s, we have $2s+1$ m_s-values

$$m_s = -s, -s+1, -s+2, ..., s-2, s-1, s \tag{7.131}$$

where

$$s = \frac{\text{integer}}{2} \geq 0 \tag{7.132}$$

There are no boundary conditions restricting the value of s, so we can have both integer and half-integer values.

We now turn our attention to a most important special case and then generalize the details.

7.3.1 Spin 1/2

We define a new operator $\vec{\sigma}_{op}$ such that

$$\vec{S}_{op} = \frac{1}{2}\hbar\vec{\sigma}_{op} \tag{7.133}$$

where $\vec{\sigma}_{op} = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ are called the *Pauli spin operators.*

It is experimentally observed that if one measures the component of this spin angular momentum along any direction, one always obtains either $\pm\hbar/2$.

If we designate the state with spin = $+\hbar/2$ or *spin up* in the $\hat{n}$ direction by the ket vectors $|\hat{n}\uparrow\rangle$ or $|\hat{n}+\rangle$, we then have

$$\vec{S}_{op}\cdot\hat{n}\,|\hat{n}\uparrow\rangle = \frac{\hbar}{2}\,|\hat{n}\uparrow\rangle \text{ and } \vec{S}_{op}\cdot\hat{n}\,|\hat{n}\downarrow\rangle = -\frac{\hbar}{2}\,|\hat{n}\downarrow\rangle \tag{7.134}$$

Any pair of eigenvectors $|\hat{n}\uparrow\rangle$ or $|\hat{n}\downarrow\rangle$ for a given direction $\hat{n}$ form a basis for the vector space associated with spin = 1/2 and we have

$\langle\hat{n}\uparrow|\,\psi\rangle =$ amplitude for finding spin "up" along $\hat{n}$ if we are in the state $|\psi\rangle$

$\langle\hat{n}\downarrow|\,\psi\rangle =$ amplitude for finding spin "down" along $\hat{n}$ if we are in the state $|\psi\rangle$

These two amplitudes exhaust all possible measurements for spin = 1/2 in the $\hat{n}$ direction and therefore completely specify the state $|\psi\rangle$. *That is what we mean physically when we say they form a basis set or representation.*

When we build the standard basis from these amplitudes, we choose it to be an eigenstate of the $\hat{S}_3$ or $\hat{S}_z$ operator, i.e., if we write

$$|\psi\rangle = \begin{pmatrix} \langle z\uparrow|\psi\rangle \\ \langle z\downarrow|\psi\rangle \end{pmatrix} = \text{ a 2 - component vector} \tag{7.135}$$

then the appropriate basis is

$$|z\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \text{ spin up in z - direction}$$
$$|z\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \text{spin down in z - direction}$$

Matrix Representations

Using this basis, the matrix representation of

$$\hat{S}_z = \frac{\hbar}{2}\hat{\sigma}_z \tag{7.136}$$

is

$$S_z = \begin{pmatrix} \langle z\uparrow|\hat{S}_z|z\uparrow\rangle & \langle z\uparrow|\hat{S}_z|z\downarrow\rangle \\ \langle z\downarrow|\hat{S}_z|z\uparrow\rangle & \langle z\downarrow|\hat{S}_z|z\downarrow\rangle \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2}\sigma_z$$
$$\rightarrow \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7.137}$$

Now

$$\hat{S}_\pm = \hat{S}_x \pm i\hat{S}_y \tag{7.138}$$
$$\rightarrow \hat{S}_x = \frac{\hat{S}_+ + \hat{S}_-}{2} \text{ and } \hat{S}_y = \frac{\hat{S}_+ - \hat{S}_-}{2i} \tag{7.139}$$

Therefore, using

$$S_x = \begin{pmatrix} \langle z\uparrow|\hat{S}_x|z\uparrow\rangle & \langle z\uparrow|\hat{S}_x|z\downarrow\rangle \\ \langle z\downarrow|\hat{S}_x|z\uparrow\rangle & \langle z\downarrow|\hat{S}_x|z\downarrow\rangle \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_x$$
$$\rightarrow \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7.140}$$

and in a similar way

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{7.141}$$

Properties of the $\hat{\sigma}_i$

From

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k$$

and

$$\vec{S}_{op} = \frac{1}{2}\hbar\vec{\sigma}_{op}$$

we get the commutation relations

$$\left[\hat{\sigma}_i, \hat{\sigma}_j\right] = 2i\varepsilon_{ijk}\hat{\sigma}_k \tag{7.142}$$

In addition, we have

$$\hat{\sigma}_i\hat{\sigma}_j = i\varepsilon_{ijk}\hat{\sigma}_k \quad i \neq j \tag{7.143}$$

$$\hat{\sigma}_i\hat{\sigma}_j + \hat{\sigma}_j\hat{\sigma}_i = \{\hat{\sigma}_i, \hat{\sigma}_j\} = 0 \text{ (called the anticommutator)} \tag{7.144}$$

$$\hat{\sigma}_{\mathrm{i}}^2 = \hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{7.145}$$

The fact that the spin = 1/2 operators anticommute is directly linked to the existence of fermions as one can see when the relativistic equation for the electron is studied.

Put all together, these relations give

$$\hat{\sigma}_i\hat{\sigma}_j = \delta_{ij} + i\varepsilon_{ijk}\hat{\sigma}_k \tag{7.146}$$

or going back to the $\hat{S}_i$

$$\hat{S}_i\hat{S}_j = \frac{\hbar^2}{4}\delta_{ij} + i\varepsilon_{ijk}\frac{\hbar}{2}\hat{S}_k \tag{7.147}$$

In the special case of spin = 1/2, we have

$$\begin{aligned}\vec{S}_{op}^2 &= \frac{\hbar^2}{4}\left(\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + \hat{\sigma}_z^2\right) = \frac{\hbar^2}{4}\left(\hat{I} + \hat{I} + \hat{I}\right) \\ &= \frac{3\hbar^2}{4}\hat{I} = \hbar^2 s(s+1)\hat{I} \text{ with } s = \frac{1}{2}\end{aligned} \tag{7.148}$$

A very useful property that we will employ many time later on is (using (7.145))

$$\begin{aligned}\left(\vec{a}\cdot\vec{\sigma}_{op}\right)\left(\vec{b}\cdot\vec{\sigma}_{op}\right) &= a_i\hat{\sigma}_i b_j\hat{\sigma}_j \\ &= \delta_{ij}a_ib_j + i\varepsilon_{ijk}a_ib_j\hat{\sigma}_k = a_ib_i + i\varepsilon_{ijk}a_ib_j\hat{\sigma}_k \\ &= \vec{a}\cdot\vec{b} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}_{op}\end{aligned} \tag{7.149}$$

Rotations in Spin Space

We have said that

$$\vec{S}_{op}\cdot\hat{n}\,|\hat{n}\pm\rangle = \pm\frac{\hbar}{2}\,|\hat{n}\pm\rangle \tag{7.150}$$

What do the states $|\hat{n}\pm\rangle$ look like in the $|\hat{z}\pm\rangle$ basis? One way to find out is the *direct approach*.

Let us choose a Cartesian basis for the unit vector (we will look at other choices afterwards)

$$\hat{n} = (n_x, n_y, n_z) \quad , \quad \text{all real} \tag{7.151}$$

We then have

$$\begin{aligned}\vec{S}_{op}\cdot\hat{n} &= \frac{\hbar}{2}\vec{\sigma}_{op}\cdot\hat{n} = \frac{\hbar}{2}(n_x\hat{\sigma}_x + n_y\hat{\sigma}_y + n_z\hat{\sigma}_z)\\ &= \frac{\hbar}{2}\left(n_x\begin{pmatrix}0&1\\1&0\end{pmatrix} + n_y\begin{pmatrix}0&-i\\i&0\end{pmatrix} + n_z\begin{pmatrix}1&0\\0&-1\end{pmatrix}\right)\\ &= \frac{\hbar}{2}\begin{pmatrix}n_z & n_x - in_y\\ n_x + in_y & -n_z\end{pmatrix}\end{aligned} \tag{7.152}$$

and

$$\vec{S}_{op}\cdot\hat{n}\,|\hat{n}+\rangle = +\frac{\hbar}{2}\,|\hat{n}+\rangle \tag{7.153}$$

$$\frac{\hbar}{2}\begin{pmatrix}n_z & n_x - in_y\\ n_x + in_y & -n_z\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix}a\\b\end{pmatrix} \tag{7.154}$$

where we have represented

$$|\hat{n}+\rangle = \begin{pmatrix}a\\b\end{pmatrix} \tag{7.155}$$

This matrix equation gives two homogeneous equations for a and b

$$\begin{aligned}(n_z - 1)a + (n_x - in_y)b &= 0\\ (n_x + in_y)a - (n_z + 1)b &= 0\end{aligned}$$

or

$$\frac{a}{b} = -\frac{n_x - in_y}{n_z - 1} = \frac{n_z + 1}{n_x + in_y} \tag{7.156}$$

The homogeneous equations have a non-trivial solution only if the determinant of the coefficients of a and b equals zero or

$$(n_z + 1)(n_z - 1) + (n_x - in_y)(n_x + in_y) = 0 \tag{7.157}$$

We assume that the vector $|\hat{n}+\rangle$ is normalized to one or

$$|a|^2 + |b|^2 = 1 \tag{7.158}$$

Putting all this together we get

$$a = \frac{1}{\sqrt{2}}\sqrt{1+n_z} \text{ and } b = \frac{1}{\sqrt{2}}\sqrt{1-n_z} \tag{7.159}$$

$$|\hat{n}+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{1+n_z} \\ \sqrt{1-n_z} \end{pmatrix} \tag{7.160}$$

We can easily check this by letting $|\hat{n}+\rangle = |\hat{z}+\rangle$ or $n_z = 1, n_x = n_y = 0$ which gives

$$|\hat{z}+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{7.161}$$

as expected. In a similar manner

$$|\hat{n}-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} -\sqrt{1-n_z} \\ \sqrt{1+n_z} \end{pmatrix} \tag{7.162}$$

Note that the two vectors $|\hat{n}+\rangle$ and $|\hat{n}-\rangle$ are orthonormal as we expect.

This calculation can also be carried out in other coordinate bases. For the spherical-polar basis we get

$$\hat{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \tag{7.163}$$

$$\vec{S}_{op}\cdot\hat{n} = \frac{\hbar}{2}\begin{pmatrix} \cos\theta & e^{-i\varphi}\sin\theta \\ e^{i\varphi}\sin\theta & -\cos\theta \end{pmatrix} \tag{7.164}$$

$$|\hat{n}+\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\varphi}\sin\frac{\theta}{2} \end{pmatrix} \text{ and } |\hat{n}-\rangle = \begin{pmatrix} \sin\frac{\theta}{2} \\ -e^{i\varphi}\cos\frac{\theta}{2} \end{pmatrix} \tag{7.165}$$

What can we say about operators in the spin $= 1/2$ vector space?

Any such operator $\hat{B}$ can be expressed as a linear combination of the four linearly independent matrices $\{\hat{I}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$ (they are, in fact, a basis for all 2×2 matrices)

$$\hat{B} = a_0\hat{I} + \mathrm{a_x}\hat{\sigma}_x + \mathrm{a_y}\hat{\sigma}_y + \mathrm{a_z}\hat{\sigma}_z = a_0\hat{I} + \mathrm{a}\cdot\vec{\sigma}_{\mathrm{op}} = \begin{pmatrix} a_0+\mathrm{a_z} & \mathrm{a_x}-i\mathrm{a_y} \\ \mathrm{a_x}+i\mathrm{a_y} & a_0-\mathrm{a_z} \end{pmatrix} \tag{7.166}$$

In particular, the density operator or the state operator $\hat{W}$ which is the same as any other operator can be written as

$$\hat{W} = \frac{1}{2}(\hat{I} + \vec{a}\cdot\vec{\sigma}_{\mathrm{op}}) \tag{7.167}$$

where the factor of 1/2 has been chosen so that we have the required property

$$Tr\hat{W} = \frac{1}{2}Tr\hat{I} + a_i Tr\sigma_i = 1 \text{ since } Tr\sigma_i = 0 \tag{7.168}$$

Since $\hat{W} = \hat{W}^\dagger$ we must also have all the a_i real. What is the *physical meaning* of the vector $\vec{a}$?

Consider the following

$$\begin{aligned}\langle\hat{\sigma}_x\rangle &= Tr(\hat{W}\hat{\sigma}_x) = \frac{1}{2}Tr((\hat{I} + a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z)\hat{\sigma}_x) \\ &= \frac{1}{2}Tr(\hat{\sigma}_x + a_x\hat{\sigma}_x^2 + a_y\hat{\sigma}_y\hat{\sigma}_x + a_z\hat{\sigma}_z\hat{\sigma}_x) \\ &= \frac{1}{2}Tr(\hat{\sigma}_x + a_x\hat{I} - ia_y\hat{\sigma}_z + ia_z\hat{\sigma}_y) \\ &= \frac{a_x}{2}Tr(\hat{I}) = a_x \end{aligned} \tag{7.169}$$

or, in general,

$$\langle\vec{\sigma}_{op}\rangle = Tr\left(\hat{W}\vec{\sigma}_{op}\right) = \vec{a} = \text{ polarization vector} \tag{7.170}$$

Now the eigenvalues of $\hat{W}$ are equal to

$$\frac{1}{2} + \text{ eigenvalues of } \vec{a}\cdot\vec{\sigma}_{op} \tag{7.171}$$

and from our earlier work the eigenvalues are $\vec{a}\cdot\vec{\sigma}_{op} = \pm 1$. Therefore the eigenvalues of $\hat{W}$ are

$$\frac{1}{2}(1 \pm |\vec{a}|) \tag{7.172}$$

But, as we showed earlier, all eigenvalues of $\hat{W}$ are ≥ 0, which says that polarization vectors have a length $|\vec{a}|$ restricted to $0 \leq |\vec{a}| \leq 1$.

Pure states have $|\vec{a}| = 1$ and this gives eigenvalues 1 and 0 for $\hat{W}$, which corresponds to *maximum* polarization.

Note that for $\vec{a} = a\hat{e}_3 = \hat{e}_3$, we have

$$\hat{W} = \frac{1}{2}(\hat{I}+\hat{\sigma}_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |z+\rangle\langle z+| \tag{7.173}$$

as it should for a pure state.

An unpolarized state has $|\vec{a}| = 0$ and this gives eigenvalues $(1/2, 1/2)$ for $\hat{W}$. This represents an *isotropic* state where $\langle\hat{S}_i\rangle = 0$.

In this case we have

$$\hat{W} = \frac{1}{2}\hat{I} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}|z+\rangle\langle z+| + \frac{1}{2}|z-\rangle\langle z-| \tag{7.174}$$

as we expect for a nonpure state or a mixture.

Let us now connect all of this stuff to rotations in spin space. For an infinitesimal rotation through angle $\alpha = |\vec{\alpha}|$ about an axis along $\hat{\alpha} = \vec{\alpha}/\alpha$, a unit vector $\hat{m}$ becomes

$$\hat{n} = \hat{m} + \vec{\alpha} \times \hat{m} \tag{7.175}$$

Following the same steps as earlier, this implies

$$\vec{S}_{op} \cdot \hat{n} = \vec{S}_{op} \cdot \hat{m} + \vec{S}_{op} \cdot (\vec{\alpha} \times \hat{m}) = \vec{S}_{op} \cdot \hat{m} + \varepsilon_{ijk} \alpha_i \hat{m}_j \hat{S}_k \tag{7.176}$$

But we have

$$\varepsilon_{ijk} \hat{S}_k = \frac{1}{i\hbar} \left[\hat{S}_i, \hat{S}_j \right] \tag{7.177}$$

which implies that

$$\begin{aligned} \vec{S}_{op} \cdot \hat{n} &= \vec{S}_{op} \cdot \hat{m} + \frac{1}{i\hbar} \left[\hat{S}_i, \hat{S}_j \right] \alpha_i \hat{m}_j \\ &= \vec{S}_{op} \cdot \hat{m} + \frac{1}{i\hbar} \left[\vec{S}_{op} \cdot \vec{\alpha}, \vec{S}_{op} \cdot \hat{m} \right] \end{aligned} \tag{7.178}$$

Using the same approximations as we did earlier, we can see that this expression, to first order in α, is equivalent to

$$\vec{S}_{op} \cdot \hat{n} = e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} \vec{S}_{op} \cdot \hat{m} e^{\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} \tag{7.179}$$

This result holds for finite rotation angles also.

Now using this result, we have

$$\begin{aligned} \vec{S}_{op} \cdot \hat{n} \left[e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} |\hat{m}+\rangle \right] &= \left[e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} \vec{S}_{op} \cdot \hat{m} \right] |\hat{m}+\rangle \\ &= \frac{\hbar}{2} \left[e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} |\hat{m}+\rangle \right] \end{aligned} \tag{7.180}$$

This says that

$$e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} |\hat{m}+\rangle = |\hat{n}+\rangle \text{ and similarly } e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} |\hat{m}-\rangle = |\hat{n}-\rangle \tag{7.181}$$

The rotation that takes $\hat{m} \to \hat{n}$ is *not* unique, however. We are free to rotate by an arbitrary amount about $\hat{n}$ after rotating $\hat{m}$ into $\hat{n}$. This freedom corresponds to adding a phase factor.

We say that the unitary operator

$$e^{-\frac{i}{\hbar} \vec{S}_{op} \cdot \vec{\alpha}} \tag{7.182}$$

has the effect of *rotating* the eigenstate of $\vec{S}_{op} \cdot \hat{m}$ into the eigenstate of $\vec{S}_{op} \cdot \hat{n}$. The operator performs *rotations* on the spin degrees of freedom. The equations are analogous to those for real rotations in space generated by $\vec{L}_{op}$.

Let us now work out a very useful identity. We can write

$$e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}} = e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} = \sum_{n=0}^{\infty}\frac{\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^n}{n!} \tag{7.183}$$

Now we have

$$\left(\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^2 = \vec{\alpha}\cdot\vec{\alpha} + i(\vec{\alpha}\times\vec{\alpha})\cdot\vec{\sigma}_{op} = \vec{\alpha}^2 = \alpha^2 \tag{7.184}$$

Therefore

$$\begin{aligned} e^{-\frac{i}{\hbar}\vec{S}_{op}\cdot\vec{\alpha}} &= \hat{I} + \frac{1}{1!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right) + \frac{1}{2!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^2 + \frac{1}{3!}\left(-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}\right)^3 + \dots \\ &= \hat{I}\left(1 - \frac{\left(\frac{\alpha}{2}\right)^2}{2!} + \frac{\left(\frac{\alpha}{2}\right)^4}{4!} - \dots.\right) - i\vec{\sigma}_{op}\cdot\hat{\alpha}\left(\frac{\left(\frac{\alpha}{2}\right)}{1!} - \frac{\left(\frac{\alpha}{2}\right)^3}{3!} + \dots.\right) \\ &= \hat{I}\cos\frac{\alpha}{2} - i(\vec{\sigma}_{op}\cdot\hat{\alpha})\sin\frac{\alpha}{2} \end{aligned} \tag{7.185}$$

Consider an example. Let $\vec{\alpha} \to -90^\circ$ rotation about the x-axis or

$$\vec{\alpha} \to -\frac{\pi}{2}\hat{x} \tag{7.186}$$

Now we have

$$\hat{\sigma}_z(\hat{\sigma}_x)^n = (-\hat{\sigma}_x)^n\hat{\sigma}_z \tag{7.187}$$

which follows from the anticommutation relations. This implies that

$$\begin{aligned} \hat{\sigma}_z f(\hat{\sigma}_x) &= f(-\hat{\sigma}_x)\hat{\sigma}_z \\ \hat{\sigma}_z f(\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z) &= f(-\hat{\sigma}_x, -\hat{\sigma}_y, \hat{\sigma}_z)\hat{\sigma}_z \end{aligned}$$

Using these relations we get

$$\begin{aligned} e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}}\hat{\sigma}_z e^{\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} &= e^{\frac{i\pi}{4}\hat{\sigma}_x}\hat{\sigma}_z e^{-\frac{i\pi}{4}\hat{\sigma}_x} = e^{\frac{i\pi}{4}\hat{\sigma}_x}e^{\frac{i\pi}{4}\hat{\sigma}_x}\hat{\sigma}_z = e^{\frac{i\pi}{2}\hat{\sigma}_x}\hat{\sigma}_z \\ &= (\cos\frac{\pi}{2} + i\hat{\sigma}_x\sin\frac{\pi}{2})\hat{\sigma}_z = i\hat{\sigma}_x\hat{\sigma}_z = \hat{\sigma}_y \end{aligned} \tag{7.188}$$

as expected for this particular rotation.

Now the spin degrees of freedom are simply additional degrees of freedom for the system. The spin degrees of freedom are independent of the spatial degrees of freedom, however. This means that we can specify them independently or that $\vec{S}_{op}$ commutes with *all* operators that depend on 3-dimensional space

$$\left[\vec{S}_{op}, \vec{r}_{op}\right] = 0 \quad , \quad \left[\vec{S}_{op}, \vec{p}_{op}\right] = 0 \quad , \quad \left[\vec{S}_{op}, \vec{L}_{op}\right] = 0 \tag{7.189}$$

To specify completely the state of a *spinning* particle or a system with internal degrees of freedom of any kind we must know

1. the amplitudes for finding the particle at points in space

2. the amplitudes for different spin orientations

By convention, we choose $\hat{z}$ as the spin quantization direction for describing state vectors. The total state vector is then a direct-product state of the form

$$|\psi\rangle = |\text{external}\rangle \otimes |\text{internal}\rangle \tag{7.190}$$

and the amplitudes are

$$\begin{aligned}\langle \vec{r}, \hat{z}+ \mid \psi\rangle = &\text{ probability amplitude for finding the particle at} \\ &\vec{r} \text{ with spin up in the } \hat{z} \text{ direction} \\ \langle \vec{r}, \hat{z}- \mid \psi\rangle = &\text{ probability amplitude for finding the particle at} \\ &\vec{r} \text{ with spin down in the } \hat{z} \text{ direction}\end{aligned}$$

where

$$\langle \vec{r}, \hat{z}+ \mid \psi\rangle = \langle \vec{r} \mid \psi_{external}\rangle \langle \hat{z}+ \mid \psi_{\text{in}ternal}\rangle \tag{7.191}$$

and so on.

The total probability density for finding a particle at $\vec{r}$ is then the sum of two terms

$$|\langle \vec{r}, \hat{z}+ \mid \psi\rangle|^2 + |\langle \vec{r}, \hat{z}- \mid \psi\rangle|^2 \tag{7.192}$$

which represents a sum over all the ways of doing it.

The total angular momentum of such a *spinning* particle is the sum of its orbital and spin angular momenta

$$\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op} \tag{7.193}$$

$\vec{J}_{op}$ is now the generator of rotations in 3-dimensional space *and* in spin space or it affects *both* external *and* internal degrees of freedom.

If we operate with $e^{-i\vec{\alpha}\cdot\vec{J}_{op}/\hbar}$ on the basis state $|\vec{r}_0, \hat{m}+\rangle$, where the particle is definitely at $\vec{r}_0$ with spin up $(+\hbar/2)$ in the $\hat{m}$ direction, then we get a new state $|\vec{r}\,'_0, \hat{n}+\rangle$ where

$$\vec{r}\,'_0 = \vec{r}_0 + \vec{\alpha} \times \vec{r}_0 \text{ and } \hat{n} = \hat{m} + \vec{\alpha} \times \hat{m} \tag{7.194}$$

Since $[\vec{S}_{op}, \vec{L}_{op}] = 0$ we have

$$\begin{aligned} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} &= e^{-\frac{i}{\hbar}(\vec{\alpha}\cdot\vec{L}_{op}+\vec{\alpha}\cdot\vec{S}_{op})} \\ &= e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot[\vec{S}_{op},\vec{L}_{op}]} \\ &= e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}} \end{aligned} \tag{7.195}$$

This implies that

$$e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}} |\vec{r}_0, \hat{m}+\rangle = e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}_{op}} e^{-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{S}_{op}} |\vec{r}_0, \hat{m}+\rangle = |\vec{r}\,'_0, \hat{n}+\rangle \tag{7.196}$$

and that $e^{-i\vec{\alpha}\cdot\vec{L}_{op}/\hbar}$ carries out the rotation of the spatial degrees of freedom, while $e^{-i\vec{\alpha}\cdot\vec{S}_{op}/\hbar}$ carries out the rotation of the spin degrees of freedom.

If

$$|\psi'\rangle = e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}}|\psi\rangle \tag{7.197}$$

then the wave function of $|\psi'\rangle$ is

$$\langle\vec{r}_0, \hat{m}+ \mid \psi'\rangle = \langle\vec{r}_0, \hat{m}+|\, e^{\frac{i}{\hbar}\vec{\alpha}\cdot\vec{J}_{op}}|\psi\rangle = \langle\vec{r}\,'_0, \hat{n}+ \mid \psi\rangle \tag{7.198}$$

This is the wavefunction of $|\psi\rangle$ evaluated at the rotated point with a rotated spin quantization direction.

The *spin representation* of rotations has a feature which is strikingly different from that of rotations in 3-dimensional space.

Consider a rotation by 2π in spin space. This implies that

$$e^{-i\pi\vec{\sigma}_{op}\cdot\vec{\alpha}} = \cos\pi - i\vec{\sigma}_{op}\cdot\vec{\alpha}\sin\pi = -\hat{I} \tag{7.199}$$

A 2π rotation in spin space is represented by $-\hat{I}$. Since the rotations $\vec{\alpha}$ and $\vec{\alpha} + 2\pi\vec{\alpha}$ are physically equivalent, we must say that the spin representation of rotations is *double-valued*, i.e.,

$$e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} \text{ and } e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot(\vec{\alpha}+2\pi\hat{\alpha})} = -e^{-\frac{i}{2}\vec{\sigma}_{op}\cdot\vec{\alpha}} \tag{7.200}$$

represent the same rotation.

7.3.2 Superselection Rules

Let us expand on this important point. The 2π rotation transformation operator is given by

$$\hat{U}_{\hat{n}}(2\pi) = e^{-\frac{2\pi i}{\hbar}\hat{n}\cdot\vec{J}_{op}} \tag{7.201}$$

When operating on the angular momentum state vectors we have

$$\hat{U}_{\hat{n}}(2\pi)\,|j,m\rangle = e^{-\frac{2\pi i}{\hbar}\hat{n}\cdot\vec{J}_{op}}\,|j,m\rangle = e^{-2\pi ij}\,|j,m\rangle = (-1)^{2j}\,|j,m\rangle \tag{7.202}$$

This says that it has no effect if j = integer and multiplies by -1 if j = half-integer.

We usually think of a rotation through 2π as a trivial operation that changes nothing in a physical system. This belief implies that we are assuming all dynamical variables are invariant under 2π rotations or that

$$\hat{U}_{\hat{n}}(2\pi)\hat{A}\hat{U}^{\dagger}_{\hat{n}}(2\pi) = \hat{A} \tag{7.203}$$

where $\hat{A}$ is any physical observable.

But, as we have seen above, $\hat{U}_{\hat{n}}(2\pi)$ is not equal to a trivial operator (not equal to the identity operator for all physical states and operators). This says that invariance under $\hat{U}_{\hat{n}}(2\pi)$ may lead to nontrivial consequences.

The consequences that arise from invariance of an observable are not identical to those that arise from invariance of a state.

Let $\hat{U}$ = a unitary operator that leaves the observable $\hat{F}$ invariant or that we have

$$\left[\hat{U}, \hat{F}\right] = 0 \tag{7.204}$$

Now consider a state that is not invariant under the transformation $\hat{U}$. If it is a pure state represented by $|\psi\rangle$, then $|\psi'\rangle = \hat{U}|\psi\rangle \neq |\psi\rangle$. The expectation value of $\hat{F}$ in the state $|\psi\rangle$ is

$$\left\langle \hat{F} \right\rangle = \langle \psi' | \hat{F} | \psi' \rangle = \langle \psi | \hat{U}^+ \hat{F} \hat{U} | \psi \rangle = \langle \psi | \hat{U}^+ \hat{U} \hat{F} | \psi \rangle = \langle \psi | \hat{F} | \psi \rangle \tag{7.205}$$

which implies that the *observable statistical properties* of $\hat{F}$ are the same in the two states $|\psi\rangle$ and $|\psi'\rangle$. This conclusion certainly holds for $\hat{U}(2\pi)$. Does anything else hold? Is there something peculiar to $\hat{U}(2\pi)$?

It turns out that $\hat{U}(2\pi)$ divides the vector space into two subspaces:

1. integer angular momentum - has states $|+\rangle$ where

$$\hat{U}(2\pi)|+\rangle = |+\rangle \tag{7.206}$$

2. half-integer angular momentum - has states $|-\rangle$ where

$$\hat{U}(2\pi)|-\rangle = -|-\rangle \tag{7.207}$$

Now for any invariant physical observable $\hat{B}$ (where $\left[\hat{U}, \hat{B}\right] = 0$), we have

$$\begin{aligned} &\langle + | \hat{U}(2\pi) \hat{B} | - \rangle = \langle + | \hat{B} \hat{U}(2\pi) | - \rangle \\ &\langle + | \hat{B} | - \rangle = - \langle + | \hat{B} | - \rangle \\ &\rightarrow \langle + | \hat{B} | - \rangle = 0 \end{aligned} \tag{7.208}$$

This says that *all* physical observable have *vanishing matrix elements* between states with integer angular momentum and states with half-integer angular momentum (states in the two subspaces).

This is called a *superselection rule.*

A superselection rule says that there is *no observable distinction* among vectors of the form

$$|\psi_\varphi\rangle = |+\rangle + e^{i\varphi}|-\rangle \tag{7.209}$$

for different values of the phase φ. This is so because

$$\langle\psi_\varphi|\,\hat{B}\,|\psi_\varphi\rangle = \langle +|\,\hat{B}\,|+\rangle + \langle -|\,\hat{B}\,|-\rangle \tag{7.210}$$

and this is independent of φ.

What about a more general state represented by the state operator

$$\hat{W} = \sum_{i,j} w_{ij}\,|i\rangle\,\langle j| \tag{7.211}$$

If we break up the basis for the space into + and − subspaces, i.e.,

$$\text{basis set } = \{\{+\text{ states}\},\{-\text{ states}\}\}$$

then the matrix representation of $\hat{W}$ partitions into four blocks

$$\hat{W} = \begin{pmatrix} \hat{W}_{++} & \hat{W}_{+-} \\ \hat{W}_{-+} & \hat{W}_{--} \end{pmatrix} \tag{7.212}$$

While, for any physical observable, the same partitioning scheme produces

$$\hat{B} = \begin{pmatrix} \hat{B}_{++} & 0 \\ 0 & \hat{B}_{--} \end{pmatrix} \tag{7.213}$$

i.e., there is no mixing between the subspaces. This gives

$$\left\langle \hat{B} \right\rangle = Tr\left(\hat{W}\hat{B}\right) = Tr_+\left(\hat{W}_{++}\hat{B}_{++}\right) + Tr_-\left(\hat{W}_{--}\hat{B}_{--}\right) \tag{7.214}$$

where $Tr_\pm$ implies a trace only over the particular subspace. The cross matrix elements $\hat{W}_{+-}$ and $\hat{W}_{-+}$ do not contribute to the expectation value of the observables, or *interference between vectors of the* $|+\rangle$ *and* $|-\rangle$ *types is not observable.*

All equations of motion decouple into two separate equations in each of the two subspaces and no cross-matrix elements of $\hat{W}$ between the two subspaces ever contribute.

If we assume that the cross matrix elements are zero initially, then they will never develop(become nonzero) in time.

What is the difference between a generator $\hat{U}(2\pi)$ of a superselection rule and a symmetry operation that is generated by a universally conserved quantity such as the displacement operator

$$e^{-\frac{i}{\hbar}\vec{a}\cdot\vec{P}_{op}}$$

which is generated by the total momentum $\vec{P}_{op}$?

The Hamiltonian of any closed system is invariant under both transformations.

Both give rise to a quantum number that must be conserved in any transition. In these examples the quantum numbers are

$$\pm 1 \text{ for } \hat{U}(2\pi) \text{ and the total momentum}$$

The difference is that there exist observables that do not commute with $\vec{P}_{op}$ and $\vec{Q}_{op}$, but there are *no observables* that do not commute with $\hat{U}(2\pi)$.

By measuring the position one can distinguish states that differ only by a displacement, but there is no way to distinguish between states that differ only by a 2π rotation.

The superselection rules from $\hat{U}(2\pi)$, which separates the integer and half-integer angular momentum states, is the only such rule in the quantum mechanics of stable particles (non-relativistic quantum mechanics).

In Quantum Field Theory(relativistic quantum mechanics), where particles can be created and annihilated, the total electric charge operator generates another superselection rule, provided that one assumes all observables are invariant under gauge transformations.

This says that no interference can be observed between states of different total charge because there are no observables that do not commute with the charge operator.

In a theory of stable particles, the charge of each particle and hence the total charge is an invariant. Thus, the total charge operator is simply a multiple of $\hat{I}$. Every operator commutes with $\hat{I}$ implying that the charge superselection rule is trivial in non-relativistic quantum mechanics.

Now back to the normal world.

The techniques that we have developed for the spin = 1/2 system can be applied to any two-state system. Here is an example of a two-sided box solved using both the Schrodinger and Heisenberg pictures.

7.3.3 A Box with 2 Sides

Let us consider a box containing a particle in a state $|\psi\rangle$. The box is divided into two halves (**R**ight and **L**eft) by a thin partition. The only property that we will assign to the particle is whether it is on the **R**ight or **L**eft side of the box.

This means that the system has only two states, namely, $|R\rangle$ and $|L\rangle$ that must

have the properties

$$\langle R \mid \psi \rangle = \text{ amplitude to find the particle on right side if in state } |\psi\rangle$$
$$\langle L \mid \psi \rangle = \text{ amplitude to find the particle on leftt side if in state } |\psi\rangle$$

We also suppose that the particle can tunnel through the thin partition and that

$$\frac{d}{dt}\langle R \mid \psi \rangle = \frac{K}{i\hbar}\langle L \mid \psi \rangle \quad , \quad K \text{ real} \tag{7.215}$$

How does this system develop in time?

We solve this problem in two ways, namely, using the Schrodinger and Heisenberg pictures.

Schrodinger Picture

We define the general system state vector

$$\begin{aligned} |\psi\rangle &= \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} \langle R \mid \psi \rangle \\ \langle L \mid \psi \rangle \end{pmatrix} \\ &= \psi_R \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi_L \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \psi_R |R\rangle + \psi_L |L\rangle \end{aligned} \tag{7.216}$$

The time-dependent Schrodinger equation is

$$i\hbar \frac{\partial}{\partial t}|\psi(t)\rangle = i\hbar \begin{pmatrix} \frac{\partial \psi_R}{\partial t} \\ \frac{\partial \psi_L}{\partial t} \end{pmatrix} = \hat{H} \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} \tag{7.217}$$

Now we are given that

$$\frac{d\psi_R}{dt} = \frac{K}{i\hbar}\psi_L \rightarrow \frac{d\psi_R^*}{dt} = -\frac{K}{i\hbar}\psi_L^* \tag{7.218}$$

The state vector must remain normalized as it develops in time so that

$$\begin{aligned} &\langle \psi \mid \psi \rangle = |\psi_R|^2 + |\psi_L|^2 = 1 \\ &\frac{\partial \langle \psi \mid \psi \rangle}{\partial t} = 0 = \frac{\partial \psi_R^*}{\partial t}\psi_R + \psi_R^* \frac{\partial \psi_R}{\partial t} + \frac{\partial \psi_L^*}{\partial t}\psi_L + \psi_L^* \frac{\partial \psi_L}{\partial t} \\ &0 = -\frac{K}{i\hbar}\psi_L^*\psi_R + \psi_R^* \frac{K}{i\hbar}\psi_L + \frac{\partial \psi_L^*}{\partial t}\psi_L + \psi_L^* \frac{\partial \psi_L}{\partial t} \\ &0 = \psi_L \left(\frac{\partial \psi_L^*}{\partial t} + \frac{K}{i\hbar}\psi_R^* \right) + \psi_L^* \left(\frac{\partial \psi_L}{\partial t} - \frac{K}{i\hbar}\psi_R \right) \end{aligned}$$

which says that

$$\frac{d\psi_L}{dt} = \frac{K}{i\hbar}\psi_R \rightarrow \frac{d\psi_L^*}{dt} = -\frac{K}{i\hbar}\psi_R^* \tag{7.219}$$

Therefore we have

$$K\begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} = \hat{H}\begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} \tag{7.220}$$

which says that

$$\hat{H} = K\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7.221}$$

The eigenvalues of $\hat{H}$ are $\pm K$ and its eigenvectors are

$$|\pm K\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \tag{7.222}$$

Note that for $\hat{\wp}$ = parity operator (switches right and left) we have

$$\hat{\wp}\,|\pm K\rangle = \pm\,|\pm K\rangle \tag{7.223}$$

so these are also states of definite parity.

If the initial state of the system is

$$|\psi(0)\rangle = \begin{pmatrix} \psi_R(0) \\ \psi_L(0) \end{pmatrix} \tag{7.224}$$

then we can write this state in terms of energy eigenstates as

$$\begin{aligned} |\psi(0)\rangle &= \begin{pmatrix} \psi_R(0) \\ \psi_L(0) \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\left(\psi_R(0) + \psi_L(0)\right)|+K\rangle + \frac{1}{\sqrt{2}}\left(\psi_R(0) - \psi_L(0)\right)|-K\rangle \end{aligned} \tag{7.225}$$

Since we know the time dependence of energy eigenstates

$$|\pm K\rangle_t = e^{\mp \frac{i}{\hbar}Kt}\,|\pm K\rangle \tag{7.226}$$

the time dependence of $|\psi(t)\rangle$ is given by

$$\begin{aligned} |\psi(t)\rangle = \frac{1}{\sqrt{2}}&\left(\psi_R(0) + \psi_L(0)\right)e^{-\frac{i}{\hbar}Kt}\,|+K\rangle \\ &+ \frac{1}{\sqrt{2}}\left(\psi_R(0) - \psi_L(0)\right)e^{+\frac{i}{\hbar}Kt}\,|-K\rangle \end{aligned} \tag{7.227}$$

or

$$\begin{aligned} |\psi(t)\rangle &= \frac{1}{2}\begin{pmatrix} \left(\psi_R(0) + \psi_L(0)\right)e^{-\frac{i}{\hbar}Kt} + \left(\psi_R(0) - \psi_L(0)\right)e^{+\frac{i}{\hbar}Kt} \\ \left(\psi_R(0) + \psi_L(0)\right)e^{-\frac{i}{\hbar}Kt} - \left(\psi_R(0) - \psi_L(0)\right)e^{+\frac{i}{\hbar}Kt} \end{pmatrix} \\ &= \left(\psi_R(0)\cos\frac{Kt}{\hbar} - i\psi_L(0)\sin\frac{Kt}{\hbar}\right)|R\rangle \\ &\quad + \left(-i\psi_R(0)\sin\frac{Kt}{\hbar} + \psi_L(0)\cos\frac{Kt}{\hbar}\right)|L\rangle \end{aligned} \tag{7.228}$$

Therefore, the probability that the particle is on the **R**ight side at time t is

$$P_R = |\langle R \mid \psi(t)\rangle|^2 = \left|\psi_R(0)\cos\frac{Kt}{\hbar} - i\psi_L(0)\sin\frac{Kt}{\hbar}\right|^2 \tag{7.229}$$

Now suppose that

$$\psi_R(0) = \frac{1}{\sqrt{2}} \text{ and } \psi_L(0) = e^{i\delta}\psi_R(0) \tag{7.230}$$

This says that the particle was equally probable to be on either side at $t = 0$, but that the amplitudes differed by a phase factor. In this case, we get

$$P_R = \frac{1}{2}\left(1 + \sin\delta\sin\frac{2Kt}{\hbar}\right) \tag{7.231}$$

Heisenberg Picture

Let us define an operator $\hat{Q}$ such that for a state vector

$$|\psi\rangle = \psi_R|R\rangle + \psi_L|L\rangle$$

we have

$$\langle\psi|\hat{Q}|\psi\rangle = |\psi_R|^2 = \text{ probability that particle is on right side} \tag{7.232}$$

This says that

$$\langle\psi|\hat{Q}|\psi\rangle = |\psi_R|^2 = \langle\psi \mid R\rangle\langle R \mid \psi\rangle \tag{7.233}$$

or

$$\hat{Q} = |R\rangle\langle R| = \text{ pure state projection operator } = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \hat{I} + \hat{\sigma}_z \tag{7.234}$$

and that the expectation value of the pure state projection operator is equal to the probability of being in that state. This agrees with our earlier discussions.

Now we have

$$\hat{H} = K\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = K\hat{\sigma}_x \tag{7.235}$$

Therefore,

$$\hat{Q}(t) = e^{\frac{i}{\hbar}\hat{H}t}\hat{Q}e^{-\frac{i}{\hbar}\hat{H}t} \tag{7.236}$$

Now as we saw earlier

$$e^{\frac{i}{\hbar}\hat{H}t} = e^{\frac{i}{\hbar}K\hat{\sigma}_x t} = \cos\frac{Kt}{\hbar}\hat{I} + i\sin\frac{Kt}{\hbar}\hat{\sigma}_x \tag{7.237}$$

Thus,

$$\begin{aligned}\hat{Q}(t) &= \left(\cos\frac{Kt}{\hbar}\hat{I} + i\sin\frac{Kt}{\hbar}\hat{\sigma}_x\right)\left(\hat{I}+\hat{\sigma}_z\right)\left(\cos\frac{Kt}{\hbar}\hat{I} - i\sin\frac{Kt}{\hbar}\hat{\sigma}_x\right)\\ &= \cos^2\frac{Kt}{\hbar} + \sin^2\frac{Kt}{\hbar} + \cos^2\frac{Kt}{\hbar}\hat{\sigma}_z - i\sin\frac{Kt}{\hbar}\cos\frac{Kt}{\hbar}\left[\hat{\sigma}_z,\hat{\sigma}_x\right] + \sin^2\frac{Kt}{\hbar}\hat{\sigma}_x\hat{\sigma}_z\hat{\sigma}_x\\ &= 1 + \cos^2\frac{Kt}{\hbar}\hat{\sigma}_z + 2\sin\frac{Kt}{\hbar}\cos\frac{Kt}{\hbar}\hat{\sigma}_y - \sin^2\frac{Kt}{\hbar}\hat{\sigma}_z\\ &= 1 + \cos\frac{2Kt}{\hbar}\hat{\sigma}_z + \sin\frac{2Kt}{\hbar}\hat{\sigma}_y \end{aligned} \tag{7.238}$$

Then

$$P_R(t) = \langle\psi(0)|\,\hat{Q}(t)\,|\psi(0)\rangle \tag{7.239}$$

where

$$|\psi(0)\rangle = \frac{1}{\sqrt{2}}|R\rangle + \frac{e^{i\delta}}{\sqrt{2}}|L\rangle \tag{7.240}$$

Now

$$\begin{aligned}\hat{\sigma}_z|R\rangle &= |R\rangle \quad , \quad \hat{\sigma}_z|L\rangle = -|L\rangle\\ \hat{\sigma}_y|R\rangle &= -i|L\rangle \quad , \quad \hat{\sigma}_y|L\rangle = i|R\rangle\end{aligned}$$

and we get

$$P_R = \frac{1}{2}(1 + \sin\delta\sin\frac{2Kt}{\hbar}) \tag{7.241}$$

as before.

7.4 Magnetic Resonance

How does an experimentalist observe the spin of a particle?

Classically, a spinning charge distribution will have an associated magnetic moment. In non-relativistic quantum mechanics particles with internal spin degrees of freedom also have magnetic moments which are connected to their angular momentum. We write for the magnetic moment operator

$$\begin{aligned}\vec{M}_{op} &= \vec{M}^{orbital}_{op} + \vec{M}^{spin}_{op}\\ &= g_\ell\frac{q}{2mc}\vec{L}_{op} + g_s\frac{q}{2mc}\vec{S}_{op} = \frac{q}{2mc}\left(g_\ell\vec{L}_{op} + g_s\vec{S}_{op}\right)\end{aligned} \tag{7.242}$$

where, as we shall derive later,

$$g_{j\ell s} = 1 + \frac{j(j+1) - \ell(\ell+1) + s(s+1)}{2j(j+1)} \tag{7.243}$$

$$g_\ell = g_{jj0} = 1 \text{ and } g_s = g_{j0j} = 2 \tag{7.244}$$

Therefore we have

$$\vec{M}_{op} = \frac{q}{2mc}\left(\vec{L}_{op} + 2\vec{S}_{op}\right) = \frac{q}{2mc}\left(\vec{J}_{op} + \vec{S}_{op}\right) \tag{7.245}$$

which says that $\vec{M}_{op}$ is *not parallel* to $\vec{J}_{op}$. The energy operator or contribution to the Hamiltonian operator from the magnetic moment in a magnetic field $\vec{B}(\vec{r},t)$ is

$$\hat{H}_M = -\vec{M}_{op}\cdot\vec{B}(\vec{r},t) \tag{7.246}$$

Spin measurement experiments are designed to detect the effects of this extra contribution to the system energy and thus detect the effects of spin angular momentum.

We will return to a full discussion of the effect of $\hat{H}_M$ on the energy levels of atoms, etc in a later chapter. For now we will restrict our attention to spin space and investigate the effect of the spin contribution to $\hat{H}_M$ on the states of a particle.

We will use

$$\hat{H}_{spin} = -\frac{g}{2}\frac{q}{mc}\vec{B}\cdot\bar{S}_{op} \tag{7.247}$$

where we have replaced g_s by g. We have not set $g = 2$ since it turns out that it is not exactly that value(due to relativistic effects).

In the Schrodinger picture we have

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = \hat{H}_{spin}|\psi(t)\rangle \tag{7.248}$$

If we ignore spatial dependences (worry only about spin effects), we have

$$|\psi(t)\rangle = \begin{pmatrix} \langle\hat{n}+\,|\,\psi(t)\rangle \\ \langle\hat{n}-\,|\,\psi(t)\rangle \end{pmatrix} \tag{7.249}$$

$$\begin{aligned} i\hbar\frac{d}{dt}\begin{pmatrix} \langle\hat{n}+\,|\,\psi(t)\rangle \\ \langle\hat{n}-\,|\,\psi(t)\rangle \end{pmatrix} &= -\frac{g}{2}\frac{q}{mc}\vec{B}\cdot\bar{S}_{op}\begin{pmatrix} \langle\hat{n}+\,|\,\psi(t)\rangle \\ \langle\hat{n}-\,|\,\psi(t)\rangle \end{pmatrix} \\ &= -\frac{g}{4}\frac{q\hbar}{mc}\vec{B}\cdot\vec{\sigma}_{op}\begin{pmatrix} \langle\hat{n}+\,|\,\psi(t)\rangle \\ \langle\hat{n}-\,|\,\psi(t)\rangle \end{pmatrix} \\ &= -\frac{g}{4}\frac{q\hbar}{mc}\begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix}\begin{pmatrix} \langle\hat{n}+\,|\,\psi(t)\rangle \\ \langle\hat{n}-\,|\,\psi(t)\rangle \end{pmatrix} \end{aligned} \tag{7.250}$$

This represents two coupled differential equations for the time dependence of the amplitudes $\langle+\,|\,\psi\rangle$ and $\langle-\,|\,\psi\rangle$. When solved, the solution tells us the time dependence of the measurable probabilities.

We can see the physics of the motion best in the Heisenberg picture, where the

operators, instead of the states, move in time. In this case, we have

$$i\hbar\frac{d\hat{S}_i(t)}{dt} = \left[\hat{S}_i(t), \hat{H}_{spin}(t)\right]$$
$$= -\frac{gq}{2mc}\left[\hat{S}_i(t), \hat{S}_j(t)\right]B_j(t) = -i\frac{gq\hbar}{2mc}\varepsilon_{ijk}\hat{S}_k(t)B_j(t) \tag{7.251}$$

for each component. The operators are all time-dependent Heisenberg picture operators. This gives

$$\frac{d\vec{S}_{op}(t)}{dt} = \frac{gq}{2mc}\vec{S}_{op}(t)\times\vec{B}(t) = \vec{M}_{spin}(t)\times\vec{B}(t) \tag{7.252}$$

The right-hand-side is the *torque* exerted by the magnetic field on the magnetic moment.

In operator language, this equation implies that the rate of change of the spin angular momentum vector equals the applied torque. This implies that the spin vector with $q < 0$ precesses in a positive sense about the magnetic field direction as shown in Figure 7.1 below.

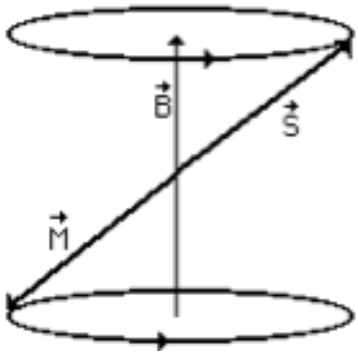


Figure 7.1: Motion of Spin

Suppose that $\vec{B}(t) = B_0\hat{z}$ (independent of t). We then have

$$\frac{d\hat{S}_z(t)}{dt} = 0 \tag{7.253}$$
$$\frac{d\hat{S}_x(t)}{dt} = \frac{gqB_0}{2mc}\hat{S}_y(t) \tag{7.254}$$
$$\frac{d\hat{S}_y(t)}{dt} = -\frac{gqB_0}{2mc}\hat{S}_x(t) \tag{7.255}$$

These coupled differential equations have the solutions

$$\hat{S}_x(t) = \hat{S}_x(0)\cos\omega_0 t + \hat{S}_y(0)\sin\omega_0 t \tag{7.256}$$

$$\hat{S}_y(t) = -\hat{S}_x(0)\sin\omega_0 t + \hat{S}_y(0)\cos\omega_0 t \tag{7.257}$$

$$\hat{S}_z(t) = \hat{S}_z(0) \tag{7.258}$$

where

$$\omega_0 = \frac{gqB_0}{2mc} \tag{7.259}$$

Now suppose that at $t = 0$, the spin is in the $+x$-direction, which says that

$$\left\langle \hat{S}_x(0) \right\rangle = \langle \hat{x}+| \hat{S}_x(0) |\hat{x}+\rangle = \frac{\hbar}{2} \tag{7.260}$$

$$\left\langle \hat{S}_y(0) \right\rangle = \left\langle \hat{S}_z(0) \right\rangle = 0 \tag{7.261}$$

Therefore, the expectation values of the solutions become

$$\left\langle \hat{S}_x(t) \right\rangle = \frac{\hbar}{2}\cos\omega_0 t \tag{7.262}$$

$$\left\langle \hat{S}_y(t) \right\rangle = -\frac{\hbar}{2}\sin\omega_0 t \tag{7.263}$$

$$\left\langle \hat{S}_z(0) \right\rangle = 0 \tag{7.264}$$

which says that the expectation value of the spin vector rotates in a negative sense in the $x - y$ plane (precession about the z-axis).

Now let us return to the Schrodinger picture to see what precession looks like there. We will do the calculation a couple of different ways.

First we use the time-development operator. For $\hat{B} = B\hat{n}$ we have

$$\hat{H} = -\frac{gq\hbar B}{4mc}\vec{\sigma}_{op} \cdot \hat{n} \tag{7.265}$$

If we define

$$\omega_L = \frac{qB}{mc} = \text{ Larmor frequency} \tag{7.266}$$

and let $g = 2$ we have

$$\hat{H} = -\frac{1}{2}\hbar\omega_L\vec{\sigma}_{op} \cdot \hat{n} \tag{7.267}$$

In the Schrodinger picture,

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = \hat{H}_{spin}(t)|\psi(t)\rangle \tag{7.268}$$

When $\hat{H}_{spin}(t)$ is time-independent, we have the solution

$$|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle \tag{7.269}$$

where

$$\hat{U}(t) = e^{-\frac{i}{\hbar}\hat{H}t} \tag{7.270}$$

or

$$|\psi(t)\rangle = e^{i\frac{\omega_L t}{2}\vec{\sigma}_{op}\cdot\hat{n}}|\psi(0)\rangle \tag{7.271}$$

Since $\hat{n}$ = a unit vector, we have $(\vec{\sigma}_{op}\cdot\hat{n})^2 = \hat{I}$, which gives, as before,

$$|\psi(t)\rangle = \left(\cos\frac{\omega_L t}{2} + i(\vec{\sigma}_{op}\cdot\hat{n})\sin\frac{\omega_L t}{2}\right)|\psi(0)\rangle \tag{7.272}$$

This is the solution to the Schrodinger equation in this case.

Now let

$$|\psi(0)\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\hat{z}\uparrow\rangle = |\hat{z}+\rangle = |+\rangle \tag{7.273}$$

From earlier we have

$$\vec{\sigma}_{op}\cdot\hat{n} = \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix} \tag{7.274}$$

We then have

$$\begin{aligned}
|\psi(t)\rangle &= \cos\frac{\omega_L t}{2}|+\rangle + i\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}|+\rangle \sin\frac{\omega_L t}{2} \\
&= \cos\frac{\omega_L t}{2}|+\rangle + i\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\sin\frac{\omega_L t}{2} \\
&= \cos\frac{\omega_L t}{2}|+\rangle + i\sin\frac{\omega_L t}{2}\begin{pmatrix} n_z \\ n_x + in_y \end{pmatrix} \\
&= \cos\frac{\omega_L t}{2}|+\rangle + i\sin\frac{\omega_L t}{2}\left(n_z|+\rangle + (n_x + in_y)|-\rangle\right) \\
&= \left(\cos\frac{\omega_L t}{2} + in_z\sin\frac{\omega_L t}{2}\right)|+\rangle + i(n_x + in_y)\sin\frac{\omega_L t}{2}|-\rangle
\end{aligned} \tag{7.275}$$

This says that as the initial state $|+\rangle$ develops in time it picks up some amplitude to be in the $|-\rangle$ state.

Special Case

Let

$$\begin{aligned}
&\hat{n} = \hat{y} \to \vec{B} \text{ is in the } y\text{- direction} \\
&n_y = 1\ ,\ n_x = n_z = 0
\end{aligned}$$

This gives

$$|\psi(t)\rangle = \cos\frac{\omega_L t}{2}|+\rangle - \sin\frac{\omega_L t}{2}|-\rangle \tag{7.276}$$

which implies that $\langle \hat{s}_z \rangle$ flips with frequency $\nu = \omega_L/4\pi$ or that the spin vector is precessing around the direction of the magnetic field.

Finally, let us just solve the Schrodinger equation directly. We have

$$i\hbar \frac{d}{dt} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t) \rangle \\ \langle \hat{n}- \mid \psi(t) \rangle \end{pmatrix} = -\frac{1}{2}\frac{q\hbar}{mc} \begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix} \begin{pmatrix} \langle \hat{n}+ \mid \psi(t) \rangle \\ \langle \hat{n}- \mid \psi(t) \rangle \end{pmatrix} \tag{7.277}$$

If we choose $\hat{n} = \hat{x}$,then we have

$$i\hbar \frac{d}{dt} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = -\frac{1}{2}\hbar\omega_L \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} \tag{7.278}$$

where $a(t) = \langle \hat{n}+ \mid \psi(t) \rangle$ and $b(t) = \langle \hat{n}- \mid \psi(t) \rangle$

and $|a|^2 + |b|^2 = 1$

This gives two coupled differential equations

$$\dot{a} = i\frac{\omega_L}{2} b \text{ and } \dot{b} = i\frac{\omega_L}{2} a \tag{7.279}$$

which are solved as follows:

$$(\dot{a} + \dot{b}) = i\frac{\omega_L}{2}(\dot{a} + \dot{b}) \rightarrow a + b = (a(0) + b(0))e^{i\frac{\omega_L}{2}t} \tag{7.280}$$

$$(\dot{a} - \dot{b}) = -i\frac{\omega_L}{2}(\dot{a} - \dot{b}) \rightarrow a - b = (a(0) - b(0))e^{-i\frac{\omega_L}{2}t} \tag{7.281}$$

or

$$a(t) = a(0)\cos\frac{\omega_L}{2}t + ib(0)\sin\frac{\omega_L}{2}t \tag{7.282}$$

$$b(t) = ia(0)\sin\frac{\omega_L}{2}t + b(0)\cos\frac{\omega_L}{2}t \tag{7.283}$$

For the initial state

$$|\psi(0)\rangle = \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{7.284}$$

we get

$$a(t) = \cos\frac{\omega_L}{2}t \quad , \quad b(t) = i\sin\frac{\omega_L}{2}t \tag{7.285}$$

and

$$a(t) = \cos\frac{\omega_L}{2}t \quad , \quad b(t) = i\sin\frac{\omega_L}{2}t \tag{7.286}$$

as before.

7.4.1 Spin Resonance

Let us now consider the addition of an oscillating magnetic field perpendicular to the applied field $\vec{B} = B_0\hat{z}$. In particular, we add the field

$$\vec{B}_1 = B_1 \cos\omega t \hat{x} \tag{7.287}$$

In the Schrodinger picture, we then have the equation of motion

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t\hat{\sigma}_x\right)|\psi(t)\rangle \tag{7.288}$$

Now we first make the transformation (shift to rotating coordinate system since we know that if the extra field were not present the spin would just be precessing about the direction of the applied field)

$$|\psi(t)\rangle = e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle \tag{7.289}$$

which gives

$$i\hbar\frac{d}{dt}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right) = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t\hat{\sigma}_x\right)e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle \tag{7.290}$$

$$\begin{aligned} &i\hbar e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\frac{d}{dt}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right) \\ &\quad = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0 e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_z e^{i\frac{\omega t}{2}\hat{\sigma}_z} + B_1\cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}\hat{\sigma}_z}\right)|\psi'(t)\rangle \end{aligned} \tag{7.291}$$

or

$$\begin{aligned} &i\hbar e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\left(e^{i\frac{\omega t}{2}\hat{\sigma}_z}\frac{d}{dt}|\psi'(t)\rangle + i\frac{\omega}{2}\hat{\sigma}_z e^{i\frac{\omega t}{2}\hat{\sigma}_z}|\psi'(t)\rangle\right) \\ &\quad = -\frac{e\hbar}{2mc}\frac{g}{2}\left(B_0\hat{\sigma}_z + B_1\cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}\hat{\sigma}_z}\right)|\psi'(t)\rangle \end{aligned} \tag{7.292}$$

Now

$$\begin{aligned} \cos\omega t e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\hat{\sigma}_x e^{i\frac{\omega t}{2}t\hat{\sigma}_z} &= \cos\omega t\hat{\sigma}_x e^{i\omega t\hat{\sigma}_z} \\ &= \hat{\sigma}_x(\cos^2\omega t + i\hat{\sigma}_z\cos\omega t\sin\omega t) \\ &= \hat{\sigma}_x(\frac{1}{2} + \frac{1}{2}\cos 2\omega t + i\hat{\sigma}_z\frac{1}{2}\sin 2\omega t) \\ &= \frac{\hat{\sigma}_x}{2} + \frac{1}{2}\left(\hat{\sigma}_x\cos 2\omega t + i\hat{\sigma}_x\hat{\sigma}_z\sin 2\omega t\right) \\ &= \frac{\hat{\sigma}_x}{2} + \frac{1}{2}\left(\hat{\sigma}_x\cos 2\omega t + \hat{\sigma}_y\sin 2\omega t\right) \end{aligned} \tag{7.293}$$

Defining

$$\omega_0 = \frac{geB_0}{4mc} \text{ and } \omega_1 = \frac{geB_1}{4mc} \tag{7.294}$$

we get

$$
\begin{aligned}
i\hbar\frac{d}{dt}|\psi'(t)\rangle & \\
&= \left(\left(\frac{\omega-\omega_0}{2}\right)\hat{\sigma}_z - \frac{\omega_1}{2}\hat{\sigma}_x\right)|\psi'(t)\rangle \\
&\qquad + \frac{\omega_1}{2}\left(\hat{\sigma}_x\cos 2\omega t + \hat{\sigma}_y \sin 2\omega t\right)|\psi'(t)\rangle
\end{aligned}
\tag{7.295}
$$

The two higher frequency terms produce high frequency wiggles in $|\psi'(t)\rangle$. Since we will be looking at the average motion of the spin vector (expectation values changing in time), these terms will average to zero and we can neglect them in this discussion.

We thus have

$$
i\hbar\frac{d}{dt}|\psi'(t)\rangle = \left(\left(\frac{\omega-\omega_0}{2}\right)\hat{\sigma}_z - \frac{\omega_1}{2}\hat{\sigma}_x\right)|\psi'(t)\rangle \tag{7.296}
$$

which has a solution

$$
|\psi'(t)\rangle = e^{-i\frac{\Omega t}{2}\hat{\sigma}}|\psi'(0)\rangle \tag{7.297}
$$

$$
\Omega = \left[(\omega-\omega_0)^2+\omega_1^2\right]^{1/2} \text{ and } \hat{\sigma} = \frac{\omega-\omega_0}{\Omega}\hat{\sigma}_z - \frac{\omega_1}{\Omega}\hat{\sigma}_x \tag{7.298}
$$

We note that

$$
\hat{\sigma}^2 = \frac{(\omega-\omega_0)^2+\omega_1^2}{\Omega^2}\hat{I} = \hat{I} \tag{7.299}
$$

The final solution to the original problem (again neglecting the higher frequency terms) is then(after leaving the rotating system)

$$
|\psi(t)\rangle = e^{-i\frac{\omega t}{2}\hat{\sigma}_z}e^{-i\frac{\Omega t}{2}\hat{\sigma}}|\psi(0)\rangle \tag{7.300}
$$

Example

Let us choose

$$
|\psi(0)\rangle = |+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{7.301}
$$

We get

$$
\begin{aligned}
|\psi(t)\rangle &= e^{-i\frac{\omega t}{2}\hat{\sigma}_z}e^{-i\frac{\Omega t}{2}\hat{\sigma}}|+\rangle = e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\left(\cos\frac{\Omega t}{2} - i\hat{\sigma}\sin\frac{\Omega t}{2}\right)|+\rangle \\
&= e^{-i\frac{\omega t}{2}\hat{\sigma}_z}\left(\cos\frac{\Omega t}{2}|+\rangle - i\frac{\omega-\omega_0}{\Omega}\sin\frac{\Omega t}{2}\hat{\sigma}_z|+\rangle + i\frac{\omega_1}{\Omega}\sin\frac{\Omega t}{2}\hat{\sigma}_x|+\rangle\right)
\end{aligned}
\tag{7.302}
$$

Now using

$$
\hat{\sigma}_z|+\rangle = |+\rangle \text{ and } \hat{\sigma}_x|+\rangle = |-\rangle \tag{7.303}
$$

$$
e^{-i\frac{\omega t}{2}\hat{\sigma}_z} = \cos\frac{\omega t}{2} - i\sin\frac{\omega t}{2}\hat{\sigma}_z \tag{7.304}
$$

we have

$$|\psi(t)\rangle = A^{(+)}\,|+\rangle + A^{(-)}\,|-\rangle \tag{7.305}$$

$$A^{(+)} = \cos\frac{\omega t}{2}\cos\frac{\Omega t}{2} - i\sin\frac{\omega t}{2}\cos\frac{\Omega t}{2} - i\frac{\omega-\omega_0}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} - \frac{\omega-\omega_0}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2} \tag{7.306}$$

$$A^{(-)} = -i\frac{\omega_1}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} + \frac{\omega_1}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2} \tag{7.307}$$

Therefore, the amplitude for spin flip to the state $|\hat{z}\downarrow\rangle = |-\rangle$ at time t is

$$\begin{aligned}\langle - \mid \psi(t)\rangle = A^{(-)} &= -i\frac{\omega_1}{\Omega}\cos\frac{\omega t}{2}\sin\frac{\Omega t}{2} + \frac{\omega_1}{\Omega}\sin\frac{\omega t}{2}\sin\frac{\Omega t}{2} \\ &= -i\frac{\omega_1}{\Omega}\sin\frac{\Omega t}{2}e^{i\frac{\omega t}{2}}\end{aligned} \tag{7.308}$$

and the probability of spin flip is

$$P_{flip}(t) = |\langle - \mid \psi(t)\rangle|^2 = \frac{\omega_1^2}{\Omega^2}\sin^2\frac{\Omega t}{2} = \frac{\omega_1^2}{2\Omega^2}(1-\cos\Omega t) \tag{7.309}$$

What is happening to the spin vector? If we plot $P_{flip}(t)$ versus t as in Figure 7.2 below we get

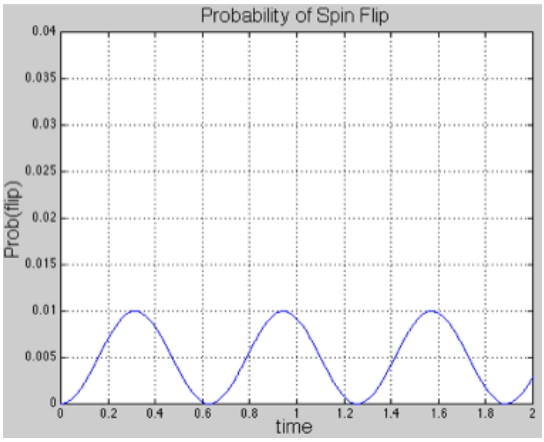


Figure 7.2: Spin=Flip Probability versus Time

where the peak values are given by ω_1^2/Ω^2. What is the value of ω_1^2/Ω^2 and can it be large? If we plot ω_1^2/Ω^2 versus ω (the frequency of the added field) we get as in Figure 7.3 the result

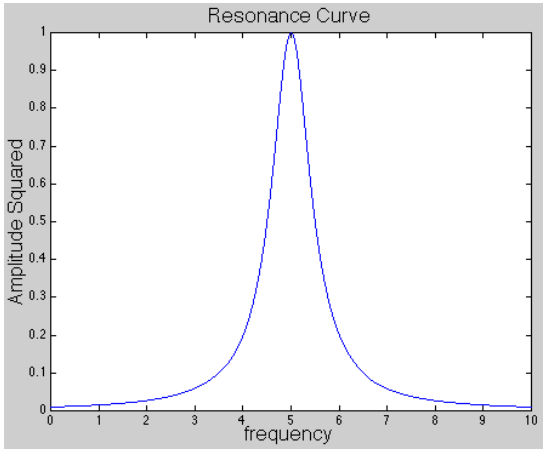


Figure 7.3: Resonance Curve

The peak occurs at $\omega - \omega_0 \approx 0$ (called a *resonance*). Therefore, if $\omega - \omega_0 \gg \omega_1$ (called *off-resonance*), the maximum probability for spin flip is small(that corresponds to the Figure 9.2). However, if $\omega - \omega_0 \approx 0$, which corresponds to resonance, the maximum probability ≈ 1 and the spin has flipped with certainty. The spin system preferentially absorbs energy (flipping spin) near resonance.

This spin resonance process is used to determine a wide variety of spin properties of systems.

7.5 Addition of Angular Momentum

A derivation of the addition process for arbitrary angular momentum values is very complex. We can, however, learn and understand all of the required steps within the context of a special case, namely, combining two spin = 1/2 systems into a new system. We will do the general derivation after the special case.

7.5.1 Addition of Two Spin = 1/2 Angular Momenta

We define

$$\vec{S}_{1,op} = \text{ spin operator for system 1} \tag{7.310}$$

$$\vec{S}_{2,op} = \text{ spin operator for system 2} \tag{7.311}$$

The operators for system 1 are assumed to be independent of the operators for system 2, which implies that

$$\left[\hat{S}_{1i}, \hat{S}_{2j}\right] = 0 \text{ for all } i \text{ and } j \tag{7.312}$$

We characterize each system by eigenvalue/eigenvector equations

$$\vec{S}^2_{1,op}\,|s_1, m_1\rangle = \hbar^2 s_1(s_1+1)\,|s_1, m_1\rangle \tag{7.313}$$

$$\hat{S}_{1,z}\,|s_1, m_1\rangle = \hbar m_1\,|s_1, m_1\rangle \tag{7.314}$$

$$\vec{S}^2_{2,op}\,|s_2, m_2\rangle = \hbar^2 s_2(s_2+1)\,|s_2, m_2\rangle \tag{7.315}$$

$$\hat{S}_{2,z}\,|s_2, m_2\rangle = \hbar m_2\,|s_2, m_2\rangle \tag{7.316}$$

where

$$s_1 = s_2 = \frac{1}{2} \text{ and } m_1 = \pm\frac{1}{2}\,,\; m_2 = \pm\frac{1}{2} \tag{7.317}$$

Since both s_1 and s_2 are fixed and unchanging during this addition process we will drop them from the arguments and subscripts to lessen the complexity of the equations.

Each space (1 and 2) is 2-dimensional and thus each has two ($2s+1=2$) basis states corresponding to the number of m-values in each case

$$\left|s_1 = \frac{1}{2}, m_1 = \frac{1}{2}\right\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle_1 = |\uparrow\rangle_1 = |+\rangle_1 \tag{7.318}$$

$$\left|s_1 = \frac{1}{2}, m_1 = -\frac{1}{2}\right\rangle = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle_1 = |\downarrow\rangle_1 = |-\rangle_1 \tag{7.319}$$

$$\left|s_2 = \frac{1}{2}, m_2 = \frac{1}{2}\right\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle_2 = |\uparrow\rangle_2 = |+\rangle_2 \tag{7.320}$$

$$\left|s_2 = \frac{1}{2}, m_2 = -\frac{1}{2}\right\rangle = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle_2 = |\downarrow\rangle_2 = |-\rangle_2 \tag{7.321}$$

This means that there are 4 possible basis states for the combined system that we can construct using the direct product procedure. We label them as

$$|\uparrow\uparrow\rangle = |{+}{+}\rangle = |+\rangle_1 \otimes |+\rangle_2 \quad , \quad |\uparrow\downarrow\rangle = |{+}{-}\rangle = |+\rangle_1 \otimes |-\rangle_2 \tag{7.322}$$

$$|\downarrow\uparrow\rangle = |{-}{+}\rangle = |-\rangle_1 \otimes |+\rangle_2 \quad , \quad |\downarrow\downarrow\rangle = |{-}{-}\rangle = |-\rangle_1 \otimes |-\rangle_2 \tag{7.323}$$

so that the first symbol in the combined states corresponds to system 1 and the second symbol to system 2. These are not the only possible basis states as we shall see. The "1" operators operate only on the "1" part of the direct product, for example,

$$\vec{S}^2_{1,op}\,|{+}{-}\rangle = \hbar^2 s_1(s_1+1)\,|{+}{-}\rangle = \frac{3}{4}\hbar^2\,|{+}{-}\rangle \tag{7.324}$$

$$\hat{S}_{1,z}\,|{+}{-}\rangle = \hbar m_1\,|{+}{-}\rangle = \frac{\hbar}{2}\,|{+}{-}\rangle \tag{7.325}$$

$$\vec{S}^2_{2,op}\,|{+}{-}\rangle = \hbar^2 s_2(s_2+1)\,|{+}{-}\rangle = \frac{3}{4}\hbar^2\,|{+}{-}\rangle \tag{7.326}$$

$$\hat{S}_{2,z}\,|{+}{-}\rangle = \hbar m_2\,|{+}{-}\rangle = -\frac{\hbar}{2}\,|{+}{-}\rangle \tag{7.327}$$

The total spin angular momentum operator for the combined system is

$$\vec{S}_{op} = \vec{S}_{1,op} + \vec{S}_{2,op} \tag{7.328}$$

It obeys the same commutation rules as the individual system operators, i.e.,

$$\left[\hat{S}_i, \hat{S}_j\right] = i\hbar\varepsilon_{ijk}\hat{S}_k \tag{7.329}$$

This tells us to look for the same kind of angular momentum eigenstates and eigenvalues

$$\vec{S}^2_{op}\left|s,m\right\rangle = \hbar^2 s(s+1)\left|s,m\right\rangle \tag{7.330}$$

$$\hat{S}_z\left|s,m\right\rangle = \hbar m\left|s,m\right\rangle \tag{7.331}$$

To proceed, we need to derive some relations. Squaring the total spin operator we have

$$\begin{aligned}\vec{S}^2_{op} &= \left(\vec{S}_{1,op} + \vec{S}_{2,op}\right)^2 = \vec{S}^2_{1,op} + \vec{S}^2_{2,op} + 2\vec{S}_{1,op}\cdot\vec{S}_{2,op} \\ &= \frac{3}{4}\hbar^2\hat{I} + \frac{3}{4}\hbar^2\hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + 2\hat{S}_{1x}\hat{S}_{2x} + 2\hat{S}_{1y}\hat{S}_{2y}\end{aligned} \tag{7.332}$$

Now using

$$\hat{S}_{1\pm} = \hat{S}_{1x} \pm i\hat{S}_{1y} \text{ and } \hat{S}_{2\pm} = \hat{S}_{2x} \pm i\hat{S}_{2y} \tag{7.333}$$

we have

$$2\hat{S}_{1x}\hat{S}_{2x} + 2\hat{S}_{1y}\hat{S}_{2y} = \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \tag{7.334}$$

and therefore

$$\vec{S}^2_{op} = \frac{3}{2}\hbar^2\hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \tag{7.335}$$

Now suppose we choose the four states $\left|++\right\rangle, \left|+-\right\rangle, \left|-+\right\rangle, \left|--\right\rangle$ as the orthonormal basis for the 4-dimensional vector space of the combined system. We then ask the following question.

Are these basis states also eigenstates of the spin operators for the combined system?

We have

$$\begin{aligned}\hat{S}_z\left|++\right\rangle &= (\hat{S}_{1z} + \hat{S}_{2z})\left|+\right\rangle_1 \otimes \left|+\right\rangle_2 = (\hat{S}_{1z}\left|+\right\rangle_1) \otimes \left|+\right\rangle_2 + \left|+\right\rangle_1 \otimes (\hat{S}_{2z}\left|+\right\rangle_2) \\ &= (\frac{\hbar}{2}\left|+\right\rangle_1) \otimes \left|+\right\rangle_2 + \left|+\right\rangle_1 \otimes (\frac{\hbar}{2}\left|+\right\rangle_2) = (\frac{\hbar}{2} + \frac{\hbar}{2})\left|+\right\rangle_1 \otimes \left|+\right\rangle_2 \\ &= \hbar\left|++\right\rangle\end{aligned} \tag{7.336}$$

and similarly

$$\hat{S}_z\left|+-\right\rangle = 0 \tag{7.337}$$

$$\hat{S}_z\left|-+\right\rangle = 0 \tag{7.338}$$

$$\hat{S}_z\left|--\right\rangle = -\hbar\left|--\right\rangle \tag{7.339}$$

This says that the direct-product basis states are eigenvectors of (total) $\hat{S}_z$ and that the eigenvalues are $m = 1, 0, -1$ and 0 (a second time).

Since each value s of the total angular momentum of the combined system must have $2s + 1$ associated m-values, these results tell us that the combined system will have

$$s = 1 \rightarrow m = 1, 0, -1 \text{ and } s = 0 \rightarrow m = 0 \tag{7.340}$$

which accounts for all of the four states. What about the $\hat{S}^2_{op}$ operator?

We have

$$\begin{aligned} \vec{S}^2_{op} |{+}{+}\rangle &= \left(\frac{3}{2}\hbar^2 \hat{I} + 2\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \right) |{+}{+}\rangle \\ &= \left(\frac{3}{2}\hbar^2 + 2\frac{\hbar}{2}\frac{\hbar}{2} + \hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+} \right) |{+}{+}\rangle \end{aligned} \tag{7.341}$$

Now using

$$\hat{S}_{1+} |{+}{+}\rangle = 0 = \hat{S}_{+} |{+}{+}\rangle \tag{7.342}$$

we get

$$\vec{S}^2_{op} |{+}{+}\rangle = 2\hbar^2 |{+}{+}\rangle = \hbar^2 1(1+1) |{+}{+}\rangle \tag{7.343}$$

or the state $|{+}{+}\rangle$ is also an eigenstate of $\hat{S}^2_{op}$ with $s = 1$

$$|{+}{+}\rangle = |s = 1, m = 1\rangle = |1, 1\rangle \tag{7.344}$$

Similarly, we have

$$\vec{S}^2_{op} |{-}{-}\rangle = 2\hbar^2 |{-}{-}\rangle = \hbar^2 1(1+1) |{-}{-}\rangle \tag{7.345}$$

or the state $|{-}{-}\rangle$ is also an eigenstate of $\hat{S}^2_{op}$ with $s = 1$

$$|{-}{-}\rangle = |s = 1, m = -1\rangle = |1, -1\rangle \tag{7.346}$$

In the same manner we can show that $|{+}{-}\rangle$ and $|{-}{+}\rangle$ are *not* eigenstates of $\vec{S}^2_{op}$. So the simple direct-product states are not appropriate to describe the combined system if we want to characterize it using

$$\vec{S}^2_{op} \text{ and } \hat{S}_z \tag{7.347}$$

However, since the direct-product states are a complete basis set, we should be able to construct the remaining two eigenstates of $\hat{S}^2_{op}$ and $\hat{S}_z$ as linear combinations of the direct-product states

$$|s, m\rangle = \sum_{m_1, m_2} a_{smm_1m_2} |m_1, m_2\rangle \tag{7.348}$$

where we have left out the s_1 and s_2 dependence in the states and coefficients.

In a formal manner, we can identify the so-called *Clebsch-Gordon coefficients*

$a_{smm_1m_2}$ by using the orthonormality of the direct-product basis states. We have

$$\begin{aligned}\langle m_1', m_2' \mid s, m\rangle &= \sum_{m_1,m_2} a_{smm_1m_2} \langle m_1', m_2' \mid m_1, m_2\rangle \\ &= \sum_{m_1,m_2} a_{smm_1m_2}\delta_{m_1'm_1}\delta_{m_2'm_2} = a_{smm_1'm_2'} \end{aligned} \tag{7.349}$$

where we have used

$$\langle m_1', m_2' \mid m_1, m_2\rangle = \delta_{m_1'm_1}\delta_{m_2'm_2} \tag{7.350}$$

This does not help us actually compute the coefficients because we do not know the states $|s, m\rangle$. A procedure that works in this case and that can be generalized, is the following.

We already found that $|{+}{+}\rangle = |1, 1\rangle$ and $|{-}{-}\rangle = |1, -1\rangle$. Now we define the operators

$$\hat{S}_\pm = \hat{S}_x \pm i\hat{S}_y = (\hat{S}_{1x} + \hat{S}_{2x}) \pm i(\hat{S}_{1y} + \hat{S}_{2y}) = \hat{S}_{1\pm} + \hat{S}_{2\pm} \tag{7.351}$$

These are the raising and lowering operators for the combined system and thus they satisfy the relations

$$\hat{S}_\pm |s, m\rangle = \hbar\sqrt{s(s+1) - m(m\pm 1)}\,|s, m\pm 1\rangle \tag{7.352}$$

This gives

$$\hat{S}_+ |1, 1\rangle = 0 = \hat{S}_- |1, -1\rangle \tag{7.353}$$

as we expect. If, however, we apply $\hat{S}_-$ to the topmost state (*maximum s and m values*) we get

$$\begin{aligned}\hat{S}_- |1, 1\rangle &= \hbar\sqrt{1(1+1) - 1(1-1)}\,|1, 0\rangle = \hbar\sqrt{2}\,|1, 0\rangle \\ &= (\hat{S}_{1-} + \hat{S}_{2-})\,|+\rangle_1 \otimes |+\rangle_2 = (\hat{S}_{1-}\,|+\rangle_1) \otimes |+\rangle_2 + |+\rangle_1 \otimes (\hat{S}_{2-}\,|+\rangle_2) \\ &= \hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1) - \frac{1}{2}(\frac{1}{2}-1)}\,|-\rangle_1 \otimes |+\rangle_2 \\ &\qquad + \hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1) - \frac{1}{2}(\frac{1}{2}-1)}\,|+\rangle_1 \otimes |-\rangle_2 \\ &= \hbar\,|{-}{+}\rangle + \hbar\,|{+}{-}\rangle \end{aligned} \tag{7.354}$$

or

$$|1, 0\rangle = |s = 1, m = 0\rangle = \frac{1}{\sqrt{2}}\,|{-}{+}\rangle + \frac{1}{\sqrt{2}}\,|{+}{-}\rangle \tag{7.355}$$

Note that the only terms that appear on the right hand side are those that have $m = m_1 + m_2$. We can easily see that this is a general property since

$$\begin{aligned}\hat{S}_z |s, m\rangle = m\hbar\,|s, m\rangle &= \sum_{m_1,m_2} a_{smm_1m_2}(\hat{S}_{1z} + \hat{S}_{2z})\,|m_1, m_2\rangle \\ &= \hbar \sum_{m_1,m_2} a_{smm_1m_2}(m_1 + m_2)\,|m_1, m_2\rangle \end{aligned} \tag{7.356}$$

The only way to satisfy this equation is for $m = m_1 + m_2$ in *every term* in the sum. Thus, our linear combination should really be written as a single sum of the form

$$|s,m\rangle = \sum_{\substack{m_1,m_2 \\ m_1+m_2=m}} a_{smm_1m_2}\,|m_1,m_2\rangle = \sum_{m_1} a_{smm_1,m-m_1}\,|m_1,m-m_1\rangle \tag{7.357}$$

These three states

$$|1,1\rangle = |{+}{+}\rangle \to a_{1,1,\frac{1}{2},\frac{1}{2}} = 1 \tag{7.358}$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}(|{-}{+}\rangle + |{+}{-}\rangle) \to a_{1,1,-\frac{1}{2},\frac{1}{2}} = a_{1,1,\frac{1}{2},-\frac{1}{2}} = \frac{1}{\sqrt{2}} \tag{7.359}$$

$$|1,-1\rangle = |{-}{-}\rangle \to a_{1,1,-\frac{1}{2},-\frac{1}{2}} = 1 \tag{7.360}$$

are called a *triplet*.

The state $|s=0,m=0\rangle = |0,0\rangle$ can then be found as follows: the state has $m = 0$, $\hat{S}_z\,|0,0\rangle = 0$ and thus each state(term) in the linear combination must have $m = m_1 + m_2 = 0$, which means we must be able to write

$$|0,0\rangle = a\,|{+}{-}\rangle + b\,|{-}{+}\rangle \tag{7.361}$$

where

$$|a|^2 + |b|^2 = 1 \text{ (state is normalized to 1)} \tag{7.362}$$

Now we must also have

$$\langle 1,0\,|\,0,0\rangle = 0 \text{ (since the } |s,m\rangle \text{ states are orthogonal)} \tag{7.363}$$

which implies that

$$\frac{1}{\sqrt{2}}a + \frac{1}{\sqrt{2}}b = 0 \to b = -a \tag{7.364}$$

We then have

$$2\,|a|^2 = 1 \to a = \frac{1}{\sqrt{2}} = -b \tag{7.365}$$

and

$$|0,0\rangle = |s=0,m=0\rangle = \frac{1}{\sqrt{2}}\,|{+}{-}\rangle - \frac{1}{\sqrt{2}}\,|{-}{+}\rangle \tag{7.366}$$

This is called a *singlet state*. That completes the construction of the angular momentum states for the combined system of two spin-1/2 systems.

We now *generalize* this procedure for the addition of any two angular momenta.

7.5.2 General Addition of Two Angular Momenta

Given two angular momenta $\vec{J}_{1,op}$ and $\vec{J}_{2,op}$, we have the operators and states for each separate system

$$\vec{J}_{1,op} \to \vec{J}^2_{1,op}, \hat{J}_{1z}, \hat{J}_{1\pm} \to |j_1, m_1\rangle$$
$$\vec{J}_{2,op} \to \vec{J}^2_{2,op}, \hat{J}_{2z}, \hat{J}_{2\pm} \to |j_2, m_2\rangle$$

with

$$\vec{J}^2_{1,op}|j_1, m_1\rangle = \hbar^2 j_1(j_1+1)|j_1, m_1\rangle, \hat{J}_{1z}|j_1, m_1\rangle = \hbar m_1|j_1, m_1\rangle \tag{7.367}$$
$$\hat{J}_{1\pm}|j_1, m_1\rangle = \hbar\sqrt{j_1(j_1+1) - m_1(m_1\pm 1)}|j_1, m_1\pm 1\rangle \tag{7.368}$$
$$\vec{J}^2_{2,op}|j_2, m_2\rangle = \hbar^2 j_2(j_2+1)|j_2, m_2\rangle, \hat{J}_{2z}|j_2, m_2\rangle = \hbar m_2|j_2, m_2\rangle \tag{7.369}$$
$$\hat{J}_{2\pm}|j_2, m_2\rangle = \hbar\sqrt{j_2(j_2+1) - m_2(m_2\pm 1)}|j_2, m_2\pm 1\rangle \tag{7.370}$$

Remember that there are $2j_1 + 1$ possible m_1 values and $2j_2 + 1$ possible m_2 values.

Since all of the "1" operators commute with all of the "2" operators, we can find a common eigenbasis for the four operators $\vec{J}^2_{1,op}, \hat{J}_{1z}, \vec{J}^2_{2,op}, \hat{J}_{2z}$ in terms of the direct-product states

$$|j_1, j_2, m_1, m_2\rangle = |j_1, m_1\rangle \otimes |j_2, m_2\rangle \tag{7.371}$$

For the combined system we define total operators as before

$$\vec{J}_{op} = \vec{J}_{1,op} + \vec{J}_{2,op} = \text{ total angular momentum} \tag{7.372}$$
$$\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}, \left[\vec{J}^2_{op}, \hat{J}_z\right] = 0 \tag{7.373}$$
$$\left[\vec{J}^2_{op}, \hat{J}^2_{1,op}\right] = 0 \quad , \quad \left[\vec{J}^2_{op}, \hat{J}^2_{2,op}\right] = 0 \tag{7.374}$$
$$\left[\hat{J}^2_{1,op}, \hat{J}_z\right] = 0 \quad , \quad \left[\hat{J}^2_{2,op}, \hat{J}_z\right] = 0 \tag{7.375}$$

These commutators imply that we can construct a common eigenbasis of $\vec{J}^2_{1,op}$, $\hat{J}_{1z}$,$\vec{J}^2_{2,op}$,$\hat{J}_{2z}$ using the states $|j_1, j_2, m_1, m_2\rangle$ where

$$\vec{J}^2_{op}|j, m\rangle = \hbar^2 j(j+1)|j, m\rangle \text{ and } \hat{J}_z|j, m\rangle = \hbar m|j, m\rangle \tag{7.376}$$

There are $2j + 1$ possible m values for each allowed j value. We cannot use the operators $\hat{J}_{1z}, \hat{J}_{2z}$ to construct the eigenbasis for the combined system because they do not commute with $\vec{J}^2_{op}$.

Remember that in order for a label to appear in a ket vector it must be one of the eigenvalues of a set of commuting observables since only such a set shares a common eigenbasis.

We now determine how to write the $|j_1, j_2, m_1, m_2\rangle$ in terms of the $|j_1, m_1\rangle \otimes |j_2, m_2\rangle$ basis. We have

$$\begin{aligned} &|j_1, j_2, j, m\rangle \\ &\quad = \sum_{j_1'} \sum_{j_2'} \sum_{m_1} \sum_{m_2} |j_1', j_2', m_1, m_2\rangle \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle \end{aligned} \tag{7.377}$$

where

$$\langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle = \text{ Clebsch - Gordon coefficients} \tag{7.378}$$

This corresponds to inserting an identity operator of the form

$$\hat{I} = \sum_{j_1'} \sum_{j_2'} \sum_{m_1} \sum_{m_2} |j_1', j_2', m_1, m_2\rangle \langle j_1', j_2', m_1, m_2| \tag{7.379}$$

Since

$$\begin{aligned} &\langle j_1', j_2', m_1, m_2| \vec{J}_{1,op}^{\,2} |j_1, j_2, j, m\rangle \\ &\qquad = \hbar^2 j_1'(j_1'+1) \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle \\ &\qquad = \hbar^2 j_1(j_1+1) \langle j_1', j_2', m_1, m_2 \mid j_1, j_2, j, m\rangle \end{aligned}$$

the Clebsch-Gordon(CG) coefficients must vanish unless $j_1' = j_1$ and similarly unless $j_2' = j_2$. Thus we have

$$|j_1, j_2, j, m\rangle = \sum_{m_1} \sum_{m_2} |j_1, j_2, m_1, m_2\rangle \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle \tag{7.380}$$

Also, as we saw earlier, since $\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}$ we must have

$$\begin{aligned} \langle j_1, j_2, m_1, m_2| \hat{J}_z |j_1, j_2, j, m\rangle &= \hbar(m_1+m_2) \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle \\ &= \hbar m \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle \end{aligned}$$

which implies that the CG coefficients must vanish unless $m_1 + m_2 = m$. Thus the only non-vanishing coefficients are

$$\langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m = m_1 + m_2\rangle \tag{7.381}$$

and we can write

$$\begin{aligned} |j_1, j_2, j, m\rangle &= \sum_{m_1} \sum_{m_2 = m - m_1} |j_1, j_2, m_1, m_2\rangle \langle j_1, j_2, m_1, m_2 \mid j_1, j_2, j, m\rangle \\ &= \sum_{m_1} |j_1, j_2, m_1, m_2 = m - m_1\rangle \langle j_1, j_2, m_1, m_2 = m - m_1 \mid j_1, j_2, j, m\rangle \end{aligned}$$

For fixed j_1 and j_2 there are $2j_1 + 1$ possible m_1 values and $2j_2 + 1$ possible m_2 values. Thus, there are $(2j_1+1)(2j_2+1)$ linearly independent states of the form

$$|j_1, j_2, m_1, m_2\rangle = |j_1, m_1\rangle \otimes |j_2, m_2\rangle$$

and hence the vector space describing the combined system is $(2j_1+1)(2j_2+1)$-dimensional.

This says that there must be $(2j_1+1)(2j_2+1)$ states of the form $|j_1, j_2, j, m\rangle$ also.

We notice that there is only *one* state with $m = m_1 + m_2 = j_1 + j_2$, namely,

$$m_1 = j_1 \text{ and } m_2 = j_2 \tag{7.382}$$

This state has the *maximum possible m value.*

There are two states with $m = m_1 + m_2 = j_1 + j_2 - 1$, namely,

$$m_1 = j_1 - 1, m_2 = j_2 \text{ and } m_1 = j_1, m_2 = j_2 - 1 \tag{7.383}$$

and so on.

For *example*

$$j_1 = 2\ ,\ j_2 = 1 \rightarrow (2(2)+1))(2(1)+1)) = 15 \text{ states} \tag{7.384}$$

If we label these states by the m-values only (since j_1 and j_2 do not change) or $|m_1, m_2\rangle$, then we have(in this example)

$$\begin{aligned}
&m = 3 \rightarrow 1 \text{ state} \rightarrow |2,1\rangle \\
&m = 2 \rightarrow 2 \text{ states} \rightarrow |2,0\rangle\ ,\ |1,1\rangle \\
&m = 1 \rightarrow 3 \text{ states} \rightarrow |1,0\rangle\ ,\ |0,1\rangle\ ,\ |2,-1\rangle \\
&m = 0 \rightarrow 3 \text{ states} \rightarrow |0,0\rangle\ ,\ |1,-1\rangle\ ,\ |-1,1\rangle \\
&m = -1 \rightarrow 3 \text{ states} \rightarrow |-1,0\rangle\ ,\ |0,-1\rangle\ ,\ |-2,1\rangle \\
&m = -2 \rightarrow 2 \text{ states} \rightarrow |-2,0\rangle\ ,\ |-1,-1\rangle \\
&m = -3 \rightarrow 1 \text{ state} \rightarrow |-2,-1\rangle
\end{aligned}$$

for a total of 15 states. The combined system, as we shall see by construction, has these states

$$\begin{aligned}
&j = 3 \rightarrow m = 3,2,1,0,-1,-2,-3 \rightarrow 7 \text{ states} \\
&j = 2 \rightarrow m = 2,1,0,-1,-2 \rightarrow 5 \text{ states} \\
&j = 1 \rightarrow m = 1,0,-1 \rightarrow 3 \text{ states}
\end{aligned}$$

for a total of 15 states.

The general rules, which follows from group theory, are

1. The combined system has allowed j values given by

$$j_1 + j_2 \geq j \geq |j_1 - j_2| \text{ in integer steps} \tag{7.385}$$

2. The total number of states is given by the total number of m-values for all the allowed j-values or

$$\sum_{j=|j_1-j_2|}^{j_1+j_2}(2j+1) = (2j_1+1)(2j_2+1) \tag{7.386}$$

We write the addition of two angular momenta symbolically as

$$j_1 \otimes j_2 = |j_1 - j_2| \oplus |j_1 - j_2| + 1 \oplus |j_1 - j_2| + 2 \oplus \ldots\ldots \oplus j_1 + j_2 - 1 \oplus j_1 + j_2 \tag{7.387}$$

Examples

Our original special case of adding two spin = 1/2 systems gives

$$j_1 = j_2 = \frac{1}{2} \rightarrow j = 0, 1 \rightarrow \frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1 \rightarrow 4 \text{ states} \tag{7.388}$$

which is the result we found earlier.

Other cases are:

$j_1 = j_2 = 1 \rightarrow j = 0, 1, 2 \rightarrow 1 \otimes 1 = 0 \oplus 1 \oplus 2 \rightarrow 9$ states
$j_1 = 2, j_2 = 1 \rightarrow j = 1, 2, 3 \rightarrow 2 \otimes 1 = 1 \oplus 2 \oplus 3 \rightarrow 15$ states
$j_1 = 2, j_2 = 3 \rightarrow j = 1, 2, 3, 4, 5 \rightarrow 2 \otimes 3 = 1 \oplus 2 \oplus 3 \oplus 4 \oplus 5 \rightarrow 35$ states

7.5.3 Actual Construction of States

Notation

1. states labeled $|7, 6\rangle$ are $|j, m\rangle$ states
2. states labeled $|3, 2\rangle_{\otimes}$ are $|m_1, m_2\rangle$ states
3. we suppress the j_1, j_2 labels everywhere

Procedure

1. choose j_1 and j_2
2. write down the direct-product $|m_1, m_2\rangle$ basis
3. determine the allowed j values
4. write down the maximum m state ($m = j_1 + j_2$); it is unique
5. the maximum m-value corresponds to the maximum j-value
6. use the lowering operator to generate all other m-states for this J-value; there are $2j + 1$, i.e.,

$$\hat{J}_- |j, m\rangle = \hbar\sqrt{j(j+1) - m(m-1)}\,|j, m-1\rangle$$

7. find the maximum m-state for the next lowest j-value; it is constructed from the same basis states as in the corresponding m-states for higher j-values; use orthonormality properties to figure out coefficients

8. repeat (6) and (7) until all j-values have been dealt with

More Detailed Examples (we must learn this process by doing it)

#1 - Individual system values

$$j_1 = j_2 = \frac{1}{2} \to m_1, m_2 = \frac{1}{2}, -\frac{1}{2} \text{ (we already did this example)} \tag{7.389}$$

The basis states are (we use the notation $|+-\rangle$ here instead of $|1/2, -1/2\rangle_{\otimes}$)

$$\begin{array}{rcccccc} & |++\rangle & , & |+-\rangle & , & |-+\rangle & , & |--\rangle \\ m = m_1 + m_2 = & 1 & & 0 & & -1 & & 0 \end{array}$$

Construction Algebra

Allowed j-values are $j = 1, 0$

$j = 1$ has $2j + 1 = 3$ m-values $= 1, 0, -1$

$j = 0$ has $2j + 1 = 1$ m value $= 0$

$|1, 1\rangle = |++\rangle$ maximum or topmost (j, m) state is always unique

$\hat{J}_- |1, 1\rangle = \sqrt{2}\hbar |1, 0\rangle = (\hat{J}_{1-} + \hat{J}_{2-}) |++\rangle = \hbar |+-\rangle + \hbar |-+\rangle$

$|1, 0\rangle = \frac{1}{\sqrt{2}} |+-\rangle + \frac{1}{\sqrt{2}} |-+\rangle$

$\hat{J}_- |1, 0\rangle = \sqrt{2}\hbar |1, -1\rangle = (\hat{J}_{1-} + \hat{J}_{2-}) \left(\frac{1}{\sqrt{2}} |+-\rangle + \frac{1}{\sqrt{2}} |-+\rangle \right) = \sqrt{2}\hbar |--\rangle$

$|1, -1\rangle = |--\rangle$

We must now have $|0, 0\rangle = a |+-\rangle + b |-+\rangle$ (Rule 7) with

$$|a|^2 + |b|^2 = 1 \text{ and } \langle 1, 0 | , 0, 0 \rangle = \frac{1}{\sqrt{2}} a + \frac{1}{\sqrt{2}} b = 0$$

which gives

$$a = -b = \frac{1}{\sqrt{2}}$$

or

$|0, 0\rangle = \frac{1}{\sqrt{2}} |+-\rangle - \frac{1}{\sqrt{2}} |-+\rangle$

All the $j-$values are now done. We end up with the Clebsch-Gordon coefficients.

#2 - Individual system values

$$j_1 = j_2 = 1 \rightarrow m_1, m_2 = 1, 0, -1 \tag{7.390}$$

The basis states are

$$\begin{array}{lccccc} & |1,1\rangle_\otimes & |1,0\rangle_\otimes & |1,-1\rangle_\otimes & |0,1\rangle_\otimes & |0,0\rangle_\otimes \\ m = m_1 + m_2 = & 2 & 1 & 0 & 1 & 0 \end{array}$$

$$\begin{array}{lcccc} & |0,-1\rangle_\otimes & |-1,1\rangle_\otimes & |-1,0\rangle_\otimes & |-1,-1\rangle_\otimes \\ m = m_1 + m_2 = & -1 & 0 & -1 & -2 \end{array}$$

Construction Algebra

Allowed j-values are $j = 2, 1, 0$

$j = 2$ has $2j + 1 = 3$ m-values $= 2, 1, 0, -1, -2$

$j = 1$ has $2j + 1 = 3$ m-values $= 1, 0, -1$

$j = 0$ has $2j + 1 = 1$ m value $= 0$

$|2,2\rangle = |1,1\rangle_\otimes$ maximum or topmost (j, m) state is always unique

$\hat{J}_- |2,2\rangle = 2\hbar |2,1\rangle = (\hat{J}_{1-} + \hat{J}_{2-}) |1,1\rangle_\otimes = \sqrt{2}\hbar |1,0\rangle_\otimes + \sqrt{2}\hbar |0,1\rangle_\otimes$

$|2,1\rangle = \frac{1}{\sqrt{2}} |1,0\rangle_\otimes + \frac{1}{\sqrt{2}} |0,1\rangle_\otimes$

$$\begin{aligned} \hat{J}_- |2,1\rangle = \sqrt{6}\hbar |2,0\rangle = (\hat{J}_{1-} + \hat{J}_{2-}) \left(\tfrac{1}{\sqrt{2}} |1,0\rangle_\otimes + \tfrac{1}{\sqrt{2}} |0,1\rangle_\otimes \right) \\ = \hbar |1,-1\rangle_\otimes + 2\hbar |0,0\rangle_\otimes + \hbar |-1,1\rangle_\otimes \end{aligned}$$

$|2,0\rangle = \frac{1}{\sqrt{6}} |1,-1\rangle_\otimes + \frac{2}{\sqrt{6}} |0,0\rangle_\otimes + \frac{1}{\sqrt{6}} |-1,1\rangle_\otimes$

Continuing we have

$$\begin{aligned} |2,-1\rangle &= \frac{1}{\sqrt{2}} |-1,0\rangle_\otimes + \frac{1}{\sqrt{2}} |0,-1\rangle_\otimes \\ |2,-2\rangle &= |-1,-1\rangle_\otimes \end{aligned}$$

which completes the five $j = 2$ states.

We must now have $|1,1\rangle = a |1,0\rangle_\otimes + b |0,1\rangle_\otimes$ with

$$|a|^2 + |b|^2 = 1 \text{ and } \langle 2,1\,|\,,1,1\rangle = \tfrac{1}{\sqrt{2}}a + \tfrac{1}{\sqrt{2}}b = 0$$

which gives

$$a = -b = \tfrac{1}{\sqrt{2}}$$

or

$$|1,1\rangle = \tfrac{1}{\sqrt{2}}\,|1,0\rangle_\otimes - \tfrac{1}{\sqrt{2}}\,|0,1\rangle_\otimes$$

We now find all of the $j = 1$ states

$$\begin{aligned}\hat{J}_-\,|1,1\rangle = \sqrt{2}\hbar\,|1,0\rangle &= (\hat{J}_{1-} + \hat{J}_{2-})(\frac{1}{\sqrt{2}}\,|1,0\rangle_\otimes - \frac{1}{\sqrt{2}}\,|0,1\rangle_\otimes)\\ &= \hbar\,|1,-1\rangle_\otimes - \hbar\,|-1,1\rangle_\otimes\\ |1,0\rangle &= \frac{1}{\sqrt{2}}\,|1,-1\rangle_\otimes - \frac{1}{\sqrt{2}}\,|-1,1\rangle_\otimes\end{aligned}$$

and continuing

$$|1,-1\rangle = \frac{1}{\sqrt{2}}\,|0,-1\rangle_\otimes - \frac{1}{\sqrt{2}}\,|-1,0\rangle_\otimes$$

which completes the three $j = 1$ states.

We must now have $|0,0\rangle = a\,|1,-1\rangle_\otimes + b\,|0,0\rangle_\otimes + c\,|-1,1\rangle_\otimes$ with

$$\begin{aligned}&|a|^2 + |b|^2 = 1 \text{ and } \langle 2,0\,|\,,0,0\rangle = \tfrac{1}{\sqrt{6}}a + \tfrac{2}{\sqrt{6}}b + \tfrac{1}{\sqrt{6}}c = 0\\ &\text{and } \langle 1,0\,|\,,0,0\rangle = \tfrac{1}{\sqrt{6}}a - \tfrac{1}{\sqrt{6}}c = 0\end{aligned}$$

which gives

$$a = -b = c = \tfrac{1}{\sqrt{3}}$$

or

$$|0,0\rangle = \tfrac{1}{\sqrt{3}}\,|1,-1\rangle_\otimes - \tfrac{1}{\sqrt{3}}\,|0,0\rangle_\otimes + \tfrac{1}{\sqrt{3}}\,|-1,1\rangle_\otimes$$

All the j-values are now done.

7.6 Two- and Three-Dimensional Systems

We now turn our attention to 2– and 3–dimensional systems that can be solved analytically.

In the position representation, the wave function

$$\psi(\vec{r}) = \langle \vec{r}\,|\,\psi\rangle \tag{7.391}$$

contains all the information that is knowable about the state of a physical system. The equation that determines the wave function is the Schrodinger equation, which we derived from the energy eigenvalue equation

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle \tag{7.392}$$

The general form of the Schrodinger equation in three dimensions is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\vec{r}) + V(\vec{r})\psi(\vec{r}) = E\psi(\vec{r}) \tag{7.393}$$

We will use a succession of concrete examples to elucidate the solution techniques and the physical principles that are involved.

7.6.1 2- and 3-Dimensional Infinite Wells

2-Dimensional Infinite Square Well - Cartesian Coordinates

The potential energy function is

$$V(x,y) = \begin{cases} 0 & |x| < \frac{a}{2} \text{ and } |y| < \frac{a}{2} \rightarrow \text{ region I} \\ \infty & \text{otherwise} \rightarrow \text{ region II} \end{cases} \tag{7.394}$$

This is a simple extension of the 1−dimensional infinite well problem, but it is useful because it illustrates all the ideas we will need for more complicated problems.

In region I

$$\frac{\partial^2\psi(x,y)}{\partial x^2} + \frac{\partial^2\psi(x,y)}{\partial y^2} = -\frac{2mE}{\hbar^2}\psi(x,y) \tag{7.395}$$

In region II

$$\psi(x,y) = 0 \tag{7.396}$$

since the potential is infinite over an extended region.

We solve this equation by the separation of variables (SOV) technique. We assume

$$\psi(x,y) = X(x)Y(y) \tag{7.397}$$

Upon substitution we get

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} = -\frac{2mE}{\hbar^2} \tag{7.398}$$

Each term on the left-hand side of the equation is a function only of a single variable and hence the only way to satisfy the equation for all x and y is to set

each function of a single variable equal to a constant. We have

$$\frac{1}{X}\frac{d^2X}{dx^2} = -\frac{2mE_x}{\hbar^2} = \text{ constant} \tag{7.399}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -\frac{2mE_y}{\hbar^2} = \text{ constant} \tag{7.400}$$

$$E = E_x + E_y \tag{7.401}$$

If we choose

$$\frac{2mE_x}{\hbar^2} = k_x^2 \quad , \quad \frac{2mE_y}{\hbar^2} = k_y^2 \quad , \quad \frac{2mE}{\hbar^2} = k^2 \tag{7.402}$$

we get the solutions

$$X(x) = A\sin k_x x + B\cos k_x x \tag{7.403}$$

$$Y(y) = C\sin k_y y + D\cos k_y y \tag{7.404}$$

with the boundary conditions

$$X\left(\pm\frac{a}{2}\right) = 0 = Y\left(\pm\frac{a}{2}\right) \tag{7.405}$$

For the function $X(x)$ each boundary condition implies two solution types

$$A = 0 \text{ or } \sin\frac{k_x a}{2} = 0 \text{ and } B = 0 \text{ or } \cos\frac{k_x a}{2} = 0 \tag{7.406}$$

These are both summarized in the solution

$$X(x) = \begin{cases} \sin\frac{n_x\pi x}{a} & n_x = \text{ even} \\ & k_x = \frac{n_x\pi}{a} \quad n_x = \text{ integer} \\ \cos\frac{n_x\pi x}{a} & n_x = \text{ odd} \end{cases} \tag{7.407}$$

and similarly for $Y(y)$

$$Y(y) = \begin{cases} \sin\frac{n_y\pi y}{a} & n_y = \text{ even} \\ & k_y = \frac{n_y\pi}{a} \quad n_y = \text{ integer} \\ \cos\frac{n_y\pi y}{a} & n_y = \text{ odd} \end{cases} \tag{7.408}$$

The corresponding energy eigenvalues are

$$E = (n_x^2 + n_y^2)\frac{\pi^2\hbar^2}{2ma^2} \quad , \quad n_{\mathrm{x}}, n_{\mathrm{y}} = 1, 2, 3, 4\ldots \tag{7.409}$$

The 1–dimensional result we found earlier was

$$E_{1\,\mathrm{dim}} = \frac{n^2\pi^2\hbar^2}{2ma^2} \quad , n = 1, 2, 3, 4\ldots \tag{7.410}$$

A plot of these levels for comparison is shown in Figure 7.4 below; we choose $\frac{\pi^2\hbar^2}{2ma^2} = 1$

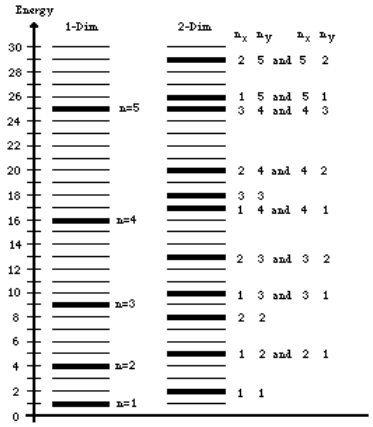


Figure 7.4: Comparison of 1D and 2D Infinite Wells

The major change is not only that the level structure gets more complex, but also that a major new feature appears, namely, *degeneracy.*

Several of the 2–dimensional well levels are degenerate, which means that *different* sets of quantum numbers give the *same energy eigenvalue.*

In the energy level diagram above the $E = 5, 10, 13, 17, 20, 25, 26$ and 29 levels are all two-fold degenerate. This degeneracy arises from the fact that $V(x, y) = V(-x, -y)$ and hence parity is conserved. This means that the correct physical eigenstates should be simultaneous eigenstates of both the parity and the energy.

The first three wave functions are

$$\psi_{11}(x, y) = \cos\frac{\pi x}{a}\cos\frac{\pi y}{a} \quad , \quad E = E_{11} \tag{7.411}$$

$$\psi_{12}(x, y) = \cos\frac{\pi x}{a}\sin\frac{2\pi y}{a} \quad , \quad E = E_{12} \tag{7.412}$$

$$\psi_{21}(x, y) = \sin\frac{2\pi x}{a}\cos\frac{\pi y}{a} \quad , \quad E = E_{21} = E_{12} \tag{7.413}$$

A simple calculation shows that we have

$$\langle 11 \mid 12\rangle = \langle 11 \mid 21\rangle = 0 \rightarrow \text{ orthogonal and } \langle 21 \mid 12\rangle \neq 0 \rightarrow \text{ not orthogonal} \tag{7.414}$$

We can construct two new eigenfunctions from the degenerate pair that are orthogonal using the Gram-Schmidt process. We get

$$\psi^{+}_{12} = \psi_{12} + \psi_{21} \text{ and } \psi^{-}_{12} = \psi_{12} - \psi_{21} \tag{7.415}$$

We then have

$$\langle 12+ \mid 12-\rangle = 0 \rightarrow \text{ orthogonal} \tag{7.416}$$

We also note that for the parity operator $\wp$ we have

$$\hat{\wp}\,|11\rangle = |11\rangle \quad , \quad \hat{\wp}\,|12+\rangle = |12+\rangle \quad , \quad \hat{\wp}\,|12-\rangle = -\,|12-\rangle \tag{7.417}$$

so that the new eigenfunctions are *simultaneous eigenstates of parity and energy.*

Two of the wave functions are plotted below in Figure 7.5 and 7.6 as $|\psi|^2$.

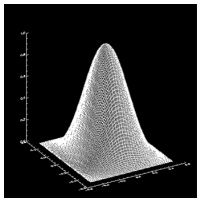


Figure 7.5: $|\psi_{11}|^2$

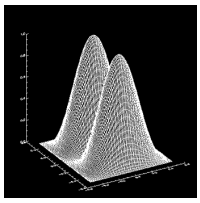


Figure 7.6: $|\psi^+_{12}|^2$

Now let us turn to the infinite circular well in two dimensions.

7.6.2 Two-Dimensional Infinite Circular Well

We consider the potential in two dimensions

$$V(r) = \begin{cases} 0 & r < a \\ \infty & r > a \end{cases} \tag{7.418}$$

The Schrodinger equation in plane-polar coordinates is

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right]\psi(r,\varphi) = E\psi(r,\varphi) \qquad r < a \tag{7.419}$$

$$\psi(r,\varphi) = 0 \qquad r < a \tag{7.420}$$

We assume(SOV)

$$\psi(r,\varphi) = R(r)\Phi(\varphi) \tag{7.421}$$

which gives

$$-\frac{\hbar^2}{2m}\left[\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}\right] = E \qquad r < a \tag{7.422}$$

We choose a separation constant

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2} = -\alpha^2 \rightarrow \Phi(\varphi) = B\sin(\alpha\varphi+\delta) \tag{7.423}$$

The requirement of single-valuedness under a φ-rotation of 2π says that

$$\begin{aligned}&\sin(\alpha\varphi+\delta) = \sin(\alpha\varphi+\delta+\alpha\pi)\\ &\rightarrow \alpha = \text{ integer } = 0,1,2,3\ldots\end{aligned}$$

Alternatively, we could write

$$\begin{aligned}&\Phi(\varphi) = Be^{i\alpha\varphi}\\ &\alpha = \text{ integer } = \ldots -3,-2,-1,0,1,2,3,\ldots\end{aligned}$$

Substitution of this solution leaves the radial differential equation

$$r^2\frac{d^2R}{dr^2}+r\frac{dR}{dr}+\left[\lambda^2r^2-\alpha^2\right]R = 0 \text{ where } \lambda^2 = \frac{2mE}{\hbar^2} \tag{7.424}$$

This is Bessel's equation. The general solution is

$$R(r) = NJ_\alpha(\lambda r)+MY_\alpha(\lambda r) \tag{7.425}$$

Now $Y_\alpha(\lambda r)\rightarrow\infty$ as $r\rightarrow 0$. Therefore, in order to have a normalizable solution in the region $r < a$ (which includes $r = 0$), we must choose $M = 0$ and thus we have

$$R(r) = NJ_\alpha(\lambda r) \tag{7.426}$$

and the complete solution is then

$$\psi_{k\alpha}(r,\varphi) = R(r)\Phi(\varphi) = NJ_\alpha(\lambda r)e^{i\alpha\varphi} \tag{7.427}$$

The continuity (or boundary) condition at $r = a$ is

$$\psi_{k\alpha}(a,\varphi) = 0 \rightarrow R(a) = 0 \rightarrow J_\alpha(\lambda a) = 0 \tag{7.428}$$

Thus, the allowed values of λ and hence the allowed values of E are given by

$$\lambda_{n\alpha}a = z_{n\alpha} = \text{ the } n^{th} \text{ zero of } J_\alpha \rightarrow E_{n\alpha} = \frac{\hbar^2}{2ma^2}z_{n\alpha}^2 \tag{7.429}$$

We compare the infinite square and circular wells in 2–dimensions using the energy level diagram in Figure 7.7 below.

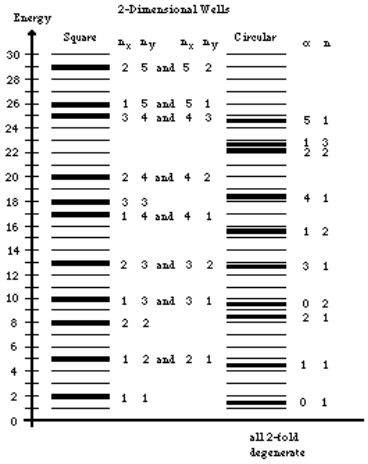


Figure 7.7: Two Dimensional Wells

Note the rather dramatic differences in both the location and degeneracies for the two sets of energy levels. Some of the wave functions for the 2–dimensional circular well are shown in Figures 7.8-7.11 below.

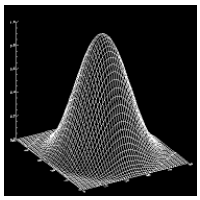


Figure 7.8: $|\psi_{01}|^2$

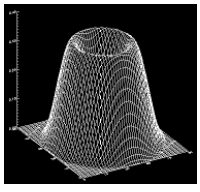


Figure 7.9: $|\psi_{11}|^2$

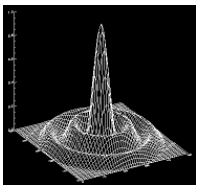


Figure 7.10: $|\psi_{03}|^2$

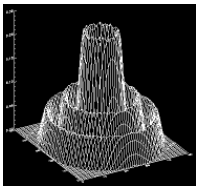


Figure 7.11: $|\psi_{23}|^2$

The 3−*dimensional infinite square well* is a simple extension of the 2−dimensional infinite square well. The result for the energies is

$$E = (n_x^2 + n_y^2 + n_z^2)\frac{\pi^2\hbar^2}{2ma^2} \quad , \quad n_x, n_y, n_z = 1, 2, 3, 4 \ldots \tag{7.430}$$

The 3−*dimensional infinite spherical well* involves the potential

$$V(r,\theta,\varphi) = \begin{cases} 0 & r < a \quad \text{region I} \\ \infty & r > a \quad \text{region II} \end{cases} \tag{7.431}$$

The Schrodinger equation is:

Region I

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} = -\frac{2m}{\hbar^2}E \tag{7.432}$$

Region II

$$\psi(r,\theta,\varphi) = 0 \tag{7.433}$$

The equation in region I can be rewritten in terms of the $\vec{L}_{op}^2$ operator as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) - \frac{\vec{L}_{op}^2\psi}{\hbar^2 r^2} = -\frac{2m}{\hbar^2}E\psi \tag{7.434}$$

Since the potential energy is spherically symmetric, we have

$$\left[\vec{L}_{op}^2, \hat{H}\right] = 0 \tag{7.435}$$

and thus the operators $\vec{L}_{op}^2$ and $\hat{H}$ have a common eigenbasis.

Earlier, we found the eigenfunctions of $\vec{L}_{op}^2$ to be the spherical harmonics $Y_{\ell m}(\theta,\varphi)$, where

$$\vec{L}_{op}^2 Y_{\ell m}(\theta,\varphi) = \hbar^2\ell(\ell+1)Y_{\ell m}(\theta,\varphi) \tag{7.436}$$

Therefore, we can write(SOV)

$$\psi(r,\theta,\varphi) = R_\ell(r)Y_{\ell m}(\theta,\varphi) \tag{7.437}$$

Substitution of this form of the solution gives the radial equation in region I

$$\frac{d^2R_\ell}{dr^2} + \frac{2}{r}\frac{dR_\ell}{dr} - \frac{\ell(\ell+1)}{r^2}R_\ell + k^2R_\ell = 0 \tag{7.438}$$

where

$$E = \frac{h^2k^2}{2m} \tag{7.439}$$

In region II we have $R_\ell(r) = 0$.

The most general solution of the radial equation(which is a different form of Bessel's equation) is

$$R_\ell(r) = Aj_\ell(kr) + B\eta_\ell(kr) \tag{7.440}$$

where

$$j_\ell(kr) = \left(\frac{\pi}{2kr}\right)^{1/2} J_{\ell+1/2}(kr) \tag{7.441}$$

$$\eta_\ell(kr) = (-1)^{\ell+1}\left(\frac{\pi}{2kr}\right)^{1/2} J_{-\ell-1/2}(kr) \tag{7.442}$$

are the spherical Bessel functions.

Now $\eta_\ell(kr) \to \infty$ as $r \to 0$. Therefore, the normalizable solution in region I (which contains $r = 0$) is

$$R_\ell(r) = Aj_\ell(kr) \tag{7.443}$$

The first few of these functions are

$$j_0(x) = \frac{\sin x}{x} \tag{7.444}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \tag{7.445}$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x \tag{7.446}$$

The boundary conditions give

$$j_\ell(k_{n\ell}a) = 0 \text{ where } k_{n\ell} = \text{ the } n^{th} \text{ zero of } j_\ell \tag{7.447}$$

The full solution is

$$\psi_{n\ell m}(\theta,\varphi) = R_{n\ell}(r)Y_{\ell m}(\theta,\varphi) = j_\ell(k_n r)P_\ell^m(\cos\theta)e^{im\varphi} \tag{7.448}$$

Some of the wavefunctions (absolute squares) are shown in Figures 7.12-7.15 below.

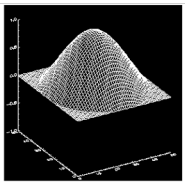


Figure 7.12: $n\ell m = 100$

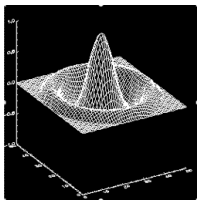


Figure 7.13: $n\ell m = 200$

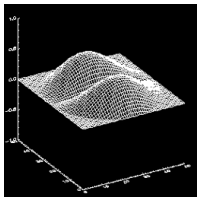


Figure 7.14: $n\ell m = 111$

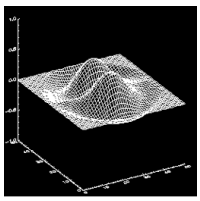


Figure 7.15: $n\ell m = 211$

7.6.3 3-Dimensional Finite Well

We now consider the potential function in three dimensions

$$V(r) = \begin{cases} -V_0 & r \le a \quad \text{region I} \\ 0 & r > a \quad \text{region II} \end{cases} \tag{7.449}$$

The Schrodinger equation is

Region I (same as infinite well)

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2}+\frac{2m}{\hbar^2}V_0\psi=-\frac{2m}{\hbar^2}E\psi \tag{7.450}$$

Region II

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2}=-\frac{2m}{\hbar^2}E\psi \tag{7.451}$$

If we choose $E=-|E|<0$ for bound states, we have

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)-\frac{\vec{L}_{op}^2\psi}{\hbar^2r^2}+\frac{2m}{\hbar^2}V_0\psi=\frac{2m}{\hbar^2}|E|\,\psi \text{ in region I} \tag{7.452}$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)-\frac{\vec{L}_{op}^2\psi}{\hbar^2r^2}=\frac{2m}{\hbar^2}|E|\,\psi \text{ in region II} \tag{7.453}$$

We then write, as before,

$$\psi(\theta,\varphi)=R_\ell(r)Y_{\ell m}(\theta,\varphi) \tag{7.454}$$

and get

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-\frac{\ell(\ell+1)R}{r^2}+\frac{2m}{\hbar^2}V_0R=\frac{2m}{\hbar^2}|E|\,R \quad \text{in region I} \tag{7.455}$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-\frac{\ell(\ell+1)R}{r^2}=\frac{2m}{\hbar^2}|E|\,R \quad \text{in region II} \tag{7.456}$$

which finally becomes

$$\frac{d^2R}{d\rho^2}+\frac{2}{\rho}\frac{dR}{d\rho}+\left[1-\frac{\ell(\ell+1)}{\rho^2}\right]R=0 \tag{7.457}$$

$$\rho=\alpha r \quad , \quad \alpha^2=\frac{2m}{\hbar^2}(V_0-|E|) \quad \text{in region I} \tag{7.458}$$

and

$$\frac{d^2R}{d\gamma^2}+\frac{2}{\gamma}\frac{dR}{d\gamma}+\left[1-\frac{\ell(\ell+1)}{\gamma^2}\right]R=0 \tag{7.459}$$

$$\gamma=i\beta r \quad , \quad \beta^2=\frac{2m}{\hbar^2}|E| \quad \text{in region II} \tag{7.460}$$

The solutions are

$$R(r)=Aj_\ell(\alpha r) \qquad r\le a \tag{7.461}$$

$$R(r)=Bh_\ell^{(1)}(i\beta r)=B\left[j_\ell(i\beta r)+i\eta_\ell(i\beta r)\right] \qquad r>a \tag{7.462}$$

where

$$j_0(x) = \frac{\sin x}{x} \tag{7.463}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \tag{7.464}$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x \tag{7.465}$$

$$j_\ell(x) = x^\ell \left(-\frac{1}{x}\frac{d}{dx}\right)^\ell \frac{\sin x}{x} \tag{7.466}$$

which are *spherical Bessel functions of the* 1^{st} *kind*, and

$$\eta_0(x) = -\frac{\cos x}{x} \tag{7.467}$$

$$\eta_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x} \tag{7.468}$$

$$\eta_2(x) = -\left(\frac{3}{x^3} - \frac{1}{x}\right)\cos x - \frac{3}{x^2}\sin x \tag{7.469}$$

$$\eta_\ell(x) = -x^\ell \left(-\frac{1}{x}\frac{d}{dx}\right)^\ell \frac{\cos x}{x} \tag{7.470}$$

which are *spherical Bessel functions of the* 2^{nd} *kind*, and

$$h_0^{(1)}(ix) = -\frac{1}{x}e^{-x} \tag{7.471}$$

$$h_1^{(1)}(ix) = i\left(\frac{1}{x} + \frac{1}{x^2}\right)e^{-x} \tag{7.472}$$

$$h_2^{(1)}(ix) = \left(\frac{1}{x} + \frac{3}{x^2} + \frac{3}{x^3}\right)e^{-x} \tag{7.473}$$

$$h_\ell^{(1)}(ix) = j_\ell(ix) + i\eta_\ell(ix) \tag{7.474}$$

which are *spherical Hankel functions of the* 1^{st} *kind*.

There is another independent solution for $r > a$, namely,

$$h_\ell^{(2)}(ix) = j_\ell(ix) - i\eta_\ell(ix) \tag{7.475}$$

which is a *spherical Hankel functions of the* 2^{nd} *kind*, but we must exclude it because it behaves like $e^{\beta r}$ as $r \to \infty$ and, hence, is not normalizable.

We have also excluded $\eta_\ell(\alpha r)$ from the solution for $r \le a$ because it diverges at $r = 0$.

We note for future reference that we have the asymptotic behaviors

$$j_\ell(x) \underset{x\to 0}{\to} \frac{x^\ell}{1\cdot 3\cdot 5\cdot\cdots\cdot(2\ell+1)} \text{ and } \eta_\ell(x) \underset{x\to 0}{\to} \frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2\ell-1)}{x^\ell} \tag{7.476}$$

$$j_\ell(x) \underset{x\to\infty}{\to} \frac{1}{x}\cos\left[x - \frac{(\ell+1)\pi}{2}\right] \text{ and } \eta_\ell(x) \underset{x\to\infty}{\to} \frac{1}{x}\sin\left[x - \frac{(\ell+1)\pi}{2}\right] \tag{7.477}$$

$$h_\ell^{(1)}(x) \underset{x\to\infty}{\to} \frac{1}{x}e^{i\left[x-\frac{1}{2}(\ell+1)\pi\right]} \text{ and } h_\ell^{(2)}(x) \underset{x\to\infty}{\to} \frac{1}{x}e^{-i\left[x-\frac{1}{2}(\ell+1)\pi\right]} \tag{7.478}$$

Since both R and dR/dr are continuous at $r = a$, we can combine the two continuity equations into one using the continuity of the so-called logarithmic derivative

$$\frac{1}{R}\frac{dR}{dr} \text{ at } r = a \tag{7.479}$$

For *each* value of ℓ this gives a transcendental equation for the energy E.

Examples:

$$\ell = 0 \qquad \xi\cot\xi = -\zeta \quad \text{and} \quad \xi^2 + \zeta^2 = \frac{2mV_0a^2}{\hbar^2} \tag{7.480}$$
$$\xi = \alpha a \quad \text{and} \quad \zeta = \beta a$$

$$\ell = 1 \qquad \frac{\cot\xi}{\xi} - \frac{1}{\xi^2} = \frac{1}{\zeta} + \frac{1}{\zeta^2}v\xi^2 + \zeta^2 = \frac{2mV_0a^2}{\hbar^2} \tag{7.481}$$
$$\xi = \alpha a \quad \text{and} \quad \zeta = \beta a$$

A graphical solution for the $\ell = 0$ case is shown in Figure 7.16 below.

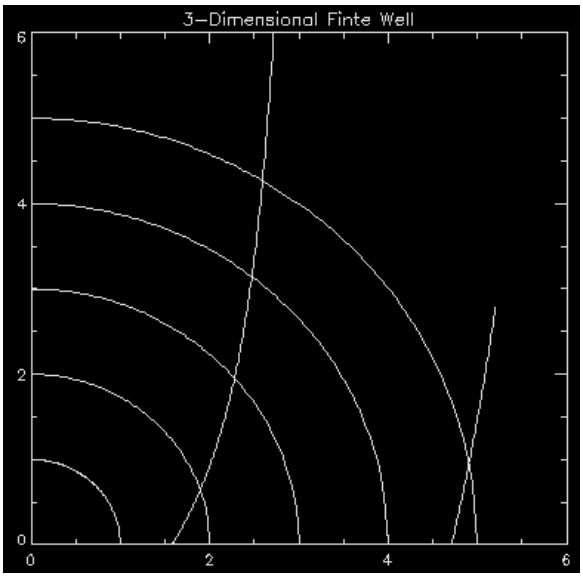


Figure 7.16: Graphical Solution

For a given well, only one circle exists on the plot. The solutions (energy eigenvalues) are given by the intersection of that circle with the cotangent curves.

The big change from the finite well in one dimension is that the quantity

$$\frac{2mV_0a^2}{\hbar^2} \tag{7.482}$$

must be larger than some minimum value before any bound state exists (this corresponds to the radius of the smallest circle that intersects the cotangent curves). In particular,

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{\pi}{2}\right)^2 \rightarrow \text{ no solution} \tag{7.483}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{3\pi}{2}\right)^2 \rightarrow \text{ 1 solution} \tag{7.484}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{5\pi}{2}\right)^2 \rightarrow \text{ 2 solutions} \tag{7.485}$$

7.6.4 Two-Dimensional Harmonic Oscillator

The 2-dimensional harmonic oscillator system has the Hamiltonian

$$\begin{aligned}\hat{H} &= \frac{1}{2m}\left(\hat{p}_x^2+\hat{p}_y^2\right)+\frac{1}{2}m\omega^2(\hat{x}^2+\hat{y}^2)\\ &= \left(\frac{\hat{p}_x^2}{2m}+\frac{1}{2}m\omega^2\hat{x}^2\right)+\left(\frac{\hat{p}_y^2}{2m}+\frac{1}{2}m\omega^2\hat{y}^2\right)\\ &= \hat{H}_x+\hat{H}_y\end{aligned} \tag{7.486}$$

where we have the commutation relations

$$\left[\hat{H}_x,\hat{H}_y\right]=0=\left[\hat{H},\hat{H}_x\right]=\left[\hat{H},\hat{H}_y\right] \tag{7.487}$$

These commutators imply that $\hat{H}$, $\hat{H}_x$ and $\hat{H}_y$ have a common eigenbasis. We label their common state vectors by

$$|E\rangle=|E_x,E_y\rangle=|E_x\rangle\,|E_y\rangle \tag{7.488}$$

with

$$\hat{H}_x\,|E_x\rangle=E_x\,|E_x\rangle \tag{7.489}$$

$$\hat{H}_y\,|E_y\rangle=E_y\,|E_y\rangle \tag{7.490}$$

$$\begin{aligned}\hat{H}\,|E\rangle &= E\,|E\rangle=(\hat{H}_x+\hat{H}_y)\,|E_x\rangle\,|E_y\rangle\\ &= (E_x+E_y)\,|E\rangle\end{aligned} \tag{7.491}$$

or

$$E=E_x+E_y \tag{7.492}$$

Now $\hat{H}_x$ (and $\hat{H}_y$) each represent a 1−dimensional oscillator. This suggests that we define new operators for the x coordinate

$$a_x=\sqrt{\frac{m\omega}{2\hbar}}\hat{x}+\frac{i}{\sqrt{2m\hbar\omega}}\hat{p}_x=\left(a_x^+\right)^+ \tag{7.493}$$

where

$$[\hat{x},\hat{p}_x]=i\hbar\rightarrow[\hat{a}_x,a_x^+]=1 \tag{7.494}$$

and

$$\hat{H}_x=\hbar\omega\left(\hat{a}_x^+\hat{a}_x+\frac{1}{2}\right)=\hbar\omega\left(\hat{N}_x+\frac{1}{2}\right) \tag{7.495}$$

As we found earlier, $\hat{N}_x$ has an eigenvalue equation

$$\hat{N}_x\,|n_x\rangle=n_x\,|n_x\rangle\quad,\quad n_x=0,1,2,3,..... \tag{7.496}$$

We then have

$$\hat{H}_x\,|n_x\rangle=\hbar\omega\left(\hat{N}_x+\frac{1}{2}\right)|n_x\rangle=\hbar\omega\left(n_x+\frac{1}{2}\right)|n_x\rangle \tag{7.497}$$

or

$$E_x = \hbar\omega\left(n_x + \frac{1}{2}\right) \tag{7.498}$$

In a similar manner, we repeat this process for the y coordinate

$$a_y = \sqrt{\frac{m\omega}{2\hbar}}\hat{y} + \frac{i}{\sqrt{2m\hbar\omega}}\hat{p}_y = \left(a_y^+\right)^+ \tag{7.499}$$

where

$$[\hat{y}, \hat{p}_y] = i\hbar \rightarrow \left[\hat{a}_y, a_y^+\right] = 1 \tag{7.500}$$

and

$$\hat{H}_y = \hbar\omega\left(\hat{a}_y^+\hat{a}_y + \frac{1}{2}\right) = \hbar\omega\left(\hat{N}_y + \frac{1}{2}\right) \tag{7.501}$$

As we found earlier, $\hat{N}_y$ has an eigenvalue equation

$$\hat{N}_y\,|n_y\rangle = n_y\,|n_y\rangle \quad , \quad n_y = 0, 1, 2, 3, ... \tag{7.502}$$

We then have

$$\hat{H}_y\,|n_y\rangle = \hbar\omega\left(\hat{N}_y + \frac{1}{2}\right)|n_y\rangle = \hbar\omega\left(n_y + \frac{1}{2}\right)|n_y\rangle \tag{7.503}$$

or

$$E_y = \hbar\omega\left(n_y + \frac{1}{2}\right) \tag{7.504}$$

Putting this all together we get

$$E = E_x + E_y = \hbar\omega(n_x + n_y + 1) = \hbar\omega(n + 1) \tag{7.505}$$

Table 7.1 below gives the resulting energy level structure.

n_x	n_y	$E/\hbar\omega$	n
0	0	1	0
1	0	2	1
0	1	2	1
0	2	3	2
2	0	3	2
1	1	3	2

Table 7.1: Energy Levels - 2D Oscillator

Each energy value, which is characterized by the quantum number n, has a *degeneracy* equal to $(n + 1)$.

The existence of degeneracy indicates (*this is a general rule*) that there is another operator that commutes with $\hat{H}$.

Since this is a central force in the $x-y$ plane, it is not difficult to guess that the other operator that commutes with $\hat{H}$ is the angular momentum about the axis perpendicular to the plane (the z–axis), $\hat{L}_z$. Since all of the x-type operators commute with all of the y-type operators we can write

$$\hat{L}_z = (\vec{r}_{op} \times \vec{p}_{op})_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x \tag{7.506}$$

Inverting the standard operator definitions we have

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}_x + \hat{a}_x^+) \quad , \quad \hat{p}_x = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}_x - \hat{a}_x^+) \tag{7.507}$$

$$\hat{y} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}_y + \hat{a}_y^+) \quad , \quad \hat{p}_y = \frac{1}{i}\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}_y - \hat{a}_y^+) \tag{7.508}$$

which gives

$$\hat{L}_z = \frac{\hbar}{i}(\hat{a}_x^+\hat{a}_y - \hat{a}_y^+ a_x) \tag{7.509}$$

Now using

$$\left[\hat{a}_x, \hat{N}_x\right] = \hat{a}_x \quad , \quad \left[\hat{a}_x^+, \hat{N}_x\right] = -\hat{a}_x^+ \tag{7.510}$$

$$\left[\hat{a}_y, \hat{N}_y\right] = \hat{a}_y \quad , \quad \left[\hat{a}_y^+, \hat{N}_y\right] = -\hat{a}_y^+ \tag{7.511}$$

we get

$$\left[\hat{H}_x, \hat{L}_z\right] = \frac{\hbar}{i}(\hat{a}_y\hat{a}_x^+ + \hat{a}_x\hat{a}_y^+) = -\left[\hat{H}_y, \hat{L}_z\right] \tag{7.512}$$

or

$$\left[\hat{H}, \hat{L}_z\right] = 0 \tag{7.513}$$

Therefore, $\hat{H}$ and $\hat{L}_z$ share a common eigenbasis. This new eigenbasis will not be an eigenbasis for $\hat{H}_x$ or $\hat{H}_y$ separately since they do not commute with $\hat{L}_z$.

This suggests that we use linear combinations of the degenerate eigenstates to find the eigenstates of $\hat{L}_z$. This works because linear combinations for fixed n remain eigenstates of $\hat{H}$ (they are no longer eigenstates of $\hat{H}_x$ or $\hat{H}_y$ however).

We define the eigenstates and eigenvalues of $\hat{L}_z$ by the equation

$$\hat{L}_z\,|m\rangle = m\hbar\,|m\rangle \tag{7.514}$$

and the common eigenstates of $\hat{H}$ and $\hat{L}_z$ by $|n, m\rangle$ where

$$\hat{H}\,|n, m\rangle = \hbar\omega(n+1)\,|n, m\rangle \tag{7.515}$$

$$\hat{L}_z\,|n, m\rangle = m\hbar\,|n, m\rangle \tag{7.516}$$

For notational clarity we will write the old states as $|n_x\rangle\,|n_y\rangle$.

For the $n = 0$ states (there is only one) we have

$$\hat{L}_z |0,0\rangle = \hat{L}_z |0\rangle |0\rangle = 0 \rightarrow m = 0 \ , \ n = 0 \tag{7.517}$$

Now we look at the $n = 1$ states. We let

$$|\psi\rangle = a |0\rangle |1\rangle + b |1\rangle |0\rangle \ \text{ where } \ |a|^2 + |b|^2 = 1 \ \text{(normalization)} \tag{7.518}$$

Since the two states that make up the linear combination are both eigenstates of $\hat{H}$ with $n = 1$, the linear combination is an eigenstate of $\hat{H}$ with $n = 1$ for any choice of a and b. We therefore choose a and b to make this state an eigenstate of $\hat{L}_z$.

We must have

$$\begin{aligned} \hat{L}_z |\psi\rangle = \hat{L}_z |1,m\rangle &= \hat{L}_z (a |0\rangle |1\rangle + b |1\rangle |0\rangle) \\ &= \frac{\hbar}{i} (\hat{a}_x^+ \hat{a}_y - \hat{a}_y^+ a_x)(a |0\rangle |1\rangle + b |1\rangle |0\rangle) \\ &= m\hbar (a |0\rangle |1\rangle + b |1\rangle |0\rangle) \end{aligned} \tag{7.519}$$

Using

$$\hat{a} |n\rangle = \sqrt{n} |n-1\rangle \quad , \quad \hat{a}^+ |n\rangle = \sqrt{n+1} |n+1\rangle \tag{7.520}$$

we get

$$a \frac{\hbar}{i} |1\rangle |0\rangle - b \frac{\hbar}{i} |0\rangle |1\rangle = m\hbar (a |0\rangle |1\rangle + b |1\rangle |0\rangle) \tag{7.521}$$

or

$$ma = ib \quad and \quad mb = -ia \tag{7.522}$$

Dividing these two equations we get

$$\frac{a}{b} = -\frac{b}{a} \tag{7.523}$$

which implies that

$$a^2 = -b^2 \rightarrow a = \pm ib \tag{7.524}$$

and

$$m = -\frac{b}{ia} = \begin{cases} +1 & a = +ib \\ -1 & a = -ib \end{cases} \tag{7.525}$$

Normalization then says that

$$m = \begin{cases} +1 & a = \frac{1}{\sqrt{2}}, b = -\frac{i}{\sqrt{2}} \\ -1 & a = \frac{1}{\sqrt{2}}, b = +\frac{i}{\sqrt{2}} \end{cases} \tag{7.526}$$

or

$$|1, \pm 1\rangle = \frac{1}{\sqrt{2}} (|0\rangle |1\rangle \mp |1\rangle |0\rangle) \tag{7.527}$$

n	m	$E/\hbar\omega$
0	0	1
1	+1	2
1	-1	2

Table 7.2: Energy Levels - (n,m) Characterization

This gives a new characterization of the first two excited energy levels as shown in Table 7.2 above.

Let us now do the $n = 2$ states. In the same way we assume

$$|\psi\rangle = a\,|2\rangle\,|0\rangle + b\,|1\rangle\,|1\rangle + c\,|0\rangle\,|2\rangle \tag{7.528}$$

where normalization gives

$$|\mathrm{a}|^2 + |\mathrm{b}|^2 + |\mathrm{c}|^2 = 1 \tag{7.529}$$

We then have

$$\begin{aligned}
\hat{L}_z\,|\psi\rangle &= \hat{L}_z\,|2,m\rangle = m\hbar\,|\psi\rangle \\
&= m\hbar a\,|2\rangle\,|0\rangle + m\hbar b\,|1\rangle\,|1\rangle + m\hbar c\,|0\rangle\,|2\rangle \\
&= \frac{\hbar}{i}(\hat{a}_x^+\hat{a}_y - \hat{a}_y^+\hat{a}_x)(a\,|2\rangle\,|0\rangle + b\,|1\rangle\,|1\rangle + c\,|0\rangle\,|2\rangle) \\
&= \frac{\hbar}{i}\left[\sqrt{2}b\,|2\rangle\,|0\rangle + \sqrt{2}(c-a)\,|1\rangle\,|1\rangle - \sqrt{2}b\,|0\rangle\,|2\rangle\right]
\end{aligned} \tag{7.530}$$

which gives

$$ma = -i\sqrt{2}b \tag{7.531}$$

$$mc = +i\sqrt{2}b \tag{7.532}$$

$$mb = -i\sqrt{2}(c-a) \tag{7.533}$$

$$\begin{aligned}
&\frac{a}{c} = -1 \to c = -a \\
&\frac{c}{b} = -\frac{b}{c-a} \to \frac{a}{b} = -\frac{b}{2a} \to b^2 \\
&\quad = -2a^2 \to b = \pm i\sqrt{2a}
\end{aligned}$$

Putting these pieces all together we have

$$a = \frac{1}{2} = -c \quad , \quad b_\pm = \pm i\frac{\sqrt{2}}{2} \tag{7.534}$$

This implies the m-values

$$m = \begin{cases} \frac{\sqrt{2}b_+}{ia} & +2 \\ \frac{\sqrt{2}b_-}{ia} & -2 \end{cases} \tag{7.535}$$

or

$$|2,0\rangle = \frac{1}{\sqrt{2}}|2\rangle|0\rangle + \frac{1}{\sqrt{2}}|0\rangle|2\rangle \tag{7.536}$$

Thus, the final energy levels are as shown in Table 7.3 below.

n	m	$E/\hbar\omega$
0	0	1
1	+1	2
1	-1	2
2	+2	3
2	0	3
2	-2	3

Table 7.3: Energy Levels - (n,m) Characterization

What a strange result? The allowed m-values are separated by $\Delta m = \pm 2$.

Let us look at this system using the Schrodinger equation to help us understand what is happening.

We have (using plane-polar coordinates)

$$-\frac{h^2}{2M}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\varphi^2}\right) + \frac{1}{2}M\omega^2 r^2\psi = E\psi \tag{7.537}$$

Choosing

$$\psi(r,\varphi) = R(r)\Phi(\varphi) \tag{7.538}$$

we get

$$-\frac{h^2}{2M}\frac{1}{R}\left(\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right) - \frac{h^2}{2Mr^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2} + \frac{1}{2}M\omega^2 r^2 = E \tag{7.539}$$

Now we must have

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2} = -m^2 = \text{constant} \tag{7.540}$$

which produces a radial equation of the form

$$-\frac{h^2}{2M}\left(\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right) + \left(\frac{h^2m^2}{2Mr^2} + \frac{1}{2}M\omega^2 r^2\right)R = ER \tag{7.541}$$

Now we change the variables using

$$r = \rho y \quad , \quad \rho = \frac{\hbar}{M\omega} \tag{7.542}$$

and get

$$\frac{d^2R}{dy^2} + \frac{1}{y}\frac{dR}{dy} + \left(\varepsilon - y^2 - \frac{m^2}{y^2}\right)R = 0 \tag{7.543}$$

where

$$\varepsilon = \frac{E}{\hbar\omega} \tag{7.544}$$

As with the solution of other differential equations, the procedure to follow is to extract out the asymptotic behavior as $y \to 0, \infty$ and solve the equation for the remaining function by recognizing the *well-known* equation that results.

As $y \to \infty$ the dominant term will be y^2 and the equation for this behavior is

$$\frac{d^2R}{dy^2} - y^2 R = 0 \tag{7.545}$$

which has a solution $R \to e^{-y^2/2}$.

As $y \to 0$ the dominant term will be $1/y^2$ and the equation for this behavior is

$$\frac{d^2R}{dy^2} - \frac{m^2}{y^2}R = 0 \tag{7.546}$$

which has a solution $R \to y^{|m|}$, where we have excluded any negative powers since that solution would diverge at $y = 0$.

Therefore we assume $R = y^{|m|}e^{-y^2/2}G(y)$. Substitution gives the equation for G as

$$\frac{d^2G}{dy^2} + \left(\frac{2\,|m|+1}{y} - 2y\right)\frac{dG}{dy} + (\varepsilon - 2 - 2\,|m|)\,G = 0 \tag{7.547}$$

Changing the variable again to $z = y^2$ we have

$$\frac{d^2G}{dz^2} + \left(\frac{|m|+1}{z} - 1\right)\frac{dG}{dz} + \left(\frac{\varepsilon - 2(|m|+1)}{4z}\right)G = 0 \tag{7.548}$$

If we are clever, we recognize this as Laguerre's equation. If not, we make a series solution substitution

$$G(z) = \sum_{s=0}^{\infty} b_s z^s \tag{7.549}$$

which gives the recursion relation

$$\frac{b_{s+1}}{b_s} = \frac{s + \frac{|m|+1}{2} - \frac{\varepsilon}{4}}{(s+1)(s+|m|+1)} \to \frac{1}{s} \text{ for large } s \tag{7.550}$$

This says that unless the series terminates (becomes a polynomial in z) it will behave like $e^z = e^{y^2}$ which implies that the solution for $R(y)$ will diverge for

large y and thus, not be normalizable.

If we choose the maximum s-value to be $s_{max} = n_r$, then we can terminate the series by choosing

$$\varepsilon = 2|m| + 2 + 4n_r \tag{7.551}$$

which then gives us the allowed energy eigenvalues

$$E_{n_r,m} = \hbar\omega(|m| + 2n_r + 1) \tag{7.552}$$

The polynomial solutions are the generalized Laguerre polynomials $L_{n_r}^{|m|}(z)$.

The first few Laguerre polynomials are

$$L_0^k(z) = 1 \quad , \quad L_1^k(z) = 1 + k - z \tag{7.553}$$

$$L_2^k(z) = \frac{1}{2}(2 + 3k + k^2 - 2z(k+2) + z^2) \tag{7.554}$$

The full wave function is

$$\psi_{n_r,m}(r,\varphi) = r^{|m|}e^{-\frac{r^2}{2\rho^2}}L_{n_r}^{|m|}\left(\frac{r^2}{\rho^2}\right)e^{im\varphi} \tag{7.555}$$

and the energy level structure is shown in Table 7.4 below.

$E/\hbar\omega$	n_r	m	degeneracy
1	0	0	1
2	0	+1	2
2	0	-1	
3	0	+2	3
3	1	0	
3	0	-2	
4	0	+3	4
4	1	+1	
4	1	-1	
4	0	-3	

Table 7.4: Energy Levels - 2D Oscillator

which is the same structure (with different labels) as in the operator solution. The fact that $\Delta m = 2$ in the 2-dimensional case is one of many peculiarities associated with two dimensions that does not appear three dimensions.

7.6.5 What happens in 3 dimensions?

In Cartesian coordinates we have a simple extension of the 2-dimensional case.

$$E = \hbar\omega\left(n_x + n_y + n_z + \frac{3}{2}\right) = \hbar\omega\left(n + \frac{3}{2}\right) \tag{7.556}$$

The degeneracy is

$$\frac{(n+1)(n+2)}{2} \tag{7.557}$$

The energy level structure is shown in Table 7.5 below.

$E/\hbar\omega$	n_x	n_y	n_z	n	degeneracy
3/2	0	0	0	0	1
5/2	1	0	0	1	3
5/2	0	1	0	1	
5/2	0	0	1	1	
7/2	2	0	0	2	6
7/2	0	2	0	2	
7/2	0	0	2	2	
7/2	1	1	0	2	
7/2	1	0	1	2	
7/2	0	1	1	2	

Table 7.5: Energy Levels - 3D Oscillator

In spherical-polar coordinates, we can follow a procedure similar to the plane-polar 2-dimensional case to get

$$\psi_{n_r,\ell,m}(y,\theta,\varphi) = y^\ell e^{-\frac{y^2}{2}} L_{n_r}^{\ell+\frac{1}{2}}(y^2) Y_{\ell m}(\theta,\varphi) \tag{7.558}$$

where

$$r = \rho y \quad , \quad \rho^2 = \frac{\hbar^2}{M\omega} \tag{7.559}$$

The corresponding energy values are

$$E_{n_r,\ell} = \hbar\omega(2n_r + \ell + \frac{3}{2}) \tag{7.560}$$

which gives Table 7.6 below.

$E/\hbar\omega$	n_r	ℓ	$n = 2n_r + \ell$	degeneracy
3/2	0	0	0	1
5/2	0	1	1	3 → m=± 1,0
7/2	1 or 0	0 or 2	2	6 → m=± 2, ± 1

Table 7.6: Energy Levels - 3D Oscillator

Finally, we look at a case we skipped over(because it is the most difficult example of this type).

7.6.6 Two-Dimensional Finite Circular Well

We consider the potential in two dimensions

$$V(r) = \begin{cases} -V_0 & r < a \\ 0 & r > a \end{cases} \tag{7.561}$$

The Schrodinger equation in plane-polar coordinates is

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right]\psi(r,\varphi) - V_0\psi(r,\varphi) = E\psi(r,\varphi) \qquad r < a \tag{7.562}$$

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right]\psi(r,\varphi) = E\psi(r,\varphi) \qquad r > a \tag{7.563}$$

We assume(SOV)

$$\psi(r,\varphi) = R(r)\Phi(\varphi) \tag{7.564}$$

which gives

$$-\frac{\hbar^2}{2m}\left[\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}\right] = E + V_0 \qquad r < a \tag{7.565}$$

$$-\frac{\hbar^2}{2m}\left[\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2}\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2}\right] = E \qquad r > a \tag{7.566}$$

We choose a separation constant

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\varphi^2} = -\alpha^2 \rightarrow \Phi(\varphi) = B\sin(\alpha\varphi+\delta) \tag{7.567}$$

The requirement of single-valuedness under a φ-rotation of 2π says that

$$\sin(\alpha\varphi+\delta) = \sin(\alpha\varphi+\delta+\alpha\pi)$$
$$\rightarrow \alpha = \text{ integer } = 0,1,2,3\ldots$$

Alternatively, we could write

$$\Phi(\varphi) = Be^{i\alpha\varphi}$$
$$\alpha = \text{ integer } = \cdots -3,-2,-1,0,1,2,3,\ldots$$

Substitution of this solution leaves the radial differential equations

$$r^2\frac{d^2R}{dr^2}+r\frac{dR}{dr}+\left[\beta^2r^2-\alpha^2\right]R = 0 \quad r < a \quad where \quad \beta^2 = \frac{2m(E+V_0)}{\hbar^2} \tag{7.568}$$

$$r^2\frac{d^2R}{dr^2}+r\frac{dR}{dr}+\left[\lambda^2r^2-\alpha^2\right]R = 0 \quad r > a \quad where \quad \lambda^2 = \frac{2mE}{\hbar^2} \tag{7.569}$$

These are Bessel's equations. The general solutions are

$$R(r) = NJ_\alpha(\beta r) + MY_\alpha(\beta r) \quad r < a \tag{7.570}$$

$$R(r) = PJ_\alpha(\lambda r) \quad r > a \tag{7.571}$$

and the complete solutions are then

$$\psi_{k\alpha}(r,\varphi) = \left\{ \begin{array}{c} R(r)\Phi(\varphi) = (NJ_\alpha(\beta r) + MY_\alpha(\beta r))e^{i\alpha\varphi} \quad r < a \\ R(r)\Phi(\varphi) = PJ_\alpha(\lambda r)e^{i\alpha\varphi} \quad r > a \end{array} \right\} \tag{7.572}$$

The continuity (or boundary) conditions at $r = a$ are

$$NJ_\alpha(\beta a) + MY_\alpha(\beta a) = PJ_\alpha(\lambda a) \tag{7.573}$$

$$N\beta \left.\frac{dJ_\alpha(\beta r)}{d(\beta r)}\right|_{r=a} + M\beta \left.\frac{dY_\alpha(\beta r)}{d(\beta r)}\right|_{r=a} = P\lambda \left.\frac{dJ_\alpha(\lambda r)}{d(\lambda r)}\right|_{r=a} \tag{7.574}$$

Let us consider the case $\alpha = 0$. We have

$$NJ_0(\beta a) + MY_0(\beta a) = PJ_0(\lambda a) \tag{7.575}$$

$$N\beta \left.\frac{dJ_0(\beta r)}{d(\beta r)}\right|_{r=a} + M\beta \left.\frac{dY_0(\beta r)}{d(\beta r)}\right|_{r=a} = P\lambda \left.\frac{dJ_0(\lambda r)}{d(\lambda r)}\right|_{r=a} \tag{7.576}$$

where

$$J_0(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 4} - \frac{x^6}{2^2 4^2 6} + ... \tag{7.577}$$

$$\frac{dJ_0(x)}{dx} = -J_1(x) = -\left(\frac{x}{2} - \frac{x^3}{2^2 4} + \frac{x^5}{2^2 4^2 6} - \frac{x^7}{2^2 4^2 6^2 8} + ...\right) \tag{7.578}$$

$$Y_0(x) = \frac{2}{\pi}\left[\ell n\left(\frac{x}{2}\right) + \gamma\right] J_0(x) \gamma = 0.5772156 \tag{7.579}$$

$$\frac{dY_0(x)}{dx} = \frac{2}{\pi}\left[\ell n\left(\frac{x}{2}\right) + \gamma\right]\frac{dJ_0(x)}{dx} + \frac{2}{\pi x}J_0(x) \tag{7.580}$$

Clearly, the 2-dimensional finite well is very difficult.

7.6.7 The 3-Dimensional Delta Function

We now consider a particle is moving in 3-dimensions under the action of the attractive delta function potential given by

$$V(r) = -\frac{\hbar^2}{2M\alpha}\delta(r-a) \tag{7.581}$$

Since this is a central force we know that we can write the solutions in the form

$$\psi(r,\theta,\varphi) = R(r)Y_{\ell m}(\theta,\varphi) \tag{7.582}$$

The Schrodinger equation is

$$-\frac{\hbar^2}{2M}\left(\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} - \frac{\ell(\ell+1)}{r^2}R\right) - \frac{\hbar^2}{2M\alpha}\delta(r-a)R - \frac{\hbar^2 k^2}{2M}R = 0 \tag{7.583}$$

where, since we are looking for bound states, we have set

$$E = -\frac{\hbar^2 k^2}{2M} < 0 \tag{7.584}$$

Now the function $R(r)$ must be continuous everywhere. But as we saw earlier in the 1-dimensional case, the derivative of R is not continuous for delta functions. The discontinuity at $r = a$ is found by integrating the radial equation around the point $r = a$.

$$-\frac{\hbar^2}{2M}\int_{a^-}^{a^+} r^2 dr\left(\frac{d^2R(r)}{dr^2} + \frac{2}{r}\frac{dR(r)}{dr} - \frac{\ell(\ell+1)}{r^2}R(r)\right)$$
$$-\frac{\hbar^2}{2M\alpha}\int_{a^-}^{a^+} r^2 dr\delta(r-a)R(r) - \frac{\hbar^2 k^2}{2M}\int_{a^-}^{a^+} r^2 dr R(r) = 0$$

which gives the second boundary (continuity) condition

$$\alpha\left[\frac{dR(a^+)}{dr} - \frac{dR(a^-)}{dr}\right] = -R(a) \tag{7.585}$$

For $r > a$ and $r < a$ we then have the equation

$$-\frac{\hbar^2}{2M}\left(\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} - \frac{\ell(\ell+1)}{r^2}R\right) - \frac{\hbar^2 k^2}{2M}R = 0 \tag{7.586}$$

which has as a solution for the case $\ell = 0$

$$R(r) = \frac{1}{r}(Ae^{-kr} + Be^{kr}) \tag{7.587}$$

For $R(r)$ to be well behaved at $r = 0$ we must choose $B = -A$. Therefore, the solution for $r < a$ is

$$R(r) = \frac{c}{r}\sinh kr \tag{7.588}$$

For $R(r)$ to be well behaved as $r \to \infty$ we must choose $B = 0$. Therefore the solution for $r > a$ is

$$R(r) = \frac{b}{r}e^{-kr} \tag{7.589}$$

The boundary conditions at $r = a$ then give the equations

$$\frac{c}{a}\sinh ka = \frac{b}{a}e^{-ka} \to c\sinh ka = be^{-ka} \tag{7.590}$$

and

$$0 = -k\alpha(be^{-ka} + c\cosh ka) + c\sinh ka \tag{7.591}$$

Eliminating b and c we get a transcendental equation for k (or E)

$$\frac{\alpha}{a} = \frac{1 - e^{-2ka}}{2ka} \tag{7.592}$$

The right-hand side of this equation has a range

$$0 < \frac{1 - e^{-2ka}}{2ka} < 1 \tag{7.593}$$

Therefore, the allowed range of the parameter α, in order that an $\ell = 0$ bound state exist, is $0 < \alpha < a$.

Just this once let us ask .... what can we say about $\ell \neq 0$? This will illustrate some properties of the spherical Bessel functions. We have the radial equation

$$r^2 \frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + \left(k^2r^2 - \ell(\ell+1)\right) R = 0 \tag{7.594}$$

which as we have seen before is the spherical Bessel function equation with the general solution

$$R(r) = Aj_\ell(kr) + B\eta_\ell(kr) \tag{7.595}$$

For $r < a$ we must choose $B = 0$ ($\eta_\ell(kr)$ diverges at $r = 0$) and for $r > a$ we must choose $B = iA$, which leads to a solution that drops off exponentially as $r \to \infty$. So we finally have

$$r < a \qquad R(r) = Gj_\ell(kr) \tag{7.596}$$

$$r > a \qquad R(r) = Hh_\ell^{(1)}(kr) = H(j_\ell(kr) + i\eta_\ell(kr)) \tag{7.597}$$

The boundary conditions then give

$$Gj_\ell(ka) = Hh_\ell^{(1)}(ka) = H(j_\ell(ka) + i\eta_\ell(ka)) \tag{7.598}$$

$$\begin{aligned} -\frac{G}{\alpha} j_\ell(ka) &= H\frac{dh_\ell^{(1)}(ka)}{dr} - G\frac{dj_\ell(ka)}{dr} \\ &= \frac{dj_\ell(ka)}{dr}(H - G) + iG\frac{d\eta_\ell(ka)}{dr} \end{aligned} \tag{7.599}$$

Now

$$\frac{dj_\ell(kr)}{dr} = \frac{k}{2\ell+1}\left[\ell j_{\ell-1}(kr) - (\ell+1)j_{\ell+1}(kr)\right] \tag{7.600}$$

and similarly for η_ℓ and $h_\ell^{(1)}$. Therefore, we get

$$Gj_\ell(ka) = (j_\ell(ka) + i\eta_\ell(ka))H \tag{7.601}$$

$$\begin{aligned} &\frac{Gk}{2\ell+1}\left[\ell j_{\ell-1}(ka) - (\ell+1)j_{\ell+1}(ka)\right] \\ &\qquad - i\frac{Gk}{2\ell+1}\left[\ell\eta_{\ell-1}(ka) - (\ell+1)\eta_{\ell+1}(ka)\right] - \frac{G}{\alpha}j_\ell(ka) \\ &= \frac{k}{2\ell+1}\left[\ell j_{\ell-1}(ka) - (\ell+1)j_{\ell+1}(ka)\right] H \end{aligned} \tag{7.602}$$

or

$$\frac{\eta_\ell(ka)}{j_\ell(ka) + i\eta_\ell(ka)} = \frac{\left[\ell\eta_{\ell-1}(ka) - (\ell+1)\eta_{\ell+1}(ka)\right] - \frac{i(2\ell+1)}{k\alpha}j_\ell(ka)}{\left[\ell j_{\ell-1}(ka) - (\ell+1)j_{\ell+1}(ka)\right]}$$

This is the *transcendental equation* for the energy!!! That is enough for nonzero angular momentum!

7.6.8 The Hydrogen Atom

Schrodinger Equation Solution

The potential energy function

$$V(r) = -\frac{Ze^2}{r} \tag{7.603}$$

represents the attractive Coulomb interaction between an atomic nucleus of charge $+Ze$ and an electron of charge $-e$.

The Schrodinger equation we have been using describes the motion of a single particle in an external field. In the hydrogen atom, however, we are interested in the motion of two particles (nucleus and electron) that are attracted to each other via the potential above (where r is the *separation distance* between the two particles).

We start by writing the Schrodinger equation for the two particle system. It involves six coordinates (three for each particle). We use Cartesian coordinates to start the discussion. We have

$$\hat{H} = \frac{\vec{p}_{1,op}^{\,2}}{2m_1} + \frac{\vec{p}_{2,op}^{\,2}}{2m_2} + V(\vec{r}) \quad where \quad \vec{r} = \vec{r}_1 - \vec{r}_2 \tag{7.604}$$

which gives

$$\begin{aligned} -\frac{\hbar^2}{2m_1}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right)\phi(x_1, y_1, z_1, x_2, y_2, z_2) \\ -\frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}\right)\phi(x_1, y_1, z_1, x_2, y_2, z_2) \\ + V(x_1 - x_2, y_1 - y_2, z_1 - z_2)\phi(x_1, y_1, z_1, x_2, y_2, z_2) \\ = E\phi(x_1, y_1, z_1, x_2, y_2, z_2) \end{aligned} \tag{7.605}$$

We now introduce relative and center-of-mass(CM) coordinates by

$$\begin{aligned} &x = x_1 - x_2 \quad , \quad y = y_1 - y_2 \quad , \quad z = z_1 - z_2 \\ &\vec{r} = (x, y, z) \\ &MX = m_1 x_1 + m_2 x_2 \quad , \quad M\mathrm{Y} = m_1 y_1 + m_2 y_2 \quad , \quad M\mathrm{Z} = m_1 z_1 + m_2 z_2 \\ &\vec{R} = (X, Y, Z) \\ &M = m_1 + m_2 = \text{ total mass of the system} \end{aligned}$$

Substitution gives

$$\begin{aligned} -\frac{\hbar^2}{2M}\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}\right)\phi(x, y, z, X, Y, Z) \\ -\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\phi(x, y, z, X, Y, Z) \\ + V(x, y, z)\phi(x, y, z, X, Y, Z) = E\phi(x, y, z, X, Y, Z) \end{aligned} \tag{7.606}$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \text{ the reduced mass} \tag{7.607}$$

We can now separate the variables by assuming a solution of the form

$$\phi(x, y, z, X, Y, Z) = \psi(x, y, z)\Psi(X, Y, Z) \tag{7.608}$$

which gives the two equations

$$-\frac{\hbar^2}{2\mu}\nabla^2_{\vec{r}}\psi(\vec{r}) + V(r)\psi(\vec{r}) = E\psi(\vec{r}) \tag{7.609}$$

$$-\frac{\hbar^2}{2M}\nabla^2_{\vec{R}}\Psi(\vec{R}) = E'\Psi(\vec{R}) \tag{7.610}$$

The second equation says that the CM of the two particles is like a free particle of mass M.

The first equation describes the relative motion of the two particles and is the same as the equation of motion of a particle of mass μ in an external potential energy $V(r)$.

In the hydrogen atom problem we are only interested in the energy levels E associated with the relative motion. In addition, since the nuclear mass is so much larger than the electron mass, we have

$$\mu \approx m_e = m_{electron} \tag{7.611}$$

This is a central force so we assume a solution of the form

$$\psi(\vec{r}) = R(r)Y_{\ell m}(\theta, \varphi) \tag{7.612}$$

and obtain the radial equation

$$-\frac{\hbar^2}{2\mu}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{Ze^2}{r}R + \frac{\ell(\ell+1)\hbar^2}{2\mu r^2}R = ER \tag{7.613}$$

where $E < 0$ for a bound state. We follow the same approach as before. We change the variables so that the equation is in dimensionless form by introducing $\rho = \alpha r$ where

$$\alpha^2 = \frac{8\mu|E|}{\hbar^2} \quad and \quad \lambda = \frac{2\mu Ze^2}{\alpha\hbar^2} = \frac{Ze^2}{\hbar}\left(\frac{\mu}{2|E|}\right)^{1/2} \tag{7.614}$$

We get

$$\frac{1}{\rho^2}\frac{d}{d\rho}\left(\rho^2\frac{dR}{dr}\rho\right) + \left(\frac{\lambda}{\rho} - \frac{1}{4} - \frac{\ell(\ell+1)}{\rho^2}\right)R = 0 \tag{7.615}$$

For $\rho \to \infty$ the equation becomes

$$\frac{1}{\rho^2}\frac{d}{d\rho}\left(\rho^2\frac{dR}{dr}\rho\right) - \frac{1}{4}R = 0 \tag{7.616}$$

which has the solution

$$R \to \rho^n e^{\pm\frac{1}{2}\rho} \tag{7.617}$$

where n can have any finite value. Since we want a normalizable solution, we will look for a solution of the form

$$R \to F(\rho) e^{-\frac{1}{2}\rho} \tag{7.618}$$

Substitution gives an equation for F

$$\frac{d^2F}{d\rho^2} + \left(\frac{2}{\rho} - 1\right)\frac{dF}{d\rho} + \left[\frac{\lambda - 1}{\rho} - \frac{\ell(\ell+1)}{\rho^2}\right]F = 0 \tag{7.619}$$

We solve this equation (it is Laguerre's equation) by series substitution.

$$F(\rho) = \rho^s(a_0 + a_1\rho + a_2\rho^2 + ...) = \rho^s L(\rho) \tag{7.620}$$

where $a_0 \neq 0$ and $s \geq 0$. We must also have that $F(0)$ is finite. Substitution gives an equation for L.

$$\rho^2\frac{d^2L}{d\rho^2} + \rho\left[2(s+1) - \rho\right]\frac{dL}{d\rho} + \left[\rho(\lambda - s - 1) + s(s+1) - \ell(\ell+1)\right]L = 0 \tag{7.621}$$

If we set $\rho = 0$ in this equation and use the fact that L is a power series we get the condition

$$s(s+1) - \ell(\ell+1) = 0 \tag{7.622}$$

or

$$s = \ell \text{ or } s = -(\ell+1) \tag{7.623}$$

Since R must be finite at the origin we exclude the second possibility and choose $s = \ell$. This gives

$$\rho^2\frac{d^2L}{d\rho^2} + \rho\left[2(\ell+1) - \rho\right]\frac{dL}{d\rho} + \left[\rho(\lambda - \ell - 1)\right]L = 0 \tag{7.624}$$

Substituting the power series gives a recursion relation for the coefficients

$$a_{\nu+1} = \frac{\nu + \ell + 1 - \lambda}{(\nu+1)(\nu+2\ell+2)}a_\nu \tag{7.625}$$

If the series does not terminate, then this recursion relation behaves like

$$\frac{a_{\nu+1}}{a_\nu} \to \frac{1}{\nu} \tag{7.626}$$

for large ν. This corresponds to the series for $\rho^n e^\rho$. Since this will give a non-normalizable result, we must terminate the series by choosing

$$\lambda = n = \text{ positive integer } = n' + \ell + 1 = \text{ total quantum number} \tag{7.627}$$

where n'(= radial quantum number) is the largest power of ρ in the solution L. n' and n are non-negative integers, $n = 1, 2, 3, 4, \ldots$

We thus obtain the solution for the energies

$$E_n = -|E_n| = -\frac{\mu Z^2 e^4}{2\hbar^2 n^2} = -\frac{Z^2 e^2}{2a_0 n^2} \tag{7.628}$$

where

$$a_0 = \text{ Bohr radius } = \frac{\hbar^2}{\mu e^2} \tag{7.629}$$

Unlike the finite square well, where we had a finite number of bound state levels, in this case, we obtain an infinite set of discrete energies. This results from the very slow decrease of the Coulomb potential energy with distance.

The Laguerre polynomial solutions are given by the generating function

$$G(\rho, s) = \frac{e^{-\frac{\rho s}{1-s}}}{1-s} = \sum_{q=0}^{\infty} \frac{L_q(\rho)}{q!} s^q \quad , \quad s < 1 \tag{7.630}$$

By differentiating the generating function with respect to ρ and s we can show that

$$\frac{dL_q}{d\rho} - q\frac{dL_{q-1}}{d\rho} = -qL_{q-1} \tag{7.631}$$

$$L_{q+1} = (2q + 1 - \rho)L_q - q^2 L_{q-1} \tag{7.632}$$

The lowest order differential equation involving only L_q that can be constructed from these two equations is

$$\rho\frac{d^2 L_q}{d\rho^2} + (1 - \rho)\frac{dL_q}{d\rho} + qL_q = 0 \tag{7.633}$$

This is not quite our original equation. However, if we define the associated Laguerre polynomials by

$$L_q^p(\rho) = \frac{d^p}{d\rho^p} L_q(\rho) \tag{7.634}$$

then differentiating the differential equation p times we get

$$\rho\frac{d^2 L_q^p}{d\rho^2} + (p + 1 - \rho)\frac{dL_q^p}{d\rho} + (q - p)L_q^p = 0 \tag{7.635}$$

Setting $\lambda = n$ we then see that the solutions to the Schrodinger equation are

$$L_{n+\ell}^{2\ell+1}(\rho) \tag{7.636}$$

which are polynomials of order $(n + \ell) - (2\ell + 1) = n - \ell - 1$ in agreement with the earlier results.

We can differentiate the generating function p times to get

$$G_p(\rho,s) = \frac{(-s)^p e^{-\frac{\rho s}{1-s}}}{(1-s)^{p+1}} = \sum_{q=0}^{\infty} \frac{L_q^p(\rho)}{q!} s^q \quad , \quad s<1 \tag{7.637}$$

Explicitly, we then have

$$L_{n+\ell}^{2\ell+1}(\rho) = \sum_{k=0}^{n-\ell-1} (-1)^{k+1} \frac{\left[(n+\ell)!\right]^2 \rho^k}{(n-\ell-1-k)!(2\ell+1+k)!k!} \tag{7.638}$$

The normalized radial wave functions are of the form

$$R_{n\ell}(\rho) = -\left\{ \left(\frac{2Z}{na_0}\right)^3 \frac{(n-\ell-1)!}{2n\left[(n+\ell)!\right]^3} \right\}^{1/2} e^{-\frac{1}{2}\rho} \rho^\ell L_{n+\ell}^{2\ell+1}(\rho) \tag{7.639}$$

with

$$a_0 = \frac{\hbar^2}{\mu e^2} \text{ and } \rho = \frac{2Z}{na_0} r \tag{7.640}$$

and

$$\psi_{n\ell m}(r,\theta,\varphi) = R_{n\ell}(r) Y_{\ell m}(\theta,\varphi) \tag{7.641}$$

The first few radial wave functions are

$$R_{10}(r) = \left(\frac{Z}{a_0}\right)^{3/2} 2e^{-\frac{Zr}{a_0}} \tag{7.642}$$

$$R_{20}(r) = \left(\frac{Z}{2a_0}\right)^{3/2} \left(2 - \frac{Zr}{a_0}\right) e^{-\frac{Zr}{2a_0}} \tag{7.643}$$

$$R_{21}(r) = \left(\frac{Z}{2a_0}\right)^{3/2} \frac{Zr}{a_0\sqrt{3}} e^{-\frac{Zr}{2a_0}} \tag{7.644}$$

What about degeneracy since the energy values do not depend on ℓ and m?

For each value of n, ℓ can vary between 0 and $n-1$, and for each value of these ℓ values m can vary between $-\ell$ and ℓ ($2\ell+1$ values). Therefore the total degeneracy of an energy level E_n is given by

$$\sum_{\ell=0}^{n-1} (2\ell+1) = 2\frac{n(n-1)}{2} + n = n^2 \tag{7.645}$$

The degeneracy with respect to m is true for any central force (as we have seen in other examples). The ℓ degeneracy, however, is characteristic of the Coulomb potential alone. It is called an *accidental* degeneracy.

Some useful expectation values are:

$$\langle r \rangle_{n\ell m} = \frac{a_0}{2Z}\left[3n^2 - \ell(\ell+1)\right] \qquad \left\langle \frac{1}{r} \right\rangle_{n\ell m} = \frac{Z}{a_0 n^2} \tag{7.646}$$

$$\left\langle \frac{1}{r^2} \right\rangle_{n\ell m} = \frac{Z^2}{a_0^2 n^3 (\ell+\frac{1}{2})} \qquad \left\langle \frac{1}{r^3} \right\rangle_{n\ell m} = \frac{Z^3}{a_0^3 n^3 \ell(\ell+\frac{1}{2})(\ell+1)} \tag{7.647}$$

7.6.9 Algebraic Solution of the Hydrogen Atom

Review of the Classical Kepler Problem

In the classical Kepler problem of an electron of charge -e moving in the electric force field of a positive charge Ze located at the origin, the angular momentum

$$\vec{L} = \vec{r} \times \vec{p} \tag{7.648}$$

is constant. The motion takes place in a plane perpendicular to the constant angular momentum direction. Newton's second law tells us that the rate of change of the momentum is equal to the force

$$\frac{d\vec{p}}{dt} = -\frac{Ze^2}{r^2}\hat{r} \text{ where } r = |\vec{r}| \text{ and } \hat{r} = \frac{\vec{r}}{r} = \text{ unit vector} \tag{7.649}$$

The so-called *Runge-Lenz* vector is defined by

$$\vec{A} = \frac{1}{Ze^2m}(\vec{L} \times \vec{p}) + \hat{r} \tag{7.650}$$

If we take its time derivative we find

$$\frac{d\vec{L}}{dt} = 0 \text{ and } \vec{L} = m\vec{r} \times \frac{d\vec{r}}{dt} = mr^2\hat{r} \times \frac{d\hat{r}}{dt} + mr(\hat{r} \times \hat{r})\frac{dr}{dt} = mr^2\hat{r} \times \frac{d\hat{r}}{dt} \tag{7.651}$$

$$\begin{aligned}\frac{d\vec{A}}{dt} &= \frac{1}{Ze^2m}(\vec{L} \times \frac{d\vec{p}}{dt} + \frac{d\vec{L}}{dt} \times \vec{p}) + \frac{d\hat{r}}{dt} = \frac{1}{Ze^2m}(\vec{L} \times \frac{d\vec{p}}{dt}) + \frac{d\hat{r}}{dt} \\ &= -\frac{1}{m}(\vec{L} \times \frac{\hat{r}}{r^2}) + \frac{d\hat{r}}{dt} = -(\hat{r} \times \frac{d\hat{r}}{dt}) \times \hat{r} + \frac{d\hat{r}}{dt} \\ &= -(\frac{d\hat{r}}{dt}(\hat{r} \cdot \hat{r}) - (\hat{r} \cdot \frac{d\hat{r}}{dt})) + \frac{d\hat{r}}{dt} = -\frac{d\hat{r}}{dt} + \frac{d(\hat{r} \cdot \hat{r})}{dt} + \frac{d\hat{r}}{dt} = 0\end{aligned} \tag{7.652}$$

Thus, the vector $\vec{A}$ is a constant of the motion. It corresponds physically to the length and direction of the *semi-major axis* of the classical elliptical orbit. The equation of the elliptical orbit is easily found using the $\vec{A}$ vector.

$$\begin{aligned}\vec{A} \cdot \vec{r} = ar\cos\theta &= \frac{1}{Ze^2m}(\vec{L} \times \vec{p}) \cdot \vec{r} + \hat{r} \cdot \vec{r} \\ &= -\frac{1}{Ze^2m}\vec{L} \cdot (\vec{r} \times \vec{p}) + r\end{aligned} \tag{7.653}$$

$$ar\cos\theta = -\frac{L^2}{Ze^2m} + r \tag{7.654}$$

$$\rightarrow \frac{1}{\mathrm{r}} = \frac{Ze^2m}{L^2}(1 - a\cos\theta) \rightarrow \text{ orbit equation (conic sections)} \tag{7.655}$$

where

$$a = \left|\vec{A}\right| \tag{7.656}$$

= eccentricity and the direction of $\vec{A}$ is from the origin to the aphelion

The Quantum Mechanical Problem

In order to make this useful in quantum mechanics, the classical Runge-Lenz vector must be generalized. In classical physics

$$\vec{L} \times \vec{p} = -\vec{p} \times \vec{L} \tag{7.657}$$

In quantum mechanics, this identity is not valid since the components of $\vec{L}_{op}$ *do not commute*. The correct quantum mechanical generalization (remember it must be a Hermitian operator if it is a physical observable) of the vector $\vec{A}$ is

$$\vec{A}_{op} = \frac{1}{2Ze^2m}(\vec{L}_{op} \times \vec{p}_{op} - \vec{p}_{op} \times \vec{L}_{op}) + \left(\frac{\vec{r}}{r}\right)_{op} \tag{7.658}$$

This satisfies $[\hat{H}, \vec{A}_{op}] = 0$, which says $\vec{A}$ is a constant. It also satisfies $\vec{A}_{op} \cdot \vec{L}_{op} = 0$.

We can derive the following commutators:

$$[\hat{L}_i, \hat{A}_j] = i\hbar\varepsilon_{ijk}\hat{A}_k \quad , \quad [\hat{L}_i, \hat{p}_j] = i\hbar\varepsilon_{ijk}\hat{p}_k \quad , \quad [\hat{L}_i, \hat{r}_j] = i\hbar\varepsilon_{ijk}\hat{r}_k \tag{7.659}$$

which gives

$$\vec{A}_{op} = \frac{1}{2Ze^2m}(2\vec{L}_{op} \times \vec{p}_{op} + i\hbar\vec{p}_{op}) + \left(\frac{\vec{r}}{r}\right)_{op} \tag{7.660}$$

We now derive two important properties of $\vec{A}$. The first property follows from the commutators

$$\left[\left(\vec{L}_{op} \times \vec{p}_{op}\right)_i, \hat{p}_j\right] + \left[\hat{p}_i, \left(\vec{L}_{op} \times \vec{p}_{op}\right)_j\right] = 0 \tag{7.661}$$

$$\left[\left(\vec{L}_{op} \times \vec{p}_{op}\right)_i, \left(\vec{L}_{op} \times \vec{p}_{op}\right)_j\right] = -i\hbar\varepsilon_{ijk}\hat{L}_k p^2 = -i\hbar\varepsilon_{ijk}p^2\hat{L}_k \tag{7.662}$$

$$\left[\left(\vec{L}_{op} \times \vec{p}_{op}\right)_i, \frac{\hat{r}_j}{r}\right] + \left[\frac{\hat{r}_i}{r}, \left(\vec{L}_{op} \times \vec{p}_{op}\right)_j\right] = 2i\hbar\varepsilon_{ijk}\frac{\hat{L}_k}{r} \tag{7.663}$$

which leads to

$$\left[\hat{A}_i, \hat{A}_j\right] = i\hbar\left(\frac{-2\hat{H}}{Z^2e^4m}\right)\varepsilon_{ijk}\hat{L}_k \tag{7.664}$$

where

$$\hat{H} = \text{ Hamiltonian } = \frac{\vec{p}_{op}^2}{2m} - Ze^2\left(\frac{1}{r}\right)_{op} \tag{7.665}$$

The second property follows from the relations

$$(\vec{L}_{op} \times \vec{p}_{op}) \cdot (\vec{L}_{op} \times \vec{p}_{op}) = \vec{L}_{op}^2\vec{p}_{op}^2 \tag{7.666}$$

$$\left(\frac{\vec{r}}{r}\right)_{op} \cdot (\vec{L}_{op} \times \vec{p}_{op}) + (\vec{L}_{op} \times \vec{p}_{op}) \cdot \left(\frac{\vec{r}}{r}\right)_{op} = -\frac{2\vec{L}_{op}^2}{r} + 2i\hbar\left(\frac{\vec{r}}{r}\right)_{op} \cdot \vec{p}_{op} \tag{7.667}$$

$$\vec{p}_{op} \cdot \left(\frac{\vec{r}}{r}\right)_{op} = \left(\frac{\vec{r}}{r}\right)_{op} \cdot \vec{p}_{op} - 2i\hbar\left(\frac{1}{r}\right)_{op} \tag{7.668}$$

$$\vec{p}_{op} \cdot \vec{L}_{op} = \vec{L}_{op} \cdot \vec{p}_{op} = 0 \tag{7.669}$$

$$\vec{p}_{op} \cdot (\vec{L}_{op} \times \vec{p}_{op}) + (\vec{L}_{op} \times \vec{p}_{op}) \cdot \vec{L}_{op} = 2i\hbar\vec{p}_{op}^2 \tag{7.670}$$

which lead to

$$\vec{A}_{op}^2 = \vec{A}_{op} \cdot \vec{A}_{op} = 1 + \frac{2\hat{H}}{Z^2 e^4 m}(\vec{L}_{op}^2 + \hbar^2) \tag{7.671}$$

We now define the two new operators (they are *ladder operators*) by

$$\vec{I}_{op}^{\pm} = \frac{1}{2}\left[\vec{L}_{op} \pm \left(-\frac{2\hat{H}}{Z^2 e^4 m}\right)^{1/2} \vec{A}_{op}\right] \tag{7.672}$$

Using the relations we have derived for $\left[\hat{A}_i, \hat{A}_j\right]$ and $\left[\hat{L}_i, \hat{A}_j\right]$ we have

$$\left[\vec{I}_{op}^{+}, \vec{I}_{op}^{-}\right] = 0 \tag{7.673}$$

which means they have a common eigenbasis. We also have

$$\left[\hat{I}_{op,i}^{\pm}, \hat{I}_{op,j}^{\pm}\right] = i\hbar\varepsilon_{ijk}\hat{I}_{op,k}^{\pm} \tag{7.674}$$

which are the standard angular momentum component commutation relations. Since they each commute with $\hat{H}$, they also have a common eigenbasis with $\hat{H}$.

Therefore, we can find a set of states such that

$$(\vec{I}_{op}^{\pm})^2 |\psi\rangle = i_{\pm}(i_{\pm} + 1)\hbar^2 |\psi\rangle \text{ and } \hat{H}|\psi\rangle = E|\psi\rangle \tag{7.675}$$

We can show that $(\vec{I}_{op}^{+})^2 = (\vec{I}_{op}^{-})^2$, which implies that $i_+ = i_-$. We also have(as before)

$$i_+ = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \ldots \tag{7.676}$$

Since $\vec{A}_{op} \cdot \vec{L}_{op} = 0$ we get

$$2\left[(\vec{I}_{op}^{+})^2 + (\vec{I}_{op}^{-})^2\right] + \hbar^2 = -\frac{Z^2 e^4 m}{2\hat{H}} \tag{7.677}$$

$$\left[4i_+(i_+ + 1) + 1\right]\hbar^2 = -\frac{Z^2 e^4 m}{2E} \tag{7.678}$$

$$E = -\frac{Z^2 e^4 m}{2(2i_+ + 1)^2} = -\frac{Z^2 e^4 m}{2n^2} \tag{7.679}$$

where we have set $n = 2i_+ + 1 = 1, 2, 3, 4, 5, \ldots$, which are the correct energy values for hydrogen.

While the energy depends only on the quantum number i_+ (or n) each state has a degeneracy depending on the number of z-component values for each i_+ value. This is

$$(2i_+ + 1)(2i_- + 1) = (2i_+ + 1)^2 = n^2 \tag{7.680}$$

which is the correct degeneracy.

7.6.10 The Deuteron

A deuteron is a bound state of a neutron and a proton. We can consider this system as a single particle with reduced mass

$$\mu = \frac{m_{proton} m_{neutron}}{m_{proton} + m_{neutron}} \approx \frac{m_{proton}}{2} = \frac{m_p}{2} \tag{7.681}$$

moving in a fixed potential $V(r)$, where r is the proton-neutron separation. As a first approximation, we assume that the nuclear interaction binding the proton and the neutron into a deuteron is a finite square well in 3–dimensions.

$$V(r) = \begin{cases} -V_0 & r < a \\ 0 & r \geq a \end{cases} \tag{7.682}$$

The physical properties of the deuteron system are:

1. Almost an $\ell = 0$ state (a small admixture of $\ell = 2$ is present). We will assume $\ell = 0$.

2. Only one bound state exists.

3. The depth and range of the potential is such that the deuteron is weakly bound. The energy level in the potential corresponds to the binding energy of the system, where

$$E = \text{ binding energy } = m_{deuteron} c^2 - (m_{proton} + m_{neutron}) c^2 < 0 \tag{7.683}$$

   Experimentally, it has been found that $E = -2.228\, MeV$.

4. By weakly bound we mean

$$\frac{|E|}{(m_{proton} + m_{neutron}) c^2} << 1 \tag{7.684}$$

   In fact, for the deuteron we have

$$\frac{|E|}{(m_{proton} + m_{neutron}) c^2} \approx 0.001 \tag{7.685}$$

5. This system is so weakly bound that any small decrease in the radius of the well a or small reduction in the depth of the well V_0 would cause the system to break up (no bound state exists).

We derived the solutions for this potential earlier.

$$R(r) = A j_0(\alpha r) \qquad r < a$$
$$R(r) = B h_0^{(1)}(i\beta r) \qquad r \geq a$$

where

$$\alpha^2 = \frac{2m}{\hbar^2}(V_0 - |E|) \text{ and } \beta^2 = \frac{2m}{\hbar^2}|E| \tag{7.686}$$

The transcendental equation arising from matching boundary conditions at $r = a$ gave us the equations

$$\eta = -\xi \cot \xi \text{ and } \xi^2 + \eta^2 = \frac{2mV_0a^2}{\hbar^2} \tag{7.687}$$

where

$$\xi = \alpha a \text{ and } \eta = \beta a \tag{7.688}$$

The graphical solution as shown below in Figure 7.17 plots

$$\eta = -\xi \cot \xi \text{ and } \eta^2 = \frac{2mV_0a^2}{\hbar^2} - \xi^2 \text{ versus } \xi \tag{7.689}$$

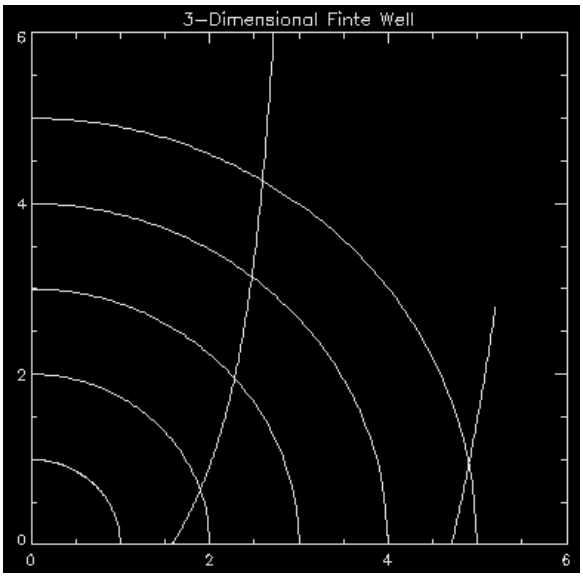


Figure 7.17: Deuteron Solution

We found a finite number of bound states for given values of the well parameters. In particular

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{\pi}{2}\right)^2 \rightarrow \text{ no solution}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{3\pi}{2}\right)^2 \rightarrow \text{ 1 solution}$$

$$\frac{2mV_0a^2}{\hbar^2} < \left(\frac{5\pi}{2}\right)^2 \rightarrow \text{ 2 solutions}$$

For the weakly bound deuteron system we expect that

$$\xi = \frac{\pi}{2} + \varepsilon \quad , \quad \varepsilon \ll \frac{\pi}{2} \tag{7.690}$$

i.e., the radius of the circle is barely large enough to get a single (weakly) bound state.

Substituting into the transcendental equation we get

$$\begin{aligned}
\eta &= -\left(\frac{\pi}{2} + \varepsilon\right)\cot\left(\frac{\pi}{2} + \varepsilon\right) = -\left(\frac{\pi}{2} + \varepsilon\right)\frac{\cos\left(\frac{\pi}{2} + \varepsilon\right)}{\sin\left(\frac{\pi}{2} + \varepsilon\right)} \\
&= -\left(\frac{\pi}{2} + \varepsilon\right)\frac{\cos\left(\frac{\pi}{2}\right)\cos(\varepsilon) - \sin\left(\frac{\pi}{2}\right)\sin(\varepsilon)}{\sin\left(\frac{\pi}{2}\right)\cos(\varepsilon) + \cos\left(\frac{\pi}{2}\right)\sin(\varepsilon)} \\
&\approx \left(\frac{\pi}{2} + \varepsilon\right)\frac{\sin(\varepsilon)}{\cos(\varepsilon)} \approx \left(\frac{\pi}{2} + \varepsilon\right)\frac{\varepsilon - \frac{\varepsilon^3}{6}}{1 - \frac{\varepsilon^2}{2}} \approx \left(\frac{\pi}{2} + \varepsilon\right)\left(\varepsilon - \frac{\varepsilon^3}{6}\right)\left(1 + \frac{\varepsilon^2}{2}\right) \\
&\approx \frac{\pi}{2}\varepsilon + (1 + \frac{\pi}{4})\varepsilon^2
\end{aligned}$$

Now substituting into the circle equation we get

$$\begin{aligned}
\left(\frac{\pi}{2} + \varepsilon\right)^2 + \left(\frac{\pi}{2}\varepsilon + (1 + \frac{\pi}{4})\varepsilon^2\right)^2 &= \frac{2mV_0a^2}{\hbar^2} \\
\frac{\pi^2}{4} + \pi\varepsilon + \left(1 + \frac{\pi^2}{4}\right)\varepsilon^2 &\approx \frac{2mV_0a^2}{\hbar^2}
\end{aligned} \tag{7.691}$$

Dropping the small quadratic terms in ε we get

$$\varepsilon \approx \frac{2mV_0a^2}{\pi\hbar^2} - \frac{\pi}{4} \tag{7.692}$$

Therefore,

$$|E| = \frac{\hbar^2}{2ma^2}\eta^2 \approx \frac{\hbar^2\pi^2}{8ma^2}\left(\frac{2mV_0a^2}{\pi\hbar^2} - \frac{\pi}{4}\right)^2 \tag{7.693}$$

A typical value for the range of the interaction is the order of 2 Fermi or $a \approx 2 \times 10^{-13}\ cm$. In order to get $|E| \approx 2.3\,MeV$, we would need a well depth of $V_0 \approx 42\,MeV$ which is reasonable (according to experimentalists).

7.6.11 The Deuteron - Another Way

Experiment (scattering) indicates that instead of a square well (very unrealistic) the actual potential is of the form

$$V(r) = -Ae^{-\frac{r}{a}} \tag{7.694}$$

where

$$A \approx 32MeV \text{ and } a \approx 2.25x10^{-13}cm \tag{7.695}$$

We can solve for this potential ($\ell = 0$ case) exactly using this clever trick I first learned from Hans Bethe at Cornell University.

The radial equation in this case is

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)R = 0 \tag{7.696}$$

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)R = 0 \tag{7.697}$$

Now let

$$R(r) = \frac{\chi(r)}{r} \tag{7.698}$$

which implies that

$$\frac{dR}{dr} = \frac{d\left(\frac{\chi}{r}\right)}{dr} = \frac{1}{r}\frac{d\chi}{dr} - \frac{1}{r^2}\chi \tag{7.699}$$

$$\frac{d^2R}{dr^2} = -\frac{2}{r^2}\frac{d\chi}{dr} + \frac{1}{r}\frac{d^2\chi}{dr^2} + \frac{2}{r^3}\chi \tag{7.700}$$

Substitution gives

$$\frac{d^2\chi}{dr^2} + \frac{2m}{\hbar^2}\left(Ae^{-\frac{r}{a}} - |E|\right)\chi = 0 \tag{7.701}$$

We now change the variables using

$$\xi = e^{-\frac{r}{2a}} \rightarrow \frac{d}{dr} = \frac{d\xi}{dr}\frac{d}{d\xi} = -\frac{\xi}{2a}\frac{d}{d\xi}$$

$$\frac{d^2}{dr^2} = \frac{d}{dr}\left(-\frac{\xi}{2a}\frac{d}{d\xi}\right) = \frac{d\xi}{dr}\frac{d}{d\xi}\left(-\frac{\xi}{2a}\frac{d}{d\xi}\right) = \left(\frac{\xi}{2a}\right)^2\frac{d^2}{d\xi^2} + \frac{\xi}{4a^2}\frac{d}{d\xi}$$

We then get the equation

$$\xi^2\frac{d^2\chi}{d\xi^2} + \xi\frac{d\chi}{d\xi} + ((\alpha a)^2\xi^2 - (ka)^2)\chi = 0 \tag{7.702}$$

where

$$(\alpha a)^2 = \frac{2mA}{\hbar^2}a^2 \text{ and } (ka)^2 = \frac{2m|E|}{\hbar^2}a^2 \tag{7.703}$$

Now Bessel's equation has the form

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \nu^2)y = 0 \tag{7.704}$$

Therefore, we have Bessel's equation with a general solution

$$\chi(r) = CJ_{ka}(\alpha a\xi) + BY_{ka}(\alpha a\xi) \tag{7.705}$$

$$R(r) = \frac{1}{r}(CJ_{ka}(\alpha a\xi) + BY_{ka}(\alpha a\xi)) \tag{7.706}$$

As $r \to \infty$, $\xi = e^{-\frac{r}{2a}} \to 0$, which implies we must choose $B = 0$ or the solution diverges.

As $r \to 0$, $\xi = e^{-\frac{r}{2a}} \to 1$, which implies we must have

$$J_{ka}(\alpha a) = 0 \tag{7.707}$$

or the solution diverges. This is then the energy eigenvalue condition. For the values of A and a given earlier we have

$$J_{ka}(6.28) = 0 \tag{7.708}$$

and if we let $|E| = qA$, this becomes

$$J_{1/2}(6.28) = 0 \tag{7.709}$$

Now

$$J_{6.4q}(6.28) = 0 \tag{7.710}$$

Therefore we have

$$6.4q = \frac{1}{2} \to q = \frac{1}{12.8} \to |E| = qA = 2.34\, MeV \tag{7.711}$$

which is an excellent result for the bound state energy of the deuteron.

7.6.12 Linear Potential

We now consider a *linear* potential energy function given by

$$V(r) = \alpha r - V_0 \tag{7.712}$$

The Schrodinger equation becomes

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) - \frac{\vec{L}_{op}^2}{\hbar^2 r^2}\psi\right) + (\alpha r - V_0 - E)\psi = 0 \tag{7.713}$$

Since it is a central potential, we can write

$$\psi = R(r)Y_{\ell m}(\theta, \varphi) \tag{7.714}$$

which implies

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{\ell(\ell+1)}{r^2}R\right) + (\alpha r - V_0 - E)R = 0 \tag{7.715}$$

If we let

$$\chi(r) = rR(r) \tag{7.716}$$

we get

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + (\alpha r - V_0 - E + \frac{\hbar^2\ell(\ell+1)}{2mr^2})\chi = 0 \tag{7.717}$$

For $\ell > 0$ there is no closed form solution for this equation and numerical methods must be applied.

For $\ell = 0$, we have

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + (\alpha r - V_0 - E)\chi = 0 \tag{7.718}$$

Now we let

$$\xi = \left(r - \frac{E + V_0}{\alpha}\right)\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3} \tag{7.719}$$

which implies that

$$\begin{aligned}
\frac{d}{dr} &= \frac{d\xi}{dr}\frac{d}{d\xi} = \left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\frac{d}{d\xi} \\
\frac{d^2}{dr^2} &= \frac{d}{dr}\left(\frac{d\xi}{dr}\frac{d}{d\xi}\right) = \frac{d\xi}{dr}\frac{d}{d\xi}\left(\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\frac{d}{d\xi}\right) \\
&= \left(\frac{2m\alpha}{\hbar^2}\right)^{2/3}\frac{d^2}{d\xi^2}
\end{aligned}$$

We then get the equation

$$\frac{d^2\chi}{d\xi^2} - \xi\chi = 0 \tag{7.720}$$

This equation is *not as simple* as it looks. Let us try a series solution of the form

$$\chi(\xi) = \sum_{n=0}^{\infty} a_n \xi^n \tag{7.721}$$

Substitution gives the relations

$$a_2 = 0 \text{ and } a_{m+2} = \frac{a_{m-1}}{(m+2)(m+1)} \tag{7.722}$$

The solution that goes to zero as $\xi \to \pm\infty$ is then of the form

$$\chi(\xi) = c_1 f(\xi) - c_2 g(\xi) = CAi(\xi) = \text{ Airy function} \tag{7.723}$$

where

$$f(x) = 1 + \frac{1}{3!}x^3 + \frac{1\cdot 4}{6!}x^6 + \frac{1\cdot 4\cdot 7}{9!}x^9 + \ldots \tag{7.724}$$

$$g(x) = x + \frac{2}{4!}x^4 + \frac{2\cdot 5}{7!}x^7 + \frac{2\cdot 5\cdot 8}{10!}x^{10} + \ldots \tag{7.725}$$

Now to insure that the wave function is normalizable we must also have

$$\chi(\xi(r=0)) = 0 \tag{7.726}$$

$$Ai\left(-\frac{E + V_0}{\alpha}\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3}\right) = 0 \tag{7.727}$$

This says that

$$-\frac{E_n + V_0}{\alpha}\left(\frac{2m\alpha}{\hbar^2}\right)^{1/3} = z_n = n^{th} \text{ zero of } Ai(\xi) \tag{7.728}$$

Thus, the allowed energies are given by

$$E_n = -z_n\left(\frac{\hbar^2\alpha^2}{2m}\right)^{1/3} - V_0 \tag{7.729}$$

The first three zeroes are

$$z_1 = -2.3381 \quad , \quad z_2 = -4.0879 \quad , \quad z_3 = -5.5209 \tag{7.730}$$

Programming the numerical method we described earlier for solving the Schrodinger equation allows us to determine E values For simplicity, we have chosen $\hbar = m = \alpha = V_0 = 1$ which gives the $\ell = 0$ energies as

$$E_n = -z_n\left(\frac{1}{2}\right)^{1/3} - 1 = -0.794z_n - 1 \tag{7.731}$$

The general ℓ equation is (in this case)

$$\frac{d^2\chi}{dr^2} - 2(r - 1 - E + \frac{\ell(\ell+1)}{2r^2})\chi = 0 \tag{7.732}$$

For $\ell = 0$ we theoretically expect the first three energy values 0.856, 2.246 and 3.384 and the program produces the values 0.855750, 2.24460, 3.38160 which is good agreement.

We now use the linear potential to look at quark-quark bound states at low energies.

7.6.13 Modified Linear Potential and Quark Bound States

Over the past three decades the quark model of elementary particles has had many successes. Some experimental results are the following:

1. Free (isolated) quarks have never been observed.
2. At small quark separations *color charge* exhibits behavior similar to that of ordinary charge.
3. Quark-Antiquark pairs form bound states.

We can explain some of the features of these experimental results by describing the quark-quark force by an effective potential of the form

$$V(r) = -\frac{A}{r} + Br \tag{7.733}$$

and using non-relativistic quantum mechanics.

The linear term corresponds to a long-range confining potential that is responsible for the fact that free, isolated quarks have not been observed.

In simple terms, a linear potential of this type means that it costs more and more energy as one of the quarks in a bound system attempts to separate from the other quark. When this extra energy is twice the rest energy of a quark, a new quark-antiquark pair can be produced. So instead of the quark getting free, one of the newly created quarks joins with one of the original quarks to recreate the bound pair (so it looks like nothing has happened) and the other new quark binds with the quark attempting to get free into a new meson. We never see a free quark! A lot of energy has been expended, but the outcome is the creation of a meson rather than the appearance of free quarks.

The other term, which resembles a Coulomb potential, reflects the fact that at small separations the so-called "color charge" forces behave like ordinary charge forces.

The original observations of quark-quark bound states was in 1974. The experiments involved bound states of charmed quarks called charmonium (named after the similar bound states of electrons and positrons called positronium). The observed bound-state energy levels were as shown in Table 7.7 below:

n	ℓ	$E(GeV)$
1	0	3.097
2	1	3.492
2	0	3.686
3	0	4.105
4	0	4.414

Table 7.7: Observed Bound-State Energy Levels

The Schrodinger equation for this potential becomes

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right)-\frac{\vec{L}_{op}^2}{\hbar^2 r^2}\psi\right)+(V(r)-E)\psi=0 \tag{7.734}$$

Since it is a central potential, we can write

$$\psi=R(r)Y_{\ell m}(\theta,\varphi) \tag{7.735}$$

which implies

$$-\frac{\hbar^2}{2m}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)-\frac{\ell(\ell+1)}{r^2}R\right)+(V(r)-E)R=0 \tag{7.736}$$

If we let

$$\chi(r) = rR(r) \tag{7.737}$$

we get

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + (V(r) - E + \frac{\hbar^2\ell(\ell+1)}{2mr^2})\chi = 0 \tag{7.738}$$

or

$$\frac{d^2\chi}{dr^2} - (aE - \frac{\ell(\ell+1)}{r^2} + \frac{b}{r} - cr)\chi = 0 \tag{7.739}$$

We must solve this system numerically. The same program as earlier works again with a modified potential function.

The results for a set of parameters ($a = 0.0385$, $b = 2.026$, $c = 34.65$) chosen to get the right relationship between the levels are shown in Table 7.8 below:

n	ℓ	$E(calculated)$	$E(rescaled)$
1	0	656	3.1
2	1	838	3.4
2	0	1160	3.6
3	0	1568	4.1
4	0	1916	4.4

Table 7.8: Quark Model-Numerical Results

which is a reasonably good result and indicates that validity of the model. The rescaled values are adjusted to correspond to theTable 7.7. A more exact parameter search produces almost exact agreement.

7.7 Problems

7.7.1 Position representation wave function

A system is found in the state

$$\psi(\theta, \varphi) = \sqrt{\frac{15}{8\pi}}\cos\theta\sin\theta\cos\varphi$$

(a) What are the possible values of $\hat{L}_z$ that measurement will give and with what probabilities?

(b) Determine the expectation value of $\hat{L}_x$ in this state.

7.7.2 Operator identities

Show that

(a) $\left[\vec{a}\cdot\vec{L},\vec{b}\cdot\vec{L}\right] = i\hbar\left(\vec{a}\times\vec{b}\right)\cdot\vec{L}$ holds under the assumption that $\vec{a}$ and $\vec{b}$ commute with each other and with $\vec{L}$.

(b) for any vector operator $\vec{V}(\hat{x},\hat{p})$ we have $\left[\vec{L}^2,\vec{V}\right] = 2i\hbar\left(\vec{V}\times\vec{L} - i\hbar\vec{V}\right)$.

7.7.3 More operator identities

Prove the identities

(a) $\left(\vec{\sigma}\cdot\vec{A}\right)\left(\vec{\sigma}\cdot\vec{B}\right) = \vec{A}\cdot\vec{B} + i\vec{\sigma}\cdot\left(\vec{A}\times\vec{B}\right)$

(b) $e^{i\phi\vec{S}\cdot\hat{n}/\hbar}\vec{\sigma}e^{-i\phi\vec{S}\cdot\hat{n}/\hbar} = \hat{n}(\hat{n}\cdot\vec{\sigma}) + \hat{n}\times\left[\hat{n}\times\vec{\sigma}\right]\cos\phi + \left[\hat{n}\times\vec{\sigma}\right]\sin\phi$

7.7.4 On a circle

Consider a particle of mass μ constrained to move on a circle of radius a. Show that

$$H = \frac{L^2}{2\mu a^2}$$

Solve the eigenvalue/eigenvector problem of H and interpret the degeneracy.

7.7.5 Rigid rotator

A rigid rotator is immersed in a uniform magnetic field $\vec{B} = B_0\hat{e}_z$ so that the Hamiltonian is

$$\hat{H} = \frac{\hat{L}^2}{2I} + \omega_0\hat{L}_z$$

where ω_0 is a constant. If

$$\langle\theta,\phi\,|\,\psi(0)\rangle = \sqrt{\frac{3}{4\pi}}\sin\theta\sin\phi$$

what is $\langle\theta,\phi\,|\,\psi(t)\rangle$? What is $\left\langle\hat{L}_x\right\rangle$ at time t?

7.7.6 A Wave Function

A particle is described by the wave function

$$\psi(\rho,\phi) = Ae^{-\rho^2/2\Delta}\cos^2\phi$$

Determine $P(L_z = 0)$, $P(L_z = 2\hbar)$ and $P(L_z = -2\hbar)$.

7.7.7 $L = 1$ System

Consider the following operators on a 3-dimensional Hilbert space

$$L_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\ L_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix},\ L_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

(a) What are the possible values one can obtain if L_z is measured?

(b) Take the state in which $L_z = 1$. In this state, what are $\langle L_x \rangle$, $\langle L_x^2 \rangle$ and $\Delta L_x = \sqrt{\langle L_x^2 \rangle - \langle L_x \rangle^2}$.

(c) Find the normalized eigenstates and eigenvalues of L_x in the L_z basis.

(d) If the particle is in the state with $L_z = -1$ and L_x is measured, what are the possible outcomes and their probabilities?

(e) Consider the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 1 \end{pmatrix}$$

in the L_z basis. If L_z^2 is measured and a result +1 is obtained, what is the state after the measurement? How probable was this result? If L_z is measured, what are the outcomes and respective probabilities?

(f) A particle is in a state for which the probabilities are $P(L_z = 1) = 1/4$, $P(L_z = 0) = 1/2$ and $P(L_z = -1) = 1/4$. Convince yourself that the most general, normalized state with this property is

$$|\psi\rangle = \frac{e^{i\delta_1}}{2}|L_z = 1\rangle + \frac{e^{i\delta_2}}{\sqrt{2}}|L_z = 0\rangle + \frac{e^{i\delta_3}}{2}|L_z = -1\rangle$$

We know that if $|\psi\rangle$ is a normalized state then the state $e^{i\theta}|\psi\rangle$ is a physically equivalent state. Does this mean that the factors $e^{i\delta_j}$ multiplying the L_z eigenstates are irrelevant? Calculate, for example, $P(L_x = 0)$.

7.7.8 A Spin-3/2 Particle

Consider a particle with spin angular momentum $j = 3/2$. The are four sublevels with this value of j, but different eigenvalues of j_z, $|m = 3/2\rangle$,$|m = 1/2\rangle$,$|m = -1/2\rangle$ and $|m = -3/2\rangle$.

(a) Show that the *raising operator* in this 4–dimensional space is

$$\hat{j}_+ = \hbar\left(\sqrt{3}\,|3/2\rangle\langle 1/2| + 2\,|1/2\rangle\langle -1/2| + \sqrt{3}\,|-1/2\rangle\langle -3/2|\right)$$

where the states have been labeled by the j_z quantum number.

(b) What is the lowering operator $\hat{j}_-$?

(c) What are the matrix representations of $\hat{J}_\pm$, $\hat{J}_x$, $\hat{J}_y$, $\hat{J}_z$ and $\hat{J}^2$ in the j_z basis?

(d) Check that the state

$$|\psi\rangle = \frac{1}{2\sqrt{2}}\left(\sqrt{3}\,|3/2\rangle + |1/2\rangle - |-1/2\rangle - \sqrt{3}\,|-3/2\rangle\right)$$

is an eigenstate of $\hat{J}_x$ with eigenvalue $\hbar/2$.

(e) Find the eigenstate of $\hat{J}_x$ with eigenvalue $3\hbar/2$.

(f) Suppose the particle describes the nucleus of an atom, which has a magnetic moment described by the operator $\vec{\mu} = g_N \mu_N \vec{j}$, where g_N is the *g-factor* and μ_N is the so-called *nuclear magneton*. At time $t = 0$, the system is prepared in the state given in (c). A magnetic field, pointing in the y direction of magnitude B, is suddenly turned on. What is the evolution of $\langle \hat{j}_z \rangle$ as a function of time if

$$\hat{H} = -\hat{\mu} \cdot \vec{B} = -g_N \mu_N \hbar \vec{J} \cdot \vec{B}\hat{y} = -g_N \mu_N \hbar B \hat{J}_y$$

where $\mu_N = e\hbar/2Mc$ = nuclear magneton? You will need to use the identity we derived earlier

$$e^{x\hat{A}}\hat{B}e^{-x\hat{A}} = \hat{B} + \left[\hat{A},\hat{B}\right]x + \left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\frac{x^2}{2} + \left[\hat{A},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\frac{x^3}{6} + ...$$

7.7.9 Arbitrary directions

Method #1

(a) Using the $|z+\rangle$ and $|z-\rangle$ states of a spin 1/2 particle as a basis, set up and solve as a problem in matrix mechanics the eigenvalue/eigenvector problem for $S_n = \vec{S} \cdot \hat{n}$ where the spin operator is

$$\vec{S} = \hat{S}_x \hat{e}_x + \hat{S}_y \hat{e}_y + \hat{S}_z \hat{e}_z$$

and

$$\hat{n} = \sin\theta\cos\varphi \hat{e}_x + \sin\theta\sin\varphi \hat{e}_y + \cos\theta \hat{e}_z$$

(b) Show that the eigenstates may be written as

$$|\hat{n}+\rangle = \cos\frac{\theta}{2}|z+\rangle + e^{i\varphi}\sin\frac{\theta}{2}|z-\rangle$$
$$|\hat{n}-\rangle = \sin\frac{\theta}{2}|z+\rangle - e^{i\varphi}\cos\frac{\theta}{2}|z-\rangle$$

Method #2

This part demonstrates another way to determine the eigenstates of $S_n = \vec{S} \cdot \hat{n}$.

The operator

$$\hat{R}(\theta \hat{e}_y) = e^{-i\hat{S}_y \theta/\hbar}$$

rotates spin states by an angle θ counterclockwise about the $y-$axis.

(a) Show that this rotation operator can be expressed in the form

$$\hat{R}(\theta \hat{e}_y) = \cos\frac{\theta}{2} - \frac{2i}{\hbar}\hat{S}_y \sin\frac{\theta}{2}$$

(b) Apply $\hat{R}$ to the states $|z+\rangle$ and $|z-\rangle$ to obtain the state $|\hat{n}+\rangle$ with $\varphi = 0$, that is, rotated by angle θ in the $x-z$ plane.

7.7.10 Spin state probabilities

The z-component of the spin of an electron is measured and found to be $+\hbar/2$.

(a) If a subsequent measurement is made of the $x-$component of the spin, what are the possible results?

(b) What are the probabilities of finding these various results?

(c) If the axis defining the measured spin direction makes an angle θ with respect to the original $z-$axis, what are the probabilities of various possible results?

(d) What is the expectation value of the spin measurement in (c)?

7.7.11 A spin operator

Consider a system consisting of a spin 1/2 particle.

(a) What are the eigenvalues and normalized eigenvectors of the operator

$$\hat{Q} = A\hat{s}_y + B\hat{s}_z$$

where $\hat{s}_y$ and $\hat{s}_z$ are spin angular momentum operators and A and B are real constants.

(b) Assume that the system is in a state corresponding to the larger eigenvalue. What is the probability that a measurement of $\hat{s}_y$ will yield the value $+\hbar/2$?

7.7.12 Simultaneous Measurement

A beam of particles is subject to a simultaneous measurement of the angular momentum observables $\hat{L}^2$ and $\hat{L}_z$. The measurement gives pairs of values

$$(\ell, m) = (0, 0) \quad and \quad (1, -1)$$

with probabilities 3/4 and 1/4 respectively.

(a) Reconstruct the state of the beam immediately before the measurements.

(b) The particles in the beam with $(\ell, m) = (1, -1)$ are separated out and subjected to a measurement of $\hat{L}_x$. What are the possible outcomes and their probabilities?

(c) Construct the spatial wave functions of the states that could arise from the second measurement.

7.7.13 Vector Operator

Consider a vector operator $\vec{V}$ that satisfies the commutation relation

$$[L_i, V_j] = i\hbar\varepsilon_{ijk}V_k$$

This is the definition of a *vector operator*.

(a) Prove that the operator $e^{-i\varphi L_x/\hbar}$ is a rotation operator corresponding to a rotation around the $x-$axis by an angle φ, by showing that

$$e^{-i\varphi L_x/\hbar}V_i e^{i\varphi L_x/\hbar} = R_{ij}(\varphi)V_j$$

where $R_{ij}(\varphi)$ is the corresponding rotation matrix.

(b) Prove that

$$e^{-i\pi L_x}|\ell, m\rangle = |\ell, -m\rangle$$

(c) Show that a rotation by π around the $z-$axis can also be achieved by first rotating around the $x-$axis by $\pi/2$, then rotating around the $y-$axis by π and, finally rotating back by $-\pi/2$ around the $x-$axis. In terms of rotation operators this is expressed by

$$e^{i\pi L_x/2\hbar}e^{-i\pi L_y/\hbar}e^{-i\pi L_x/2\hbar} = e^{-i\pi L_z/\hbar}$$

7.7.14 Addition of Angular Momentum

Two atoms with $J_1 = 1$ and $J_2 = 2$ are coupled, with an energy described by $\hat{H} = \varepsilon\vec{J}_1\cdot\vec{J}_2$, $\varepsilon > 0$. Determine all of the energies and degeneracies for the coupled system.

7.7.15 Spin = 1 system

We now consider a spin = 1 system.

(a) Use the spin = 1 states $|1,1\rangle$, $|1,0\rangle$ and $|1,-1\rangle$ (eigenstates of $\hat{S}_z$) as a basis to form the matrix representation (3×3) of the angular momentum operators $\hat{S}_x$, $\hat{S}_y$, $\hat{S}_z$, $\hat{S}^2$, $\hat{S}_+$, and $\hat{S}_-$.

(b) Determine the eigenstates of $\hat{S}_x$ in terms of the eigenstates $|1,1\rangle$, $|1,0\rangle$ and $|1,-1\rangle$ of $\hat{S}_z$.

(c) A spin = 1 particle is in the state

$$|\psi\rangle = \frac{1}{\sqrt{14}}\begin{pmatrix} 1 \\ 2 \\ 3i \end{pmatrix}$$

in the $\hat{S}_z$ basis.

(1) What are the probabilities that a measurement of $\hat{S}_z$ will yield the values $\hbar$, 0, or $-\hbar$ for this state? What is $\langle \hat{S}_z \rangle$?

(2) What is $\langle \hat{S}_x \rangle$ in this state?

(3) What is the probability that a measurement of $\hat{S}_x$ will yield the value $\hbar$ for this state?

(d) A particle with spin = 1 has the Hamiltonian

$$\hat{H} = A\hat{S}_z + \frac{B}{\hbar}\hat{S}_x^2$$

(1) Calculate the energy levels of this system.

(2) If, at $t = 0$, the system is in an eigenstate of $\hat{S}_x$ with eigenvalue $\hbar$, calculate the expectation value of the spin $\langle \hat{S}_Z \rangle$ at time t.

7.7.16 Deuterium Atom

Consider a deuterium atom (composed of a nucleus of spin = 1 and an electron). The electronic angular momentum is $\vec{J} = \vec{L} + \vec{S}$, where $\vec{L}$ is the orbital angular momentum of the electron and $\vec{S}$ is its spin. The total angular momentum of the atom is $\vec{F} = \vec{J} + \vec{I}$, where $\vec{I}$ is the nuclear spin. The eigenvalues of $\hat{J}^2$ and $\hat{F}^2$ are $J(J+1)\hbar^2$ and $F(F+1)\hbar^2$ respectively.

(a) What are the possible values of the quantum numbers J and F for the deuterium atom in the $1s(L = 0)$ ground state?

(b) What are the possible values of the quantum numbers J and F for a deuterium atom in the $2p(L = 1)$ excited state?

7.7.17 Spherical Harmonics

Consider a particle in a state described by

$$\psi = N(x + y + 2z)e^{-\alpha r}$$

where N is a normalization factor.

(a) Show, by rewriting the $Y_1^{\pm 1,0}$ functions in terms of x, y, z and r that

$$Y_1^{\pm 1} = \mp \left(\frac{3}{4\pi}\right)^{1/2} \frac{x \pm iy}{\sqrt{2}r} \quad , \quad Y_1^0 = \left(\frac{3}{4\pi}\right)^{1/2} \frac{z}{r}$$

(b) Using this result, show that for a particle described by ψ above

$$P(L_z = 0) = 2/3 \ , \ P(L_z = \hbar) = 1/6 \ , \ P(L_z = -\hbar) = 1/6$$

7.7.18 Spin in Magnetic Field

Suppose that we have a spin$-1/2$ particle interacting with a magnetic field via the Hamiltonian

$$\hat{H} = \begin{cases} -\vec{\mu} \cdot \vec{B} \ , \ \vec{B} = B\hat{e}_z & 0 \le t < T \\ -\vec{\mu} \cdot \vec{B} \ , \ \vec{B} = B\hat{e}_y & T \le t < 2T \end{cases}$$

where $\vec{\mu} = \mu_B \vec{\sigma}$ and the system is initially$(t = 0)$ in the state

$$|\psi(0)\rangle = |x+\rangle = \frac{1}{\sqrt{2}} \left(|z+\rangle + |z-\rangle\right)$$

Determine the probability that the state of the system at $t = 2T$ is

$$|\psi(2T)\rangle = |x+\rangle$$

in three ways:

(1) Using the Schrodinger equation (solving differential equations)

(2) Using the time development operator (using operator algebra)

(3) Using the density operator formalism

7.7.19 What happens in the Stern-Gerlach box?

An atom with spin = 1/2 passes through a Stern-Gerlach apparatus adjusted so as to transmit atoms that have their spins in the $+z$ direction. The atom spends time T in a magnetic field B in the $x-$direction.

(a) At the end of this time what is the probability that the atom would pass through a Stern-Gerlach selector for spins in the $-z$ direction?

(b) Can this probability be made equal to one, if so, how?

7.7.20 Spin = 1 particle in a magnetic field

[Use the results from Problem 9.15]. A particle with intrinsic spin = 1 is placed in a uniform magnetic field $\vec{B} = B_0\hat{e}_x$. The initial spin state is $|\psi(0)\rangle = |1,1\rangle$. Take the spin Hamiltonian to be $\hat{H} = \omega_0\hat{S}_x$ and determine the probability that the particle is in the state $|\psi(t)\rangle = |1,-1\rangle$ at time t.

7.7.21 Multiple magnetic fields

A spin–1/2 system with magnetic moment $\vec{\mu} = \mu_0\vec{\sigma}$ is located in a uniform time-independent magnetic field B_0 in the positive z–direction. For the time interval $0 < t < T$ an additional uniform time-independent field B_1 is applied in the positive x–direction. During this interval, the system is again in a uniform constant magnetic field, but of different magnitude and direction z' from the initial one. At and before $t = 0$, the system is in the $m = 1/2$ state with respect to the z–axis.

(a) At $t = 0+$, what are the amplitudes for finding the system with spin projections $m' = \pm 1/2$ with respect to the z'–axis?

(b) What is the time development of the energy eigenstates with respect to the z' direction, during the time interval $0 < t < T$?

(c) What is the probability at $t = T$ of observing the system in the spin state $m = -1/2$ along the original z–axis? [Express answers in terms of the angle θ between the z and z' axes and the frequency $\omega_0 = \mu_0 B_0/\hbar$]

7.7.22 Neutron interferometer

In a classic table-top experiment (neutron interferometer), a monochromatic neutron beam ($\lambda = 1.445$) is split by Bragg reflection at point A of an interferometer into two beams which are then recombined (after another reflection) at point D as in Figure 7.18 below:

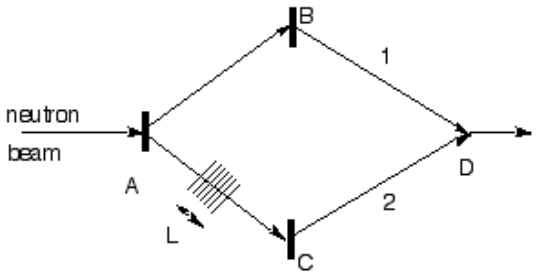


Figure 7.18: Neutron Interferometer Setup

One beam passes through a region of transverse magnetic field of strength B

(direction shown by lines)for a distance L. Assume that the two paths from A to D are identical except for the region of magnetic field.

(a) Find the explicit expression for the dependence of the intensity at point D on B, L and the neutron wavelength, with the neutron polarized parallel or anti-parallel to the magnetic field.

(b) Show that the change in the magnetic field that produces two successive maxima in the counting rates is given by

$$\Delta B = \frac{8\pi^2 \hbar c}{|e|\, g_n \lambda L}$$

where g_n (= −1.91) is the neutron magnetic moment in units of $-e\hbar/2m_n c$. This calculation was a PRL publication in 1967.

7.7.23 Magnetic Resonance

A particle of spin 1/2 and magnetic moment μ is placed in a magnetic field $\vec{B} = B_0\hat{z} + B_1\hat{x}\cos\omega t - B_1\hat{y}\sin\omega t$, which is often employed in magnetic resonance experiments. Assume that the particle has spin up along the $+z-$axis at $t = 0$ ($m_z = +1/2$). Derive the probability to find the particle with spin down ($m_z = -1/2$) at time $t > 0$.

7.7.24 More addition of angular momentum

Consider a system of two particles with $j_1 = 2$ and $j_2 = 1$. Determine the $|j, m, j_1, j_2\rangle$ states listed below in the $|j_1, m_1, j_2, m_2\rangle$ basis.

$$|3,3,j_1,j_2\rangle \ , \ |3,2,j_1,j_2\rangle \ , \ |3,1,j_1,j_2\rangle \ , \ |2,2,j_1,j_2\rangle \ , \ |2,1,j_1,j_2\rangle \ , \ |1,1,j_1,j_2\rangle$$

7.7.25 Clebsch-Gordan Coefficients

Work out the Clebsch-Gordan coefficients for the combination

$$\frac{3}{2} \otimes \frac{1}{2}$$

7.7.26 Spin−1/2 and Density Matrices

Let us consider the application of the density matrix formalism to the problem of a spin−1/2 particle in a static external magnetic field. In general, a particle with spin may carry a magnetic moment, oriented along the spin direction (by symmetry). For spin−1/2, we have that the magnetic moment (operator) is thus of the form:

$$\hat{\mu}_i = \frac{1}{2}\gamma\hat{\sigma}_i$$

where the $\hat{\sigma}_i$ are the Pauli matrices and γ is a constant giving the strength of the moment, called the gyromagnetic ratio. The term in the Hamiltonian for such a magnetic moment in an external magnetic field, $\vec{B}$ is just:

$$\hat{H} = -\vec{\mu} \cdot \vec{B}$$

The spin−1/2 particle has a spin orientation or *polarization* given by

$$\vec{P} = \langle \vec{\sigma} \rangle$$

Let us investigate the motion of the polarization vector in the external field. Recall that the expectation value of an operator may be computed from the density matrix according to

$$\left\langle \hat{A} \right\rangle = Tr\left(\hat{\rho}\hat{A}\right)$$

In addition the time evolution of the density matrix is given by

$$i\frac{\partial \hat{\rho}}{\partial t} = \left[\hat{H}(t), \hat{\rho}(t)\right]$$

Determine the time evolution $d\vec{P}/dt$ of the polarization vector. Do not make any assumption concerning the purity of the state. Discuss the physics involved in your results.

7.7.27 System of N Spin−1/2 Particle

Let us consider a system of N spin−1/2 particles per unit volume in thermal equilibrium, in an external magnetic field $\vec{B}$. In thermal equilibrium the canonical distribution applies and we have the density operator given by:

$$\hat{\rho} = \frac{e^{-\hat{H}t}}{Z}$$

where Z is the partition function given by

$$Z = Tr\left(e^{-\hat{H}t}\right)$$

Such a system of particles will tend to orient along the magnetic field, resulting in a bulk magnetization (having units of magnetic moment per unit volume), $\vec{M}$.

(a) Give an expression for this magnetization $\vec{M} = N\gamma\langle\vec{\sigma}/2\rangle$(don't work too hard to evaluate).

(b) What is the magnetization in the high-temperature limit, to lowest non-trivial order (this I want you to evaluate as completely as you can!)?

7.7.28 In a coulomb field

An electron in the Coulomb field of the proton is in the state

$$|\psi\rangle = \frac{4}{5}|1,0,0\rangle + \frac{3i}{5}|2,1,1\rangle$$

where the $|n,\ell,m\rangle$ are the standard energy eigenstates of hydrogen.

(a) What is $\langle E\rangle$ for this state? What are $\langle \hat{L}^2\rangle$, $\langle \hat{L}_x\rangle$ and $\langle \hat{L}_x\rangle$?

(b) What is $|\psi(t)\rangle$? Which, if any, of the expectation values in (a) vary with time?

7.7.29 Probabilities

(a) Calculate the probability that an electron in the ground state of hydrogen is outside the classically allowed region(defined by the classical turning points)?

(b) An electron is in the ground state of tritium, for which the nucleus is the isotope of hydrogen with one proton and two neutrons. A nuclear reaction instantaneously changes the nucleus into He^3, which consists of two protons and one neutron. Calculate the probability that the electron remains in the ground state of the new atom. Obtain a numerical answer.

7.7.30 What happens?

At the time $t = 0$ the wave function for the hydrogen atom is

$$\psi(\vec{r},0) = \frac{1}{\sqrt{10}}\left(2\psi_{100} + \psi_{210} + \sqrt{2}\psi_{211} + \sqrt{3}\psi_{21-1}\right)$$

where the subscripts are the values of the quantum numbers $(n\ell m)$. We ignore spin and any radiative transitions.

(a) What is the expectation value of the energy in this state?

(b) What is the probability of finding the system with $\ell = 1$, $m = +1$ as a function of time?

(c) What is the probability of finding an electron within $10^{-10}\,cm$ of the proton (at time $t = 0$)? A good approximate result is acceptable.

(d) Suppose a measurement is made which shows that $L = 1$, $L_x = +1$. Determine the wave function immediately after such a measurement.

7.7.31 Anisotropic Harmonic Oscillator

In three dimensions, consider a particle of mass m and potential energy

$$V(\vec{r}) = \frac{m\omega^2}{2}\left[(1-\tau)(x^2+y^2) + (1+\tau)z^2\right]$$

where $\omega \geq 0$ and $0 \leq \tau \leq 1$.

(a) What are the eigenstates of the Hamiltonian and the corresponding eigenenergies?

(b) Calculate and discuss, as functions of τ, the variation of the energy and the degree of degeneracy of the ground state and the first two excited states.

7.7.32 Exponential potential

Two particles, each of mass M, are attracted to each other by a potential

$$V(r) = -\left(\frac{g^2}{d}\right)e^{-r/d}$$

where $d = \hbar/mc$ with $mc^2 = 140\,MeV$ and $Mc^2 = 940\,MeV$.

(a) Show that for $\ell = 0$ the radial Schrodinger equation for this system can be reduced to Bessel's differential equation

$$\frac{d^2J_\rho(x)}{dx^2} + \frac{1}{x}\frac{dJ_\rho(x)}{dx} + \left(1 - \frac{\rho^2}{x^2}\right)J_\rho(x) = 0$$

by means of the change of variable $x = \alpha e^{-\beta r}$ for a suitable choice of α and β.

(b) Suppose that this system is found to have only one bound state with a binding energy of $2.2\,MeV$. Evaluate g^2/d numerically and state its units.

(c) What would the minimum value of g^2/d have to be in order to have two $\ell = 0$ bound state (keep d and M the same). A possibly useful plot is given below in Figure 7.19.

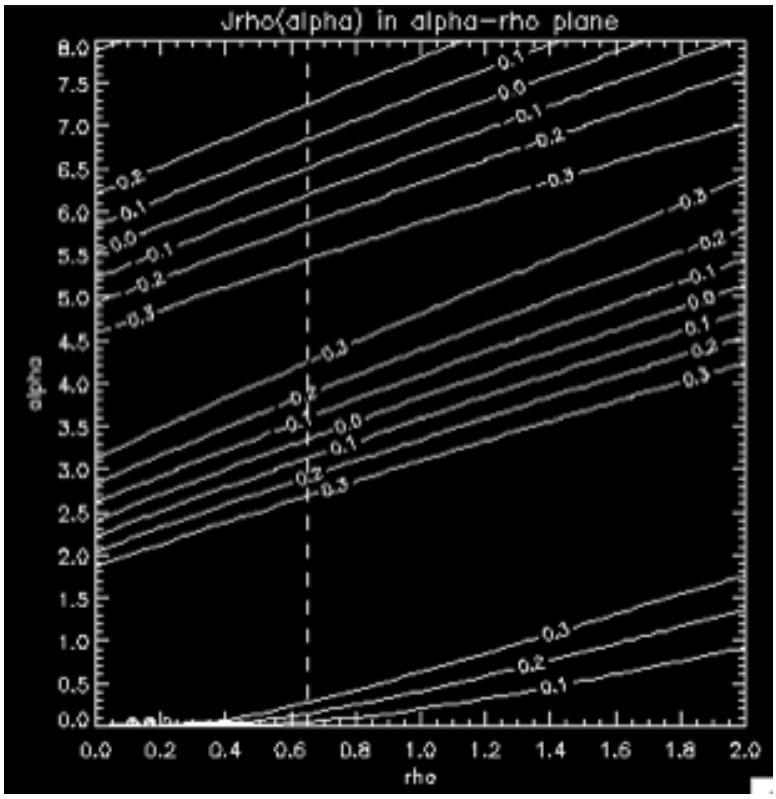


Figure 7.19: $J_\rho(\alpha)$ contours in the $\alpha - \rho$ plane

7.7.33 Bouncing electrons

An electron moves above an impenetrable conducting surface. It is attracted toward this surface by its own image charge so that classically it bounces along the surface as shown in Figure 7.20 below:

(a) Write the Schrodinger equation for the energy eigenstates and the energy eigenvalues of the electron. (Call y the distance above the surface). Ignore inertial effects of the image.

(b) What is the x and z dependence of the eigenstates?

(c) What are the remaining boundary conditions?

(d) Find the ground state and its energy? [HINT: they are closely related to those for the usual hydrogen atom]

(e) What is the complete set of discrete and/or continuous energy eigenvalues?

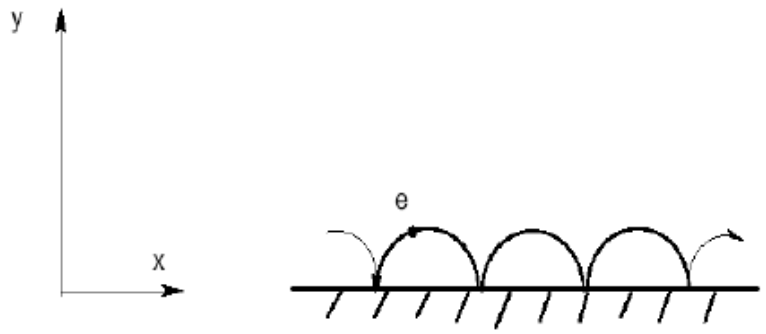


Figure 7.20: Bouncing electrons

7.7.34 Alkali Atoms

The alkali atoms have an electronic structure which resembles that of hydrogen. In particular, the spectral lines and chemical properties are largely determined by one electron(outside closed shells). A model for the potential in which this electron moves is

$$V(r) = -\frac{e^2}{r}\left(1 + \frac{b}{r}\right)$$

Solve the Schrodinger equation and calculate the energy levels.

7.7.35 Trapped between

A particle of mass m is constrained to move between two concentric impermeable spheres of radii $r = a$ and $r = b$. There is no other potential. Find the ground state energy and the normalized wave function.

7.7.36 Logarithmic potential

A particle of mass m moves in the logarithmic potential

$$V(r) = C\ell n\left(\frac{r}{r_0}\right)$$

Show that:

(a) All the eigenstates have the same mean-squared velocity. Find this mean-squared velocity. Think Virial theorem!

(b) The spacing between any two levels is independent of the mass m.

7.7.37 Spherical well

A spinless particle of mass m is subject (in 3 dimensions) to a spherically symmetric attractive square-well potential of radius r_0.

(a) What is the minimum depth of the potential needed to achieve two bound states of zero angular momentum?

(b) With a potential of this depth, what are the eigenvalues of the Hamiltonian that belong to zero total angular momentum? Solve the transcendental equation where necessary.

7.7.38 In magnetic and electric fields

A point particle of mass m and charge q moves in spatially constant crossed magnetic and electric fields $\vec{B} = B_0\hat{z}$ and $\vec{\mathcal{E}} = \mathcal{E}_0\hat{x}$.

(a) Solve for the complete energy spectrum.

(b) Find the expectation value of the velocity operator

$$\vec{v} = \frac{1}{m}\vec{p}_{mechanical}$$

in the state $\vec{p} = 0$.

7.7.39 Extra(Hidden) Dimensions

Lorentz Invariance with Extra Dimensions

If string theory is correct, we must entertain the possibility that space-time has more than four dimensions. The number of time dimensions must be kept equal to one - it seems very difficult, if not altogether impossible, to construct a consistent theory with more than one time dimension. The extra dimensions must therefore be spatial.

Can we have Lorentz invariance in worlds with more than three spatial dimensions? The answer is yes. Lorentz invariance is a concept that admits a very natural generalization to space-times with additional dimensions.

We first extend the definition of the invariant interval ds^2 to incorporate the additional space dimensions. In a world of five spatial dimensions, for example, we would write

$$ds^2 = c^2dt^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 - (dx^4)^2 - (dx^5)^2 \tag{7.740}$$

Lorentz transformations are then defined as the linear changes of coordinates that leave ds^2 invariant. This ensures that every inertial observer in the six-dimensional space-time will agree on the value of the speed of light. With more dimensions, come more Lorentz transformations. While in four-dimensional space-time we have boosts in the x^1, x^2 and x^3 directions, in this new world we have boosts along each of the five spatial dimensions. With three spatial coordinates, there are three basic spatial rotations - rotations that mix x^1 and

x^2, rotations that mix x^1 and x^3, and finally rotations that mix x^2 and x^3. The equality of the number of boosts and the number of rotations is a special feature of four-dimensional space-time. With five spatial coordinates, we have ten rotations, which is twice the number of boosts.

The higher-dimensional Lorentz invariance includes the lower-dimensional one. If nothing happens along the extra dimensions, then the restrictions of lower-dimensional Lorentz invariance apply. This is clear from equation (9.1). For motion that does not involve the extra dimensions, $dx^4 = dx^5 = 0$, and the expression for ds^2 reduces to that used in four dimensions.

Compact Extra Dimensions

It is possible for additional spatial dimensions to be undetected by low energy experiments if the dimensions are curled up into a compact space of small volume. At this point let us first try to understand what a compact dimension is. We will focus mainly on the case of one dimension. Later we will explain why small compact dimensions are hard to detect.

Consider a one-dimensional world, an infinite line, say, and let x be a coordinate along this line. For each point P along the line, there is a unique real number $x(P)$ called the $x-$coordinate of the point P. A good coordinate on this infinite line satisfies two conditions:

(1) Any two distinct points $P_1 \neq P_2$ have different coordinates $x(P_1) \neq x(P_2)$.

(2) The assignment of coordinates to points are continuous - nearby points have nearly equal coordinates.

If a choice of origin is made for this infinite line, then we can use distance from the origin to define a good coordinate. The coordinate assigned to each point is the distance from that point to the origin, with sign depending upon which side of the origin the point lies.

Imagine you live in a world with one spatial dimension. Suppose you are walking along and notice a strange pattern - the scenery repeats each time you move a distance $2\pi R$ for some value of R. If you meet your friend Phil, you see that there are Phil clones at distances $2\pi R$, $4\pi R$, $6\pi R$, ... down the line as shown in Figure 7.21 below.

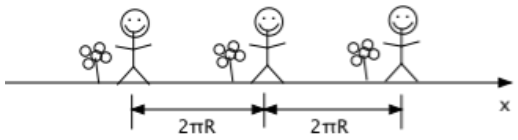


Figure 7.21: Multiple friends

In fact, there are clones up the line, as well, with the same spacing.

There is no way to distinguish an infinite line with such properties from a circle with circumference $2\pi R$. Indeed, saying that this strange line is a circle *explains* the peculiar property - there really are no Phil clones - you meet the same Phil again and again as you go around the circle!

How do we express this mathematically? We can think of the circle as an open line with an identification, that is, we declare that points with coordinates that differ by $2\pi R$ are the same point. More precisely, two points are declared to be the same point if their coordinates differ by an integer number of $2\pi R$:

$$P_1 \sim P_2 \leftrightarrow x(P_1) = x(P_2) + 2\pi R n \quad , \quad n \in \mathrm{Z} \tag{7.741}$$

This is precise, but somewhat cumbersome, notation. With no risk of confusion, we can simply write

$$x \sim x + 2\pi R \tag{7.742}$$

which should be read as *identify any two points whose coordinates differ by* $2\pi R$. With such an identification, the open line becomes a circle. The identification has turned a non-compact dimension into a compact one. It may seem to you that a line with identifications is only a complicated way to think about a circle. We will se, however, that many physical problems become clearer when we view a compact dimension as an extended one with identifications.

The interval $0 \leq x \leq 2\pi R$ is a *fundamental domain* for the identification as shown in Figure 7.22 below.

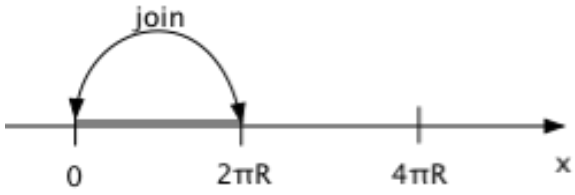


Figure 7.22: Fundamental domain

A fundamental domain is a subset of the entire space that satisfies two conditions:

(1) no two points in are identified

(2) any point in the entire space is related by the identification to some point in the fundamental domain

Whenever possible, as we did here, the fundamental domain is chosen to be a connected region. To build the space implied by the identification, we take the fundamental domain together with its boundary, and implement the identifications on the boundary. In our case, the fundamental domain together with its boundary is the segment $0 \leq x \leq 2\pi R$. In this segment we identify the point $x = 0$ with the point $x = 2\pi R$. The result is the circle.

A circle of radius R can be represented in a two-dimensional plane as the set of points that are a distance R from a point called the center of the circle. Note that the circle obtained above has been constructed directly, without the help of any two-dimensional space. For our circle, there is no point, anywhere, that represents the center of the circle. We can still speak, figuratively, of the radius R of the circle, but in our case, the radius is simply the quantity which multiplied by 2π gives the total length of the circle.

On the circle, the coordinate x is no longer a good coordinate. The coordinate x is now either multi-valued or discontinuous. This is a problem with any coordinate on a circle. Consider using angles to assign coordinates on the unit circle as shown in Figure 7.23 below.

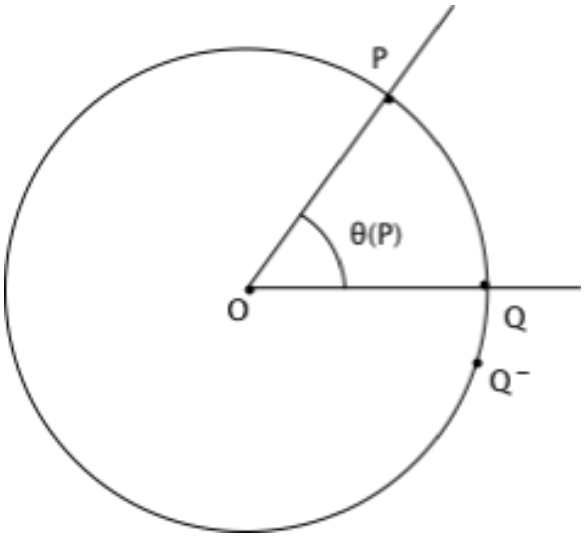


Figure 7.23: Unit circle identification

Fix a reference point Q on the circle, and let O denote the center of the circle. To any point P on the circle we assign as a coordinate the angle $\theta(P) = angle(POQ)$. This angle is naturally multi-valued. The reference point Q, for example, has $\theta(Q) = 0°$ and $\theta(Q) = 360°$. If we force angles to be single-valued by restricting $0° \leq \theta \leq 360°$, for example, then they become discontinuous. Indeed, two nearby points, Q and Q^-, then have very different angles $\theta(Q) = 0°$, while $\theta(Q^-) \sim 360°$. It is easier to work with multi-valued coordinates than it is to work with discontinuous ones.

If we have a world with several open dimensions, then we can apply the identification (9.3) to one of the dimensions, while doing nothing to the others. The dimension described by x turns into a circle, and the other dimensions remain open. It is possible, of course, to make more than one dimension compact.

Consider the example, the (x, y) plane, subject to two identifications,

$$x \sim x + 2\pi R \quad , \quad y \sim y + 2\pi R$$

It is perhaps clearer to show both coordinates simultaneously while writing the identifications. In that fashion, the two identifications are written as

$$(x, y) \sim (x + 2\pi R, y) \quad , \quad (x, y) \sim (x, y + 2\pi R) \tag{7.743}$$

The first identification implies that we can restrict our attention to $0 \leq x \leq 2\pi R$, and the second identification implies that we can restrict our attention to $0 \leq y \leq 2\pi R$. Thus, the fundamental domain can be taken to be the square region $0 \leq x, y < 2\pi R$ as shown in Figure 7.24 below.

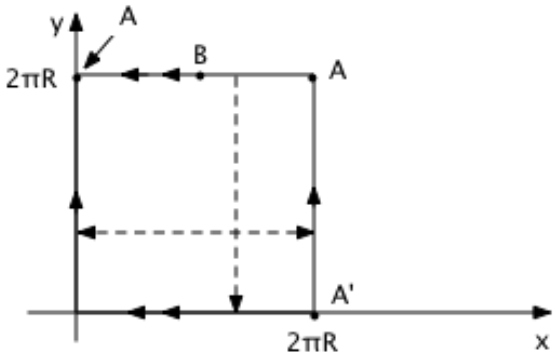


Figure 7.24: Fundamental domain = square

The identifications are indicated by the dashed lines and arrowheads. To build the space implied by the identifications, we take the fundamental domain together with its boundary, forming the full square $0 \leq x, y < 2\pi R$, and implement the identifications on the boundary. The vertical edges are identified because

they correspond to points of the form $(0, y)$ and $(2\pi R, y)$. This results in the cylinder shown in Figure 7.25 below.

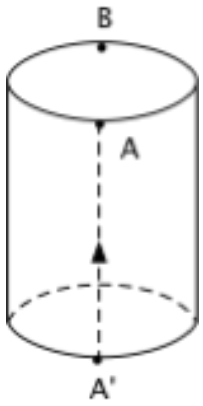


Figure 7.25: Square → cylinder

The horizontal edges are identified because they correspond to points of the form $(x, 0)$ and $(x, 2\pi R)$. The resulting space is a two-dimensional torus.

We can visualize this process in Figure 7.26 below.



Figure 7.26: 2-dimensional torus

?or in words, the torus is visualized by taking the fundamental domain (with its boundary) and gluing the vertical edges as their identification demands. The result is first (vertical) cylinder shown above (the gluing seam is the dashed line). In this cylinder, however, the bottom circle and the top circle must also be glued, since they are nothing other than the horizontal edges of the fundamental domain. To do this with paper, you must flatten the cylinder and then roll it up and glue the circles. The result looks like a flattened doughnut. With a flexible piece of garden hose, you could simply identify the two ends and obtain the familiar picture of a torus.

We have seen how to compactify coordinates using identifications. Some compact spaces are constructed in other ways. In string theory, however, compact spaces that arise from identifications are particularly easy to work with.

Sometimes identifications have fixed points, points that are related to themselves by the identification. For example, consider the real line parameterized by the coordinate x and subject to the identification $x \sim -x$. The point $x = 0$ is the unique fixed point of the identification. A fundamental domain can be chosen to be the half-line $x \geq 0$. Note that the boundary point $x = 0$ must be included in the fundamental domain. The space obtained by the above identification is in fact the fundamental domain $x \geq 0$. This is the simplest example of an *orbifold*, a space obtained by identifications that have fixed points. This orbifold is called an R^1/Z_2 orbifold. Here R^1 stands for the (one-dimensional) real line, and Z_2 describes a basic property of the identification when it is viewed as the transformation $x \to -x$ - if applied twice, it gives back the original coordinate.

Quantum Mechanics and the Square Well

The fundamental relation governing quantum mechanics is

$$[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij}$$

In three spatial dimensions the indices i and j run from 1 to 3. The generalization of quantum mechanics to higher dimensions is straightforward. With d spatial dimensions, the indices simply run over the d possible values.

To set the stage for for the analysis of small extra dimensions, let us review the standard quantum mechanics problem involving and infinite potential well.

The time-independent Schrodinger equation(in one-dimension) is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$$

In the infinite well system we have

$$V(x) = \begin{cases} 0 & \text{if } x \in (0,a) \\ \infty & \text{if } x \notin (0,a) \end{cases}$$

When $x \in (0,a)$, the Schrodinger equation becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x)$$

The boundary conditions $\psi(0) = \psi(a) = 0$ give the solutions

$$\psi_k(x) = \sqrt{\frac{2}{a}}\sin\left(\frac{k\pi x}{a}\right) \quad , \quad k = 1, 2, ..., \infty$$

The value $k = 0$ is not allowed since it would make the wave-function vanish everywhere. The corresponding energy values are

$$E_k = \frac{\hbar^2}{2m}\left(\frac{k\pi}{a}\right)^2$$

Square Well with Extra Dimensions

We now add an extra dimension to the square well problem. In addition to x, we include a dimension y that is curled up into a small circle of radius R. In other words, we make the identification

$$(x, y) \sim (x, y + 2\pi R)$$

The original dimension x has not been changed(see Figure 7.27 below). In the figure, on the left we have the original square well potential in one dimension. Here the particle lives on the the line segment shown and on the right, in the (x, y) plane the particle must remain in $0 < x < a$. The direction y is identified as $y \sim y + 2\pi R$.

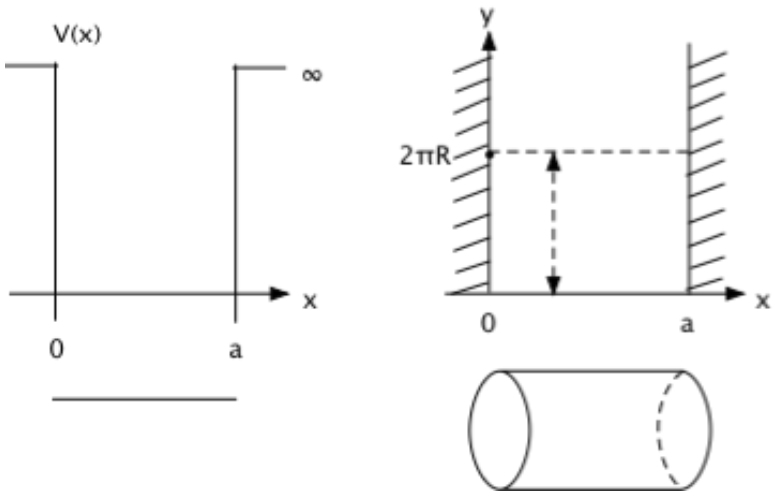


Figure 7.27: Square well with compact hidden dimension

The particle lives on a cylinder, that is, since the y direction has been turned into a circle of circumference $2\pi R$, the space where the particle moves is a cylinder. The cylinder has a length a and a circumference $2\pi R$. The potential energy $V(x,y)$ is given by

$$V(x) = \begin{cases} 0 & \text{if } x \in (0,a) \\ \infty & \text{if } x \notin (0,a) \end{cases}$$

that is, is independent of y.

We want to investigate what happens when R is small and we only do experiments at low energies. Now the only length scale in the one-dimensional infinite well system is the size a of the segment, so small R means $R << a$.

(a) Write down the Schrodinger equation for two Cartesian dimensions.

(b) Use separation of variables to find $x-$dependent and $y-$dependent solutions.

(c) Impose appropriate boundary conditions, namely, and an infinite well in the x dimension and a circle in the y dimension, to determine the allowed values of parameters in the solutions.

(d) Determine the allowed energy eigenvalues and their degeneracy.

(e) Show that the new energy levels contain the old energy levels plus additional levels.

(f) Show that when $R << a$ (a very small (compact) hidden dimension) the first new energy level appears at a very high energy. What are the experimental consequences of this result?

7.7.40 Spin$-1/2$ Particle in a D-State

A particle of spin$-1/2$ is in a D-state of orbital angular momentum. What are its possible states of total angular momentum? Suppose the single particle Hamiltonian is

$$H = A + B\vec{L}\cdot\vec{S} + C\vec{L}\cdot\vec{L}$$

What are the values of energy for each of the different states of total angular momentum in terms of the constants A, B, and C?

7.7.41 Two Stern-Gerlach Boxes

A beam of spin$-1/2$ particles traveling in the $y-$direction is sent through a Stern-Gerlach apparatus, which is aligned in the $z-$direction, and which divides the incident beam into two beams with $m = \pm 1/2$. The $m = 1/2$ beam is allowed to impinge on a second Stern-Gerlach apparatus aligned along the direction given by

$$\hat{e} = \sin\theta\hat{x} + \cos\theta\hat{z}$$

(a) Evaluate $\hat{S} = (\hbar/2)\,\vec{\sigma}\cdot\hat{e}$, where $\vec{\sigma}$ is represented by the Pauli matrices:

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

Calculate the eigenvalues of $\vec{S}$.

(b) Calculate the normalized eigenvectors of $\vec{S}$.

(c) Calculate the intensities of the two beams which emerge from the second Stern-Gerlach apparatus.

7.7.42 A Triple-Slit experiment with Electrons

A beam of spin-1/2 particles are sent into a triple slit experiment according to the figure below.

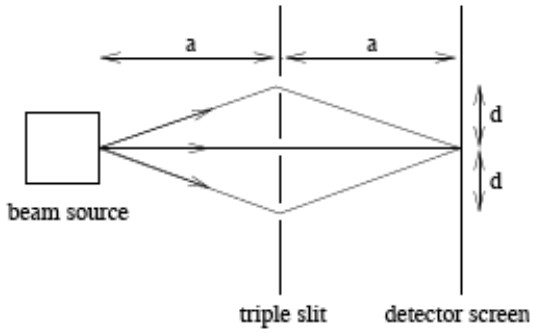


Figure 7.28: Triple-Slit Setup

Calculate the resulting intensity pattern recorded at the detector screen.

7.7.43 Cylindrical potential

The Hamiltonian is given by

$$\hat{H} = \frac{\hat{\vec{p}}^2}{2\mu} + V(\hat{\rho})$$

where $\rho = \sqrt{x^2 + y^2}$.

(a) Use symmetry arguments to establish that both $\hat{p}_z$ and $\hat{L}_z$, the $z-$component of the linear and angular momentum operators, respectively, commute with $\hat{H}$.

(b) Use the fact that $\hat{H}$, $\hat{p}_z$ and $\hat{L}_z$ have eigenstates in common to express the position space eigenfunctions of the Hamiltonian in terms of those of $\hat{p}_z$ and $\hat{L}_z$.

(c) What is the radial equation? Remember that the Laplacian in cylindrical coordinates is

$$\nabla^2\psi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\varphi^2} + \frac{\partial^2\psi}{\partial z^2}$$

A particle of mass μ is in the cylindrical potential well

$$V(\rho) = \begin{cases} 0 & \rho < a \\ \infty & \rho > a \end{cases}$$

(d) Determine the three lowest energy eigenvalues for states that also have $\hat{p}_z$ and $\hat{L}_z$ equal to zero.

(e) Determine the three lowest energy eigenvalues for states that also have $\hat{p}_z$ equal to zero. The states can have nonzero $\hat{L}_z$.

7.7.44 Crazy potentials.....

(a) A nonrelativistic particle of mass m moves in the potential

$$V(x,y,z) = A(x^2 + y^2 + 2\lambda xy) + B(z^2 + 2\mu z)$$

where $A > 0$, $B > 0$, $|\lambda| < 1$. μ is arbitrary. Find the energy eigenvalues.

(b) Now consider the following modified problem with a new potential

$$V_{new} = \begin{cases} V(x,y,z) & z > -\mu \text{ and any x and y} \\ +\infty & z < -\mu \text{ and any x and y} \end{cases}$$

Find the ground state energy.

7.7.45 Stern-Gerlach Experiment for a Spin-1 Particle

A beam of spin−1 particles, moving along the y-axis, passes through a sequence of two SG devices. The first device has its magnetic field along the z−axis and the second device has its magnetic field along the z'−axis, which points in the $x-z$ plane at an angle θ relative to the z−axis. Both devices only transmit the uppermost beam. What fraction of the particles entering the second device will leave the second device?

7.7.46 Three Spherical Harmonics

As we see, often we need to calculate an integral of the form

$$\int d\Omega Y^*_{\ell_3 m_3}(\theta,\varphi) Y_{\ell_2 m_2}(\theta,\varphi) Y_{\ell_1 m_1}(\theta,\varphi)$$

This can be interpreted as the matrix element $\langle \ell_3 m_3 | \hat{Y}^{(\ell_2)}_{m_2} | \ell_1 m_1 \rangle$, where $\hat{Y}^{(\ell_2)}_{m_2}$ is an irreducible tensor operator.

(a) Use the Wigner-Eckart theorem to determine the restrictions on the quantum numbers so that the integral does not vanish.

(b) Given the *addition rule* for Legendre polynomials:

$$P_{\ell_1}(\mu) P_{\ell_2}(\mu) = \sum_{\ell_3} \langle \ell_3 0 \mid \ell_1 0 \ell_2 0 \rangle^2 P_{\ell_3}(\mu)$$

where $\langle \ell_3 0 \mid \ell_1 0 \ell_2 0 \rangle$ is a Clebsch-Gordon coefficient. Use the Wigner-Eckart theorem to prove

$$\int d\Omega Y^*_{\ell_3 m_3}(\theta,\varphi) Y_{\ell_2 m_2}(\theta,\varphi) Y_{\ell_1 m_1}(\theta,\varphi)$$
$$= \sqrt{\frac{(2\ell_2+1)(2\ell_1+1)}{4\pi(2\ell_3+1)}} \langle \ell_3 0 \mid \ell_1 0 \ell_2 0 \rangle \langle \ell_3 m_3 \mid \ell_2 m_2 \ell_1 m_1 \rangle$$

HINT: Consider $\langle \ell_3 0 | \hat{Y}^{(\ell_2)}_0 | \ell_1 0 \rangle$.

7.7.47 Spin operators ala Dirac

Show that

$$\hat{S}_z = \frac{\hbar}{2} |z+\rangle \langle z+| - \frac{\hbar}{2} |z-\rangle \langle z-|$$
$$\hat{S}_+ = \hbar |z+\rangle \langle z-| \quad , \quad \hat{S}_- = \hbar |z-\rangle \langle z+|$$

7.7.48 Another spin = 1 system

A particle is known to have spin one. Measurements of the state of the particle yield $\langle S_x \rangle = 0 = \langle S_y \rangle$ and $\langle S_z \rangle = a$ where $0 \le a \le 1$. What is the most general possibility for the state?

7.7.49 Properties of an operator

An operator $\hat{f}$ describing the interaction of two spin−1/2 particles has the form $\hat{f} = a + b\vec{\sigma}_1 \cdot \vec{\sigma}_2$ where a and b are constants and $\vec{\sigma}_j = \sigma_{xj}\hat{\mathrm{x}} + \sigma_{yj}\hat{\mathrm{y}} + \sigma_{zj}\hat{\mathrm{z}}$ are Pauli matrix operators. The total spin angular momentum is

$$\vec{j} = \vec{j}_1 + \vec{j}_2 = \frac{\hbar}{2}(\vec{\sigma}_1 + \vec{\sigma}_2)$$

(a) Show that $\hat{f}$, $\vec{j}^2$ and $\hat{j}_z$ can be simultaneously measured.

(b) Derive the matrix representation of $\hat{f}$ in the $|j, m, j_1, j_2\rangle$ basis.

(c) Derive the matrix representation of $\hat{f}$ in the $|j_1, j_2, m_1, m_2\rangle$ basis.

7.7.50 Simple Tensor Operators/Operations

Given the tensor operator form of the particle coordinate operators

$$\vec{r} = (x, y, z);\ R_1^0 = z\ ,\ R_1^{\pm} = \mp\frac{x \pm iy}{\sqrt{2}}$$

(the subscript "1" indicates it is a rank 1 tensor), and the analogously defined particle momentum rank 1 tensor P_1^q, $q = 0, \pm 1$, calculate the commutator between each of the components and show that the results can be written in the form

$$[R_1^q, P_1^m] = \text{ simple expression}$$

7.7.51 Rotations and Tensor Operators

Using the rank 1 tensor coordinate operator in Problem 9.7.50, calculate the commutators

$$[L_{\pm}, R_1^q] \text{ and } [L_z, R_1^q]$$

where $\vec{L}$ is the standard angular momentum operator.

7.7.52 Spin Projection Operators

Show that $\mathcal{P}_1 = \frac{3}{4}\hat{I} + (\vec{S}_1 \cdot \vec{S}_2)/\hbar^2$ and $\mathcal{P}_0 = \frac{1}{4}\hat{I} - (\vec{S}_1 \cdot \vec{S}_2)/\hbar^2$ project onto the spin−1 and spin−0 spaces in $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$. Start by giving a mathematical statement of just what must be shown.

7.7.53 Two Spins in a magnetic Field

The Hamiltonian of a coupled spin system in a magnetic field is given by

$$H = A + J\frac{\vec{S}_1 \cdot \vec{S}_2}{\hbar^2} + B\frac{S_{1z} + S_{2z}}{\hbar}$$

where factors of $\hbar$ have been tossed in to make the constants A, J, B have units of energy. [J is called the *exchange constant* and B is proportional to the magnetic field].

(a) Find the eigenvalues and eigenstates of the system when one particle has spin 1 and the other has spin 1/2.

(b) Give the ordering of levels in the low field limit $J \gg B$ and the high field limit $B \gg J$ and interpret physically the result in each case.

7.7.54 Hydrogen d States

Consider the $\ell = 2$ states (for some given principal quantum number n, which is irrelevant) of the H atom, taking into account the electron spin= 1/2 (Neglect nuclear spin!).

(a) Enumerate all states in the J, M representation arising from the $\ell = 2$, s= 1/2 states.

(b) Two states have $m_j = M = +1/2$. Identify them and write them precisely in terms of the product space kets $|\ell, m_\ell; s, m_s\rangle$ using the Clebsch-Gordon coefficients.

7.7.55 The Rotation Operator for Spin–1/2

We learned that the operator

$$R_n(\Theta) = e^{-i\Theta(\vec{e}_n \cdot \hat{\mathbf{J}})/\hbar}$$

is a *rotation operator*, which rotates a vector about an axis $\vec{e}_n$ by and angle Θ. For the case of spin 1/2,

$$\hat{\mathbf{J}} = \hat{\mathbf{S}} = \frac{\hbar}{2}\hat{\vec{\sigma}} \rightarrow R_n(\Theta) = e^{-i\Theta\hat{\sigma}_n/2}$$

(a) Show that for spin 1/2

$$R_n(\Theta) = \cos\left(\frac{\Theta}{2}\right)\hat{I} - i\sin\left(\frac{\Theta}{2}\right)\hat{\sigma}_n$$

(b) Show $R_n(\Theta = 2\pi) = -\hat{I}$; Comment.

(c) Consider a series of rotations. Rotate about the y–axis by θ followed by a rotation about the z–axis by ϕ. Convince yourself that this takes the unit vector along $\vec{e}_z$ to $\vec{e}_n$. Show that up to an overall phase

$$|\uparrow_n\rangle = R_z(\phi)R_y|\uparrow_z\rangle$$

7.7.56 The Spin Singlet

Consider the entangled state of two spins

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle_A \otimes |\downarrow_z\rangle_B - |\downarrow_z\rangle_A \otimes |\uparrow_z\rangle_B\right)$$

(a) Show that (up to a phase)

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_n\rangle_A \otimes |\downarrow_n\rangle_B - |\downarrow_n\rangle_A \otimes |\uparrow_n\rangle_B\right)$$

where $|\uparrow_n\rangle$, $|\downarrow_n\rangle$ are spin spin-up and spin-down states along the direction $\vec{e}_n$. Interpret this result.

(b) Show that $\langle\Psi_{AB}|\,\hat{\sigma}_n \otimes \hat{\sigma}_{n'}\,|\Psi_{AB}\rangle = -\vec{e}_n \cdot \vec{e}_{n'}$

7.7.57 A One-Dimensional Hydrogen Atom

Consider the one-dimensional Hydrogen atom, such that the electron confined to the x axis experiences an attractive force e^2/r^2.

(a) Write down Schrodinger's equation for the electron wavefunction $\psi(x)$ and bring it to a convenient form by making the substitutions

$$a = \frac{\hbar^2}{me^2} \quad , \quad E = -\frac{\hbar^2}{2ma^2\alpha^2} \quad , \quad z = \frac{2x}{\alpha a}$$

(b) Solve the Schrodinger equation for $\psi(z)$. (You might need Mathematica, symmetry arguments plus some properties of the Confluent Hypergeometric functions or just remember earlier work).

(c) Find the three lowest allowed values of energy and the corresponding bound state wavefunctions. Plot them for suitable parameter values.

7.7.58 Electron in Hydrogen p–orbital

(a) Show that the solution of the Schrodinger equation for an electron in a p_z–orbital of a hydrogen atom

$$\psi(r,\theta,\phi) = \sqrt{\frac{3}{4\pi}} R_{n\ell}(r)\cos\theta$$

is also an eigenfunction of the square of the angular momentum operator, $\hat{L}^2$, and find the corresponding eigenvalue. Use the fact that

$$\hat{L}^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]$$

Given that the general expression for the eigenvalue of $\hat{L}^2$ is $\ell(\ell+1)\hbar^2$, what is the value of the ℓ quantum number for this electron?

(b) In general, for an electron with this ℓ quantum number, what are the allowed values of m_ℓ? (NOTE: you should not restrict yourself to a p_z electron here). What are the allowed values of s and m_s?

(c) Write down the 6 possible pairs of m_s and m_ℓ values for a single electron in a p–orbital. Given the Clebsch-Gordon coefficients shown in the table below write down all allowed coupled states $|j, m_j\rangle$ in terms of the uncoupled states $|m_\ell, m_s\rangle$. To get started here are the first three:

$$|3/2, 3/2\rangle = |1, 1/2\rangle$$
$$|3/2, 1/2\rangle = \sqrt{2/3}\,|0, 1/2\rangle + \sqrt{1/3}\,|1, -1/2\rangle$$
$$|1/2, 1/2\rangle = -\sqrt{1/3}\,|0, 1/2\rangle + \sqrt{2/3}\,|1, -1/2\rangle$$

					$\|j, m_j\rangle$		
m_{j_1}	m_{j_2}	$\|3/2, 3/2\rangle$	$\|3/2, 1/2\rangle$	$\|1/2, 1/2\rangle$	$\|3/2, -1/2\rangle$	$\|1/2, -1/2\rangle$	$\|3/2, -3/2\rangle$
1	1/2	1					
1	-1/2		$\sqrt{1/3}$	$\sqrt{2/3}$			
0	1/2		$\sqrt{2/3}$	$-\sqrt{1/3}$			
0	-1/2				$\sqrt{2/3}$	$\sqrt{1/3}$	
-1	1/2				$\sqrt{1/3}$	$-\sqrt{2/3}$	
-1	-1/2						1

Table 7.9: Clebsch-Gordon coefficients for $j_1 = 1$ and $j_2 = 1/2$

(d) The spin-orbit coupling Hamiltonian, $\hat{H}_{so}$ is given by

$$\hat{H}_{so} = \xi(\vec{r})\hat{\ell} \cdot \hat{s}$$

Show that the states with $|j, m_j\rangle$ equal to $|3/2, 3/2\rangle$, $|3/2, 1/2\rangle$ and $|1/2, 1/2\rangle$ are eigenstates of the spin-orbit coupling Hamiltonian and find the corresponding eigenvalues. Comment on which quantum numbers determine the spin-orbit energy. (HINT: there is a rather quick and easy way to do this, so if you are doing something long and tedious you might want to think again .....).

(e) The radial average of the spin-orbit Hamiltonian

$$\int_0^\infty \xi(r)|R_{n\ell}(r)|^2 r^2 dr$$

is called the spin-orbit coupling constant. It is important because it gives the average interaction of an electron in some orbital with its own spin. Given that for hydrogenic atoms

$$\xi(r) = \frac{Ze^2}{8\pi\varepsilon_0 m_e^2 c^2}\frac{1}{r^3}$$

and that for a $2p-$orbital

$$R_{n\ell}(r) = \left(\frac{Z}{a_0}\right)^{3/2}\frac{1}{2\sqrt{6}}\rho e^{-\rho/2}$$

(where $\rho = Zr/a_0$ and $a_0 = 4\pi\varepsilon_0\hbar^2/m_e c^2$) derive an expression for the spin-orbit coupling constant for an electron in a $2p-$orbital. Comment on the dependence on the atomic number Z.

(f) In the presence of a small magnetic field, B, the Hamiltonian changes by a small perturbation given by

$$\hat{H}^{(1)} = \mu_B B(\hat{\ell}_z + 2\hat{s}_z)$$

The change in energy due to a small perturbation is given in first-order perturbation theory by

$$E^{(1)} = \langle 0 | \hat{H}^{(1)} | 0 \rangle$$

where $|0\rangle$ is the unperturbed state (i.e., in this example, the state in the absence of the applied field). Use this expression to show that the change in the energies of the states in part (d) is described by

$$E^{(1)} = \mu_B B g_j m_j \tag{7.744}$$

and find the values of g_j. We will prove the perturbation theory result in the Chapter 10.

(g) Sketch an energy level diagram as a function of applied magnetic field increasing from $B = 0$ for the case where the spin-orbit interaction is stronger than the electron's interaction with the magnetic field. You can assume that the expressions you derived above for the energy changes of the three states you have been considering are applicable to the other states.

7.7.59 Quadrupole Moment Operators

The quadrupole moment operators can be written as

$$Q^{(+2)} = \sqrt{\frac{3}{8}}(x + iy)^2$$

$$Q^{(+1)} = -\sqrt{\frac{3}{2}}(x + iy)z$$

$$Q^{(0)} = \frac{1}{2}(3z^2 - r^2)$$

$$Q^{(-1)} = \sqrt{\frac{3}{2}}(x - iy)z$$

$$Q^{(-2)} = \sqrt{\frac{3}{8}}(x - iy)^2$$

Using the form of the wave function $\psi_{\ell m} = R(r)Y_m^\ell(\theta, \phi)$,

(a) Calculate $\langle \psi_{3,3} | Q^{(0)} | \psi_{3,3} \rangle$

(b) Predict all others $\langle \psi_{3,m'} | Q^{(0)} | \psi_{3,m} \rangle$ using the Wigner-Eckart theorem in terms of Clebsch-Gordon coefficients.

(c) Verify them with explicit calculations for $\langle \psi_{3,1} | Q^{(0)} | \psi_{3,0} \rangle$, $\langle \psi_{3,-1} | Q^{(0)} | \psi_{3,1} \rangle$ and $\langle \psi_{3,-2} | Q^{(0)} | \psi_{3,-3} \rangle$.

Note that we leave $\langle r^2 \rangle = \int_0^\infty r^2 \, dr R^2(r) r^2$ as an overall constant that drops out from the ratios.

7.7.60 More Clebsch-Gordon Practice

Add angular momenta $j_1 = 3/2$ and $j_2 = 1$ and work out all the Clebsch-Gordon coefficients starting from the state $|j, m\rangle = |5/2, 5/2\rangle = |3/2, 3/2\rangle \otimes |1, 1\rangle$.

7.7.61 Spherical Harmonics Properties

(a) Show that L_+ annihilates $Y_2^2 = \sqrt{15/32\pi}\sin^2\theta e^{2i\phi}$.

(b) Work out all of Y_2^m using successive applications of L_- on Y_2^2.

(c) Plot the *shapes* of Y_2^m in 3-dimensions (r, θ, ϕ) using $r = Y_2^m(\theta, \phi)$.

7.7.62 Starting Point for Shell Model of Nuclei

Consider a three-dimensional isotropic harmonic oscillator with Hamiltonian

$$H = \frac{\vec{p}^{\,2}}{2m} + \frac{1}{2}m\omega^2\vec{r}^{\,2} = \hbar\omega\left(\vec{a}^{+}\cdot\vec{a} + \frac{3}{2}\right)$$

where $\vec{p} = (\hat{p}_1, \hat{p}_2, \hat{p}_3)$, $\vec{r} = (\hat{x}_1, \hat{x}_2\hat{x}_2)$, $\vec{a} = (\hat{a}_1, \hat{a}_2, \hat{a}_3)$. We also have the commutators $[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij}$, $[\hat{x}_i, \hat{x}_j] = 0$, $[\hat{p}_i, \hat{p}_j] = 0$, $[\hat{a}_i, \hat{a}_j] = 0$, $[\hat{a}_i^+, \hat{a}_j^+] = 0$, and $[\hat{a}_i, \hat{a}_j^+] = \delta_{ij}$ Answer the following questions.

(a) Clearly, the system is spherically symmetric, and hence there is a conserved angular momentum vector. Show that $\vec{L} = \vec{r}\times\vec{p}$ commutes with the Hamiltonian.

(b) Rewrite $\vec{L}$ in terms of creation and annihilation operators.

(c) Show that $|0\rangle$ belongs to the $\ell = 0$ representation. It is called the $1S$ state.

(d) Show that the operators $\mp(a_1^+ \pm a_2^+)$ and a_3^+ form spherical tensor operators.

(e) Show that $N = 1$ states, $|1, 1, \pm1\rangle = \mp(a_1^+ \pm a_2^+)|0\rangle/\sqrt{2}$ and $|1, 1, 0\rangle = a_3^+|0\rangle$, form the $\ell = 1$ representation. (Notation is $|N, \ell, m\rangle$) It is called a $1P$ state because it is the first $P-$state.

(f) Calculate the expectation values of the quadrupole moment $Q = (3z^2 - r^2)$ for $N = \ell = 1$, $m = -1, 0, 1$ states, and verify the Wigner-Eckart theorem.

(g) There are six possible states at the $N = 2$ level. Construct the states $|2, \ell, m\rangle$ with definite $\ell = 0, 2$ and m. They are called $2S$ (because it is second $S-$state) and $1D$ (because it is the first $D-$state).

(h) How many possible states are there at the $N = 3, 4$ levels? What ℓ representations do they fall into?

(i) What can you say about general N?

(j) Verify that the operator $\Pi = e^{i\pi \vec{a}^+ \cdot \vec{a}}$ has the correct property as the parity operator by showing that $\Pi\, \vec{x}\, \Pi^+ = -\vec{x}$ and $\Pi\, \vec{p}\, \Pi^+ = -\vec{p}$.

(k) Show that $\Pi = (-1)^N$

(l) Without calculating it explicitly, show that there are no dipole transitions from the $2P$ state to the $1P$ state. As we will see in Chapter 11, this means, show that $\langle 1P | \vec{r} | 2P \rangle = 0$.

7.7.63 The Axial-Symmetric Rotor

Consider an axially symmetric object which can rotate about any of its axes but is otherwise rigid and fixed. We take the axis of symmetry to be the $z-$axis, as shown below.

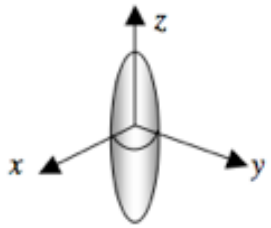


Figure 7.29: Axially Symmetric Rotor

The Hamiltonian for this system is

$$\hat{H} = \frac{\hat{L}_x^2 + \hat{L}_y^2}{2I_\perp} + \frac{\hat{L}_z^2}{2I_\parallel}$$

where $I_\perp$ and $I_\parallel$ are the moments of inertia about the principle axes.

(a) Show that the energy eigenvalues and eigenfunctions are respectively

$$E_{\ell,m} = \frac{\hbar^2}{2I_\perp}\left(\ell(\ell+1) - m^2\left(1 - \frac{I_\perp}{I_\parallel}\right)\right) \quad , \quad \psi_{\ell,m} = Y_\ell^m(\theta,\phi)$$

What are the possible values for ℓ and m? What are the degeneracies?

At $t = 0$, the system is prepared in the state

$$\psi_{\ell,m}(t=0) = \sqrt{\frac{3}{4\pi}}\frac{x}{r} = \sqrt{\frac{3}{4\pi}}\sin\theta\cos\phi$$

(b) Show that the state is normalized.

(c) Show that

$$\psi_{\ell,m}(t=0) = \frac{1}{\sqrt{2}}\left(-Y_1^1(\theta,\phi) + Y_1^{-1}(\theta,\phi)\right)$$

(d) From (c) we see that the initial state is *NOT* a single spherical harmonic (the eigenfunctions given in part (a)). Nonetheless, *show* that the wavefunction is an eigenstate of $\hat{H}$ (and thus a stationary state) and find the energy eigenvalue. Explain this.

(e) If one were to measure the observable $\hat{L}^2$ (magnitude of the angular momentum squared) and $\hat{L}_z$, what values could one find and with what probabilities?

7.7.64 Charged Particle in 2-Dimensions

Consider a charged particle on the $x-y$ plane in a constant magnetic field $\vec{B} = (0,0,B)$ with the Hamiltonian (assume $eB > 0$)

$$H = \frac{\Pi_x^2 + \Pi_y^2}{2m} \quad , \quad \Pi_i = p_i - \frac{e}{c}A_i$$

(a) Use the so-called *symmetric gauge* $\vec{A} = B(-y,x)/2$, and simplify the Hamiltonian using two annihilation operators $\hat{a}_x$ and $\hat{a}_y$ for a suitable choice of ω.

(b) Further define $\hat{a}_z = (\hat{a}_x + i\hat{a}_y)/2$ and $\hat{a}_{\bar{z}} = (\hat{a}_x - i\hat{a}_y)/2$ and then rewrite the Hamiltonian using them. General states are given in the form

$$|n,m\rangle = \frac{(\hat{a}_z^+)^n}{\sqrt{n!}}\frac{(\hat{a}_{\bar{z}}^+)^m}{\sqrt{m!}}|0,0\rangle$$

starting from the ground state where $\hat{a}_z|0,0\rangle = \hat{a}_{\bar{z}}|0,0\rangle = 0$. Show that they are Hamiltonian eigenstates of energies $\hbar\omega(2n+1)$.

(c) For an electron, what is the excitation energy when $B = 100\,kG$?

(d) Work out the wave function $\langle x,y\,|\,0,0\rangle$ in position space.

(e) $|0,m\rangle$ are all ground states. Show that their position-space wave functions are given by

$$\psi_{0,m}(z,\bar{z}) = Nz^m e^{-eB\bar{z}z/4\hbar c}$$

where $z = x + iy$ and $\bar{z} = x - iy$. Determine N.

(f) Plot the probability density of the wave function for $m = 0,3,10$ on the same scale (use ContourPlot or Plot3D in Mathematica).

(g) Assuming that the system is a circle of finite radius R, show that there are only a finite number of ground states. Work out the number approximately for large R.

(h) Show that the coherent state $e^{f\hat{a}_z^+}|0,0\rangle$ represents a near-classical cyclotron motion in position space.

7.7.65 Particle on a Circle Again

A particle of mass m is allowed to move only along a circle of radius R on a plane, $x = R\cos\theta$, $y = R\sin\theta$.

(a) Show that the Lagrangian is $L = mR^2\dot{\theta}^2/2$ and write down the canonical momentum p_θ and the Hamiltonian.

(b) Write down the Heisenberg equations of motion and solve them, (So far no representation was taken).

(c) Write down the normalized position-space wave function $\psi_k(\theta) = \langle\theta\,|\,k\rangle$ for the momentum eigenstates $\hat{p}_\theta\,|k\rangle = \hbar k\,|k\rangle$ and show that only $k = n \in \mathbb{Z}$ are allowed because of the requirement $\psi(\theta + 2\pi) = \psi(\theta)$.

(d) Show the orthonormality

$$\langle n\,|\,m\rangle = \int_0^{2\pi} \psi_n^*\psi_m\,d\theta = \delta_{nm}$$

(e) Now we introduce a constant magnetic field B inside the radius $r < d < R$ but no magnetic field outside $r > d$. Prove that the vector potential is

$$(A_x, A_y) = \begin{cases} B(-y,x)/2 & r < d \\ Bd^2(-y,x)/2r^2 & r > d \end{cases} \tag{7.745}$$

Write the Lagrangian, derive the Hamiltonian and show that the energy eigenvalues are influenced by the magnetic field even though the particle does not *see* the magnetic field directly.

7.7.66 Density Operators Redux

(a) Find a valid density operator ρ for a spin$-1/2$ system such that

$$\langle S_x\rangle = \langle S_y\rangle = \langle S_z\rangle = 0$$

Remember that for a state represented by a density operator ρ we have $\langle O_q\rangle = Tr[\rho O_q]$. Your density operator should be a 2×2 matrix with trace equal to one and eigenvalues $0 \le \lambda \le 1$. Prove that ρ you find does not correspond to a pure state and therefore cannot be represented by a state vector.

(b) Suppose that we perform a measurement of the projection operator P_i and obtain a positive result. The projection postulate (reduction postulate) for pure states says

$$|\Psi\rangle \mapsto |\Psi_i\rangle = \frac{P_i\,|\Psi\rangle}{\sqrt{\langle\Psi|\,P_i\,|\Psi\rangle}}$$

Use this result to show that in density operator notation $\rho = |\Psi\rangle\langle\Psi|$ maps to

$$\rho_i = \frac{P_i\rho P_i}{Tr[\rho P_i]}$$

7.7.67 Angular Momentum Redux

(a) Define the angular momentum operators L_x, L_y, L_z in terms of the position and momentum operators. Prove the following commutation result for these operators: $[L_x, L_y] - i\hbar L_z$.

(b) Show that the operators $L_\pm = L_x \pm iL_y$ act as raising and lowering operators for the z component of angular momentum by first calculating the commutator $[L_z, L_\pm]$.

(c) A system is in state ψ, which is an eigenstate of the operators L^2 and L_z with quantum numbers ℓ and m. Calculate the expectation values $\langle L_x \rangle$ and $\langle L_x^2 \rangle$. HINT: express L_x in terms of $L_\pm$.

(d) Hence show that L_x and L_y satisfy a general form of the uncertainty principle:

$$\langle (\Delta A)^2 \rangle \langle (\Delta B)^2 \rangle \geq -\frac{1}{4} \langle [A, B] \rangle$$

7.7.68 Wave Function Normalizability

The time-independent Schrodinger equation for a spherically symmetric potential $V(r)$ is

$$-\frac{\hbar^2}{2\mu}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) - \frac{\ell(\ell+1)}{r^2}\right] = (E - V)R$$

where $\psi = R(r)Y_\ell^m(\theta, \phi)$, so that the particle is in an eigenstate of angular momentum.

Suppose $R(r) \propto r^{-\alpha}$ and $V(r) \propto -r^{-\beta}$ near the origin. Show that $\alpha < 3/2$ is required if the wavefunction is to be normalizable, but that $\alpha < 1/2$ (or $\alpha < (3-\beta)/2$ if $\beta > 2$) is required for the expectation value of energy to be finite.

7.7.69 Currents

The quantum flux density of probability is

$$\vec{j} = \frac{i\hbar}{2m}\left(\psi \nabla \psi^* - \psi^* \nabla \psi\right)$$

It is related to the probability density $\rho = |\psi|^2$ by $\nabla \cdot \vec{j} + \dot{\rho} = 0$.

(a) Consider the case where ψ is a stationary state. Show that ρ and $\vec{j}$ are then independent of time. Show that, in one spatial dimension, $\vec{j}$ is also independent of position.

(b) Consider a 3D plane wave $\psi = Ae^{i\vec{k}\cdot\vec{x}}$. What is $\vec{j}$ in this case? Give a physical interpretation.

7.7.70 Pauli Matrices and the Bloch Vector

(a) Show that the Pauli operators

$$\sigma_x = \frac{2}{\hbar} S_x \quad , \quad \sigma_y = \frac{2}{\hbar} S_y \quad , \quad \sigma_z = \frac{2}{\hbar} S_z$$

satisfy

$$Tr[\sigma_i, \sigma_j] = 2\delta_{ij}$$

where the indices i and j can take on the values x, y or z. You will probably want to work with matrix representations of the operators.

(b) Show that the Bloch vectors for a spin−1/2 degree of freedom

$$\vec{s} = \langle S_x \rangle \hat{x} + \langle S_y \rangle \hat{y} + \langle S_z \rangle \hat{z}$$

has length$\hbar/2$ if and only if the corresponding density operator represents a pure state. You may wish to make use of the fact that an arbitrary spin−1/2 density operator can be parameterized in the following way:

$$\rho = \frac{1}{2} \left(I + \langle \sigma_x \rangle \sigma_x + \langle \sigma_y \rangle \sigma_y + \langle \sigma_z \rangle \sigma_z \right)$$

Chapter 8

Time-Independent Perturbation Theory

8.1 Nondegenerate Case

For most physically interesting systems, it is not possible to find simple, exact formulas for the energy eigenvalues and state vectors.

In many cases, however, the real system is very *similar* to another system that we can solve exactly in closed form.

Our procedure will then be to approximate the real system by the similar system and approximately calculate corrections to find the corresponding values for the real system. The approximation method that is most often used is called perturbation theory.

8.1.1 Rayleigh-Schrodinger Perturbation Theory

Consider the problem of finding the energies (eigenvalues) and state vectors (eigenvectors) for a system with a Hamiltonian $\hat{H}$ that can be written in the form

$$\hat{H} = \hat{H}_0 + \hat{V} \tag{8.1}$$

where we have already solved the system described by $\hat{H}_0$, i.e., we know that

$$\hat{H}_0 |n\rangle = \varepsilon_n |n\rangle \tag{8.2}$$

with $\langle m | n \rangle = \delta_{mn}$ (remember that the eigenvectors of a Hermitian operator always form a complete orthonormal set ... or we can make them so using the Gram-Schmidt process if degeneracy exists).

We call this solvable system the *unperturbed* or *zero-order* system.

We then assume that the extra term $\hat{V}$ is a *small correction* to $\hat{H}_0$ (that is what we mean by *similar* systems).

This says that the real physical system will have a solution given by

$$\hat{H}\,|N\rangle = E_n\,|N\rangle \tag{8.3}$$

where the real physical state vectors $|N\rangle$ are only *slightly different* from the unperturbed state vectors $|n\rangle$ and the real physical energies E_n are only *slightly different* from the unperturbed energies ε_n. Mathematically, we can express this situation by writing the perturbation in the form

$$\hat{V} = g\hat{U} \tag{8.4}$$

where g is some small ($\ll 1$) constant factor pulled out of the correction term $\hat{V}$ that characterizes its strength (or effect on the system described by $\hat{H}_0$) of the perturbation.

As $g \to 0$, each eigenvector $|N\rangle$ of $\hat{H}$ must approach the corresponding eigenvector $|n\rangle$ of $\hat{H}_0$ and each energy eigenvalue E_n of $\hat{H}$ must approach the corresponding energy eigenvalue ε_n of $\hat{H}_0$.

We can guarantee that this property is true by assuming that power series expansions in the small parameter g exist for all physically relevant quantities of the real system, i.e.,

$$\hat{H} = \hat{H}_0 + \hat{V} = \hat{H}_0 + g\hat{U} \tag{8.5}$$

$$|N\rangle = |n\rangle + g\left|N^{(1)}\right\rangle + g^2\left|N^{(2)}\right\rangle + \ldots \tag{8.6}$$

$$E_n = \varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} + \ldots \tag{8.7}$$

where the terms $\left|N^{(i)}\right\rangle$ and $E_n^{(i)}$ are called the i^{th}-order correction to the unperturbed or zero-order solution. This is a major assumption, that we cannot, in general, prove is true a priori, i.e., we cannot prove that the power series converge and therefore make sense.

The usual normalization condition we might impose would be $\langle N\,|\,N\rangle = 1$. Since the results of any calculation are independent of the choice of normalization (remember the expectation value and density operator definitions all include the norm in the denominator), we choose instead to use the normalization condition

$$\langle n\,|\,N\rangle = 1 \tag{8.8}$$

which will greatly simplify our derivations and subsequent calculations.

Substituting the power series expansion into the normalization condition we get

$$\langle n\,|\,N\rangle = 1 = \langle n\,|\,n\rangle + g\left\langle n\,\middle|\,N^{(1)}\right\rangle + g^2\left\langle n\,\middle|\,N^{(2)}\right\rangle + \ldots \tag{8.9}$$

But since we already have assumed that $\langle n\,|\,n\rangle = 1$, we must have

$$0 = g\left\langle n\,\middle|\,N^{(1)}\right\rangle + g^2\left\langle n\,\middle|\,N^{(2)}\right\rangle + ... \tag{8.10}$$

Now the only way for a power series to be *identically zero* is for the coefficient of each power of g to be separately equal to zero. This gives the result

$$\left\langle n \mid N^{(i)} \right\rangle = 0 \quad , \quad i = 1, 2, 3, 4, \ldots \tag{8.11}$$

as a direct consequence of the normalization condition, i.e., all corrections to the state vector are orthogonal to the unperturbed state vector.

We now substitute all of these power series into the original energy eigenvalue equation for $\hat{H}$:

$$\begin{aligned}
&\hat{H}\,|N\rangle = E_n\,|N\rangle \\
&\left(\hat{H}_0 + g\hat{U}\right)\left(|n\rangle + g\left|N^{(1)}\right\rangle + g^2 + \left|N^{(2)}\right\rangle ....\right) \\
&\qquad = \left(\varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} + ...\right)\left(|n\rangle + g\left|N^{(1)}\right\rangle + g^2 + \left|N^{(2)}\right\rangle ....\right)
\end{aligned}$$

We now multiply everything out and collect terms in a single power series in g. We get

$$\begin{aligned}
0 = &\left(\hat{H}_0\,|n\rangle - \varepsilon_n\,|n\rangle\right)g^0 + \left(\hat{H}_0\left|N^{(1)}\right\rangle + \hat{U}\,|n\rangle - \varepsilon_n\left|N^{(1)}\right\rangle - E_n^{(1)}\,|n\rangle\right)g^1 \\
&+ ..... \\
&+ \left(\hat{H}_0\left|N^{(k)}\right\rangle + \hat{U}\left|N^{(k-1)}\right\rangle - \varepsilon_n\left|N^{(k)}\right\rangle - E_n^{(1)}\left|N^{(k-1)}\right\rangle - .... - E_n^{(k)}\,|n\rangle\right)g^k \\
&+ .....
\end{aligned}$$

Since the power series is equal to zero, the coefficient of each power of g must be equal to zero. We get (labelling the equation by the corresponding power of g)

$$0^{th} - order \qquad \hat{H}_0\,|n\rangle = \varepsilon_n\,|n\rangle \tag{8.12}$$

which is just our original assumption = unperturbed solution.

$$1^{st} - order \qquad \hat{H}_0\left|N^{(1)}\right\rangle + \hat{U}\,|n\rangle = \varepsilon_n\left|N^{(1)}\right\rangle + E_n^{(1)}\,|n\rangle \tag{8.13}$$

.....

.....

$$\begin{aligned}
k^{th} - order \qquad &\hat{H}_0\left|N^{(k)}\right\rangle + \hat{U}\left|N^{(k-1)}\right\rangle \\
&= \varepsilon_n\left|N^{(k)}\right\rangle + E_n^{(1)}\left|N^{(k-1)}\right\rangle + ... + E_n^{(k)}\left|N^{(0)}\right\rangle
\end{aligned} \tag{8.14}$$

.....

where we have used the notation $|n\rangle = \left|N^{(0)}\right\rangle$.

Let us consider the $1^{st} - order$ equation. If we apply the linear functional $\langle n|$ we get

$$\langle n|\,\hat{H}_0\left|N^{(1)}\right\rangle + \langle n|\,\hat{U}\,|n\rangle = \langle n|\,\varepsilon_n\left|N^{(1)}\right\rangle + \langle n|\,E_n^{(1)}\,|n\rangle \tag{8.15}$$

$$\varepsilon_n\left\langle n \mid N^{(1)}\right\rangle + \langle n|\,\hat{U}\,|n\rangle = \varepsilon_n\left\langle n \mid N^{(1)}\right\rangle + E_n^{(1)}\,\langle n \mid n\rangle \tag{8.16}$$

or since $\left\langle n \middle| N^{(1)} \right\rangle = 0$ we get

$$E_n^{(1)} = \langle n|\hat{U}|n\rangle = 1^{st} - order \text{ correction to the energy}$$

$$= \text{ diagonal matrix element of } \hat{U} \text{ in the } |n\rangle \text{ (unperturbed) basis}$$

$$\text{or the expectation value of } \hat{U} \text{ in the state } |n\rangle$$

$$\text{or in the } n^{th} \text{ unperturbed state}$$

Therefore, to first order in g we have

$$\begin{aligned} E_n &= \varepsilon_n + gE_n^{(1)} = \varepsilon_n + g\langle n|\hat{U}|n\rangle \\ &= \varepsilon_n + \langle n|\hat{V}|n\rangle \end{aligned} \tag{8.17}$$

where we have reabsorbed the factor g back into the original potential energy function.

In the same manner, if we apply the linear functional $\langle n|$ to the $k^{th} - order$ equation we get

$$E_n^{(k)} = \langle n|\hat{U}\left|N^{(k-1)}\right\rangle \tag{8.18}$$

This says that, if we know the correction to the eigenvector to order $(k-1)$, then we can calculate the correction to the energy eigenvalue to order k (the next order).

Now the $k^{th} - order$ correction to the eigenvector, $\left|N^{(k)}\right\rangle$ is just another vector in the space and, hence, we can expand it as a linear combination of the $|n\rangle$ states (since they are a basis).

$$\left|N^{(k)}\right\rangle = \sum_{m\neq n} |m\rangle\left\langle m \middle| N^{(k)}\right\rangle \tag{8.19}$$

The state $|n\rangle$ is not included because $\left\langle n \middle| N^{(i)}\right\rangle = 0$ by our choice of normalization.

In order to evaluate this sum, we must find an expression for the coefficients $\left\langle m \middle| N^{(k)}\right\rangle$.

This can be done by applying the linear functional $\langle m|$, $m \neq n$ to the $k^{th} - order$ equation. We get

$$\begin{aligned} &\langle m|\hat{H}_0\left|N^{(k)}\right\rangle + \langle m|\hat{U}\left|N^{(k-1)}\right\rangle \\ &\qquad = \langle m|\varepsilon_n\left|N^{(k)}\right\rangle + \langle m|E_n^{(1)}\left|N^{(k-1)}\right\rangle + ... + \langle m|E_n^{(k)}\left|N^{(0)}\right\rangle \\ &\varepsilon_m\left\langle m \middle| N^{(k)}\right\rangle + \langle m|\hat{U}\left|N^{(k-1)}\right\rangle \\ &\qquad = \varepsilon_n\left\langle m \middle| N^{(k)}\right\rangle + E_n^{(1)}\left\langle m \middle| N^{(k-1)}\right\rangle + ... + E_n^{(k)}\left\langle m \middle| N^{(0)}\right\rangle \end{aligned}$$

If we assume that $\varepsilon_m \neq \varepsilon_n$ (we have *nondegenerate levels*) we get

$$\begin{aligned} &\left\langle m \middle| N^{(k)}\right\rangle \\ &= \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\hat{U}\left|N^{(k-1)}\right\rangle - E_n^{(1)}\left\langle m \middle| N^{(k-1)}\right\rangle - ..... - E_n^{(k)}\left\langle m \middle| N^{(0)}\right\rangle\right) \end{aligned} \tag{8.20}$$

This formula allows us to find the $k^{th}-order$ correction to the eigenvector in terms of lower order corrections to $|N\rangle$ and E_n as long as $|n\rangle$ corresponds to a nondegenerate level.

To see how this works, we will calculate the corrections to second order.

For first order, we let $k = 1$ and get

$$\begin{aligned}\left\langle m \mid N^{(1)}\right\rangle &= \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\hat{U}\left|N^{(0)}\right\rangle - E_n^{(1)}\left\langle m \mid N^{(0)}\right\rangle\right) \\ &= \frac{1}{\varepsilon_n - \varepsilon_m}\left(\langle m|\hat{U}|n\rangle - E_n^{(1)}\langle m \mid n\rangle\right) \\ &= \frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle \end{aligned} \tag{8.21}$$

which gives

$$\left|N^{(1)}\right\rangle = \sum_{m\neq n}|m\rangle\left\langle m \mid N^{(1)}\right\rangle = \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle \tag{8.22}$$

Therefore, to first order in g we have

$$|N\rangle = |n\rangle + g\left|N^{(1)}\right\rangle = |n\rangle + \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{V}|n\rangle \tag{8.23}$$

We then use $N^{(1)}$ to calculate $E_n^{(2)}$, the second order correction to the energy, using

$$\begin{aligned}E_n^{(2)} &= \langle n|\hat{U}\left|N^{(1)}\right\rangle = \langle n|\hat{U}\left(\sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{U}|n\rangle\right) \\ &= \sum_{m\neq n}\frac{\left|\langle n|\hat{U}|m\rangle\right|^2}{\varepsilon_n - \varepsilon_m}\end{aligned} \tag{8.24}$$

Therefore, to second order in g we have

$$\begin{aligned}E_n &= \varepsilon_n + gE_n^{(1)} + g^2E_n^{(2)} \\ &= \varepsilon_n + \langle n|\hat{V}|n\rangle + \sum_{m\neq n}\frac{\left|\langle n|\hat{V}|m\rangle\right|^2}{\varepsilon_n - \varepsilon_m}\end{aligned} \tag{8.25}$$

We then obtain the second order correction to the state vector in the same way.

$$\begin{aligned}
\left\langle m \,\middle|\, N^{(2)} \right\rangle &= \frac{1}{\varepsilon_n - \varepsilon_m} \left(\langle m| \hat{U} \left| N^{(1)} \right\rangle - E_n^{(1)} \left\langle m \,\middle|\, N^{(1)} \right\rangle - E_n^{(2)} \left\langle m \,\middle|\, N^{(0)} \right\rangle \right) \\
&= \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} \left| N^{(1)} \right\rangle - \frac{1}{\varepsilon_n - \varepsilon_m} \langle n| \hat{U} |n\rangle \left\langle m \,\middle|\, N^{(1)} \right\rangle \\
&= \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} \left(\sum_{k \neq n} |k\rangle \frac{1}{\varepsilon_n - \varepsilon_k} \langle k| \hat{U} |n\rangle \right) \\
&\qquad - \frac{1}{\varepsilon_n - \varepsilon_m} \langle n| \hat{U} |n\rangle \frac{1}{\varepsilon_n - \varepsilon_m} \langle m| \hat{U} |n\rangle \\
&= \sum_{k \neq n} \frac{\langle m| \hat{U} |k\rangle \langle k| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)(\varepsilon_n - \varepsilon_k)} - \frac{\langle n| \hat{U} |n\rangle \langle m| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)^2}
\end{aligned} \tag{8.26}$$

Therefore,

$$\left| N^{(2)} \right\rangle = \sum_{m \neq n} \sum_{k \neq n} |m\rangle \frac{\langle m| \hat{U} |k\rangle \langle k| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)(\varepsilon_n - \varepsilon_k)} - \sum_{m \neq n} |m\rangle \frac{\langle n| \hat{U} |n\rangle \langle m| \hat{U} |n\rangle}{(\varepsilon_n - \varepsilon_m)^2} \tag{8.27}$$

and so on.

An Example

We will now do an example where we know the exact answer so that we can compare it to the perturbation results.

We consider a 1-dimensional system represented by a perturbed harmonic oscillator where

$$\hat{H} = \hat{H}_0 + \hat{V} \tag{8.28}$$

with

$$\hat{H}_0 = \hbar\omega(\hat{a}^+ \hat{a} + \frac{1}{2}) \rightarrow \text{ harmonic oscillator} \tag{8.29}$$

$$\hat{H}_0 |n\rangle = \varepsilon_n |n\rangle = \hbar\omega(n + \frac{1}{2}) |n\rangle \tag{8.30}$$

In standard operator notation

$$\hat{H}_0 = \frac{\hat{p}^2}{2m} + \frac{1}{2} k x^2 \quad , \quad k = m\omega^2 \tag{8.31}$$

We now perturb the system with the potential energy term

$$\hat{V} = \frac{1}{2} k' x^2 \quad , \quad k' << k \tag{8.32}$$

Therefore,

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}(k + k') x^2 \tag{8.33}$$

We still have a harmonic oscillator (with a changed spring constant). This says that the new energies are given by

$$E_n = \hbar\tilde{\omega}(n + \frac{1}{2}) \tag{8.34}$$

where

$$\tilde{\omega} = \sqrt{\frac{k+k'}{m}} = \sqrt{\frac{k}{m}}\sqrt{1+\frac{k'}{k}} = \omega\sqrt{1+\frac{k'}{k}} \tag{8.35}$$

Therefore, the exact energies for the perturbed system are

$$E_n = \hbar\tilde{\omega}(n + \frac{1}{2}) = \hbar\omega\sqrt{1+\frac{k'}{k}}(n + \frac{1}{2}) \tag{8.36}$$

For $k' \ll k$, we can expand this as

$$E_n = \hbar\omega(n + \frac{1}{2})\left(1 + \frac{1}{2}\frac{k'}{k} - \frac{1}{8}\left(\frac{k'}{k}\right)^2 + \dots\right) \tag{8.37}$$

which should correspond to the perturbation calculated energy calculated to 2^{nd} order in perturbation theory. We now do the perturbation calculation.

Our earlier derivation gives

$$E_n = \varepsilon_n + \langle n|\hat{V}|n\rangle + \sum_{m\neq n}\frac{\left|\langle n|\hat{V}|m\rangle\right|^2}{\varepsilon_n - \varepsilon_m} \tag{8.38}$$

$$|N\rangle = |n\rangle + \sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n - \varepsilon_m}\langle m|\hat{V}|n\rangle \tag{8.39}$$

where

$$\hat{V} = \frac{1}{2}k'x^2 = \frac{1}{4}\frac{k'\hbar}{m\omega}(\hat{a}+\hat{a}^+)^2 \tag{8.40}$$

We need to calculate this matrix element

$$\begin{aligned}\langle m|\hat{V}|n\rangle &= \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|(\hat{a}+\hat{a}^+)^2|n\rangle = \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|\hat{a}^2+\hat{a}\hat{a}^+ + \hat{a}^+\hat{a} + (\hat{a}^+)^2|n\rangle \\ &= \frac{1}{4}\frac{k'\hbar}{m\omega}\langle m|\left(\sqrt{n(n-1)}|n-2\rangle + (n+1)|n\rangle + n|n\rangle + \sqrt{(n+1)(n+2)}|n+2\rangle\right) \\ &= \frac{1}{4}\frac{k'\hbar}{m\omega}\left(\sqrt{n(n-1)}\delta_{m,n-2} + (2n+1)\delta_{m,n} + \sqrt{(n+1)(n+2)}\delta_{m,n+2}\right)\end{aligned} \tag{8.41}$$

where we have used

$$\langle m|\hat{a}|n\rangle = \sqrt{n}\langle m \mid n-1\rangle = \sqrt{n}\delta_{m,n-1} \tag{8.42}$$

$$\langle m|\hat{a}^+|n\rangle = \sqrt{n+1}\langle m \mid n+1\rangle = \sqrt{n+1}\delta_{m,n+1} \tag{8.43}$$

Therefore,

$$\langle n|\hat{V}|n\rangle = \frac{1}{4}\frac{k'\hbar}{m\omega}(2n+1) = \hbar\omega(n+\frac{1}{2})\left(\frac{1}{2}\frac{k'}{k}\right) \tag{8.44}$$

and

$$\begin{aligned}\sum_{m\neq n}\frac{\left|\langle n|\hat{V}|m\rangle\right|^2}{\varepsilon_n-\varepsilon_m} &= \left(\frac{1}{4}\frac{k'\hbar}{m\omega}\right)^2\left[\frac{n(n-1)}{2\hbar\omega}-\frac{(n+1)(n+2)}{(-2\hbar\omega)}\right]\\ &= -\hbar\omega\left(n+\frac{1}{2}\right)\left[\frac{1}{8}\left(\frac{k'}{k}\right)^2\right]\end{aligned} \tag{8.45}$$

which then gives

$$E_n = \hbar\omega(n+\frac{1}{2})\left(1+\frac{1}{2}\frac{k'}{k}-\frac{1}{8}\left(\frac{k'}{k}\right)^2+\ldots\right) \tag{8.46}$$

in agreement with the exact result (to 2^{nd} order).

To calculate the new state vector to first order we need

$$\begin{aligned}&\sum_{m\neq n}|m\rangle\frac{1}{\varepsilon_n-\varepsilon_m}\langle m|\hat{V}|n\rangle\\ &\quad=\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{n(n-1)}}{2\hbar\omega}|n-2\rangle+\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{(n+1)(n+2)}}{(-2\hbar\omega)}|n+2\rangle\end{aligned} \tag{8.47}$$

which gives

$$|N\rangle = |n\rangle+\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{n(n-1)}}{2\hbar\omega}|n-2\rangle-\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{(n+1)(n+2)}}{2\hbar\omega}|n+2\rangle \tag{8.48}$$

What does the new ground state wave function look like? We have

$$|N=0\rangle = |0\rangle-\frac{\sqrt{2}}{2\hbar\omega}|2\rangle \tag{8.49}$$

and

$$\langle x\mid N=0\rangle = \langle x\mid 0\rangle-\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\langle x\mid 2\rangle \tag{8.50}$$

$$\psi_{N=0}(x) = \psi_0(x)-\frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\psi_2(x) \tag{8.51}$$

Now we found earlier that

$$\langle x\mid 0\rangle = \psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}e^{-\frac{m\omega x^2}{2\hbar}} \tag{8.52}$$

and

$$\langle x\mid 2\rangle = \psi_2(x) = \left(\frac{m\omega}{4\pi\hbar}\right)^{1/4}(2\frac{m\omega}{\hbar}x^2-1)e^{-\frac{m\omega x^2}{2\hbar}} \tag{8.53}$$

which gives

$$\psi_{N=0}(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}} \left(1 + \frac{1}{4}\frac{k'\hbar}{m\omega}\frac{\sqrt{2}}{2\hbar\omega}\frac{1}{\sqrt{2}}\left(1 - 2\frac{m\omega}{\hbar}x^2\right)\right)$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}} \left(1 + \frac{k'}{8k} - \frac{m\omega k'}{4\hbar k}x^2\right) \tag{8.54}$$

Since we are only changing the spring constant we should have

$$\psi_{N=0}(x) = \left(\frac{m\tilde{\omega}}{\pi\hbar}\right)^{1/4} e^{-\frac{m\tilde{\omega}x^2}{2\hbar}} = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1 + \frac{k'}{k}\right)^{1/8} e^{-\frac{m\omega x^2}{2\hbar}\sqrt{1+\frac{k'}{k}}}$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1 + \frac{k'}{8k}\right) e^{-\frac{m\omega x^2}{2\hbar}\left(1+\frac{k'}{2k}\right)}$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \left(1 + \frac{k'}{8k}\right)\left(1 - \frac{m\omega k'}{4\hbar k}x^2\right) e^{-\frac{m\omega x^2}{2\hbar}}$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}} \left(1 + \frac{k'}{8k} - \frac{m\omega k'}{4\hbar k}x^2\right) \tag{8.55}$$

which agrees with the perturbation result to this order.

The perturbation theory we have developed so far breaks down if there are any states where

$$\varepsilon_n = \varepsilon_m \text{ but } \langle m|\hat{V}|n\rangle \neq 0 \tag{8.56}$$

i.e., degenerate states with nonzero matrix elements of the perturbing potential between them.

8.2 Degenerate Case

We handle this case as follows. Suppose we have a group of k states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle \tag{8.57}$$

that are degenerate states of the unperturbed Hamiltonian $\hat{H}_0$, i.e.,

$$\hat{H}_0|n_i\rangle = \varepsilon_{n_1}|n_i\rangle \ , \ i = 1, 2, 3, 4, 5, \ldots, k \tag{8.58}$$

If $\langle n_i|\hat{V}|n_j\rangle \neq 0$ for $i \neq j$ within this set, the previous perturbation formulas will fail because the energy denominators $\varepsilon_{n_i} - \varepsilon_{n_j} \to 0$.

Remember, however, that any linear combination of the degenerate states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle \tag{8.59}$$

is also an eigenstate of $\hat{H}_0$ with the *same* energy ε_{n_1}.

Therefore, if we can choose a different set of basis states (start off with a *new set of zero-order states*) within this degenerate subspace, i.e., choose a new set of k orthogonal states (linear combinations of the old set of degenerate states)

$$|n_\alpha\rangle = \sum_{i=1}^{k} C_{\alpha i} |n_i\rangle \tag{8.60}$$

such that we have

$$\langle n_\alpha | \hat{V} | n_\beta \rangle = 0 \text{ for } \alpha \neq \beta \tag{8.61}$$

then we can use the perturbation formulas as derived earlier.

This procedure will work because the terms with zero denominators will have zero numerators and if one looks at the derivation, this means that these terms do not even appear in the final results, i.e., the zero numerators take effect before the zero denominators appear.

This condition says that the correct choice of zero-order states *within the degenerate subspace (the set of degenerate vectors)* for doing degenerate perturbation theory is that set which diagonalizes the matrix representation of $\hat{V}$ *within each* group of degenerate states.

The problem of diagonalizing $\hat{V}$ within a group of k states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \ldots \ , \ |n_k\rangle \tag{8.62}$$

is that of finding the eigenvectors and eigenvalues of the $k \times k$ matrix

$$\begin{pmatrix} \langle n_1 | \hat{V} | n_1 \rangle & \langle n_1 | \hat{V} | n_2 \rangle & \cdot & \langle n_1 | \hat{V} | n_k \rangle \\ \langle n_2 | \hat{V} | n_1 \rangle & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \langle n_k | \hat{V} | n_1 \rangle & \cdot & \cdot & \langle n_k | \hat{V} | n_k \rangle \end{pmatrix} \tag{8.63}$$

We now show that if the coefficients $C_{\alpha i}$ of the new zero-order states are just the components of the eigenvectors of this matrix, then it will be diagonalized in the degenerate subspace.

Suppose that we represent the eigenvector by the column vector

$$|n_\alpha\rangle = \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix} \tag{8.64}$$

then the statement that $|n_\alpha\rangle$ is an eigenvector of the $k \times k$ submatrix of $\hat{V}$ with eigenvalues that we write as $E^{(1)}_{n_\alpha}$ is equivalent to writing

$$\hat{V} |n_\alpha\rangle = E^{(1)}_{n_\alpha} |n_\alpha\rangle \tag{8.65}$$

or

$$\begin{pmatrix} \langle n_1|\hat{V}|n_1\rangle & \langle n_1|\hat{V}|n_2\rangle & \cdot & \langle n_1|\hat{V}|n_k\rangle \\ \langle n_2|\hat{V}|n_1\rangle & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \langle n_k|\hat{V}|n_1\rangle & \cdot & \cdot & \langle n_k|\hat{V}|n_k\rangle \end{pmatrix} \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix} = E^{(1)}_{n_\alpha} \begin{pmatrix} C_{\alpha 1} \\ C_{\alpha 2} \\ \cdot \\ C_{\alpha k} \end{pmatrix} \tag{8.66}$$

or finally

$$\sum_i \langle n_j|\hat{V}|n_i\rangle C_{\alpha i} = E^{(1)}_{n_\alpha} C_{\alpha i} \tag{8.67}$$

All of these calculations take place within the k-dimensional degenerate subspace.

We will assume that the eigenvectors are normalized, which implies that

$$\sum_i |C_{\alpha i}|^2 = 1 \tag{8.68}$$

i.e., vectors are normalized to one.

Now consider another of the new vectors given by

$$|n_\beta\rangle = \sum_{j=1}^{k} C_{\beta j}|n_j\rangle \tag{8.69}$$

We then have

$$\langle n_\beta| = \sum_{j=1}^{k} C^*_{\beta j}\langle n_j| \tag{8.70}$$

Applying the linear functional $\langle n_\beta|$ to the eigenvector/eigenvalue equation we get

$$\sum_j \sum_i \langle n_j|C^*_{\beta j}\hat{V}C_{\alpha i}|n_i\rangle = E^{(1)}_{n_\alpha} \sum_j C^*_{\beta j}C_{\alpha i} \tag{8.71}$$

Now, since the eigenvectors of any Hermitian matrix are always a complete orthonormal set (or can always be made so using the Gram-Schmidt process), the orthonormality of the new vectors says that

$$\sum_j C^*_{\beta j}C_{\alpha i} = \langle n_\beta \mid n_\alpha\rangle = \delta_{\beta\alpha} \tag{8.72}$$

Therefore the vectors

$$|n_\alpha\rangle = \sum_{i=1}^{k} C_{\alpha i}|n_i\rangle \tag{8.73}$$

satisfy

$$\left(\sum_j \langle n_j| C^*_{\beta j}\right)\hat{V}\left(\sum_i C_{\alpha i}|n_i\rangle\right) = E^{(1)}_{n_\alpha}\delta_{\alpha\beta} \tag{8.74}$$

$$\langle n_\beta|\hat{V}|n_\alpha\rangle = E^{(1)}_{n_\alpha}\delta_{\alpha\beta} \tag{8.75}$$

This says that the corresponding eigenvalue $E^{(1)}_{n_\alpha}$ of one of the new vectors is the first order energy corrections for the state $|N_\alpha\rangle$. The states

$$|n_\alpha\rangle \ , \ |n_\beta\rangle \ , \ |n_\gamma\rangle \ , \ \dots \ , \ |n_\kappa\rangle \tag{8.76}$$

are called the *new zeroth-order state vectors.*

Thus the group of states

$$|n_1\rangle \ , \ |n_2\rangle \ , \ |n_3\rangle \ , \ \dots \ , \ |n_k\rangle \tag{8.77}$$

in the presence of the perturbation $\hat{V}$ split(rearrange) into the k states

$$|n_\alpha\rangle \ , \ |n_\beta\rangle \ , \ |n_\gamma\rangle \ , \ \dots \ , \ |n_\kappa\rangle \tag{8.78}$$

which are given to first order by

$$|N_\alpha\rangle = |n_\alpha\rangle + \sum_{m\neq\alpha,\beta,..,\kappa} \frac{|m\rangle\langle m|\hat{V}|n_\alpha\rangle}{\varepsilon_{n_1} - \varepsilon_m} \tag{8.79}$$

and the energy shift to second order is

$$E_{n_\alpha} = \varepsilon_{n_1} + \langle n_\alpha|\hat{V}|n_\alpha\rangle + \sum_{m\neq\alpha,\beta,..,\kappa} \frac{\left|\langle m|\hat{V}|n_\alpha\rangle\right|^2}{\varepsilon_{n_1} - \varepsilon_m} \tag{8.80}$$

where

$$\langle n_\alpha|\hat{V}|n_\alpha\rangle = E^{(1)}_{n_\alpha} \tag{8.81}$$

is an eigenvalue of the $\hat{V}$ matrix in the degenerate subspace.

An Example

We now consider a 2–dimensional oscillator that is perturbed by a potential of the form

$$\hat{V} = \lambda\hat{x}\hat{y} \tag{8.82}$$

We then have

$$\hat{H} = \hat{H}_0 + \hat{V} \tag{8.83}$$

where

$$\hat{H}_0 = \frac{\hat{p}_x^2}{2m} + \frac{\hat{p}_y^2}{2m} + \frac{1}{2}k(x^2 + y^2) \tag{8.84}$$

As we showed earlier, using the $\hat{a}_x$ and $\hat{a}_y$ operators we get

$$\hat{H}_0 = \hbar\omega(\hat{a}_x^+\hat{a}_x + \hat{a}_y^+\hat{a}_y + 1) \tag{8.85}$$

$$\hat{H}_0|n_x, n_y\rangle = \varepsilon_{n_x,n_y}|n_x, n_y\rangle = \hbar\omega(n_x + n_y + 1)|n_x, n_y\rangle \tag{8.86}$$

$$\text{degeneracy} \ = n_x + n_y + 1 \tag{8.87}$$

and

$$\hat{V} = \lambda \frac{\hbar}{2m\omega}(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+) \tag{8.88}$$

The unperturbed ground-state is $|0,0\rangle$ with $\varepsilon_{0,0} = \hbar\omega$. It is a nondegenerate level, so we can apply the standard perturbation theory to get

$$E_0 = \varepsilon_{0,0} + \langle 0,0|\,\hat{V}\,|0,0\rangle + \sum_{\substack{m\neq 0\\ n\neq 0}} \frac{\left|\langle m,n|\,\hat{V}\,|0,0\rangle\right|}{\varepsilon_{0,0} - \varepsilon_{m,n}} \tag{8.89}$$

Now

$$\langle 0,0|\,\hat{V}\,|0,0\rangle = \frac{\lambda\hbar}{2m\omega}\,\langle 0,0|\,(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+)\,|0,0\rangle = 0 \tag{8.90}$$

and

$$\begin{aligned}\langle m,n|\,\hat{V}\,|0,0\rangle &= \frac{\lambda\hbar}{2m\omega}\,\langle m,n|\,(\hat{a}_x + \hat{a}_x^+)(\hat{a}_y + \hat{a}_y^+)\,|0,0\rangle \\ &= \frac{\lambda\hbar}{2m\omega}\,\langle m,n|\,\hat{a}_x^+\hat{a}_y^+\,|0,0\rangle = \frac{\lambda\hbar}{2m\omega}\,\langle m,n\mid 1,1\rangle \\ &= \frac{\lambda\hbar}{2m\omega}\delta_{m,1}\delta_{n,1}\end{aligned} \tag{8.91}$$

Thus, the correction to first order is zero. Calculating to second order we get

$$E_0 = \hbar\omega - \left(\frac{\lambda\hbar}{2m\omega}\right)^2 \frac{1}{2\hbar\omega} = \hbar\omega\left(1 - \frac{\lambda^2}{8m^2\omega^4}\right) \tag{8.92}$$

The next unperturbed level is 2–fold degenerate, i.e.,

$$\begin{aligned} n_x = 0, n_y = 1 &\rightarrow \varepsilon_{0,1} = 2\hbar\omega \\ n_x = 1, n_y = 0 &\rightarrow \varepsilon_{1,0} = 2\hbar\omega \end{aligned}$$

For ease of notation we will sometimes denote

$$|1,0\rangle = |a\rangle \;\text{ and }\; |0,1\rangle = |b\rangle \tag{8.93}$$

We now use degenerate perturbation theory.

The procedure is to evaluate the $\hat{V}$ matrix in the 2×2 degenerate subspace, diagonalize it and obtain the first order corrections.

The 2×2 matrix is

$$V = \begin{pmatrix} \hat{V}_{aa} & \hat{V}_{ab} \\ \hat{V}_{ba} & \hat{V}_{bb} \end{pmatrix} = \begin{pmatrix} \langle a|\,\hat{V}\,|a\rangle & \langle a|\,\hat{V}\,|b\rangle \\ \langle b|\,\hat{V}\,|a\rangle & \langle b|\,\hat{V}\,|b\rangle \end{pmatrix} \tag{8.94}$$

Now

$$\hat{V}_{aa} = \langle 1,0|\,\hat{V}\,|1,0\rangle = 0 = \langle 0,1|\,\hat{V}\,|0,1\rangle = \hat{V}_{bb} \tag{8.95}$$

$$\begin{aligned}\hat{V}_{ab} = \hat{V}_{ba} &= \langle 1,0|\,\hat{V}\,|0,1\rangle = \frac{\lambda\hbar}{2m\omega}\,\langle 1,0|\,\hat{a}_x^+\hat{a}_y\,|0,1\rangle \\ &= \frac{\lambda\hbar}{2m\omega}\end{aligned} \tag{8.96}$$

Therefore the 2×2 submatrix is

$$V = \frac{\lambda\hbar}{2m\omega}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{8.97}$$

This is simple to diagonalize. We get these results

$$|a'\rangle = \frac{1}{\sqrt{2}}(|a\rangle + |b\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \text{ eigenvalue } = +\frac{\lambda\hbar}{2m\omega} \tag{8.98}$$

$$|b'\rangle = \frac{1}{\sqrt{2}}(|a\rangle - |b\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightarrow \text{ eigenvalue } = -\frac{\lambda\hbar}{2m\omega} \tag{8.99}$$

$|a'\rangle$ and $|b'\rangle$ are the *new zeroth order state vectors* (eigenvectors of the 2×2 submatrix) and

$$\pm\frac{\lambda\hbar}{2m\omega} \tag{8.100}$$

are the corresponding first order energy corrections.

Thus, the 2–fold degenerate level splits into two nondegenerate levels as shown in Figure 8.1 below.

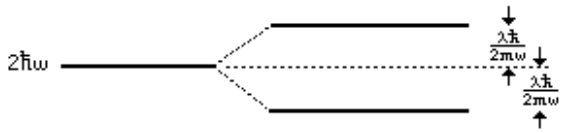


Figure 8.1: Splitting of a degenerate Level

where

$$E_{a'} = 2\hbar\omega - \frac{\lambda\hbar}{2m\omega} \tag{8.101}$$

$$E_{b'} = 2\hbar\omega + \frac{\lambda\hbar}{2m\omega} \tag{8.102}$$

$$\Delta E = \text{levelsplitting} = \frac{\lambda\hbar}{m\omega} \tag{8.103}$$

Another Example

Now let us consider a system of two spin–1/2 particles in a magnetic field. We also assume that there exists a direct spin-spin interaction so that the Hamiltonian takes the form

$$\hat{H} = \left(\alpha\vec{S}_{1,op} + \beta\vec{S}_{2,op}\right)\cdot\vec{B} + \gamma\vec{S}_{1,op}\cdot\vec{S}_{2,op} \tag{8.104}$$

If we choose $\vec{B} = B\hat{z}$ and $\gamma \gg \alpha, \beta$, then we can write

$$\hat{H} = \hat{H}_0 + \hat{V} \tag{8.105}$$

$$\hat{H}_0 = \gamma \vec{S}_{1,op} \cdot \vec{S}_{2,op} \tag{8.106}$$

$$\hat{V} = \alpha B \hat{S}_{1z} + \beta B \hat{S}_{2z} \tag{8.107}$$

We define

$$\vec{S}_{op} = \vec{S}_{1,op} + \vec{S}_{2,op} = \text{ total spin angular momentum} \tag{8.108}$$

and then we have

$$\begin{aligned} \vec{S}_{op}^2 = \vec{S}_{op} \cdot \vec{S}_{op} &= (\vec{S}_{1,op} + \vec{S}_{2,op}) \cdot (\vec{S}_{1,op} + \vec{S}_{2,op}) \\ &= \vec{S}_{1,op}^2 + \vec{S}_{2,op}^2 + 2\vec{S}_{1,op} \cdot \vec{S}_{2,op} \\ &= \frac{3}{4}\hbar^2 \hat{I} + \frac{3}{4}\hbar^2 \hat{I} + 2\vec{S}_{1,op} \cdot \vec{S}_{2,op} \end{aligned} \tag{8.109}$$

or

$$\hat{H}_0 = \gamma \vec{S}_{1,op} \cdot \vec{S}_{2,op} = \frac{\gamma}{2}\left(\vec{S}_{op}^2 - \frac{3}{2}\hbar^2 \hat{I}\right) \tag{8.110}$$

Our earlier discussion of the addition of angular momentum says that when we add two spin−1/2 angular momenta we get the resultant total angular momentum values 0 and 1, i.e.,

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1 \tag{8.111}$$

Each separate spin−1/2 system has the eigenvectors/eigenvalues

$$\vec{S}_{1,op}^2 \,|\pm\rangle = \frac{3}{4}\hbar \,|\pm\rangle \quad , \quad \vec{S}_{1z} \,|\pm\rangle = \pm\frac{\hbar}{2} \,|\pm\rangle \tag{8.112}$$

The corresponding direct-product states are

$$|{+}{+}\rangle \ , \ |{+}{-}\rangle \ , \ |{-}{+}\rangle \ , \ |{-}{-}\rangle \tag{8.113}$$

where the symbols mean

$$|{+}{-}\rangle = |{+}\rangle_1 \,|{-}\rangle_2 \tag{8.114}$$

and so on.

The total angular momentum states are (we derived them earlier) labeled as $|s, m\rangle$ where

$$\vec{S}_{op}^2 \,|s, m\rangle = \hbar^2 s(s+1) \,|s, m\rangle \tag{8.115}$$

$$\hat{S}_z \,|s, m\rangle = (\hat{S}_{1z} + \hat{S}_{2z}) \,|s, m\rangle = \pm m\hbar \,|s, m\rangle \tag{8.116}$$

They are given in terms of the direct product states by

$$|1,1\rangle = |{+}{+}\rangle = |1\rangle \tag{8.117}$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}|{+}{-}\rangle + \frac{1}{\sqrt{2}}|{-}{+}\rangle = |2\rangle \tag{8.118}$$

$$|1,-1\rangle = |{-}{-}\rangle = |3\rangle \tag{8.119}$$

$$|0,0\rangle = \frac{1}{\sqrt{2}}|{+}{-}\rangle - \frac{1}{\sqrt{2}}|{-}{+}\rangle = |4\rangle \tag{8.120}$$

The total angular momentum states are eigenstates of $\hat{H}_0$ and we use them as the unperturbed or zero-order states.

$$\hat{H}_0|1,1\rangle = \frac{\gamma\hbar^2}{4}|1,1\rangle = \varepsilon_1|1,1\rangle \tag{8.121}$$

$$\hat{H}_0|1,0\rangle = \frac{\gamma\hbar^2}{4}|1,0\rangle = \varepsilon_2|1,0\rangle \tag{8.122}$$

$$\hat{H}_0|1,-1\rangle = \frac{\gamma\hbar^2}{4}|1,-1\rangle = \varepsilon_3|1,-1\rangle \tag{8.123}$$

$$\hat{H}_0|0,0\rangle = -\frac{3\gamma\hbar^2}{4}|0,0\rangle = \varepsilon_4|0,0\rangle \tag{8.124}$$

We thus have one nondegenerate level and one 3–fold degenerate level. Now using

$$\hat{V} = \alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z} \tag{8.125}$$

we do perturbation theory on these levels.

Nondegenerate Level

First order:

$$\begin{aligned} E_4^{(1)} &= \langle 4|\hat{V}|4\rangle \\ &= \left(\frac{1}{\sqrt{2}}\langle{+}{-}| + \frac{1}{\sqrt{2}}\langle{-}{+}|\right)(\alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z})\left(\frac{1}{\sqrt{2}}|{+}{-}\rangle - \frac{1}{\sqrt{2}}|{-}{+}\rangle\right) = 0 \end{aligned}$$

Second order:

$$\begin{aligned} E_4^{(2)} = \sum_{m\neq 4}\frac{\left|\langle m|\hat{V}|4\rangle\right|^2}{\varepsilon_4-\varepsilon_m} &= \frac{\left|\langle{+}{+}|\,\alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|{+}{-}\rangle - \frac{1}{\sqrt{2}}|{-}{+}\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_1} \\ &+ \frac{\left|\left(\frac{1}{\sqrt{2}}\langle{+}{-}| + \frac{1}{\sqrt{2}}\langle{-}{+}|\right)\alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|{+}{-}\rangle - \frac{1}{\sqrt{2}}|{-}{+}\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_2} \\ &+ \frac{\left|\langle{-}{-}|\,\alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|{+}{-}\rangle - \frac{1}{\sqrt{2}}|{-}{+}\rangle\right)\right|^2}{\varepsilon_4-\varepsilon_3} \end{aligned}$$

$$E_4^{(2)} = \frac{\left|\left(\frac{1}{\sqrt{2}}\langle +-| + \frac{1}{\sqrt{2}}\langle -+|\right)\alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z}\left(\frac{1}{\sqrt{2}}|+-\rangle - \frac{1}{\sqrt{2}}|-+\rangle\right)\right|^2}{\varepsilon_4 - \varepsilon_2}$$
$$= -\frac{B^2(\alpha-\beta)^2}{\gamma}$$

Therefore the energy to second order for the non-degenerate level is

$$E_4 = \varepsilon_4 + E_4^{(1)} + E_4^{(2)} = -\frac{3\gamma\hbar^2}{4} - \frac{B^2(\alpha-\beta)^2}{\gamma} \tag{8.126}$$

Degenerate Level

In this case, the 3×3 degenerate submatrix of $\hat{V}$ is

$$\begin{pmatrix} \langle 1|\hat{V}|1\rangle & \langle 1|\hat{V}|2\rangle & \langle 1|\hat{V}|3\rangle \\ \langle 2|\hat{V}|1\rangle & \langle 2|\hat{V}|2\rangle & \langle 2|\hat{V}|3\rangle \\ \langle 3|\hat{V}|1\rangle & \langle 3|\hat{V}|2\rangle & \langle 3|\hat{V}|3\rangle \end{pmatrix} = \frac{(\alpha+\beta)\hbar B}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{8.127}$$

which is already diagonal. Since the diagonal elements are the first order energy corrections, we have (to first order)

$$E_1 = \frac{\gamma\hbar^2}{4} + \frac{(\alpha+\beta)\hbar B}{2} \tag{8.128}$$
$$E_2 = \frac{\gamma\hbar^2}{4} \tag{8.129}$$
$$E_3 = \frac{\gamma\hbar^2}{4} - \frac{(\alpha+\beta)\hbar B}{2} \tag{8.130}$$

Exact Solution

We can, in fact, solve this problem exactly and compare it to the perturbation result. We do this by choosing a new basis set (arbitrary choice made to simplify calculations) and rewriting $\hat{H}$ in terms of operators appropriate to the basis choice(that is what is meant by simplify calculations).

We then use the new basis to construct the 4×4 $\hat{H}$ matrix and then diagonalize the matrix. This method always works for a system with a small number of states.

Choose the direct product states as a basis

$$|++\rangle = |1\rangle\ ,\ |+-\rangle = |2\rangle\ ,\ |-+\rangle = |3\rangle\ ,\ |--\rangle = |4\rangle \tag{8.131}$$

Write $\hat{H}$ as (choose operators appropriate(easy to calculate) to the basis or the HOME space)

$$\hat{H} = \alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z} + \gamma\left(\hat{S}_{1z}\hat{S}_{2z} + \hat{S}_{1x}\hat{S}_{2x} + \hat{S}_{1y}\hat{S}_{2y}\right)$$
$$= \alpha B\hat{S}_{1z} + \beta B\hat{S}_{2z} + \gamma\left(\hat{S}_{1z}\hat{S}_{2z} + \frac{1}{2}\left(\hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+}\right)\right) \tag{8.132}$$

Construct the 4×4 $\hat{H}$ matrix

$$\begin{pmatrix} \langle 1|\hat{H}|1\rangle & \langle 1|\hat{H}|2\rangle & \langle 1|\hat{H}|3\rangle & \langle 1|\hat{H}|4\rangle \\ \langle 2|\hat{H}|1\rangle & \langle 2|\hat{H}|2\rangle & \langle 2|\hat{H}|3\rangle & \langle 2|\hat{H}|4\rangle \\ \langle 3|\hat{H}|1\rangle & \langle 3|\hat{H}|2\rangle & \langle 3|\hat{H}|3\rangle & \langle 3|\hat{H}|4\rangle \\ \langle 4|\hat{H}|1\rangle & \langle 4|\hat{H}|2\rangle & \langle 4|\hat{H}|3\rangle & \langle 4|\hat{H}|4\rangle \end{pmatrix} \tag{8.133}$$

Using

$$\hat{S}_z|\pm\rangle = \pm\frac{\hbar}{2}|\pm\rangle \tag{8.134}$$

$$\hat{S}_+|+\rangle = 0 = \hat{S}_-|-\rangle \tag{8.135}$$

$$\hat{S}_+|-\rangle = \hbar|+\rangle \text{ and } \hat{S}_-|+\rangle = \hbar|-\rangle \tag{8.136}$$

we get

$$\begin{pmatrix} B\hbar(\alpha+\beta)+\frac{\gamma\hbar^2}{4} & 0 & 0 & 0 \\ 0 & B\hbar(\alpha-\beta)-\frac{\gamma\hbar^2}{4} & \frac{\gamma\hbar^2}{2} & 0 \\ 0 & \frac{\gamma\hbar^2}{2} & -B\hbar(\alpha-\beta)-\frac{\gamma\hbar^2}{4} & 0 \\ 0 & 0 & 0 & -B\hbar(\alpha+\beta)+\frac{\gamma\hbar^2}{4} \end{pmatrix}$$

Diagonalizing to get the eigenvalues we find the exact energies

$$E_1 = \frac{\gamma\hbar^2}{4} + B\hbar(\alpha+\beta) \quad , \quad E_2 = -\frac{\gamma\hbar^2}{4} + \frac{1}{2}\sqrt{\gamma^2\hbar^4 + 4B^2\hbar^2(\alpha-\beta)^2}$$
$$E_3 = -\frac{\gamma\hbar^2}{4} - \frac{1}{2}\sqrt{\gamma^2\hbar^4 + 4B^2\hbar^2(\alpha-\beta)^2} \quad , \quad E_4 = \frac{\gamma\hbar^2}{4} - B\hbar(\alpha+\beta)$$

To compare to the perturbation calculation we let $B \to 0$ and we get the approximation

$$E_1 = \frac{\gamma\hbar^2}{4} + B\hbar(\alpha+\beta) \quad , \quad E_2 = \frac{\gamma\hbar^2}{4} + \frac{B^2(\alpha-\beta)^2}{\gamma}$$
$$E_3 = -\frac{3\gamma\hbar^2}{4} - \frac{B^2(\alpha-\beta)^2}{\gamma} \quad , \quad E_4 = \frac{\gamma\hbar^2}{4} - B\hbar(\alpha+\beta)$$

which agrees with the perturbation results.

8.2.1 More Ideas about Perturbation Methods

The main problem with Rayleigh-Schrodinger perturbation theory ($RSPT$) is that the form of the higher order terms becomes increasingly complex and, hence, the series is difficult to evaluate.

The Brillouin-Wigner Method

This technique allows us to see the higher order structure of the perturbation series more clearly.

Consider the energy eigenvalue equation

$$\hat{H}\left|N\right\rangle = E_n\left|N\right\rangle = \left(\hat{H}_0 + g\hat{U}\right)\left|N\right\rangle \tag{8.137}$$

Applying the linear functional $\langle m|$ we get

$$\begin{aligned}
&\left\langle m\right|\hat{H}\left|N\right\rangle = E_n\left\langle m \mid N\right\rangle = \left\langle m\right|\hat{H}_0\left|N\right\rangle + g\left\langle m\right|\hat{U}\left|N\right\rangle \\
&E_n\left\langle m \mid N\right\rangle = \varepsilon_m\left\langle m \mid N\right\rangle + g\left\langle m\right|\hat{U}\left|N\right\rangle \\
&\left(E_n - \varepsilon_m\right)\left\langle m \mid N\right\rangle = g\left\langle m\right|\hat{U}\left|N\right\rangle
\end{aligned} \tag{8.138}$$

We will use the normalization $\langle n \mid N\rangle = 1$ once again.

Now since the $|m\rangle$ states are a complete orthonormal basis we can always write

$$\begin{aligned}
\left|N\right\rangle &= \sum_m \left|m\right\rangle\left\langle m \mid N\right\rangle = \left|n\right\rangle\left\langle n \mid N\right\rangle + \sum_{m\neq n}\left|m\right\rangle\left\langle m \mid N\right\rangle \\
&= \left|n\right\rangle + \sum_{m\neq n}\left|m\right\rangle\left\langle m \mid N\right\rangle
\end{aligned} \tag{8.139}$$

Using the results above (8.139) we get

$$\left|N\right\rangle = \left|n\right\rangle + \sum_{m\neq n}\left|m\right\rangle\frac{g}{E_n - \varepsilon_m}\left\langle m\right|\hat{U}\left|N\right\rangle \tag{8.140}$$

We now develop a series expansion of $|N\rangle$ in powers of g as follows:

$0^{th} - order$:

$$\left|N\right\rangle = \left|n\right\rangle$$

$1^{st} - order$: (substitute $0^{th} - order$ result for $|N\rangle$ into general formula (8.141))

$$\left|N\right\rangle = \left|n\right\rangle + \sum_{m\neq n}\left|m\right\rangle\frac{g}{E_n - \varepsilon_m}\left\langle m\right|\hat{U}\left|n\right\rangle$$

This is *not the same result* as in RSPT since the full energy E_n remains in the denominator.

$2^{nd} - order$: (substitute $1^{st} - order$ result for $|N\rangle$ into general formula (8.141))

$$\begin{aligned}
\left|N\right\rangle &= \left|n\right\rangle + \sum_{m\neq n}\left|m\right\rangle\frac{g}{E_n - \varepsilon_m}\left\langle m\right|\hat{U}\left(\left|n\right\rangle + \sum_{j\neq n}\left|j\right\rangle\frac{g}{E_n - \varepsilon_j}\left\langle j\right|\hat{U}\left|n\right\rangle\right) \\
&= \left|n\right\rangle + \sum_{m\neq n}\left|m\right\rangle\frac{g}{E_n - \varepsilon_m}\left\langle m\right|\hat{U}\left|n\right\rangle \\
&\quad + g^2\sum_{m\neq n}\sum_{j\neq n}\left|m\right\rangle\frac{g}{E_n - \varepsilon_m}\left\langle m\right|\hat{U}\left|j\right\rangle\frac{g}{E_n - \varepsilon_j}\left\langle j\right|\hat{U}\left|n\right\rangle
\end{aligned}$$

and so on.

This is a complex power series in g since the full energy E_n remains in the denominator.

If we let $|m\rangle = |n\rangle$ in (8.139) we get

$$E_n \langle n \mid N\rangle = \varepsilon_n \langle n \mid N\rangle + g \langle n|\hat{U}|N\rangle \tag{8.141}$$

$$E_n = \varepsilon_n + g \langle n|\hat{U}|N\rangle \tag{8.142}$$

This can be expanded to give E_n as a series in powers of g, i.e., substituting the $0^{th} - order$ approximation for $|N\rangle$ gives the $1^{st} - order$ approximation for E_n and substituting the $1^{st} - order$ approximation for $|N\rangle$ gives the $2^{nd} - order$ approximation for E_n and so on.

If we substitute the $1^{st} - order$ approximation for $|N\rangle$ we get

$$\begin{aligned} E_n &= \varepsilon_n + g \langle n|\hat{U}|N\rangle \\ &= \varepsilon_n + g \langle n|\hat{U}\left(|n\rangle + \sum_{m\neq n} |m\rangle \frac{g}{E_n - \varepsilon_m} \langle m|\hat{U}|n\rangle\right) \\ &= \varepsilon_n + g \langle n|\hat{U}|n\rangle + g^2 \sum_{m\neq n} \frac{\left|\langle m|\hat{U}|n\rangle\right|^2}{E_n - \varepsilon_m} \end{aligned} \tag{8.143}$$

which is the second-order energy.

So the BWPT and the RSPT agree at each order of perturbation theory, as they must. The structure of the equations, however, is very different.

A simple example shows the very different properties of the two methods.

Consider the Hamiltonian given by $\hat{H} = \hat{H}_0 + \hat{V}$ where

$$\hat{H}_0 = \begin{pmatrix} \varepsilon_1 & 0 \\ 0 & \varepsilon_2 \end{pmatrix} \rightarrow \text{ eigenvectors } |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{and} |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8.144}$$

and

$$\hat{V} = \begin{pmatrix} 0 & \alpha \\ \alpha^* & 0 \end{pmatrix} \tag{8.145}$$

The exact energy eigenvalues are obtained by diagonalizing the $\hat{H}$ matrix

$$\hat{H} = \begin{pmatrix} \varepsilon_1 & \alpha \\ \alpha^* & \varepsilon_2 \end{pmatrix} \tag{8.146}$$

to get the characteristic equation

$$\det\begin{bmatrix} \varepsilon_1 - E & \alpha \\ \alpha^* & \varepsilon_2 - E \end{bmatrix} = 0 = E^2 - (\varepsilon_1 + \varepsilon_2)E + (\varepsilon_1\varepsilon_2 - |\alpha|^2) \tag{8.147}$$

which has solutions

$$E_1 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) + \frac{1}{2}\sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4\left|\alpha\right|^2} \tag{8.148}$$

$$E_2 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}\sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4\left|\alpha\right|^2} \tag{8.149}$$

In the *degenerate limit*, $\varepsilon_1 = \varepsilon_2 = \varepsilon$, we have the *exact* solutions

$$E_1 = \varepsilon + |\alpha| \text{ and } E_2 = \varepsilon - |\alpha| \tag{8.150}$$

Now, BWPT gives

$$\begin{aligned} E_n &= \varepsilon_n + \langle n|\hat{V}|N\rangle \\ &= \varepsilon_n + \langle n|\hat{V}|n\rangle + \sum_{m\neq n} \frac{\left|\langle m|\hat{V}|n\rangle\right|^2}{E_n - \varepsilon_m} \end{aligned} \tag{8.151}$$

or

$$E_1 = \varepsilon_n + \langle 1|\hat{V}|1\rangle + \frac{\left|\langle 2|\hat{V}|1\rangle\right|^2}{E_1 - \varepsilon_2} = \varepsilon_1 + \frac{|\alpha|^2}{E_1 - \varepsilon_2} \tag{8.152}$$

Rearranging we get

$$E_1^2 - (\varepsilon_1 + \varepsilon_2)E_1 + (\varepsilon_1\varepsilon_2 - |\alpha|^2) = 0 \tag{8.153}$$

which is the same eigenvalue equation as the exact solution. In the degenerate limit, $\varepsilon_1 = \varepsilon_2 = \varepsilon$, we have

$$E_1 = \varepsilon + |\alpha| \tag{8.154}$$

or BWPT gives the exact answer for this simple system, *even in the degenerate case.*

On the other hand, RSPT gives to second-order

$$E_1 = \varepsilon_1 + \langle 1|\hat{V}|1\rangle + \frac{|\alpha|^2}{\varepsilon_1 - \varepsilon_2} \tag{8.155}$$

This is equivalent to the exact formula to second order only!

In the degenerate limit we get nonsense since the denominator vanishes. As we saw earlier, RSPT requires an *entirely different procedure in the degenerate case.*

Notice that RSPT is in trouble even if $\varepsilon_1 \approx \varepsilon_2$ which implies that

$$\frac{|\alpha|^2}{\varepsilon_1 - \varepsilon_2} \text{ is very large} \tag{8.156}$$

and thus, that the perturbation expansion makes no sense(the terms are supposed to get smaller!).

A clever trick for handling these *almost degenerate* cases using RSPT goes as follows.

Almost Degenerate Perturbation Theory

Given $\hat{H} = \hat{H}_0 + \hat{V}$, suppose, as in the last example, we have two states $|n\rangle$ and $|m\rangle$ of the unperturbed (zero order) Hamiltonian $\hat{H}_0$ that have energies that are approximately equal.

This is a troublesome situation for RSPT because it is an expansion that includes increasing numbers of

$$\frac{1}{\varepsilon_n - \varepsilon_m} \tag{8.157}$$

terms. This implies that successive terms in the perturbation series might decrease slowly or not at all.

To develop a more rapidly converging perturbation expansion we rearrange the calculation as follows. We use the definition of the identity operator in terms of projection operators to write

$$\hat{V} = \hat{I}\hat{V}\hat{I} = \sum_{i,j} |i\rangle \langle i| \hat{V} |j\rangle \langle j| \tag{8.158}$$

We then break up $\hat{V}$ into two parts

$$\hat{V} = \hat{V}_1 + \hat{V}_2 \tag{8.159}$$

where we separate out the m and n terms into $\hat{V}_1$

$$\begin{aligned} \hat{V}_1 = {} & |m\rangle \langle m| \hat{V} |m\rangle \langle m| + |m\rangle \langle m| \hat{V} |n\rangle \langle n| \\ & + |n\rangle \langle n| \hat{V} |m\rangle \langle m| + |n\rangle \langle n| \hat{V} |n\rangle \langle n| \end{aligned} \tag{8.160}$$

and $\hat{V}_2$ = the rest of the terms. We then write

$$\hat{H} = \hat{H}_0 + \hat{V} = \hat{H}_0 + \hat{V}_1 + \hat{V}_2 = \hat{H}_0' + \hat{V}_2 \tag{8.161}$$

This new procedure then finds exact eigenvectors/eigenvalues of $\hat{H}_0'$ and treats $\hat{V}_2$ by ordinary perturbation theory.

Since the basis is orthonormal, we have from the definition of $\hat{v}_2$, i.e.,

$$\hat{V}_2 = \sum_{i,j \neq m,n} |i\rangle \langle i| \hat{V} |j\rangle \langle j| \tag{8.162}$$

which gives

$$0 = \langle n| \hat{V}_2 |n\rangle = \langle n| \hat{V}_2 |m\rangle = \langle m| \hat{V}_2 |n\rangle = \langle m| \hat{V}_2 |m\rangle \tag{8.163}$$

Thus, the closeness of the levels ε_n and ε_m will not prevent us from applying standard perturbation theory to $\hat{V}_2$, i.e., the numerators of terms with very small energy denominators, which might cause the series to diverge, all vanish

identically!

Now if $|i\rangle$ is an eigenvector of $\hat{H}_0$ (not $|m\rangle$ or $|n\rangle$), then it is also an eigenvector of $\hat{H}_0'$ since, by the orthonormality condition,

$$\hat{V}_1 |i\rangle = 0 \tag{8.164}$$

Neither $|m\rangle$ nor $|n\rangle$ is an eigenvector of $\hat{H}_0'$ however.

Now, the $\hat{H}_0$ matrix is diagonal since we are using its eigenvectors as a basis. The $\hat{H}_0'$ matrix is diagonal also except for the 2×2 submatrix

$$\begin{pmatrix} \langle m| \hat{H}_0' |m\rangle & \langle m| \hat{H}_0' |n\rangle \\ \langle n| \hat{H}_0' |m\rangle & \langle n| \hat{H}_0' |n\rangle \end{pmatrix} \tag{8.165}$$

Therefore, we can finish the solution of the problem of $\hat{H}_0'$ by diagonalizing this 2×2 matrix.

Diagonalizing the 2×2 matrix is equivalent to finding the linear combinations (or new zero order eigenvectors)

$$\alpha |n\rangle + \beta |m\rangle \tag{8.166}$$

that diagonalize the 2×2 matrix.

We must have

$$\hat{H}_0' (\alpha |n\rangle + \beta |m\rangle) = \left(\hat{H}_0 + \hat{V}_1\right) (\alpha |n\rangle + \beta |m\rangle) = E' (\alpha |n\rangle + \beta |m\rangle) \tag{8.167}$$

Now

$$\begin{aligned} \hat{H}_0' |n\rangle &= \hat{H}_0 |n\rangle + \hat{V}_1 |n\rangle \\ &= \varepsilon_n |n\rangle + |m\rangle \langle m| \hat{V} |m\rangle \langle m \mid n\rangle + |m\rangle \langle m| \hat{V} |n\rangle \langle n \mid n\rangle \\ &\qquad + |n\rangle \langle n| \hat{V} |m\rangle \langle m \mid n\rangle + |n\rangle \langle n| \hat{V} |n\rangle \langle n \mid n\rangle \\ &= \varepsilon_n |n\rangle + |m\rangle \langle m| \hat{V} |n\rangle + |n\rangle \langle n| \hat{V} |n\rangle \\ &= \left(\varepsilon_n + \langle n| \hat{V} |n\rangle\right) |n\rangle + |m\rangle \langle m| \hat{V} |n\rangle \\ &= E_n^{(1)} |n\rangle + \langle m| \hat{V} |n\rangle |m\rangle \end{aligned} \tag{8.168}$$

and similarly

$$\hat{H}_0' |m\rangle = E_m^{(1)} |m\rangle + \langle n| \hat{V} |m\rangle |n\rangle \tag{8.169}$$

Therefore, we get

$$\alpha \left(E_n^{(1)} |n\rangle + \langle m| \hat{V} |n\rangle |m\rangle\right) + \beta \left(E_m^{(1)} |m\rangle + \langle n| \hat{V} |m\rangle |n\rangle\right) = E' (\alpha |n\rangle + \beta |m\rangle) \tag{8.170}$$

Since the state vectors $|m\rangle$ and $|n\rangle$ are orthogonal, we must then have

$$E_n^{(1)}\alpha + \langle n|\hat{V}|m\rangle\beta = E'\alpha \tag{8.171}$$

$$\langle m|\hat{V}|n\rangle\alpha + E_m^{(1)}\beta = E'\beta \tag{8.172}$$

These equations have two solutions, namely,

$$\alpha = \langle n|\hat{V}|m\rangle \tag{8.173}$$

$$\beta_\pm = \frac{E_m^{(1)} - E_n^{(1)}}{2} \pm \sqrt{\left(\frac{E_m^{(1)} - E_n^{(1)}}{2}\right)^2 + \left|\langle n|\hat{V}|m\rangle\right|^2} \tag{8.174}$$

which then give the results

$$E_\pm = \frac{E_m^{(1)} + E_n^{(1)}}{2} \pm \sqrt{\left(\frac{E_m^{(1)} - E_n^{(1)}}{2}\right)^2 + \left|\langle n|\hat{V}|m\rangle\right|^2} \tag{8.175}$$

We then know all the eigenvectors/eigenvalues of $\hat{H}_0'$ (we know all of the unperturbed states) and we can deal with $\hat{V}_2$ by perturbation theory.

Finally, let us introduce another interesting idea.

Fake Degenerate Perturbation Theory

Consider the problem of finding the energy eigenvalues and state vectors for a system with a Hamiltonian $\hat{H} = \hat{H}_0 + \hat{V}$ where we know the solution to the zero-order system

$$\hat{H}_0|n\rangle = \varepsilon_n|n\rangle \tag{8.176}$$

We will assume that the unperturbed states are nondegenerate.

Now define

$$E_{average} = E_{av} = \frac{1}{n}\sum_n \varepsilon_n \tag{8.177}$$

and redefine

$$\hat{H} = E_{av}\hat{I} + \hat{U} \tag{8.178}$$

where

$$\hat{U} = \hat{H}_0 - E_{av}\hat{I} + \hat{V} \tag{8.179}$$

If the energies associated with $\hat{U}$ are small corrections to E_{av}, then we can use degenerate perturbation theory to solve this problem, i.e., the new unperturbed Hamiltonian is

$$\hat{H}_0' = E_{av}\hat{I} \tag{8.180}$$

and all of its levels are degenerate in zero order.

The problem is then solved by diagonalizing the $\hat{U}$ matrix in the basis of $\hat{H}_0$ states.

8.2.2 Thoughts on Degeneracy and Position Representation

When we derived the energy spectrum of the hydrogen atom we found that the states were labeled by three quantum numbers

$$|\psi\rangle = |n\ell m\rangle \tag{8.181}$$

where

$n =$ the radial quantum number
$\ell =$ orbital angular momentum quantum number
$m = z -$ component of orbital angular momentum quantum number

and we found that

$n = 1, 2, 3, ...$
$\ell = 0, 1, 2, ..., n-1$ for a given value of n
$m = -\ell, -\ell+1, ...\ell-1, \ell$ for a given value of ℓ

The energy eigenvalues, however, did not depend on ℓ or m. We found that

$$E_{n\ell m} = E_n = -\frac{e^2}{2a_0 n^2} \tag{8.182}$$

Therefore, each energy level had a degeneracy given by

$$\begin{aligned} g &= \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{\ell} 1 = \sum_{\ell=0}^{n-1} (2\ell+1) = 2\sum_{\ell=0}^{n-1} \ell + \sum_{\ell=0}^{n-1} 1 \\ &= 2\frac{n(n-1)}{2} + n = n^2 \end{aligned} \tag{8.183}$$

The degeneracy with respect to m is understandable since no direction is explicitly preferred in the Hamiltonian. We expect that this degeneracy will disappear as soon as a preferred direction is added to the Hamiltonian, as in the case of external electric(Stark effect) or magnetic(Zeeman effect) fields.

The degeneracy with respect to ℓ is a property peculiar to the pure $1/r$ Coulomb potential. Since no other atom except hydrogen has a pure Coulomb potential, we expect this degeneracy to vanish in other atoms.

Such a degeneracy is called an *accidental degeneracy.*

Now the electron and proton making up the hydrogen atom also have spin angular momentum. The presence of these extra(internal) degrees of freedom should change the Hamiltonian.

The Schrodinger equation was derived from the eigenvalue equation for the Hamiltonian

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle \tag{8.184}$$

by re-expressing that equation in the position representation. The associated Schrodinger wave functions were given by the scalar product(linear functional) relation

$$\psi(\vec{r}) = \langle \vec{r} \mid \psi \rangle \tag{8.185}$$

The single particle Schrodinger equation is relevant for problems where the Hamiltonian contains terms dependent on ordinary 3–dimensional space(for many-particle systems we must use a multi-dimensional configuration space which bears no simple relationship to ordinary three-dimensional space). Spin is an internal degree of freedom that has no representation in the 3–dimensional space of the Schrodinger wave equation.

The Schrodinger picture, however, does not choose a particular representation and, therefore, we can include spin within the context of solving the Schrodinger equation in the following ad hoc manner. A more rigorous treatment requires relativity.

If there are spin-dependent terms in the Hamiltonian, then we expand the Hilbert space used to solved the problem by constructing a new basis that is made up of direct product states of the following type

$$|\psi_{new}\rangle = |\psi\rangle \otimes |s, m_s\rangle \tag{8.186}$$

where $|\psi\rangle$ depends on only ordinary 3–dimensional space and $|s, m_s\rangle$ is an eigenvector of $\vec{S}^2_{op}$ and $\hat{S}_z$.

The energy eigenvalue equation becomes

$$\begin{aligned}\hat{H}\,|\psi_{new}\rangle &= E\,|\psi_{new}\rangle \\ &= (((3-\text{space operators}))\,|\psi\rangle) \otimes (((\text{spin - dependent operators}))\,|s, m_s\rangle)\end{aligned}$$

and the corresponding *wave function* is

$$\langle \vec{r} \mid \psi_{new}\rangle = \langle \vec{r} \mid \psi\rangle\,|s, m_s\rangle = \psi(\vec{r})\,|s, m_s\rangle \tag{8.187}$$

where abstract spin vector is stuck onto the wave function in some way (maybe with superglue).

Let us now investigate what happens in atoms when we add in spin, some aspects of relativity and external fields. We restrict our attention to one-electron atoms like hydrogen at this point.

8.3 Spin-Orbit Interaction - Fine Structure

The proton in hydrogen generates an electric field

$$\vec{\mathcal{E}} = \frac{e}{r^2}\hat{r} = \frac{e}{r^3}\vec{r} \tag{8.188}$$

that acts on the moving electron. This result is approximately true (to first order) in most atoms. Now special relativity says that an electron moving with a velocity $\vec{v}$ through an electric field $\vec{\mathcal{E}}$ also behaves as if it is interacting with a magnetic field given by

$$\vec{B} = -\frac{1}{c}\vec{v} \times \vec{\mathcal{E}} \tag{8.189}$$

to first order in v/c.

This magnetic field interacts with the spin (actually with its associated magnetic moment) to produce an additional contribution to the energy of the form

$$E = -\vec{M}_{spin} \cdot \vec{B} \tag{8.190}$$

where

$$\vec{M}_{spin} = -\frac{e}{mc}\vec{S} \tag{8.191}$$

Substituting everything in we get

$$\begin{aligned} E &= -\frac{e}{mc^2}\vec{S} \cdot \left(\vec{v} \times \vec{\mathcal{E}}\right) = -\frac{e}{mc^2}\vec{S} \cdot \left(\vec{v} \times \frac{e}{r^3}\vec{r}\right) \\ &= \frac{1}{m^2c^2}\vec{S} \cdot \vec{L}\frac{e^2}{r^3} \end{aligned} \tag{8.192}$$

Now

$$\frac{e^2}{r^3} = \frac{1}{r}\frac{dV}{dr} \text{ for } V(r) = -\frac{e^2}{r} = \text{ potential energy of the electron} \tag{8.193}$$

so that we finally obtain the so-called spin-orbit energy contribution

$$E = \left[\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S} \cdot \vec{L} = E_{spin-orbit} = E_{so} \tag{8.194}$$

This corresponds to an additional term in the Hamiltonian of the form

$$\hat{H}_{so} = \left[\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op} \cdot \vec{L}_{op} \tag{8.195}$$

This term couples the orbital and spin angular momentum degrees of freedom (hence the label spin-orbit energy) and mixes 3–dimensional space with *spin* space. That is why we had to expand the Hilbert space as we discussed earlier.

Another way to think about this interaction is that the electron spin magnetic

moment vector (or spin vector) is precessing about the direction of the magnetic field. The equations for such a precessional motion are

$$\vec{M}_{spin} \times \vec{B} = \frac{d\vec{S}}{dt} = \vec{\Omega}_{L(armor)} \times \vec{S} \tag{8.196}$$

where

$$\left|\vec{\Omega}_L\right| = \frac{eB}{m} \tag{8.197}$$

Now

$$\vec{B} = -\frac{1}{c}\vec{v} \times \vec{\mathcal{E}} = \frac{1}{emc^2}\frac{1}{r}\frac{dV}{dr}\vec{L} \tag{8.198}$$

which implies that

$$\left|\vec{\Omega}_L\right| = \frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\left|\vec{L}\right| \tag{8.199}$$

It turns out that this is exactly a factor of 2 too large. There is another relativistic effect, which gives another precession (called Thomas precession) effect, that cancels exactly one-half of this spin-orbit effect.

8.3.1 Thomas Precession

This is a relativistic kinematic effect. It results from the time dilation between the rest frames of the electron and the proton. This causes observers in these two frames to disagree on the time required for one of the particles to a make a complete revolution about the other particle.

If an observer on the electron measures a time interval T, then the observer on the proton measures

$$T' = \gamma T \text{ where } \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \quad , \quad v = \text{ speed of the electron} \tag{8.200}$$

We assume uniform circular motion for simplicity.

The orbital angular velocities measured by the observers are

$$\frac{2\pi}{T} \text{ and } \frac{2\pi}{T'} \tag{8.201}$$

respectively.

In the rest frame of the electron, the spin angular momentum vector maintains its direction in space. This implies that an observer on the proton sees this spin vector precessing at a rate equal to the *difference* of the two angular velocities,

i.e., the precessional frequency is

$$\begin{aligned}\Omega_{Thomas} &= \frac{2\pi}{T} - \frac{2\pi}{T'} = \frac{2\pi}{T'}\left(\frac{T'}{T} - 1\right) \\ &= \frac{2\pi}{T'}\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1\right) \\ &\approx \frac{2\pi}{T'}\left(\frac{v^2}{2c^2}\right)\end{aligned} \tag{8.202}$$

But we also have

$$\frac{2\pi}{T'} = \omega = \frac{\left|\vec{L}\right|}{mr^2} \text{ and } \frac{mv^2}{r} = -\frac{dV}{dr} \tag{8.203}$$

for circular motion.

Thus, we get

$$\left|\vec{\Omega}_T\right| = -\frac{1}{2}\frac{1}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\left|\vec{L}\right| = -\frac{1}{2}\left|\vec{\Omega}_L\right| \tag{8.204}$$

Therefore, the combined precession is reduced by a factor of two and we get the result

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op} \tag{8.205}$$

The energy levels arising from this correction are called the *atomic fine structure*.

8.4 Another Relativity Correction

The correct relativistic kinetic energy term is

$$\begin{aligned}K = (\gamma-1)mc^2 &= \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1\right)mc^2 \\ &= \left(\left(1 + \frac{1}{2}\frac{v^2}{c^2} - \frac{1}{8}\frac{v^4}{c^4} + ...\right) - 1\right)mc^2 \\ &= \left(\frac{1}{2}\frac{v^2}{c^2} - \frac{1}{8}\frac{v^4}{c^4}\right)mc^2 = \frac{\vec{p}_{op}^{\,2}}{2m} - \frac{\vec{p}_{op}^{\,4}}{8m^3c^2}\end{aligned} \tag{8.206}$$

Therefore, we must correct the $\vec{p}_{op}^{\,2}/2m$ we have already included in the Hamiltonian by adding a terms of the form

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2} \tag{8.207}$$

Thus, if no external field are present we have the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so} \tag{8.208}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} - e^2 \left(\frac{1}{r}\right)_{op} \tag{8.209}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2} \tag{8.210}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op} \tag{8.211}$$

8.5 External Fields - Zeeman and Stark Effects; Hyperfine Structure

8.5.1 Zeeman Effect

If an external magnetic field exists, then it interacts with the total magnetic moment of the electron, where

$$\vec{M}_{total} = \vec{M}_{orbital} + \vec{M}_{spin} = -\frac{e}{2mc}\left(g_\ell \vec{L} + g_s \vec{S}\right) \tag{8.212}$$

as we derived earlier. If we define

$$\mu_B = \text{ Bohr magneton } = \frac{e\hbar}{mc} \tag{8.213}$$

and let $\vec{B}_{ext} = B\hat{z}$, then we have, using $g_\ell = 1$ and $g_s = 2$, the result

$$E_{Zeeman} = -\vec{M}_{total}\cdot\vec{B}_{ext} = \frac{\mu_B B}{\hbar}(L_z + 2S_z) \tag{8.214}$$

Thus, we must add a term of the form

$$\hat{H}_{Zeeman} = \frac{\mu_B B}{\hbar}(\hat{L}_z + 2\hat{S}_z) \tag{8.215}$$

to the Hamiltonian when an external magnetic field is present.

We can see directly how the orbital angular momentum part of this energy arises. We saw earlier that if we had a Hamiltonian

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} + V(\vec{r}_{op}) \tag{8.216}$$

and we add an electromagnetic field characterized by a vector potential $\vec{A}$, where $\vec{B} = \nabla \times \vec{A}$, then the momentum operator changes to

$$\vec{p}_{em} = \vec{p} - \frac{e}{c}\vec{A}(\vec{r}) \tag{8.217}$$

and the Hamiltonian changes to

$$\begin{aligned}\hat{H} &= \frac{\vec{p}_{em,op}^2}{2m} + V(\vec{r}_{op}) = \frac{\left(\vec{p}_{op} - \frac{e}{c}\vec{A}(\vec{r}_{op})\right)^2}{2m} + V(\vec{r}_{op}) \\ &= \hat{H}_0 - \frac{e}{2mc}\left(\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op}) + \vec{A}(\vec{r}_{op})\cdot\vec{p}_{op}\right) + \frac{e^2}{2mc^2}\vec{A}^2(\vec{r}_{op}) \end{aligned} \tag{8.218}$$

The magnetic field has to be enormous or the radial quantum number n very large for the $\vec{A}^2$ term to have any effect, so we will neglect it for now. Let us look at the term

$$\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op}) + \vec{A}(\vec{r}_{op})\cdot\vec{p}_{op} \tag{8.219}$$

For a uniform (constant in magnitude and direction) external field $\vec{B}$, we have

$$\vec{A} = -\frac{1}{2}\vec{r}\times\vec{B} \tag{8.220}$$

I will prove this so we get a chance to see the use of ϵ_{ijk} in vector algebra.

$$\begin{aligned}\nabla\times\vec{A} &= -\frac{1}{2}\nabla\times\left(\vec{r}\times\vec{B}\right) = -\frac{1}{2}\sum_{ijk}\varepsilon_{ijk}\frac{\partial}{\partial x_j}\left(\vec{r}\times\vec{B}\right)_k\hat{e}_i \\ &= -\frac{1}{2}\sum_{ijk}\varepsilon_{ijk}\frac{\partial}{\partial x_j}\left(\sum_{mn}\varepsilon_{kmn}x_m B_n\right)\hat{e}_i \\ &= -\frac{1}{2}\sum_{ij}\sum_{mn}\left(\sum_k\varepsilon_{ijk}\varepsilon_{mnk}\right)\frac{\partial}{\partial x_j}\left(x_m B_n\right)\hat{e}_i \\ &= -\frac{1}{2}\sum_{ij}\sum_{mn}\left(\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}\right)\frac{\partial}{\partial x_j}\left(x_m B_n\right)\hat{e}_i \\ &= -\frac{1}{2}\sum_{ij}\left[\frac{\partial}{\partial x_j}\left(x_i B_j\right)\hat{e}_i - \frac{\partial}{\partial x_j}\left(x_j B_i\right)\hat{e}_i\right] \\ &= -\frac{1}{2}\sum_{ij}\left[\frac{\partial x_i}{\partial x_j}B_j + x_i\frac{\partial B_j}{\partial x_j} - \frac{\partial x_j}{\partial x_j}B_i - x_j\frac{\partial B_i}{\partial x_j}\right]\hat{e}_i\end{aligned}$$

Now

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} \text{ and } \frac{\partial B_i}{\partial x_j} = 0 \tag{8.221}$$

so we get

$$\begin{aligned}\nabla\times\vec{A} &= -\frac{1}{2}\sum_{ij}\left[\delta_{ij}B_j - \delta_{jj}B_i\right]\hat{e}_i \\ &= -\frac{1}{2}\left[\sum_i B_i\hat{e}_i - 3\sum_i B_i\hat{e}_i\right] \\ &= -\frac{1}{2}\left[\vec{B} - 3\vec{B}\right] = \vec{B}\end{aligned}$$

Therefore, we have

$$\begin{aligned}\vec{p}_{op}\cdot\vec{A}(\vec{r}_{op})+\vec{A}(\vec{r}_{op})\cdot\vec{p}_{op} &= -\frac{1}{2}\left[\vec{p}_{op}\cdot\left(\vec{r}_{op}\times\vec{B}\right)+\left(\vec{r}_{op}\times\vec{B}\right)\cdot\vec{p}_{op}\right] \\ &= -\frac{1}{2}\left[\sum_{ijk}\varepsilon_{ijk}B_k\left[\hat{p}_i\hat{x}_j+\hat{x}_j\hat{p}_i\right]\right] = -\frac{1}{2}\left[\sum_{ijk}\varepsilon_{ijk}B_k\left[2\hat{x}_j\hat{p}_i-i\hbar\delta_{ij}\right]\right] \\ &= \sum_{ijk}\varepsilon_{kji}\hat{x}_j\hat{p}_iB_k = \left(\vec{r}_{op}\times\vec{p}_{op}\right)\cdot\vec{B} = \vec{L}_{op}\cdot\vec{B} \end{aligned} \tag{8.222}$$

which then gives

$$\hat{H} = \hat{H}_0 - \frac{e}{2mc}\vec{L}_{op}\cdot\vec{B} + \frac{e^2}{2mc^2}\vec{A}^2(\vec{r}_{op}) \tag{8.223}$$

which accounts for the orbital angular momentum part of the Zeeman energy.

The spin angular momentum part of the Zeeman energy cannot be derived from the non-relativistic Schrodinger equation. When one derives the Dirac *relativistic* equation for the electron, the $\vec{S}_{op}\cdot\vec{B}$ term appears naturally.

8.5.2 Stark Effect

If a hydrogen atom is placed in an external electric field $\vec{\mathcal{E}}$ which is constant is space and time (uniform and static), then an additional energy appears. It corresponds to an interaction between and electric dipole made up of the electron and proton separated by a distance and the external electric field. We introduce the electric dipole moment operator

$$\vec{d}_{op} = -e\vec{r}_{op} \tag{8.224}$$

where $\vec{r}$ is the position vector of the electron relative to the proton. We then write the extra energy term as

$$\hat{H}_{dipole} = -\vec{d}_{op}\cdot\vec{\mathcal{E}} \tag{8.225}$$

If we choose $\vec{\mathcal{E}} = \mathcal{E}\hat{z}$, then we have $\hat{H}_{dipole} = -ez\mathcal{E}$. The full Hamiltonian is then

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so} + \hat{H}_{Zeeman} + \hat{H}_{dipole} \tag{8.226}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} - e^2\left(\frac{1}{r}\right)_{op} \tag{8.227}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2} \tag{8.228}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op} \tag{8.229}$$

$$\hat{H}_{Zeeman} = \frac{\mu_B}{\hbar}(\vec{L}_{op}+2\vec{S}_{op})\cdot\vec{B} \tag{8.230}$$

$$\hat{H}_{dipole} = -e\vec{r}_{op}\cdot\vec{\mathcal{E}} \tag{8.231}$$

8.5.3 Hyperfine Structure

The nuclear magnetic dipole moment also generates a magnetic field. If we assume that it is a point dipole $\vec{M}_N$, then the magnetic field is given by

$$\vec{B}(\vec{r}) = \left(\frac{3\left(\vec{M}_N \cdot \vec{r}\right)\vec{r}}{r^5} - \frac{\vec{M}_N}{r^3} \right) + \frac{8\pi}{3}\vec{M}_N \delta(\vec{r}) \tag{8.232}$$

where the first two terms are the standard result of the magnetic field due to a loop of current as seen from very far away (approximates dipole as a point) and the last term is peculiar to a point dipole. The last term will give a contribution only for spherically symmetric states ($\ell = 0$). The extra energy is then

$$\begin{aligned} \hat{H}_{hyperfine} &= -\vec{M}_e \cdot \vec{B} \\ &= -\left(\frac{3\left(\vec{M}_N \cdot \vec{r}\right)\left(\vec{M}_e \cdot \vec{r}\right)}{r^5} - \frac{\vec{M}_N \cdot \vec{M}_e}{r^3} \right) - \frac{8\pi}{3}\vec{M}_N \cdot \vec{M}_e \delta(\vec{r}) \end{aligned} \tag{8.233}$$

where

$$\vec{M}_N = g_N \frac{Ze}{2m_N c}\vec{S}_{N,op} \text{ and } \vec{M}_e = \frac{e}{mc}\vec{S}_{e,op} \tag{8.234}$$

This is clearly due to spin-spin interactions between the electron and the nucleus and gives rise to the so-called hyperfine level splitting.

8.6 Examples

Now that we have identified all of the relevant corrections to the Hamiltonian for atoms, let us illustrate the procedures for calculation of the new energy levels via perturbation theory. We look at the simplest atom first.

8.6.1 Spin-Orbit, Relativity, Zeeman Effect in Hydrogen Atom

The Hamiltonian is

$$\hat{H} = \hat{H}_0 + \hat{H}_{relativity} + \hat{H}_{so} + \hat{H}_{Zeeman} \tag{8.235}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^{\,2}}{2m} - e^2\left(\frac{1}{r}\right)_{op} \tag{8.236}$$

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2} \tag{8.237}$$

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op} \cdot \vec{L}_{op} \tag{8.238}$$

$$\hat{H}_{Zeeman} = \frac{\mu_B}{\hbar}(\vec{L}_{op} + 2\vec{S}_{op}) \cdot \vec{B} \tag{8.239}$$

The first step is to calculate all relevant commutators so that we can find those operators that have a common eigenbasis.

$$\left[\vec{J}_{op}^2, \hat{H}_0\right] = 0 = \left[\hat{J}_z, \hat{H}_0\right] = \left[\vec{L}_{op}^2, \hat{H}_0\right] = \left[\vec{S}_{op}^2, \hat{H}_0\right] \tag{8.240}$$

$$\left[\vec{L}_{op}^2, \vec{S}_{op}^2\right] = 0 = \left[\vec{L}_{op}^2, \vec{J}_{op}^2\right] = \left[\vec{S}_{op}^2, \vec{J}_{op}^2\right] \tag{8.241}$$

$$\left[\vec{J}_{op}^2, \hat{J}_z\right] = 0 = \left[\vec{L}_{op}^2, \hat{J}_z\right] = \left[\vec{S}_{op}^2, \hat{J}_z\right] \tag{8.242}$$

This says that there exists a common set of eigenvectors in the unperturbed system for the set of operators

$$\hat{H}_0, \vec{J}_{op}^2, \vec{L}_{op}^2, \vec{S}_{op}^2, \hat{J}_z \tag{8.243}$$

We label these states by the corresponding eigenvalues of the commuting set of observables (these are called *good quantum numbers*)

$$|n, \ell, s, j, m_j\rangle \tag{8.244}$$

We also have

$$\left[\hat{S}_z, \hat{H}_0\right] = 0 = \left[\hat{L}_z, \hat{H}_0\right] = \left[\vec{L}_{op}^2, \hat{H}_0\right] = \left[\vec{S}_{op}^2, \hat{H}_0\right] \tag{8.245}$$

$$\left[\vec{L}_{op}^2, \vec{S}_{op}^2\right] = 0 = \left[\vec{L}_{op}^2, \hat{S}_z\right] = \left[\vec{S}_{op}^2, \hat{S}_z\right] = \left[\vec{L}_{op}^2, \hat{L}_z\right] = \left[\vec{S}_{op}^2, \hat{L}_z\right] \tag{8.246}$$

which says that there exists another common set of eigenvectors in the unperturbed system for the operators

$$\hat{H}_0, \vec{L}_{op}^2, \vec{S}_{op}^2, \hat{L}_z, \hat{S}_z \tag{8.247}$$

We label these states by the corresponding eigenvalues of this commuting set of observables (again these are called *good quantum numbers*)

$$|n, \ell, s, m_\ell, m_s\rangle \tag{8.248}$$

In this latter basis, the unperturbed or zero-order Hamiltonian has solutions represented by

$$\hat{H}_0 |n, \ell, m_\ell, s, m_s\rangle = E_n^{(0)} |n, \ell, m_\ell, s, m_s\rangle \quad , \quad E_n^{(0)} = -\frac{Ze^2}{2a_0 n^2}$$

$$Ze = \text{ nucleus charge } (Z = 1 \text{ for hydrogen }) \, , \; a_0 = \text{ Bohr radius } = \frac{\hbar^2}{me^2}$$

$$\begin{aligned}\psi_{n\ell m_\ell s m_s}(r, \theta, \phi) &= \langle \vec{r}| \left(|nlm_\ell s m_s\rangle\right) = \langle \vec{r}| \left(|nlm_\ell\rangle |sm_s\rangle\right) \\ &= \langle \vec{r} \mid nlm_\ell\rangle |sm_s\rangle = \psi_{n\ell m_\ell}(r, \theta, \phi) |sm_s\rangle\end{aligned}$$

and first few unperturbed wave functions are

$$\psi_{100}(r,\theta,\phi) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\frac{Zr}{a_0}} \tag{8.249}$$

$$\psi_{200}(r,\theta,\phi) = \frac{1}{\sqrt{32\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\left(2-\frac{Zr}{a_0}\right) e^{-\frac{Zr}{2a_0}} \tag{8.250}$$

$$\psi_{210}(r,\theta,\phi) = \frac{1}{\sqrt{32\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\frac{Zr}{a_0} e^{-\frac{Zr}{2a_0}}\cos\theta \tag{8.251}$$

$$\psi_{21\pm1}(r,\theta,\phi) = \frac{1}{\sqrt{64\pi}}\left(\frac{Z}{a_0}\right)^{3/2}\frac{Zr}{a_0} e^{-\frac{Zr}{2a_0}}\sin\theta e^{\pm i\phi} \tag{8.252}$$

We also have the relations below for the unperturbed states.

$$\vec{L}_{op}^2\,|n,\ell,s,j,m_j\rangle = \hbar^2\ell(\ell+1)\,|n,\ell,s,j,m_j\rangle \tag{8.253}$$

$$\vec{S}_{op}^2\,|n,\ell,s,j,m_j\rangle = \hbar^2 s(s+1)\,|n,\ell,s,j,m_j\rangle \tag{8.254}$$

$$\vec{J}_{op}^2\,|n,\ell,s,j,m_j\rangle = \hbar j(j+1)\,|n,\ell,s,j,m_j\rangle \tag{8.255}$$

$$\hat{J}_z\,|n,\ell,s,j,m_j\rangle = \hbar^2 m_j\,|n,\ell,s,j,m_j\rangle \tag{8.256}$$

$$\vec{L}_{op}^2\,|n,\ell,s,m_\ell,m_s\rangle = \hbar^2\ell(\ell+1)\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.257}$$

$$\vec{S}_{op}^2\,|n,\ell,s,m_\ell,m_s\rangle = \hbar^2 s(s+1)\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.258}$$

$$\hat{L}_z\,|n,\ell,s,m_\ell,m_s\rangle = \hbar m_\ell\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.259}$$

$$\hat{S}_z\,|n,\ell,s,m_\ell,m_s\rangle = \hbar m_s\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.260}$$

$$\hat{J}_z\,|n,\ell,s,m_\ell,m_s\rangle = (\hat{L}_z+\hat{S}_z)\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.261}$$

$$= \hbar(m_\ell+m_s)\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.262}$$

$$= \hbar m_j\,|n,\ell,s,m_\ell,m_s\rangle \tag{8.263}$$

Since the total angular momentum is given by

$$\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op} \tag{8.264}$$

the rules we developed for the addition of angular momentum say that

$$j = \ell+s, \ell+s-1, ..., |\ell-s|+1, |\ell-s| \tag{8.265}$$

and

$$m_j = j, j-1, j-2, ..., -j+1, -j \tag{8.266}$$

In the case of hydrogen, where $s = 1/2$, we have only two allowed total j values for each ℓ value, namely,

$$j = \ell \pm \frac{1}{2} \tag{8.267}$$

We can use either of the two sets of basis states (both are an orthonormal basis)

$$|n,\ell,s,j,m_j\rangle \text{ or } |nlm_\ell sm_s\rangle \tag{8.268}$$

as the zero-order states for a perturbation theory development of the energies. The choice depends on the specific perturbations we are trying to calculate.

Let us start off by using the $|n, \ell, s, j, m_j\rangle$ states.

If we use the potential energy function

$$V(r) = -\frac{e^2}{r} \tag{8.269}$$

for hydrogen, then the spin-orbit correction to the Hamiltonian becomes

$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op} = \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\vec{S}_{op}\cdot\vec{L}_{op} \tag{8.270}$$

Now $\vec{J}_{op} = \vec{L}_{op} + \vec{S}_{op}$ implies that

$$\vec{J}_{op}^2 = \vec{L}_{op}^2 + \vec{S}_{op}^2 + 2\vec{L}_{op}\cdot\vec{S}_{op} \tag{8.271}$$

$$\rightarrow \vec{L}_{op}\cdot\vec{S}_{op} = \frac{1}{2}\left(\vec{J}_{op}^2 - \vec{L}_{op}^2 - \vec{S}_{op}^2\right) \tag{8.272}$$

and therefore

$$\hat{H}_{so} = \frac{e^2}{4m^2c^2}\left(\frac{1}{r^3}\right)\left(\vec{J}_{op}^2 - \vec{L}_{op}^2 - \vec{S}_{op}^2\right) \tag{8.273}$$

Therefore,

$$\left[\vec{J}_{op}^2, \hat{H}_{so}\right] = 0 = \left[\hat{J}_z, \hat{H}_{so}\right] = \left[\vec{L}_{op}^2, \hat{H}_{so}\right] = \left[\vec{S}_{op}^2, \hat{H}_{so}\right] = \left[\hat{H}_0, \hat{H}_{so}\right] \tag{8.274}$$

which implies that the state vectors $|n, \ell, s, j, m_j\rangle$ are also eigenvectors of $\hat{H}_{so}$. This means that the matrix representation of $\hat{H}_{so}$ in this basis will be diagonal and we can apply standard non-degenerate perturbation theory.

Applying our rules for first order perturbation theory we have

$$\begin{aligned} E_{n\ell sjm_j} &= E_n^{(0)} + \langle n\ell sjm_j|\,\hat{H}_{so}\,|n\ell sjm_j\rangle \\ &= E_n^{(0)} + \frac{e^2\hbar^2}{4m^2c^2}\left(j(j+1) - \ell(\ell+1) - s(s+1)\right)\langle n\ell sjm_j|\,\frac{1}{r_{op}^3}\,|n\ell sjm_j\rangle \end{aligned} \tag{8.275}$$

We now evaluate

$$\langle n\ell sjm_j|\,\frac{1}{r_{op}^3}\,|n\ell sjm_j\rangle = \int d^3\vec{r}\,' \int d^3\vec{r}\,\langle n\ell sjm_j \mid \vec{r}\,'\rangle\langle\vec{r}\,'|\,\frac{1}{r_{op}^3}\,|\vec{r}\rangle\langle\vec{r}\mid n\ell sjm_j\rangle \tag{8.276}$$

Now

$$\langle\vec{r}\,'|\,\frac{1}{r_{op}^3}\,|\vec{r}\rangle = \frac{1}{r^3}\langle\vec{r}\,'\mid\vec{r}\rangle = \frac{1}{r^3}\delta(\vec{r}\,' - \vec{r}) \tag{8.277}$$

which gives

$$\langle n\ell sjm_j| \frac{1}{r_{op}^3} |n\ell sjm_j\rangle = \int d^3\vec{r} \frac{1}{r^3} |\langle \vec{r} \mid n\ell sjm_j\rangle|^2 = \int d^3\vec{r} \frac{1}{r^3} \left|\psi_{n\ell sjm_j}(\vec{r})\right|^2$$

Therefore, we can calculate the energy corrections once we know $|n, \ell, s, j, m_j\rangle$.

We first consider the trivial case of the $n = 1$ level in hydrogen. We have

$$E_1^{(0)} = -\frac{e^2}{2a_0} \tag{8.278}$$

and the corresponding states are shown in Table 8.1 below.

n	ℓ	s	m_ℓ	m_s	j	m_j
1	0	1/2	0	+1/2	1/2	+1/2
1	0	1/2	0	-1/2	1/2	-1/2

Table 8.1: $n = 1$ level quantum numbers

or

$$\left|1, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \text{ and } \left|1, 0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\rangle_{jm} \tag{8.279}$$

where we have added the label jm to distinguish them from the $|n, \ell, s, m_\ell, m_s\rangle$ states which we label with $m_\ell m_s = mm$. We are able to specify m_ℓ and m_s also in this case because when $\ell = 0$ we must have $m_\ell = 0$ and $j = s$ which says that $m_j = m_s$.

This is a *two-fold degenerate* ground state for the atom in *zeroth order*.

Since $\ell = 0$, which implies that $j = s = 1/2$, the expectation value $\langle \hat{H}_{so} \rangle = 0$. Thus, there is no spin orbit correction for this state to first order, In fact, there is no spin orbit correction to any order for an $\ell = 0$ state.

Now in general, we can write

$$|n, \ell, s, j, m_j\rangle = \sum_{\substack{m_\ell, m_s \\ m_\ell + m_s = m_j}} a_{n\ell sjm_j m_\ell m_s} |n, \ell, s, m_\ell, m_s\rangle \tag{8.280}$$

where the $a_{n\ell sjm_j m_\ell m_s}$ are the relevant Clebsch-Gordon(CG) coefficients.

For the $n = 1$ level we have the simple cases where

$$\left|1, 0, \frac{1}{2}, \frac{1}{2}, \pm\frac{1}{2}\right\rangle_{jm} = \left|1, 0, \frac{1}{2}, 0, \pm\frac{1}{2}\right\rangle_{mm} \tag{8.281}$$

i.e., the CG coefficients are equal to 1,

$$a_{1,0,\frac{1}{2},\frac{1}{2},\pm\frac{1}{2},0,\pm\frac{1}{2}} = 1 \tag{8.282}$$

which is always true for the (maximum j, maximum(minimum) m_j) states. There is always only one such state.

The next level is the $n = 2$ level of hydrogen and the complexity of the calculation increases fast. We have

$$E_2^{(0)} = -\frac{e^2}{8a_0} \tag{8.283}$$

It is always the case that the direct-product states $|n, \ell, s, m_\ell, m_s\rangle$ are easier to write down. For this level the $|n, \ell, s, j, m_j\rangle$ states need to be constructed from the $|n, \ell, s, m_\ell, m_s\rangle$ states. Before we proceed, we can enumerate the states in both schemes. The degeneracy is given by

$$\text{degeneracy} = 2\sum_{\ell=0}^{n-1}(2\ell+1) = 2n^2 = 8 = \sum_{\ell=0}^{n-1}\sum_{j=|\ell-s|}^{\ell+s}(2j+1) = 2\sum_{\ell=0}^{n-1}(2\ell+1)$$

The states are shown in Tables 8.2 and 8.3 below.

n	ℓ	s	m_ℓ	m_s	ket
2	1	1/2	1	+1/2	$\|2,1,1/2,1,1/2\rangle_{mm}$
2	1	1/2	0	+1/2	$\|2,1,1/2,0,1/2\rangle_{mm}$
2	1	1/2	-1	+1/2	$\|2,1,1/2,-1,1/2\rangle_{mm}$
2	1	1/2	1	-1/2	$\|2,1,1/2,1,-1/2\rangle_{mm}$
2	1	1/2	0	-1/2	$\|2,1,1/2,0,-1/2\rangle_{mm}$
2	1	1/2	-1	-1/2	$\|2,1,1/2,-1,-1/2\rangle_{mm}$
2	0	1/2	0	1/2	$\|2,0,1/2,0,1/2\rangle_{mm}$
2	0	1/2	0	-1/2	$\|2,0,1/2,0,-1/2\rangle_{mm}$

Table 8.2: $n = 2$ level quantum numbers $m_\ell m_s$ states

n	ℓ	s	j	m_j	ket
2	1	1/2	3/2	3/2	$\|2,1,1/2,1,1/2\rangle_{mm}$
2	1	1/2	3/2	1/2	$\|2,1,1/2,0,1/2\rangle_{mm}$
2	1	1/2	3/2	-1/2	$\|2,1,1/2,-1,1/2\rangle_{mm}$
2	1	1/2	3/2	-3/2	$\|2,1,1/2,1,-1/2\rangle_{mm}$
2	1	1/2	1/2	1/2	$\|2,1,1/2,0,-1/2\rangle_{mm}$
2	1	1/2	1/2	-1/2	$\|2,1,1/2,-1,-1/2\rangle_{mm}$
2	0	1/2	1/2	1/2	$\|2,0,1/2,0,1/2\rangle_{mm}$
2	0	1/2	1/2	-1/2	$\|2,0,1/2,0,-1/2\rangle_{mm}$

Table 8.3: $n = 2$ level quantum numbers jm_j states

In the first set of states, we could have also included the m_j label since we must

have $m_j = m_\ell + m_s$.

In order to learn all the intricate details of this type of calculation, we shall proceed in two ways using the spin-orbit correction as an example.

In method #1, we will construct the $|n, \ell, s, j, m_j\rangle$ states (the zero-order state vectors) from the $|n, \ell, s, m_\ell, m_s\rangle$ and then calculate the first-order energy corrections. In this basis, the $\langle \hat{H}_{so} \rangle$ matrix will be diagonal.

In method #2, we will construct the $\langle \hat{H}_{so} \rangle$ matrix using the easiest states to write down, namely the $|n, \ell, s, m_\ell, m_s\rangle$ states, and then diagonalize it to find the correct first order energies and new zero-order state vectors, which should be the $|n, \ell, s, j, m_j\rangle$ states.

Method #1

We start with the state with maximum j and m_j values. This state always has a CG coefficient equal to 1, i.e., there is only one way to construct it from the other angular momenta.

$$\left| n = 2, \ell = 1, s = \frac{1}{2}, j = \frac{3}{2}, m_j = \frac{3}{2} \right\rangle_{jm} = \left| n = 2, \ell = 1, s = \frac{1}{2}, m_\ell = 1, m_s = \frac{1}{2} \right\rangle_{mm}$$

where we have shown all the labels explicitly. From now on we will write such equations as

$$\left| 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{3}{2} \right\rangle_{jm} = \left| 2, 1, \frac{1}{2}, 1, \frac{1}{2} \right\rangle_{mm}$$

We then use the lowering operators to obtain

$$\begin{aligned}
\hat{J}_- \left| 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{3}{2} \right\rangle_{jm} &= \hbar \sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right) - \frac{3}{2}\left(\frac{3}{2}-1\right)} \left| 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2} \right\rangle_{jm} \\
&= \hbar\sqrt{3} \left| 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2} \right\rangle_{jm} = \left(\hat{L}_- + \hat{S}_-\right) \left| 2, 1, \frac{1}{2}, 1, \frac{1}{2} \right\rangle_{mm} \\
&= \hbar\sqrt{1(1+1) - 1(1-1)} \left| 2, 1, \frac{1}{2}, 0, \frac{1}{2} \right\rangle_{mm} \\
&\qquad + \hbar \sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}-1\right)} \left| 2, 1, \frac{1}{2}, 1, -\frac{1}{2} \right\rangle_{mm} \\
&= \hbar\sqrt{2} \left| 2, 1, \frac{1}{2}, 0, \frac{1}{2} \right\rangle_{mm} + \hbar\sqrt{1} \left| 2, 1, \frac{1}{2}, 1, -\frac{1}{2} \right\rangle_{mm}
\end{aligned}$$

or

$$\left| 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2} \right\rangle_{jm} = \sqrt{\frac{2}{3}} \left| 2, 1, \frac{1}{2}, 0, \frac{1}{2} \right\rangle_{mm} + \sqrt{\frac{1}{3}} \left| 2, 1, \frac{1}{2}, 1, -\frac{1}{2} \right\rangle_{mm} \tag{8.284}$$

Notice that we use the total J operators on the left and the L and S operators on the right.

The result is a linear combination of the $|n, \ell, s, m_\ell, m_s\rangle$ states all with $m_j = 1/2$ as we expected.

Continuing this process we have

$$\begin{aligned}
\hat{J}_- \left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2}\right\rangle_{jm} &= \hbar\sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}-1\right)}\left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{1}{2}\right\rangle_{jm} \\
&= 2\hbar\left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{1}{2}\right\rangle_{jm} \\
&= \left(\hat{L}_- + \hat{S}_-\right)\left(\sqrt{\frac{2}{3}}\left|2, 1, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} + \sqrt{\frac{1}{3}}\left|2, 1, \frac{1}{2}, 1, -\frac{1}{2}\right\rangle_{mm}\right) \\
&= \hbar\sqrt{\frac{4}{3}}\left|2, 1, \frac{1}{2}, -1, \frac{1}{2}\right\rangle_{mm} + \hbar\sqrt{\frac{8}{3}}\left|2, 1, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm}
\end{aligned}$$

or

$$\begin{aligned}
&\left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{1}{2}\right\rangle_{jm} \\
&\qquad = \sqrt{\frac{1}{3}}\left|2, 1, \frac{1}{2}, -1, \frac{1}{2}\right\rangle_{mm} + \sqrt{\frac{2}{3}}\left|2, 1, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm}
\end{aligned} \tag{8.285}$$

and finally

$$\left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{3}{2}\right\rangle_{jm} = \left|2, 1, \frac{1}{2}, -1, -\frac{1}{2}\right\rangle_{mm} \tag{8.286}$$

We now need to construct the maximum state for then next lowest value of j, namely,

$$\left|2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \tag{8.287}$$

This state has $m_j = 1/2$ so it must be constructed out of the same states that make up

$$\left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2}\right\rangle_{jm} \tag{8.288}$$

or it can be written as

$$\left|2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} = a\left|2, 1, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} + b\left|2, 1, \frac{1}{2}, 1, -\frac{1}{2}\right\rangle_{mm} \tag{8.289}$$

Now we must have

$$\left\langle 2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2} \,\middle|\, 2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} = 0 \text{ orthogonality} \tag{8.290}$$

$$\left\langle 2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \,\middle|\, 2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} = 1 \text{ normalization} \tag{8.291}$$

Using the orthonormality of the $|n,\ell,s,m_\ell,m_s\rangle$ states we get

$$\sqrt{\frac{2}{3}}a+\sqrt{\frac{1}{3}}b=0 \text{ and } a^2+b^2=1 \tag{8.292}$$

The solution is

$$a=\sqrt{\frac{1}{3}} \text{and} b=-\sqrt{\frac{2}{3}} \tag{8.293}$$

and therefore

$$\left|2,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle_{jm}=\sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm}-\sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},1,-\frac{1}{2}\right\rangle_{mm} \tag{8.294}$$

In a similar manner, we find

$$\begin{aligned}&\left|2,1,\frac{1}{2},\frac{1}{2},-\frac{1}{2}\right\rangle_{jm}\\&\quad=\sqrt{\frac{2}{3}}\left|2,1,\frac{1}{2},-1,\frac{1}{2}\right\rangle_{mm}-\sqrt{\frac{1}{3}}\left|2,1,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm}\end{aligned} \tag{8.295}$$

Finally, we construct the other $j=1/2$ states with $\ell=0$. They are

$$\left|2,0,\frac{1}{2},0,\frac{1}{2}\right\rangle_{jm}=\left|2,0,\frac{1}{2},0,\frac{1}{2}\right\rangle_{mm} \tag{8.296}$$

$$\left|2,0,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{jm}=\left|2,0,\frac{1}{2},0,-\frac{1}{2}\right\rangle_{mm} \tag{8.297}$$

We can now calculate the first-order energy corrections. We do not actually need the detailed construction of the states to do this, but we will need these states to compare with the results of Method #2 later. We found earlier (8.276) that in the $|n,\ell,s,j,m_j\rangle$ basis

$$E_{n\ell sjm_j}=E_n^{(0)}+\frac{e^2\hbar^2A_{n\ell sjm_j}}{4m^2c^2}\left(j(j+1)-\ell(\ell+1)-s(s+1)\right) \tag{8.298}$$

$$A_{n\ell sjm_j}=\int d^3\vec{r}\frac{1}{r^3}\left|\psi_{n\ell sjm_j}(\vec{r})\right|^2 \tag{8.299}$$

or

$$\Delta E_{n,\ell,s,j,m_j}=E_{n,\ell,s,j,m_j}-E_n^{(0)}=\frac{e^2\hbar^2A_\pm}{4m^2c^2}\begin{cases}\ell & j=\ell+1/2\\-(\ell+1) & j=\ell-1/2\\0 & \ell=0\end{cases} \tag{8.300}$$

$$A_\pm=\int d^3\vec{r}\frac{1}{r^3}\left|\psi_{n,\ell,s,j=\ell\pm\frac{1}{2},m_j}(\vec{r})\right|^2 \tag{8.301}$$

Evaluating the integrals we get

$$\Delta E_{n,\ell,s,j,m_j} = \frac{Z^2 \left|E_n^{(0)}\right| \alpha^2}{n(2\ell+1)(\ell+1)} \qquad j = \ell + \frac{1}{2} \tag{8.302}$$

$$\Delta E_{n,\ell,s,j,m_j} = -\frac{Z^2 \left|E_n^{(0)}\right| \alpha^2}{n\ell(2\ell+1)} \qquad j = \ell - \frac{1}{2} \tag{8.303}$$

$$\Delta E_{n,\ell,s,j,m_j} = 0 \qquad \ell = 0 \tag{8.304}$$

where

$$\alpha = \frac{e^2}{\hbar c} = \text{ fine structure constant} \tag{8.305}$$

Therefore, for the $n = 2$ level we have

$$\Delta E_{2,1,\frac{1}{2},\frac{3}{2},m_j} = \frac{Z^2 \left|E_2^{(0)}\right| \alpha^2}{12} \qquad j = \ell + \frac{1}{2} = \frac{3}{2} \tag{8.306}$$

$$\Delta E_{2,1,\frac{1}{2},\frac{1}{2},m_j} = -\frac{Z^2 \left|E_2^{(0)}\right| \alpha^2}{6} \qquad j = \ell - \frac{1}{2} = \frac{1}{2} \tag{8.307}$$

$$\Delta E_{2,0,\frac{1}{2},\frac{1}{2},m_j} = 0 \qquad j = \ell + \frac{1}{2} = \frac{1}{2} \tag{8.308}$$

We note that for hydrogen $Z^2\alpha^2 \approx 10^{-4}$ and thus, the fine structure splitting is significantly smaller than the zero-order energies.

The relativity correction is the same order of magnitude as the spin-orbit correction. We found

$$\hat{H}_{relativity} = -\frac{\vec{p}_{op}^{\,4}}{8m^3c^2} \tag{8.309}$$

This gives the correction

$$\begin{aligned} \Delta E_{rel} &= -\frac{\hbar^4}{8m^3c^2} \int_0^\infty r^2 dr \psi^*_{n\ell m}(r) \nabla^4 \psi_{n\ell m}(r) \\ &= \frac{Z^2 \left|E_n^{(0)}\right| \alpha^2}{4n^2} \left(3 - \frac{4n}{\ell + \frac{1}{2}}\right) \end{aligned} \tag{8.310}$$

Combining these two correction terms(spin-orbit and relativity) gives

$$\Delta E_{fine\ structure} = \frac{Z^2 \left|E_2^{(0)}\right| \alpha^2}{4n^2} \left(3 - \frac{4n}{j + \frac{1}{2}}\right) \tag{8.311}$$

The result is independent of ℓ. It turns out that this result is valid for $\ell = 0$ also. There is an additional term that must be added to the Hamiltonian which contributes only in $\ell = 0$ states. It is called the Darwin term.

The Darwin term comes from the relativistic equation for the electron and takes the form

$$\hat{H}_{Darwin} = \frac{\hbar^2}{8m^2c^2}\nabla^2 V = -\frac{\hbar^2}{8m^2c^2}\left(4\pi e Q_{nuclear}(\vec{r})\right) = \frac{\pi\hbar^2 Ze^2}{2m^2c^2}\delta(\vec{r}) \qquad (8.312)$$

where

$$Q_{nuclear}(\vec{r}) = \text{ the nuclear charge density} \qquad (8.313)$$

Because of the delta function, a contribution

$$\left\langle \hat{H}_{Darwin} \right\rangle_{nj\ell} = \frac{\pi\hbar^2 Ze^2}{2m^2c^2}\left|\psi_{n\ell}(0)\right|^2 = \frac{mc^2 Z^4\alpha^4}{2n^3}\delta_{\ell,0} \qquad (8.314)$$

arises for $\ell = 0$ states *only*. This is identical to the contribution $\langle \hat{H}_{so} + \hat{H}_{rel} \rangle$ for $\ell = 0, j = 1/2$.

Method #2

We use the $|n, \ell, s, m_\ell, m_s\rangle$ basis. In this case, the best form of the operators to use (means we know how to evaluate them with these states) are

$$\begin{aligned}\hat{H}_{so} &= \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\vec{S}_{op}\cdot\vec{L}_{op} \\ &= \frac{e^2}{2m^2c^2}\left(\frac{1}{r^3}\right)\left(\hat{L}_z\hat{S}_z + \frac{1}{2}\left(\hat{L}_+\hat{S}_- + \hat{L}_-\hat{S}_+\right)\right)\end{aligned} \qquad (8.315)$$

If we label the rows and columns of the matrix representation by

$$\begin{aligned}
|1\rangle &= \left|2, 1, \frac{1}{2}, 1, \frac{1}{2}\right\rangle_{mm} & m_j &= \frac{3}{2} \\
|2\rangle &= \left|2, 1, \frac{1}{2}, 1, -\frac{1}{2}\right\rangle_{mm} & m_j &= \frac{1}{2} \\
|3\rangle &= \left|2, 1, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} & m_j &= \frac{1}{2} \\
|4\rangle &= \left|2, 1, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm} & m_j &= -\frac{1}{2} \\
|5\rangle &= \left|2, 1, \frac{1}{2}, -1, \frac{1}{2}\right\rangle_{mm} & m_j &= -\frac{1}{2} \\
|6\rangle &= \left|2, 1, \frac{1}{2}, -1, -\frac{1}{2}\right\rangle_{mm} & m_j &= -\frac{3}{2} \\
|7\rangle &= \left|2, 0, \frac{1}{2}, 0, \frac{1}{2}\right\rangle_{mm} & m_j &= \frac{1}{2} \\
|8\rangle &= \left|2, 0, \frac{1}{2}, 0, -\frac{1}{2}\right\rangle_{mm} & m_j &= -\frac{1}{2}
\end{aligned}$$

then we get the matrix for $\langle \hat{H}_{so} \rangle$ as shown in Table 8.4 below.

I have used a table rather than an equation format so that I could clearly label the rows and columns by the state index.

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	b	c	0	0	0	0	0
3	0	c	d	0	0	0	0	0
4	0	0	0	e	f	0	0	0
5	0	0	0	f	g	0	0	0
6	0	0	0	0	0	h	0	0
7	0	0	0	0	0	0	p	0
8	0	0	0	0	0	0	0	q

Table 8.4: $\langle \hat{H}_{so} \rangle$ matrix

We have marked the non-zero elements. Using the operator properties derived earlier we get

$$
\begin{aligned}
a &= \langle 1 | \hat{H}_{so} | 1 \rangle = \left\langle 2,1,\frac{1}{2},1,\frac{1}{2} \right| \hat{H}_{so} \left| 2,1,\frac{1}{2},1,\frac{1}{2} \right\rangle_{mm} \\
&= \left\langle 2,1,\frac{1}{2},1,\frac{1}{2} \right| \frac{e^2}{2m^2c^2} \left(\frac{1}{r^3} \right) \left(\hat{L}_z \hat{S}_z + \frac{1}{2} \left(\hat{L}_+ \hat{S}_- + \hat{L}_- \hat{S}_+ \right) \right) \left| 2,1,\frac{1}{2},1,\frac{1}{2} \right\rangle_{mm} \\
&= \frac{e^2}{2m^2c^2} \langle 2,1 | \frac{1}{r^3} | 2,1 \rangle_{mm} \left\langle 2,1,\frac{1}{2},1,\frac{1}{2} \right| \left(\hat{L}_z \hat{S}_z + \frac{1}{2} \left(\hat{L}_+ \hat{S}_- + \hat{L}_- \hat{S}_+ \right) \right) \left| 2,1,\frac{1}{2},1,\frac{1}{2} \right\rangle_{mm} \\
&= \frac{e^2}{2m^2c^2} \langle 2,1 | \frac{1}{r^3} | 2,1 \rangle_{mm} \left\langle 2,1,\frac{1}{2},1,\frac{1}{2} \right| \hat{L}_z \hat{S}_z \left| 2,1,\frac{1}{2},1,\frac{1}{2} \right\rangle \\
&= \frac{e^2\hbar^2}{4m^2c^2} \int_0^\infty \frac{1}{r} R_{n\ell}^2(r) dr = \frac{e^2\hbar^2}{4m^2c^2} \left[\frac{Z^3}{a_0^3 n^3 \ell(\ell+\frac{1}{2})(\ell+1)} \right]
\end{aligned}
\tag{8.316}
$$

Similar calculations give $\langle \hat{H}_{so} \rangle$ as shown in Table 8.5 below.

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	-a	$\sqrt{2}$a	0	0	0	0	0
3	0	$\sqrt{2}$a	0	0	0	0	0	0
4	0	0	0	0	$\sqrt{2}$a	0	0	0
5	0	0	0	$\sqrt{2}$a	-a	0	0	0
6	0	0	0	0	0	a	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

Table 8.5: $\langle \hat{H}_{so} \rangle$ matrix - revised

This says that (only diagonal elements)

$$E_1^{(1)} = a = E_6^{(1)} \quad , \quad E_7^{(1)} = 0 = E_8^{(1)} \tag{8.317}$$

and

$$|1\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{3}{2}\right\rangle_{jm} \text{ and } |6\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{3}{2}\right\rangle_{jm}$$
$$|7\rangle = \left|2, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \text{ and } |8\rangle = \left|2, 0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\rangle_{jm}$$

In order to find $E_{2'}^{(1)}, E_{3'}^{(1)}, E_{4'}^{(1)}, E_{5'}^{(1)}$, corresponding to the new zero-order state vectors $|2'\rangle, |3'\rangle, |4'\rangle, |5'\rangle$, we must diagonalize the two 2×2 submatrices.

We begin with the submatrix involving states $|2\rangle$ and $|3\rangle$ as shown in Table 8.6, namely,

	2	3
2	-a	$\sqrt{2}$a
3	$\sqrt{2}$a	0

Table 8.6: $\langle \hat{H}_{so} \rangle$ 2-3 submatrix

The characteristic equation is

$$(-a - E)(-E) - 2a^2 = 0 = E^2 + aE - 2a^2 \tag{8.318}$$

or

$$E_{2'}^{(1)} = a \text{ and } E_{3'}^{(1)} = -2a \tag{8.319}$$

Notice that these energies are

$$E_{2'}^{(1)} = \ell a \text{ and } E_{3'}^{(1)} = -(\ell + 1)a \tag{8.320}$$

as expected for the

$$j = \ell + \frac{1}{2} \text{ and } j = \ell - \frac{1}{2} \tag{8.321}$$

states respectively.

We find the eigenvectors using the eigenvalue equations. For $|2'\rangle$ we have

$$\begin{pmatrix} -a & a\sqrt{2} \\ a\sqrt{2} & 0 \end{pmatrix} |2'\rangle = \begin{pmatrix} -a & a\sqrt{2} \\ a\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = E_{2'}^{(1)} \begin{pmatrix} u \\ v \end{pmatrix} = a \begin{pmatrix} u \\ v \end{pmatrix}$$

or

$$-u + \sqrt{2}v = u \text{ and } \sqrt{2}u = v \tag{8.322}$$

Using the normalization condition $u^2 + v^2 = 1$ we get

$$u = \sqrt{\frac{1}{3}} \text{ and } v = \sqrt{\frac{2}{3}} \tag{8.323}$$

or

$$|2'\rangle = \sqrt{\frac{1}{3}}|2\rangle + \sqrt{\frac{2}{3}}|3\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, \frac{1}{2}\right\rangle_{jm} \tag{8.324}$$

and similarly

$$|3'\rangle = \sqrt{\frac{2}{3}}|2\rangle - \sqrt{\frac{1}{3}}|3\rangle = \left|2, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right\rangle_{jm} \tag{8.325}$$

We then deal with the submatrix involving states $|4\rangle$ and $|5\rangle$ as shown in Table 8.7, namely,

	4	5
4	0	$\sqrt{2}$a
5	$\sqrt{2}$a	-a

Table 8.7: $\langle \hat{H}_{so} \rangle$ 4-5 submatrix

The characteristic equation is

$$(-a - E)(-E) - 2a^2 = 0 = E^2 + aE - 2a^2 \tag{8.326}$$

or

$$E^{(1)}_{4'} = a \text{ and } E^{(1)}_{5'} = -2a \tag{8.327}$$

and the eigenvectors are

$$|4'\rangle = \sqrt{\frac{2}{3}}|4\rangle + \sqrt{\frac{1}{3}}|5\rangle = \left|2, 1, \frac{1}{2}, \frac{3}{2}, -\frac{1}{2}\right\rangle_{jm} \tag{8.328}$$

$$|5'\rangle = \sqrt{\frac{1}{3}}|4\rangle - \sqrt{\frac{2}{3}}|5\rangle = \left|2, 1, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\rangle_{jm} \tag{8.329}$$

So including the spin-orbit correction we end up with the energy levels

$E^{(0)}_2 + a$ for $1, 2', 4', 6 \rightarrow$ 4 - fold degenerate

$E^{(0)}_2$ for $7, 8 \rightarrow$ 2 - fold degenerate

$E^{(0)}_2 - 2a$ for $3', 5' \rightarrow$ 2 - fold degenerate

8.6.2 Spin-Orbit and Arbitrary Magnetic Field

Now let us add on the Zeeman correction (for $\vec{B} = B\hat{z}$)

$$\hat{H}_{Zeeman} = \frac{\mu_B}{\hbar}(\vec{L}_{op} + 2\vec{S}_{op}) \cdot \vec{B} = \frac{\mu_B B}{\hbar}(\hat{L}_z + 2S_z) \tag{8.330}$$

We can solve this problem for *arbitrary magnetic field* by repeating Method #2 using the correction term as the sum of spin-orbit and Zeeman effects.

The zero-order Hamiltonian is $\hat{H}_0$ and the zero-order state vectors are the $|n, \ell, s, m_\ell, m_s\rangle$ states. The eight zero-order $n = 2$ states are all degenerate with energy

$$E_2^{(0)} = -\frac{e^2}{8a_0} \tag{8.331}$$

so we must use degenerate perturbation theory.

We have already calculated the $\langle \hat{H}_{so} \rangle$ in this basis. It is shown in Table 8.8 below.

	1	2	3	4	5	6	7	8
1	a	0	0	0	0	0	0	0
2	0	-a	$\sqrt{2}$a	0	0	0	0	0
3	0	$\sqrt{2}$a	0	0	0	0	0	0
4	0	0	0	0	$\sqrt{2}$a	0	0	0
5	0	0	0	$\sqrt{2}$a	-a	0	0	0
6	0	0	0	0	0	a	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

Table 8.8: $\langle \hat{H}_{so} \rangle$ matrix

where

$$a = \frac{e^2\hbar^2}{96m^2a_0^3c^2} \tag{8.332}$$

The $\langle \hat{L}_z + 2\hat{S}_z \rangle$ matrix is diagonal in this representation and its diagonal elements are given by $m_\ell + 2m_s$ and so the Zeeman contribution $\langle \hat{H}_{Zeeman} \rangle$ is shown in Table 8.9 below.

	1	2	3	4	5	6	7	8
1	2b	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	b	0	0	0	0	0
4	0	0	0	-b	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	-2b	0	0
7	0	0	0	0	0	0	b	0
8	0	0	0	0	0	0	0	-b

Table 8.9: $\langle \hat{H}_{Zeeman} \rangle$ matrix

where $b = \mu_B B$. The combined perturbation matrix $\langle \hat{V} \rangle$ is then given in Table 8.10 below.

	1	2	3	4	5	6	7	8
1	a+2b	0	0	0	0	0	0	0
2	0	-a	$a\sqrt{2}$	0	0	0	0	0
3	0	$a\sqrt{2}$	b	0	0	0	0	0
4	0	0	0	-b	$a\sqrt{2}$	0	0	0
5	0	0	0	$a\sqrt{2}$	-a	0	0	0
6	0	0	0	0	0	a-2b	0	0
7	0	0	0	0	0	0	b	0
8	0	0	0	0	0	0	0	-b

Table 8.10: $\langle \hat{V} \rangle$ matrix

After diagonalizing, the new energies are

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B$$
$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + \sqrt{9a^2 + 2ab + b^2})$$
$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - \sqrt{9a^2 + 2ab + b^2})$$
$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + \sqrt{9a^2 - 2ab + b^2})$$

$$E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - \sqrt{9a^2 - 2ab + b^2})$$
$$E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$
$$E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$
$$E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

If we let B be small so that $b \ll a$, we then get the approximate energies

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B$$
$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + 3a(1 + \frac{b}{9a}))$$
$$= -\frac{e^2}{8a_0} + a + \frac{2}{3}b = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + \frac{2}{3}\mu_B B$$
$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - 3a(1 + \frac{b}{9a}))$$
$$= -\frac{e^2}{8a_0} - 2a + \frac{1}{3}b = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0}\frac{\alpha^2}{6} + \frac{1}{3}\mu_B B$$
$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + 3a(1 - \frac{b}{9a}))$$
$$= -\frac{e^2}{8a_0} + a - \frac{2}{3}b = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - \frac{2}{3}\mu_B B$$
$$E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - 3a(1 - \frac{b}{9a}))$$
$$= -\frac{e^2}{8a_0} - 2a - \frac{1}{3}b = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0}\frac{\alpha^2}{6} - \frac{1}{3}\mu_B B$$
$$E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$
$$E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$
$$E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

This is clearly a perturbation of the spin-orbit energy levels. We assume that the new state vectors become the zero-order vectors in the spin-orbit case for low fields. In this case the Zeeman effect corrections(to the fine structure energies) are given by Table 8.11 below.

ℓ	s	j	m_j	ΔE_{Zeeman}	State
1	1/2	3/2	3/2	$2\mu_B B$	$1'$
1	1/2	3/2	1/2	$2\mu_B B/3$	$2'$
1	1/2	1/2	1/2	$\mu_B B/3$	$3'$
1	1/2	3/2	-1/2	$-2\mu_B B/3$	$4'$
1	1/2	1/2	-1/2	$-\mu_B B/3$	$5'$
1	1/2	3/2	-3/2	$-2\mu_B B$	$6'$
0	1/2	1/2	1/2	$\mu_B B$	$7'$
0	1/2	1/2	-1/2	$-\mu_B B$	$8'$

Table 8.11: $n = 2$ energy corrections for small B

A little bit of study shows the general relation

$$\Delta E_{Zeeman} = g\mu_B B m_j \tag{8.333}$$

where

$$g = \text{ Lande g - factor } = 1 + \frac{j(j+1) - \ell(\ell+1) + s(s+1)}{2j(j+1)} \tag{8.334}$$

This is called the *Zeeman effect*.

We can prove this result in general. The general method uses the *Wigner-Eckart Theorem*.

8.6.3 Wigner-Eckart Theorem

Consider a vector operator $\vec{A}_{op}$. We have already shown that the Cartesian components of any vector operator has the following commutation relations with the Cartesian components of the angular momentum operator

$$\left[\hat{A}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{A}_k \tag{8.335}$$

We will now prove the following powerful theorem:

In a basis that diagonalizes $\vec{J}^2_{op}$ and $\hat{J}_z$ (i.e., the $|\lambda, \ell, s, j, m_j\rangle$ states, where λ signifies other operators that commute with $\vec{J}^2_{op}$ and $\hat{J}_z$), the matrix elements of $\vec{A}_{op}$ between states with the *same* j–value are proportional to the matrix elements of $\vec{J}_{op}$ and the proportionality factor is independent of m_j.

The algebra involved in the proof is simpler if we work in the so-called *spherical basis* instead of the Cartesian basis. The spherical basis uses

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y \quad , \quad \hat{J}_0 = \hat{J}_z \tag{8.336}$$

$$\hat{A}_\pm = \hat{A}_x \pm i\hat{A}_y \quad , \quad \hat{A}_0 = \hat{A}_z \tag{8.337}$$

The corresponding commutators are

$$\left[\hat{A}_\pm, \hat{J}_0\right] = \mp\hbar\hat{A}_\pm \quad , \quad \left[\hat{A}_\pm, \hat{J}_\pm\right] = 0 \tag{8.338}$$

$$\left[\hat{A}_\pm, \hat{J}_\mp\right] = \pm 2\hbar\hat{A}_0 \quad , \quad \left[\hat{A}_0, \hat{J}_0\right] = 0 \tag{8.339}$$

$$\left[\hat{A}_0, \hat{J}_\pm\right] = \pm\hbar\hat{A}_\pm \tag{8.340}$$

which all follow from the original commutator for the arbitrary vector operator. Now, by definition of the operators, we have

$$\hat{J}_0 |\lambda, j, m_j\rangle = \hbar m_j |\lambda, j, m_j\rangle = \langle j, m_j| \hat{J}_0 |j, m_j\rangle |\lambda, j, m_j\rangle \tag{8.341}$$

$$\begin{aligned}\hat{J}_\pm |\lambda, j, m_j\rangle &= \hbar\sqrt{j(j+1) - m_j(m_j \pm 1)}\, |\lambda, j, m_j \pm 1\rangle \\ &= \langle j, m_j \pm 1| \hat{J}_\pm |j, m_j\rangle |\lambda, j, m_j \pm 1\rangle\end{aligned} \tag{8.342}$$

$$\begin{aligned}\langle\lambda, j, m_j| \hat{J}_\pm &= \left(\hat{J}_\mp |\lambda, j, m_j\rangle\right)^+ \\ &= \langle\lambda, j, m_j \mp 1| \hbar\sqrt{j(j+1) - m_j(m_j \mp 1)} \\ &= \langle\lambda, j, m_j \mp 1| \langle j, m_j| \hat{J}_\pm |j, m_j \mp 1\rangle\end{aligned} \tag{8.343}$$

We now work with the matrix elements of some of the commutators and use the defining relations above to prove the theorem.

First, we have

$$\begin{aligned}\mp\hbar \left\langle\lambda', j, m_j'\right| \hat{A}_\pm |\lambda, j, m_j\rangle &= \left\langle\lambda', j, m_j'\right| \left[\hat{A}_\pm, \hat{J}_0\right] |\lambda, j, m_j\rangle \\ &= \left\langle\lambda', j, m_j'\right| \left[\hat{A}_\pm \hat{J}_0 - \hat{J}_0 \hat{A}_\pm\right] |\lambda, j, m_j\rangle \\ &= (m_j - m_j')\hbar \left\langle\lambda', j, m_j'\right| \hat{A}_\pm |\lambda, j, m_j\rangle\end{aligned} \tag{8.344}$$

or

$$0 = (m_j - m_j' \pm 1)\hbar \left\langle\lambda', j, m_j'\right| \hat{A}_\pm |\lambda, j, m_j\rangle \tag{8.345}$$

This says that either $m_j' = m_j \pm 1$ or the matrix element

$$\left\langle\lambda', j, m_j'\right| \hat{A}_\pm |\lambda, j, m_j\rangle = 0 \tag{8.346}$$

Since we have an identical property for the matrix elements of $\hat{J}_\pm$ this implies that the matrix elements of $\hat{A}_\pm$ are proportional to those of $\hat{J}_\pm$ and we can write the proportionality constant as

$$\frac{\langle\lambda', j, m_j \pm 1| \hat{A}_\pm |\lambda, j, m_j\rangle}{\langle j, m_j \pm 1| \hat{J}_\pm |j, m_j\rangle} \tag{8.347}$$

Second, we have

$$\left\langle\lambda', j, m_j'\right| \left[\hat{A}_\pm, \hat{J}_\pm\right] |\lambda, j, m_j\rangle = 0 \tag{8.348}$$

$$\left\langle\lambda', j, m_j'\right| \hat{A}_\pm \hat{J}_\pm |\lambda, j, m_j\rangle = \left\langle\lambda', j, m_j'\right| \hat{J}_\pm \hat{A}_\pm |\lambda, j, m_j\rangle \tag{8.349}$$

$$\begin{aligned}&\left\langle\lambda', j, m_j'\right| \hat{A}_\pm |\lambda, j, m_j \pm 1\rangle \langle\lambda', j, m_j \pm 1| \hat{J}_\pm |\lambda, j, m_j\rangle \\ &\quad = \left\langle\lambda', j, m_j' \mp 1\right| \hat{A}_\pm |\lambda, j, m_j\rangle \langle\lambda, j, m_j \pm 1| \hat{J}_\pm \left|\lambda, j, m_j' \mp 1\right\rangle\end{aligned} \tag{8.350}$$

Using the result from the first commutator this says that $m'_j = m_j \pm 2$, which, in turn, implies that

$$\frac{\langle \lambda', j, m_j \pm 2 | \hat{A}_\pm | \lambda, j, m_j \pm 1 \rangle}{\langle j, m_j \pm 2 | \hat{J}_\pm | j, m_j \pm 1 \rangle} = \frac{\langle \lambda', j, m_j \pm 1 | \hat{A}_\pm | \lambda, j, m_j \rangle}{\langle j, m_j \pm 1 | \hat{J}_\pm | j, m_j \rangle} \tag{8.351}$$

This says that the proportionality constant is independent of m_j.

We define a new symbol for the proportionality constant

$$\langle \lambda', j || A || \lambda, j \rangle_\pm = \text{ the reduced matrix element} \tag{8.352}$$

which gives the relation

$$\langle \lambda', j, m'_j | \hat{A}_\pm | \lambda, j, m_j \rangle = \langle \lambda', j || A || \lambda, j \rangle_\pm \langle j, m'_j | \hat{J}_\pm | j, m_j \rangle \tag{8.353}$$

To complete the proof we need to show that the same result holds for $\hat{A}_0$ and that

$$\langle \lambda', j || A || \lambda, j \rangle_+ = \langle \lambda', j || A || \lambda, j \rangle_- \tag{8.354}$$

We have

$$\begin{aligned}
\pm 2\hbar \langle \lambda', j, m'_j | \hat{A}_0 | \lambda, j, m_j \rangle &= \langle \lambda', j, m'_j | \left[\hat{A}_\pm, \hat{J}_\mp \right] | \lambda, j, m_j \rangle \\
&= \langle \lambda', j, m'_j | \left[\hat{A}_\pm \hat{J}_\mp - \hat{J}_\mp \hat{A}_\pm \right] | \lambda, j, m_j \rangle \\
&= \langle \lambda', j, m'_j | \hat{A}_\pm | \lambda, j, m_j \mp 1 \rangle \langle j, m_j \mp 1 | \hat{J}_\mp | j, m_j \rangle \\
&\quad - \langle \lambda', j, m'_j \pm 1 | \hat{A}_\pm | \lambda, j, m_j \rangle \langle j, m'_j | \hat{J}_\mp | j, m'_j \pm 1 \rangle
\end{aligned} \tag{8.355}$$

Now substituting in the matrix element of $\hat{A}_\pm$ we get

$$\begin{aligned}
&\pm 2\hbar \langle \lambda', j, m'_j | \hat{A}_0 | \lambda, j, m_j \rangle \\
&\quad = \langle \lambda', j || A || \lambda, j \rangle_\pm \left[\langle j, m'_j | \hat{J}_\pm | j, m_j \mp 1 \rangle \langle j, m_j \mp 1 | \hat{J}_\mp | j, m_j \rangle \right. \\
&\qquad \left. - \langle j, m'_j \pm 1 | \hat{J}_\pm | j, m_j \rangle \langle j, m'_j | \hat{J}_\mp | j, m'_j \pm 1 \rangle \right]
\end{aligned} \tag{8.356}$$

This says that $\hat{A}_0$ has non-vanishing matrix elements only when $m'_j = m_j$. We then get

$$\begin{aligned}
&\pm 2\hbar \langle \lambda', j, m'_j | \hat{A}_0 | \lambda, j, m_j \rangle \\
&\quad = \langle \lambda', j || A || \lambda, j \rangle_\pm \left[\left| \langle j, m_j \mp 1 | \hat{J}_\mp | j, m_j \rangle \right|^2 - \left| \langle j, m_j \pm 1 | \hat{J}_\mp | j, m_j \rangle \right|^2 \right] \\
&\quad = \pm 2\hbar m_j = \pm 2\hbar \langle \lambda', j, m'_j | \hat{J}_0 | \lambda, j, m_j \rangle
\end{aligned} \tag{8.357}$$

Putting it all together we get

$$\langle \lambda', j, m'_j | \hat{A}_0 | \lambda, j, m_j \rangle = \langle \lambda', j || A || \lambda, j \rangle_\pm \langle j, m'_j | \hat{J}_0 | j, m_j \rangle \tag{8.358}$$

Since no operator has a $\pm$ subscript, this also says that

$$\langle \lambda', j || A || \lambda, j \rangle_+ = \langle \lambda', j || A || \lambda, j \rangle_- = \langle \lambda', j || A || \lambda, j \rangle \tag{8.359}$$

and we finally have

$$\left\langle \lambda', j, m_j' \right| \vec{A}_{op} \left| \lambda, j, m_j \right\rangle = \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle_{\pm} \left\langle j, m_j' \right| \vec{J}_{op} \left| j, m_j \right\rangle \tag{8.360}$$

This completes the proof of the *Wigner-Eckart theorem*.

A very important extension of this theorem is the following result:

$$\begin{aligned} \left\langle \lambda', j, m_j' \right| \vec{A}_{op} \cdot \vec{J}_{op} \left| \lambda, j, m_j \right\rangle &= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle j, m_j' \right| \vec{J}_{op}^2 \left| j, m_j \right\rangle \\ &= \delta_{m_j', m_j} \hbar^2 j(j+1) \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \end{aligned} \tag{8.361}$$

which says that the scalar product is diagonal in m_j. This result follows directly from the Wigner-Eckart theorem

$$\begin{aligned} \left\langle \lambda', j, m_j' \right| \vec{A}_{op} \cdot \vec{J}_{op} \left| \lambda, j, m_j \right\rangle &= \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} J_{op,k} \left| \lambda, j, m_j \right\rangle \\ &= \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} \left(\sum_{m_j''} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| \right) J_{op,k} \left| \lambda, j, m_j \right\rangle \\ &= \sum_{\lambda'', j'', m_j''} \sum_k \left\langle \lambda', j, m_j' \right| A_{op,k} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| J_{op,k} \left| \lambda, j, m_j \right\rangle \\ &= \sum_{m_j''} \sum_k \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle \lambda', j, m_j' \right| J_{op,k} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| J_{op,k} \left| \lambda, j, m_j \right\rangle \\ &= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \sum_k \left\langle \lambda', j, m_j' \right| J_{op,k} \left(\sum_{m_j''} \left| \lambda, j, m_j'' \right\rangle \left\langle \lambda, j, m_j'' \right| \right) J_{op,k} \left| \lambda, j, m_j \right\rangle \\ &= \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \left\langle j, m_j' \right| \vec{J}_{op}^2 \left| j, m_j \right\rangle = \delta_{m_j', m_j} \hbar^2 j(j+1) \left\langle \lambda', j \right| |A| \left| \lambda, j \right\rangle \end{aligned}$$

Now back to the Zeeman effect. In the low field limit, we need to evaluate the diagonal matrix elements

$$\begin{aligned} \left\langle \ell s j m_j \right| (\hat{L}_z + 2\hat{S}_z) \left| \ell s j m_j \right\rangle &= \left\langle \ell s j m_j \right| (\hat{J}_z + \hat{S}_z) \left| \ell s j m_j \right\rangle \\ &= \hbar m_j + \left\langle \ell s j m_j \right| \hat{S}_z \left| \ell s j m_j \right\rangle \end{aligned} \tag{8.362}$$

Now the Wigner-Eckart theorem says that

$$\begin{aligned} \left\langle \ell s j m_j \right| \hat{S}_z \left| \ell s j m_j \right\rangle &= \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle \left\langle \ell s j m_j \right| \hat{J}_z \left| \ell s j m_j \right\rangle \\ &= \hbar m_j \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle \end{aligned} \tag{8.363}$$

The scalar product matrix element formula gives

$$\begin{aligned} \left\langle \ell s j m_j \right| \vec{S}_{op} \cdot \vec{J}_{op} \left| \ell s j m_j \right\rangle &= \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle \left\langle j m_j \right| \vec{J}_{op}^2 \left| j m_j \right\rangle \\ &= \hbar^2 j(j+1) \left\langle \ell s j \right| |S| \left| \ell s j \right\rangle \end{aligned} \tag{8.364}$$

But we also have

$$(\vec{J}_{op} - \vec{S}_{op})^2 = \vec{L}_{op}^2 = \vec{J}_{op}^2 + \vec{S}_{op}^2 - 2\vec{S}_{op} \cdot \vec{J}_{op} \tag{8.365}$$

$$\langle \ell s j m_j | \vec{S}_{op} \cdot \vec{J}_{op} | \ell s j m_j \rangle = \frac{1}{2} \langle \ell s j m_j | (\vec{J}_{op}^2 + \vec{S}_{op}^2 - \vec{L}_{op}^2) | \ell s j m_j \rangle$$
$$= \frac{1}{2} \hbar^2 (j(j+1) + s(s+1) - \ell(\ell+1)) \tag{8.366}$$

or

$$\langle \ell s j | |S| | \ell s j \rangle = \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)} \tag{8.367}$$

and thus

$$\langle \ell s j m_j | (\hat{L}_z + 2\hat{S}_z) | \ell s j m_j \rangle$$
$$= \hbar m_j + \hbar m_j \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)}$$
$$= \hbar m_j g_{j\ell s} \tag{8.368}$$
$$g_{j\ell s} = 1 + \frac{j(j+1) + s(s+1) - \ell(\ell+1)}{2j(j+1)} = \text{ Lande g - factor} \tag{8.369}$$

Finally, we have

$$\langle \ell s j m_j | \hat{H}_{Zeeman} | \ell s j m_j \rangle = \mu_B B m_j g_{j\ell s} \tag{8.370}$$

and the result we found earlier in the special example case is now proved in general.

8.6.4 Paschen-Bach Effect

When B is large enough such that $\Delta E_{Zeeman} \gg \Delta E_{so}$, but not large enough so that the $\vec{B}^2$ term we neglected earlier is important. We have the so-called *Paschen-Bach* effect. If the $\vec{B}^2$ term is dominant we have the so-called *quadratic Zeeman effect*.

The best way to see what is happening for all magnetic field values is a plot. In CGS Gaussian units

$$\mu_B = 5.7884 \times 10^{-9} \frac{eV}{gauss} \ , \ a_0 = 5.2918 \times 10^{-8} \, cm \ , \ e = 4.80 \times 10^{-10} \, esu$$
$$\frac{e^2}{a_0} = 27.2 \, eV a = 1.509 \times 10^{-5} \, eV b = 5.7884 \times 10^{-9} B \, eV$$

Using our earlier results we then have

$$E_{1'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} + 2\mu_B B$$
$$E_{2'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) + \sqrt{9a^2 + 2ab + b^2})$$
$$E_{3'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a-b) - \sqrt{9a^2 + 2ab + b^2})$$
$$E_{4'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) + \sqrt{9a^2 - 2ab + b^2})$$
$$E_{5'} = -\frac{e^2}{8a_0} + \frac{1}{2}(-(a+b) - \sqrt{9a^2 - 2ab + b^2})$$
$$E_{6'} = -\frac{e^2}{8a_0} + \frac{e^2}{8a_0}\frac{\alpha^2}{12} - 2\mu_B B$$
$$E_{7'} = -\frac{e^2}{8a_0} + \mu_B B$$
$$E_{8'} = -\frac{e^2}{8a_0} - \mu_B B$$

A plot of

$$\left(E + \frac{e^2}{8a_0}\right) \times 10^5\, eV \text{ versus } \log_e(B(gauss))$$

looks like Figure 8.2 below.

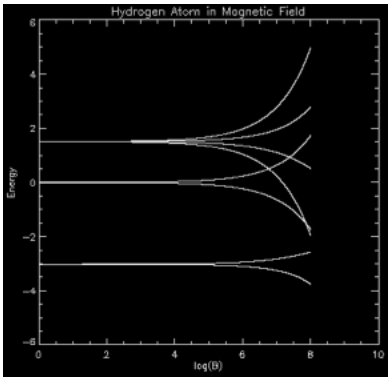


Figure 8.2: Hydrogen Atom In a Magnetic Field - Zeeman Effect

This plot for fields below 400 gauss ($\log_e(B) \approx 6$) shows the characteristic level structure of the Zeeman effect.

The very large magnetic field Paschen-Bach effect is illustrated in Figure 8.3 below.

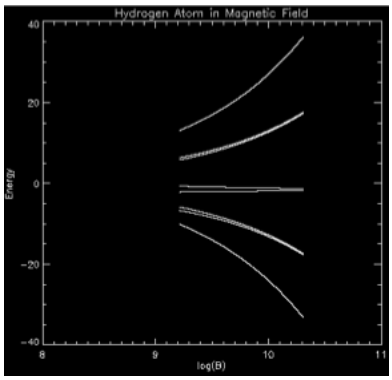


Figure 8.3: Hydrogen Atom In a Magnetic Field - Paschen-Bach Effect

Notice the equally-spaced level signature of the Paschen-Bach effect.

We now define some notation that will be important later as we study atomic spectra. For small magnetic fields we found that the approximate state vectors are the $|n\ell sjm_j\rangle$ states. The energy levels including spin-orbit effects are

$$\begin{aligned} E_n &= E_n^{(0)} + \Delta E_{so} \\ &= -\frac{Ze^2}{2a_0 n^2} \\ &\quad + \frac{Z^2\alpha^2 \left|E_n^{(0)}\right|}{n\ell(2\ell+1)(\ell+1)}(1-\delta_{\ell,0})\left(\ell\delta_{j,\ell+1/2} - (\ell+1)\delta_{j,\ell-1/2}\right) \end{aligned} \tag{8.371}$$

We define a spectroscopic notation to label the energy levels using the scheme shown below:

$$|n\ell sjm_j\rangle \to n^{2S+1}L(symbol)_J \tag{8.372}$$

so that

$$\left|21\frac{1}{2}\frac{3}{2}m_j\right\rangle \to 2^2P_{\frac{3}{2}}\ ,\ \left|21\frac{1}{2}\frac{1}{2}m_j\right\rangle \to 2^2P_{\frac{1}{2}}\ ,\ \left|20\frac{1}{2}\frac{1}{2}m_j\right\rangle \to 2^2S_{\frac{1}{2}}$$

The $L(symbols)$ are defined by Table 8.12 below.

L	0	1	2	3
Symbol	S	P	D	F

Table 8.12: Spectroscopic Labels

The energy level diagram for $n = 1$ and $n = 2$ is shown in Figure 8.4 below.

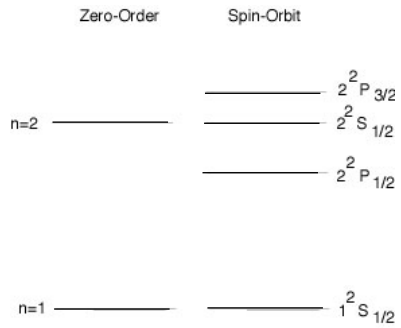


Figure 8.4: Spin-Orbit Energy Levels

Earlier we calculated the relativistic correction and found that it was the same order of magnitude as the spin-orbit correction for hydrogen. We found

$$\Delta E_{rel} = -\frac{Z^2\alpha^2\left|E_n^{(0)}\right|}{n}\left(\frac{2}{2\ell+1}-\frac{3}{4n}\right) \tag{8.373}$$

Combining these two corrections we have

$$\Delta E_{fs} = \Delta E_{so} + \Delta E_{rel} = -\frac{Z^2\alpha^2\left|E_n^{(0)}\right|}{n}\left(\frac{1}{j+\frac{1}{2}}-\frac{3}{4n}\right) \quad j = \ell \pm \frac{1}{2} \tag{8.374}$$

which is independent of ℓ. This changes the energy level structure to that shown in Figure 8.5 below.

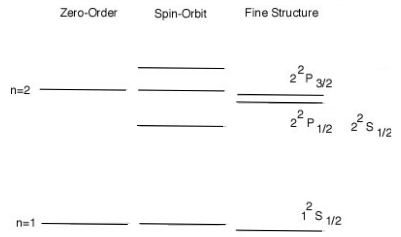


Figure 8.5: Fine Structure Energy Levels

The observed spectral lines result from an electron making a transition between these levels. We will discuss this topic later.

8.6.5 Stark Effect

When a hydrogen atom is placed in an external electric field $\vec{\mathcal{E}}_0$, the potential energy of the proton and the electron is given by

$$\begin{aligned} V_{dipole}(\vec{r}_e, \vec{r}_p) &= -e\vec{\mathcal{E}}_0 \cdot \vec{r}_p + e\vec{\mathcal{E}}_0 \cdot \vec{r}_e \\ &= e\mathcal{E}_0(z_e - z_p) = e\mathcal{E}_0 z \end{aligned} \tag{8.375}$$

where

$$z = z_e - z_p = z_{relative} \tag{8.376}$$

Therefore, we consider the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_{dipole} \tag{8.377}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{op}^2}{2m} - e^2 \left(\frac{1}{r}\right)_{op} \tag{8.378}$$

$$\hat{H}_{dipole} = e\mathcal{E}_0 z_{op} \tag{8.379}$$

For weak electric fields, we can apply perturbation theory (we ignore spin in this calculation). First, we apply perturbation theory to the $n = 1$ ground state of the hydrogen atom.

For the ground state, the wave function is $\psi_{100}(\vec{r})$ and the first-order correction to the energy is

$$\begin{aligned} E_1^{(1)} &= \langle 100 | e\mathcal{E}_0 z_{op} | 100 \rangle \\ &= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' \langle 100 \mid \vec{r} \rangle \langle \vec{r} | z_{op} | \vec{r}' \rangle \langle \vec{r}' \mid 100 \rangle \\ &= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' z \psi_{100}^*(\vec{r}) \langle \vec{r} \mid \vec{r}' \rangle \psi_{100}(\vec{r}') \\ &= e\mathcal{E}_0 \int d^3\vec{r} d^3\vec{r}' z \psi_{100}^*(\vec{r}) \delta(\vec{r} - \vec{r}') \psi_{100}(\vec{r}') \\ &= e\mathcal{E}_0 \int d^3\vec{r} z |\psi_{100}(\vec{r})|^2 \end{aligned} \tag{8.380}$$

This equals zero since the integrand is the product of an even and odd functions. Thus, the first-order correction is zero for the ground state.

The second-order correction is given by non-degenerate perturbation theory as

$$E_1^{(2)} = \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{\ell} \frac{|\langle n\ell m | e\mathcal{E}_0 z_{op} | 100 \rangle|^2}{E_1^{(0)} - E_m^{(0)}} \tag{8.381}$$

Using $z = r\cos\theta$ we have

$$\langle n\ell m | z_{op} | 100 \rangle = \int d^3\vec{r} \left[R_{n\ell}(r) Y_{\ell m}^*(\theta, \phi)\right] [r\cos\theta] R_{10}(r) Y_{00}(\theta, \phi) \tag{8.382}$$

Now

$$Y_{00} = \frac{1}{\sqrt{4\pi}} \text{ and } z = \sqrt{\frac{4\pi}{3}} Y_{10} \tag{8.383}$$

Therefore,

$$\langle n\ell m | z_{op} | 100 \rangle = \int r^3 dr R_{n\ell}(r) R_{10}(r) \frac{1}{\sqrt{3}} \int d\Omega Y^*_{\ell m}(\theta,\phi) Y_{10}(\theta,\phi) \tag{8.384}$$

Now

$$\int d\Omega Y^*_{\ell m}(\theta,\phi) Y_{10}(\theta,\phi) = \delta_{\ell,1}\delta_{m,0} \tag{8.385}$$

by the orthonormality of the $(\vec{L}^2_{op}, \hat{L}_z)$ eigenfunctions. Therefore,

$$\langle n\ell m | z_{op} | 100 \rangle = \frac{1}{\sqrt{3}} \delta_{\ell,1}\delta_{m,0} \int_0^\infty r^3 dr R_{n1}(r) R_{10}(r) \tag{8.386}$$

and

$$|\langle n10 | z_{op} | 100 \rangle|^2 = \frac{1}{3} \frac{2^8 n^7 (n-1)^{2n-5}}{(n+1)^{2n+5}} a_0^2 = \beta(n) a_0^2 \tag{8.387}$$

Finally,

$$E_1^{(2)} = (e\mathcal{E}_0 a_0)^2 \sum_{n=2}^{\infty} \frac{\beta(n)}{\frac{e^2}{2a_0}\left(1 - \frac{1}{n^2}\right)} = -2F\mathcal{E}_0^2 a_0^3 \tag{8.388}$$

where

$$F = \sum_{n=2}^{\infty} \frac{n^2 \beta(n)}{(n^2-1)} \approx 1.125 \tag{8.389}$$

Therefore, the ground state exhibits a *quadratic* Stark effect.

The $n = 2$ level, which is the first excited state of hydrogen, has 4 degenerate states.

$$\begin{aligned} n = 2 \to \ell = 0 \to \psi_{200} &= \psi_1 \\ \to \ell = 1 \to m = &\begin{cases} 1 \to \psi_{211} = \psi_2 \\ 0 \to \psi_{210} = \psi_3 \\ -1 \to \psi_{21-1} = \psi_4 \end{cases} \end{aligned}$$

We must use degenerate perturbation theory. We construct the 4×4 $\langle e\mathcal{E}_0 z_{op} \rangle$ matrix and then diagonalize it. We have

$$\langle e\mathcal{E}_0 z_{op} \rangle = e\mathcal{E}_0 \begin{pmatrix} \langle 1 | z_{op} | 1 \rangle & \langle 1 | z_{op} | 2 \rangle & \langle 1 | z_{op} | 3 \rangle & \langle 1 | z_{op} | 4 \rangle \\ \langle 2 | z_{op} | 1 \rangle & \langle 2 | z_{op} | 2 \rangle & \langle 2 | z_{op} | 3 \rangle & \langle 2 | z_{op} | 4 \rangle \\ \langle 3 | z_{op} | 1 \rangle & \langle 3 | z_{op} | 2 \rangle & \langle 3 | z_{op} | 3 \rangle & \langle 3 | z_{op} | 4 \rangle \\ \langle 4 | z_{op} | 1 \rangle & \langle 4 | z_{op} | 2 \rangle & \langle 4 | z_{op} | 3 \rangle & \langle 4 | z_{op} | 4 \rangle \end{pmatrix} \tag{8.390}$$

Now z has no φ dependence and therefore,

$$\langle j | z_{op} | k \rangle = 0 \text{ if } m_j \neq m_k \tag{8.391}$$

Thus,

$$\begin{aligned}
\langle 1|\, z_{op}\, |2\rangle &= 0 = \langle 1|\, z_{op}\, |4\rangle \\
\langle 2|\, z_{op}\, |1\rangle &= 0 = \langle 2|\, z_{op}\, |3\rangle = \langle 2|\, z_{op}\, |4\rangle \\
\langle 3|\, z_{op}\, |2\rangle &= 0 = \langle 3|\, z_{op}\, |4\rangle \\
\langle 4|\, z_{op}\, |1\rangle &= 0 = \langle 4|\, z_{op}\, |2\rangle = \langle 4|\, z_{op}\, |3\rangle
\end{aligned}$$

and the matrix becomes

$$\langle e\mathcal{E}_0 z_{op}\rangle = e\mathcal{E}_0 \begin{pmatrix} \langle 1|\, z_{op}\, |1\rangle & 0 & \langle 1|\, z_{op}\, |3\rangle & 0 \\ 0 & \langle 2|\, z_{op}\, |2\rangle & 0 & 0 \\ \langle 3|\, z_{op}\, |1\rangle & 0 & \langle 3|\, z_{op}\, |3\rangle & 0 \\ 0 & 0 & 0 & \langle 4|\, z_{op}\, |4\rangle \end{pmatrix} \tag{8.392}$$

We also have

$$\langle 1|\, z_{op}\, |1\rangle = 0 = \langle 2|\, z_{op}\, |2\rangle = \langle 3|\, z_{op}\, |3\rangle = \langle 4|\, z_{op}\, |4\rangle \tag{8.393}$$

since these integrands involve the product of even and odd functions.

Finding out which matrix elements are equal to zero without actually evaluating the integrals corresponds to finding what are called *selection rules*. We will elaborate on the idea of selection rules in the next section on the Van der Waal's interaction.

Thus, the matrix finally becomes (after relabeling the rows and columns)

$$\langle e\mathcal{E}_0 z_{op}\rangle = e\mathcal{E}_0 \begin{pmatrix} 0 & \langle 1|\, z_{op}\, |3\rangle & 0 & 0 \\ \langle 3|\, z_{op}\, |1\rangle & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{8.394}$$

where

$$\langle 1|\, z_{op}\, |3\rangle = \langle 3|\, z_{op}\, |1\rangle = \int \psi_{200}(\vec{r}) z \psi_{210}(\vec{r}) d^3\vec{r} = -3e\mathcal{E}_0 a_0 \tag{8.395}$$

Diagonalizing the 2×2 submatrix gives eigenvalues $\pm 3e\mathcal{E}_0 a_0$. The first-order energies and new zero-order wave functions are

$$\psi_{211}(\vec{r}) \to E_{211} = E_2^{(0)} \text{ remains degenerate} \tag{8.396}$$

$$\psi_{21-1}(\vec{r}) \to E_{21-1} = E_2^{(0)} \text{ remains degenerate} \tag{8.397}$$

$$\begin{aligned}
\psi_{(+)}(\vec{r}) &= \frac{1}{\sqrt{2}} \left(\psi_{200}(\vec{r}) - \psi_{210}(\vec{r})\right) \to E_+ = E_2^{(0)} + 3e\mathcal{E}_0 a_0 \\
\psi_{(-)}(\vec{r}) &= \frac{1}{\sqrt{2}} \left(\psi_{200}(\vec{r}) + \psi_{210}(\vec{r})\right) \to E_- = E_2^{(0)} - 3e\mathcal{E}_0 a_0
\end{aligned}$$

The degeneracy is broken for the $m = 0$ levels and we see a *linear* Stark effect. The linear Stark effect only appears for degenerate levels.

8.6.6 Van der Waal's Interaction

We now consider a system consisting of two widely separated atoms. In particular, we consider the interaction between two hydrogen atoms, where we treat the two protons as fixed point charges separated by a vector $\vec{R}$ and we define

$$\begin{aligned}\vec{r}_1 &= \text{ vector from first proton to its electron}\\ \vec{r}_2 &= \text{ vector from second proton to its electron}\end{aligned}$$

as shown in Figure 8.6 below.

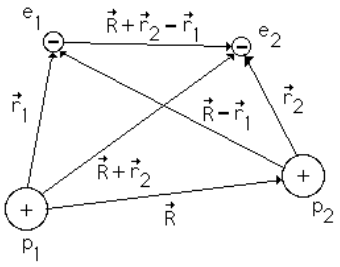


Figure 8.6: Van der Wall's System

The Hamiltonian is given by

$$\hat{H} = \hat{H}_0 + V \tag{8.398}$$

where

$$\hat{H}_0 = \frac{\vec{p}_{1,op}^{\,2}}{2\mu} - \frac{e^2}{r_1} + \frac{\vec{p}_{2,op}^{\,2}}{2\mu} - \frac{e^2}{r_2} = \text{ 2 non-interacting hydrogen atoms} \tag{8.399}$$

and

$$\begin{aligned}\hat{V} &= \text{ rest of the Coulomb interactions}\\ &= V_{p_1p_2} + V_{e_1e_2} + V_{e_1p_2} + V_{e_2p_1}\\ &= e^2\left(\frac{1}{R} + \frac{1}{\left|\vec{R}+\vec{r}_2-\vec{r}_1\right|} - \frac{1}{\left|\vec{R}+\vec{r}_2\right|} - \frac{1}{\left|\vec{R}-\vec{r}_1\right|}\right)\end{aligned} \tag{8.400}$$

This is the perturbation potential.

We know the zero-order solution for two non-interacting hydrogen atoms. It is

$$\text{zero-order states : } |n_1\ell_1m_1\rangle\,|n_2\ell_2m_2\rangle \tag{8.401}$$

with

$$\text{zero-order energies : } E^{(0)}_{n_1n_2} = -\frac{e^2}{2a_0}\left(\frac{1}{n_1^2} + \frac{1}{n_2^2}\right) \tag{8.402}$$

where

$$\hat{H}_0 \left|n_1\ell_1 m_1\right\rangle \left|n_2\ell_2 m_2\right\rangle = -\frac{e^2}{2a_0}\left(\frac{1}{n_1^2}+\frac{1}{n_2^2}\right)\left|n_1\ell_1 m_1\right\rangle \left|n_2\ell_2 m_2\right\rangle \tag{8.403}$$

This expression for the perturbation potential is too complicated to calculate. We will need to make an approximation. We make the reasonable assumption that

$$R >> r_2 \text{ and } R >> r_1 \tag{8.404}$$

We have two useful mathematical results that we can apply. In general, we can write

$$\frac{1}{\left|\vec{R}+\vec{a}\right|} = \frac{1}{\left[R^2+2\vec{R}\cdot\vec{a}+a^2\right]^{1/2}} = \frac{1}{R}\left[1+\frac{2\vec{R}\cdot\vec{a}}{R^2}+\frac{a^2}{R^2}\right]^{-1/2} \tag{8.405}$$

and for small x we have

$$\left[1+x\right]^{-1/2} \approx 1-\frac{1}{2}x+\frac{3}{8}x^2-\frac{5}{16}x^3+..... \tag{8.406}$$

Using

$$x = \frac{2\vec{R}\cdot\vec{a}}{R^2}+\frac{a^2}{R^2} \tag{8.407}$$

we get the general result

$$\begin{aligned}\frac{1}{\left|\vec{R}+\vec{a}\right|} &= \frac{1}{R}\left[1-\frac{1}{2}\left(\frac{2\vec{R}\cdot\vec{a}}{R^2}+\frac{a^2}{R^2}\right)+\frac{3}{8}\left(\frac{2\vec{R}\cdot\vec{a}}{R^2}+\frac{a^2}{R^2}\right)^2-.....\right] \\ &= \frac{1}{R}\left[1-\frac{\vec{a}\cdot\vec{R}}{R^2}-\frac{1}{2}\frac{a^2}{R^2}+\frac{3}{2}\frac{\left(\vec{a}\cdot\vec{R}\right)^2}{R^4}+.....\right]\end{aligned} \tag{8.408}$$

Therefore, we have

$$\frac{1}{\left|\vec{R}+\vec{r}_2\right|} = \frac{1}{R}-\frac{\vec{r}_2\cdot\vec{R}}{R^3}-\frac{1}{2}\frac{r_2^2}{R^3}+\frac{3}{2}\frac{\left(\vec{r}_2\cdot\vec{R}\right)^2}{R^5} \tag{8.409}$$

$$\frac{1}{\left|\vec{R}-\vec{r}_1\right|} = \frac{1}{R}+\frac{\vec{r}_1\cdot\vec{R}}{R^3}-\frac{1}{2}\frac{r_1^2}{R^3}+\frac{3}{2}\frac{\left(\vec{r}_1\cdot\vec{R}\right)^2}{R^5} \tag{8.410}$$

$$\begin{aligned}\frac{1}{\left|\vec{R}+\vec{r}_2-\vec{r}_1\right|} &= \frac{1}{R}-\frac{\left(\vec{r}_2-\vec{r}_1\right)\cdot\vec{R}}{R^2}-\frac{1}{2}\frac{\left(\vec{r}_2-\vec{r}_1\right)^2}{R^2}+\frac{3}{2}\frac{\left(\left(\vec{r}_2-\vec{r}_1\right)\cdot\vec{R}\right)^2}{R^4} \\ &= \frac{1}{R}-\frac{\vec{r}_2\cdot\vec{R}}{R^3}+\frac{\vec{r}_1\cdot\vec{R}}{R^3}-\frac{1}{2}\frac{r_1^2}{R^3}-\frac{1}{2}\frac{r_2^2}{R^3}+\frac{\vec{r}_1\cdot\vec{r}_2}{R^3} \\ &\quad +\frac{3}{2}\frac{\left(\vec{r}_2\cdot\vec{R}\right)^2}{R^5}+\frac{3}{2}\frac{\left(\vec{r}_1\cdot\vec{R}\right)^2}{R^5}-3\frac{\left(\left(\vec{r}_1\cdot\vec{R}\right)\left(\vec{r}_2\cdot\vec{R}\right)\right.}{R^5}\end{aligned} \tag{8.411}$$

Putting all this together we get

$$V = \frac{e^2}{R^3}\left[\vec{r}_1 \cdot \vec{r}_2 - 3\frac{\left((\vec{r}_1 \cdot \vec{R})(\vec{r}_2 \cdot \vec{R})\right)}{R^2}\right] \tag{8.412}$$

Physically, this says that for large separations, the interaction between the atoms is the same as that between two dipoles $e\vec{r}_1$ and $e\vec{r}_2$ separated by $\vec{R}$.

To simplify the algebra, we now choose the vector $\vec{R}$ to lie along the $z-$axis

$$\vec{R} = R\hat{z} \tag{8.413}$$

which gives

$$\begin{aligned} V &= \frac{e^2}{R^3}\left[(x_1x_2 + y_1y_2 + z_1z_2) - 3\frac{z_1z_2R^2}{R^2}\right] \\ &= \frac{e^2}{R^3}(x_1x_2 + y_1y_2 - 2z_1z_2) \end{aligned} \tag{8.414}$$

We now specialize to consider the case $n_1 = n_2 = 2$. When $n = 2$, there are 4 electron states for each atom

$$\begin{aligned} &\ell = 0, m = 0 \\ &\ell = 1, m = 1, 0, -1 \end{aligned}$$

Therefore, there are $16 = (4 \times 4)$ degenerate unperturbed zero-order states, each with energy

$$E_0 = -\frac{e^2}{8a_0} - \frac{e^2}{8a_0} = -\frac{e^2}{4a_0} \tag{8.415}$$

We use degenerate perturbation theory. To carry out degenerate perturbation theory, we must construct the 16×16 matrix representation of $\langle\hat{V}\rangle$ and diagonalize it to find the energies corrected to first-order.

The typical matrix element is (leaving off the n labels)

$$\begin{aligned} &\langle \ell_1 m_1 \ell_2 m_2 | \hat{V} | \ell_1 m_1 \ell_2 m_2 \rangle \\ &\quad = \frac{e^2}{R^3}\langle \ell_1 m_1 | \hat{x}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{x}_2 | \ell_2 m_2 \rangle \\ &\qquad + \frac{e^2}{R^3}\langle \ell_1 m_1 | \hat{y}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{y}_2 | \ell_2 m_2 \rangle \\ &\qquad - 2\frac{e^2}{R^3}\langle \ell_1 m_1 | \hat{z}_1 | \ell_1 m_1 \rangle \langle \ell_2 m_2 | \hat{z}_2 | \ell_2 m_2 \rangle \end{aligned} \tag{8.416}$$

We have

$$x = r\sin\theta\cos\phi = -r\sqrt{\frac{2\pi}{3}}\left(Y_{1,1} - Y_{1,-1}\right) \tag{8.417}$$

$$y = r\sin\theta\sin\phi = +ir\sqrt{\frac{2\pi}{3}}\left(Y_{1,1} + Y_{1,-1}\right) \tag{8.418}$$

$$z = r\cos\theta = r\sqrt{\frac{4\pi}{3}}Y_{10} \tag{8.419}$$

and

$$\begin{aligned}&\langle n\ell m|\, x\,|n\ell' m'\rangle \\ &\quad = -\sqrt{\frac{2\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r)R_{n\ell'}(r)dr\right]\left[\int d\Omega Y^*_{\ell m}(Y_{1,1} - Y_{1,-1})Y_{\ell' m'}\right]\end{aligned} \tag{8.420}$$

$$\begin{aligned}&\langle n\ell m|\, y\,|n\ell' m'\rangle \\ &\quad = i\sqrt{\frac{2\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r)R_{n\ell'}(r)dr\right]\left[\int d\Omega Y^*_{\ell m}(Y_{1,1} + Y_{1,-1})Y_{\ell' m'}\right]\end{aligned} \tag{8.421}$$

$$\begin{aligned}&\langle n\ell m|\, z\,|n\ell' m'\rangle \\ &\quad = \sqrt{\frac{4\pi}{3}}\left[\int_0^\infty r^3 R_{n\ell}(r)R_{n\ell'}(r)dr\right]\left[\int d\Omega Y^*_{\ell m}Y_{10}Y_{\ell' m'}\right]\end{aligned} \tag{8.422}$$

Now let us return to the subject of *selection rules*.

We will just begin the discussion here and then elaborate and finish it later when we cover the topic of time-dependent perturbation theory.

Consider the integrals involving the spherical harmonics above. We have

$$\int d\Omega Y^*_{\ell m}Y_{1m''}Y_{\ell' m'} = 0 \text{unless} \begin{cases} \ell + \ell' + 1 = even \\ m = m' + m'' \end{cases} \tag{8.423}$$

These rules follow from doing the integrations over the θ and φ variables.

In particular, when the perturbation involves x, y, or z we have

$$\begin{aligned} \text{for } x \text{ and } y \qquad & m = m' \pm 1 \\ \text{for } z \qquad & m = m' \end{aligned}$$

which is the so-called

$$\Delta m = \pm 1, 0 \tag{8.424}$$

selection rule for this type of perturbation.

In addition, we have the

$$\Delta\ell = \pm 1 \tag{8.425}$$

selection rule for this type of perturbation.

These two rules will enable us to say many matrix elements are equal to zero by inspection.

We can derive two more very useful selection rules as follows. We know that

$$\left[\hat{L}_i, r_j\right] = i\hbar\varepsilon_{ijk} r_k \tag{8.426}$$

This allows us to write (after much algebra)

$$\begin{aligned}\left[\hat{L}_z, \hat{V}\right] &= \left[(\hat{L}_{1z} + \hat{L}_{2z}), \hat{V}\right] = \left[\hat{L}_{1z}, \hat{V}\right] + \left[\hat{L}_{2z}, \hat{V}\right] \\ &= \frac{e^2}{R^3}\left[\hat{L}_{1z}, (x_1x_2 + y_1y_2 - 2z_1z_2)\right] + \frac{e^2}{R^3}\left[\hat{L}_{2z}, (x_1x_2 + y_1y_2 - 2z_1z_2)\right] \\ &= 0\end{aligned}$$

This implies that $\left[\hat{L}_z, \hat{H}\right] = 0$ or that the $z-$component of the total angular momentum of the electrons is not changed by this perturbation (it is conserved).

This gives the selection rule

$$m_1 + m_2 = m_1' + m_2' \tag{8.427}$$

Summarizing the selection rules we have

$$\begin{aligned}&\ell_1 + \ell_1' + 1 = even \qquad (\text{ = reason b for a zero}) \\ &\ell_2 + \ell_2' + 1 = even \qquad (\text{ = reason c for a zero}) \\ &m_1 - m_1' = \pm 1, 0 \qquad (\text{ = reason d for a zero}) \\ &m_2 - m_2' = \pm 1, 0 \qquad (\text{ = reason d for a zero}) \\ &m_1 + m_2 = m_1' + m_2' \qquad (\text{ = reason a for a zero})\end{aligned}$$

and we have also given reason labels for each.

The unperturbed states are (using the format $|\ell_1 m_1\rangle\,|\ell_2 m_2\rangle$ are

$$\begin{aligned}&|1\rangle = |00\rangle\,|00\rangle\,,\ |2\rangle = |00\rangle\,|11\rangle\,,\ |3\rangle = |00\rangle\,|10\rangle\,,\ |4\rangle = |00\rangle\,|1,-1\rangle \\ &|5\rangle = |11\rangle\,|00\rangle\,,\ |6\rangle = |11\rangle\,|11\rangle\,,\ |7\rangle = |11\rangle\,|10\rangle\,,\ |8\rangle = |11\rangle\,|1,-1\rangle \\ &|9\rangle = |10\rangle\,|00\rangle\,,\ |10\rangle = |10\rangle\,|11\rangle\,,\ |11\rangle = |10\rangle\,|10\rangle\,,\ |12\rangle = |10\rangle\,|1,-1\rangle \\ &|13\rangle = |1,-1\rangle\,|00\rangle\,,\ |14\rangle = |1,-1\rangle\,|11\rangle\,,\ |15\rangle = |1,-1\rangle\,|10\rangle\,,\ |16\rangle = |1,-1\rangle\,|1,-1\rangle\end{aligned}$$

The $\langle\hat{V}\rangle$ matrix looks like (using labels (VALUE) or (0=reason)) and labeling the rows/columns in order as 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

0bc	0ba	0a	0a	0a	0a	0a	A	0a	0a	B	0a	0a	C	0a	0a
0ba	0bc	0a	0a	D	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0bc	0a	0a	0a	0a	E	0a	0a	0a	0a	0a	0a	0a
0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	F	0a	0a	0a
0a	D	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a	0a
A	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a	0a
0a	0a	E	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a	0a	0a
B	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0bc	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a	0a
0a	0a	0a	F	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0a
C	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a	0a	0bc	0a	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc	0a
0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0a	0bc

Table 8.13: $\langle \hat{V} \rangle$ matrix entries

There are only 12 nonzero elements(out of 256) and because the matrix is Hermitian we only have 6 to calculate (one side of the diagonal). It should now be clear why finding the relevant selection rules is so important!!!

The 10 nonzero elements are given by the expressions

$$A = C = \frac{e^2}{R^3} \langle 200| \langle 200| \left(x_1 x_2 + y_1 y_2\right) |211\rangle |21, -1\rangle \tag{8.428}$$

$$B = E = -2\frac{e^2}{R^3} \langle 200| \langle 200| z_1 z_2 |210\rangle |21, 0\rangle \tag{8.429}$$

$$D = \frac{e^2}{R^3} \langle 200| \langle 211| \left(x_1 x_2 + y_1 y_2\right) |211\rangle |200\rangle \tag{8.430}$$

$$F = \frac{e^2}{R^3} \langle 200| \langle 21, -1| \left(x_1 x_2 + y_1 y_2\right) |21, -1\rangle |200\rangle \tag{8.431}$$

If we define

$$\alpha = \sqrt{\frac{8\pi}{3}} \int_0^\infty r^3 R_{20} R_{21} dr \tag{8.432}$$

we have

$$\langle 200| x |211\rangle = \frac{\alpha}{2} = -\langle 200| x |21, -1\rangle \tag{8.433}$$

$$\langle 200| y |211\rangle = i\frac{\alpha}{2} = \langle 200| x |21, -1\rangle \tag{8.434}$$

$$\langle 200| z |211\rangle = \frac{\sqrt{2}}{2}\alpha \tag{8.435}$$

and

$$A = C = \frac{E}{2} = \frac{B}{2} = -D = -F = -\frac{1}{2}\alpha^2 \frac{e^2}{R^3} \tag{8.436}$$

Now we rearrange the row/column labels(the original choice was arbitrary) to create a *Jordan canonical form* with blocks on the diagonal. We choose

	1	8	11	14	2	5	3	9	4	13	6	7	10	12	15	16
0	A	2A	A	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	-A	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	-A	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2A	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	2A	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	-A	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	-A	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 8.14: Jordan Canonical Form

The is called the *block-diagonalized form*. We have one 4×4 and three 2×2 matrices to diagonalize. We get the eigenvalues

$$\begin{aligned} &4 \times 4 \to 0, 0, \pm\sqrt{6}A \qquad , \qquad 2 \times 2 \to \pm A \\ &2 \times 2 \to \pm 2A \qquad , \qquad 2 \times 2 \to \pm A \end{aligned}$$

Therefore, the energies correct to first-order are

$$E = \begin{cases} E_0 + \sqrt{6}A & \text{degeneracy} = 1 \\ E_0 + 2A & \text{degeneracy} = 1 \\ E_0 + A & \text{degeneracy} = 2 \\ E_0 & \text{degeneracy} = 8 \\ E_0 - A & \text{degeneracy} = 2 \\ E_0 - 2A & \text{degeneracy} = 1 \\ E_0 - \sqrt{6}A & \text{degeneracy} = 1 \end{cases} \tag{8.437}$$

That was a *real* problem!

8.7 Variational Methods

All perturbation methods rely on our ability to make the separation $\hat{H} = \hat{H}_0 + \hat{V}$ where $\hat{H}_0$ is solvable exactly and $\hat{V}$ is a small correction. The Rayleigh-Ritz variational method is not subject to any such restrictions. This method is based on the following mathematical results.

We can always write

$$\hat{H} = \hat{I}\hat{H} = \sum_N |N\rangle \langle N| \hat{H} = \sum_N E_n |N\rangle \langle N| \tag{8.438}$$

where

$$\hat{H} |N\rangle = E_n |N\rangle \tag{8.439}$$

This is just the spectral decomposition of $\hat{H}$ in terms of its eigenvectors and eigenvalues. Now, if we choose some arbitrary state vector $|\psi\rangle$ (called a *trial* vector), then have

$$\begin{aligned} \langle\psi| \hat{H} |\psi\rangle &= \sum_N E_n \langle\psi \mid N\rangle \langle N \mid \psi\rangle \\ &\geq \sum_N E_0 \langle\psi \mid N\rangle \langle N \mid \psi\rangle = E_0 \sum_N \langle\psi \mid N\rangle \langle N \mid \psi\rangle \\ &\geq E_0 \langle\psi| \left(\sum_N |N\rangle \langle N| \right) |\psi\rangle = E_0 \langle\psi| \hat{I} |\psi\rangle = E_0 \langle\psi \mid \psi\rangle \end{aligned} \tag{8.440}$$

or

$$\frac{\langle\psi| \hat{H} |\psi\rangle}{\langle\psi \mid \psi\rangle} \geq E_0 \tag{8.441}$$

for *any choice* of the trial vector $|\psi\rangle$, where E_0 is the ground state energy (the lowest energy). Equality holds only if $|\psi\rangle$ is the true ground state vector. This result says that

$$\frac{\langle\psi| \hat{H} |\psi\rangle}{\langle\psi \mid \psi\rangle} \text{ is an upper bound for } E_0 \tag{8.442}$$

Procedure

1. Pick a trial vector $|\psi\rangle$ that contains unknown parameters $\{\alpha_k\}$
2. Calculate
$$\frac{\langle\psi| \hat{H} |\psi\rangle}{\langle\psi \mid \psi\rangle} = E_0 (\{\alpha_k\}) \tag{8.443}$$
3. Since $E_0 (\{\alpha_k\})$ is an upper bound, we then minimize it with respect to all of the parameters $\{\alpha_k\}$. This gives a least upper bound for that choice of the functional form for the trial vector.

4. We perform the minimization by setting

$$\frac{\partial E_0}{\partial \alpha_k} = 0 \text{ for all } k \tag{8.444}$$

5. The more complex the trial vector, i.e., the more parameters we incorporate allows us to better approximate the true functional form of the ground state vector and we will get closer and closer to the true energy value.

What about states other than the ground state? If the ground state has different symmetry properties than the first excited state, i.e.,

$$\begin{aligned} \text{ground state } &\rightarrow \ell = 0 \rightarrow \text{ contains } Y_{00} \\ 1^{st} \text{ excited state } &\rightarrow \ell = 1 \rightarrow \text{ contains } Y_{1m} \end{aligned}$$

then if we choose a trial vector with the symmetry of an $\ell = 0$ state we obtain an approximation to the ground state energy. If, however, we choose a trial vector with the symmetry of an $\ell = 1$ state, then we obtain an approximation to the first-excited state energy and so on.

In other words, the variational method always gives the least upper bound for the energy of the state with the *same symmetry* as the trial vector.

Example

Let us choose the harmonic oscillator Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2 \tag{8.445}$$

and a trial wave function

$$\psi(x,a) = \begin{cases} (a^2 - x^2)^2 & |x| < a \\ 0 & |x| \geq a \end{cases} \tag{8.446}$$

where a is an unknown parameter. The variational principle says that

$$\frac{\langle \psi(a) | \hat{H} | \psi(a) \rangle}{\langle \psi(a) \mid \psi(a) \rangle} = E_0(a) \geq E_0 = \text{ true ground state energy}$$

We get a *best* value for this choice of trial function by minimizing with respect to a using

$$\frac{dE_0(a)}{da} = 0 \tag{8.447}$$

Now we need to calculate the integrals. We have for the denominator (the normalization integral)

$$\begin{aligned}\langle\psi(a)\mid\psi(a)\rangle &= \int_{-a}^{a}\psi^2(x,a)dx = \int_{-a}^{a}(a^2-x^2)^4dx \\ &= 2\int_{0}^{a}(a^2-x^2)^4dx = \frac{336}{315}a^9\end{aligned} \tag{8.448}$$

and for the numerator

$$\begin{aligned}\langle\psi(a)|\hat{H}|\psi(a)\rangle &= -\frac{\hbar^2}{2m}2\int_0^a\psi(x,a)\frac{d^2\psi(x,a)}{dx^2}dx+\frac{1}{2}k2\int_0^a\psi^2(x,a)x^2dx \\ &= -\frac{\hbar^2}{2m}\left(-\frac{80}{21}a^7\right)+\frac{1}{2}k\left(\frac{336}{3465}a^{11}\right)\end{aligned} \tag{8.449}$$

Therefore,

$$E_0(a) = 1.786\frac{\hbar^2}{ma^2}+0.045ka^2 \tag{8.450}$$

The minimum condition gives

$$a^2 = 6.300\left(\frac{\hbar^2}{mk}\right)^{1/2} \tag{8.451}$$

which then says that

$$0.566\hbar\omega \geq E_0 \tag{8.452}$$

The true value is $0.500\hbar\omega$. This is an excellent result considering that the trial function does not look at all like the correct ground-state wave function (it is a Gaussian function). This points out clearly how powerful the variational technique can be for many problems.

8.8 2^{nd}-Order Degenerate Perturbation Theory

Suppose that the first order correction in perturbation theory is zero for some degenerate states so that the states remain degenerate. In this case, second-order degenerate perturbation theory must be applied. This is complex. We follow the derivation in Schiff (using our notation).

We assume that

$$\varepsilon_m = \varepsilon_k \ , \ V_{km} = 0 \text{ and } V_{kk} = V_{mm} \tag{8.453}$$

so that the degeneracy is not removed in first order.

We assume the equations (to second order)

$$\hat{H} = \hat{H}_0 + \hat{V} = \hat{H}_0 + g\hat{U} \tag{8.454}$$

$$|M\rangle = a_m |m\rangle + a_k |k\rangle + g \sum_{l\neq m,k} a_l^{(1)} |l\rangle + g^2 \sum_{l\neq m,k} a_l^{(2)} |l\rangle \tag{8.455}$$

$$|K\rangle = b_m |m\rangle + b_k |k\rangle + g \sum_{l\neq m,k} b_l^{(1)} |l\rangle + g^2 \sum_{l\neq m,k} b_l^{(2)} |l\rangle \tag{8.456}$$

$$|N\rangle = |n\rangle + g \sum_{l\neq m,k} a_{nl}^{(1)} |l\rangle + g^2 \sum_{l\neq m,k} a_{nl}^{(2)} |l\rangle \quad , \quad n \neq m,k \tag{8.457}$$

$$E_p = \varepsilon_p + gE_p^{(1)} + g^2 E_p^{(2)} \tag{8.458}$$

$$\hat{H} |M\rangle = (\hat{H}_0 + \hat{V}) |M\rangle = E_m |M\rangle = (\varepsilon_m + gE_m^{(1)} + g^2 E_m^{(2)}) |M\rangle \tag{8.459}$$

where the degenerate states are labeled by k, m and we assume the degenerate zero-order states are linear combinations of the two zero order degenerate states as shown.

A very important point is being made here.

If the system remains degenerate after first-order correction, then one must redevelop the equations for degenerate perturbation theory, using the correct zero order state vectors, i.e., linear combinations of the degenerate states. Even in the special case where the degenerate basis is uncoupled, i.e.,

$$\langle k| V |m\rangle = V_{km} = 0 \tag{8.460}$$

we *must not use* non-degenerate perturbation theory, *as one might do* if many textbooks are to be *believed*.

Remember that the primary object of degenerate perturbation theory is not only to eliminate the divergent terms, but to determine the appropriate linear combinations to use for zero-order state vectors. If we start with the wrong linear combinations, then we would have a *discontinuous change* of states in the limit of vanishing interactions, which says that the perturbation expansion is not valid.

We then have

$$\begin{aligned} &ga_m U |m\rangle + ga_k U |k\rangle + g \sum_{l\neq m,k} a_l^{(1)} \varepsilon_l |l\rangle \\ &\qquad + g^2 \sum_{l\neq m,k} a_l^{(2)} \varepsilon_l |l\rangle + g^2 \sum_{l\neq m,k} a_l^{(1)} U |l\rangle \\ &= \left(gE_m^{(1)} + g^2 E_m^{(2)}\right) (a_m |m\rangle + a_k |k\rangle) + g \sum_{l\neq m,k} a_l^{(1)} \varepsilon_m |l\rangle \\ &\qquad + g^2 \sum_{l\neq m,k} a_l^{(2)} \varepsilon_m |l\rangle + g^2 \sum_{l\neq m,k} a_l^{(1)} E_m^{(1)} |l\rangle \end{aligned}$$

Applying the linear functional $\langle m|$ we get

$$ga_m U_{mm} + g^2 \sum_{l\neq m,k} a_l^{(1)} U_{ml} = gE_m^{(1)} a_m + g^2 E_m^{(2)} a_m \tag{8.461}$$

Applying the linear functional $\langle k|$ we get

$$ga_k U_{kk} + g^2 \sum_{l\neq m,k} a_l^{(1)} U_{kl} = gE_m^{(1)} a_k + g^2 E_m^{(2)} a_k \tag{8.462}$$

Applying the linear functional $\langle n|$ we get

$$\begin{aligned} &ga_m U_{nm} + ga_k U_{nk} + g\varepsilon_n a_n^{(1)} + g^2 \varepsilon_n a_n^{(2)} + g^2 \sum_{l\neq m,k} a_l^{(1)} U_{nl} \\ &= g\varepsilon_m a_m^{(1)} + g^2 \varepsilon_m a_m^{(2)} + g^2 E_m^{(1)} a_m^{(2)} \end{aligned} \tag{8.463}$$

The first-order terms in (8.462) and (8.463) give the expected result

$$E_m^{(1)} = U_{mm} = U_{kk} \tag{8.464}$$

The second-order terms in (8.462) and (8.463) give (equation (8.465))

$$\sum_{l\neq m,k} a_l^{(1)} U_{ml} = E_m^{(2)} a_m \quad , \quad \sum_{l\neq m,k} a_l^{(1)} U_{kl} = E_m^{(2)} a_k \tag{8.465}$$

The first-order terms in (8.464) give an expression for $a_l^{(1)}$ when $n = l \neq m, k$ (8.467)

$$a_l^{(1)} (\varepsilon_m - \varepsilon_l) = a_m U_{lm} + a_k U_{lk} \tag{8.466}$$

Substituting (8.467) into (8.466) we get a pair of homogeneous algebraic equations for a_m and a_k.

These equations have a nonvanishing solution if and only if the determinant of the coefficients of a_m and a_k is zero or

$$\det \begin{bmatrix} \sum\limits_{l\neq m,k} \frac{U_{ml}U_{lm}}{\varepsilon_m - \varepsilon_l} - E_m^{(2)} & \sum\limits_{l\neq m,k} \frac{U_{ml}U_{lk}}{\varepsilon_m - \varepsilon_l} \\ \sum\limits_{l\neq m,k} \frac{U_{kl}U_{lm}}{\varepsilon_m - \varepsilon_l} & \sum\limits_{l\neq m,k} \frac{U_{kl}U_{lk}}{\varepsilon_m - \varepsilon_l} - E_m^{(2)} \end{bmatrix} = 0 \tag{8.467}$$

or

$$\det \begin{bmatrix} \sum\limits_{l\neq m,k} \frac{V_{ml}V_{lm}}{\varepsilon_m - \varepsilon_l} - g^2 E_m^{(2)} & \sum\limits_{l\neq m,k} \frac{V_{ml}V_{lk}}{\varepsilon_m - \varepsilon_l} \\ \sum\limits_{l\neq m,k} \frac{V_{kl}V_{lm}}{\varepsilon_m - \varepsilon_l} & \sum\limits_{l\neq m,k} \frac{V_{kl}V_{lk}}{\varepsilon_m - \varepsilon_l} - g^2 E_m^{(2)} \end{bmatrix} = 0 \tag{8.468}$$

The two roots of this equation are $g^2 E_m^{(2)}$ and $g^2 E_k^{(2)}$ and the two pairs of solutions of (8.466) are a_m,a_k and b_m,b_k. We thus obtain perturbed energy levels in which the degeneracy has been removed in second order and we also find the correct linear combinations of the unperturbed degenerate state vectors

$|m\rangle$ and $|k\rangle$.

Example

This is a tricky problem because the degeneracy between the first and second state is not removed in first order degenerate perturbation theory.

A system that has three unperturbed states can be represented by the perturbed Hamiltonian matrix

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{V} \\ &= \begin{pmatrix} E_1 & 0 & 0 \\ 0 & E_1 & 0 \\ 0 & 0 & E_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ a^* & b^* & 0 \end{pmatrix} = \begin{pmatrix} E_1 & 0 & a \\ 0 & E_1 & b \\ a^* & b^* & E_2 \end{pmatrix}\end{aligned} \tag{8.469}$$

where $E_2 > E_1$. The quantities a and b are to be regarded as perturbations that are of same order and are small compared to $E_2 - E_1$. The procedure is:

1. Diagonalize the matrix to find the exact eigenvalues.
2. Use second-order nondegenerate perturbation theory to calculate the perturbed energies. Is this procedure correct?
3. Use second-order degenerate perturbation theory to calculate the perturbed energies.
4. Compare the three results obtained.

Solution - We have

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{V} \\ &= \begin{pmatrix} E_1 & 0 & 0 \\ 0 & E_1 & 0 \\ 0 & 0 & E_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ a^* & b^* & 0 \end{pmatrix} = \begin{pmatrix} E_1 & 0 & a \\ 0 & E_1 & b \\ a^* & b^* & E_2 \end{pmatrix}\end{aligned}$$

with $E_2 > E_1$ and $E_2 - E_1 \gg a = b$.

1. For an exact solution we need to find the eigenvalues of

$$\begin{pmatrix} E_1 & 0 & a \\ 0 & E_1 & b \\ a^* & b^* & E_2 \end{pmatrix} \tag{8.470}$$

This leads to the characteristic equation

$$(E_1 - E)(E_1 - E)(E_2 - E) - (E_1 - E)\,|b|^2 - (E_1 - E)\,|a|^2 = 0 \tag{8.471}$$

This says that one of the eigenvalues is $E = E_1$ and the remaining quadratic equation is

$$E^2 - (E_1 + E_2)E + (E_1 E_2 - |b|^2 - |a|^2) = 0 \tag{8.472}$$

or the other two eigenvalues are

$$E = \frac{1}{2}\left((E_1 + E_2) \pm \sqrt{(E_1 + E_2)^2 - 4(E_1E_2 - |b|^2 - |a|^2)}\right) \tag{8.473}$$

The exact energy values are

$$E_1$$
$$\frac{1}{2}\left((E_1 + E_2) + \sqrt{(E_1 + E_2)^2 - 4(E_1E_2 - |b|^2 - |a|^2)}\right) \approx E_1 + \frac{|a|^2 + |b|^2}{E_1 - E_2}$$
$$E = \frac{1}{2}\left((E_1 + E_2) - \sqrt{(E_1 + E_2)^2 - 4(E_1E_2 - |b|^2 - |a|^2)}\right) \approx E_2 - \frac{|a|^2 + |b|^2}{E_1 - E_2}$$

2. Apply non-degenerate second-order perturbation theory. The unperturbed system has

$$\hat{H}_0 = \begin{pmatrix} E_1 & 0 & 0 \\ 0 & E_1 & 0 \\ 0 & 0 & E_2 \end{pmatrix} \tag{8.474}$$

Since this is diagonal we have

$$E_1^{(0)} = E_1 = E_2^{(0)} \quad , \quad E_3^{(0)} = E_2 \text{ (levels 1 and 2 are degenerate)}$$

and unperturbed eigenvectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{8.475}$$

The perturbation is (in the unperturbed basis)

$$\hat{V} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ a^* & b^* & 0 \end{pmatrix} \tag{8.476}$$

Since the diagonal matrix elements of the perturbation are zero we have

$$E_1^{(1)} = E_2^{(1)} = E_3^{(1)} = 0 \text{ or no first - order corrections} \tag{8.477}$$

Thus, levels 1 and 2 remain degenerate.

If we formally (and incorrectly) apply non-degenerate second-order perturbation theory to this system we get

$$E_n^{(2)} = \sum_{m \neq n} \frac{|V_{mn}|^2}{E_n^{(0)} - E_m^{(0)}} \tag{8.478}$$

Now $V_{12} = 0, V_{13} = a, V_{23} = b$ and so we get

$$E_1^{(2)} = \sum_{m\neq 1} \frac{|V_{m1}|^2}{E_1^{(0)} - E_m^{(0)}}$$
$$= \frac{0}{0} + \frac{|V_{13}|^2}{E_1^{(0)} - E_3^{(0)}} \stackrel{?}{=} \frac{|a|^2}{E_1 - E_2} \text{ incorrect because of } 0/0 \text{ term}$$

$$E_2^{(2)} = \sum_{m\neq 2} \frac{|V_{m2}|^2}{E_2^{(0)} - E_m^{(0)}}$$
$$= \frac{0}{0} + \frac{|V_{23}|^2}{E_2^{(0)} - E_3^{(0)}} \stackrel{?}{=} \frac{|b|^2}{E_1 - E_2} \text{ incorrect because of } 0/0 \text{ term}$$

$$E_3^{(2)} = \sum_{m\neq 3} \frac{|V_{m3}|^2}{E_3^{(0)} - E_m^{(0)}}$$
$$= \frac{|V_{13}|^2}{E_3^{(0)} - E_1^{(0)}} + \frac{|V_{23}|^2}{E_3^{(0)} - E_2^{(0)}} = \frac{|a|^2 + |b|^2}{E_2 - E_1} \text{ agrees with exact solution}$$

3. Now we apply second-order degenerate perturbation theory for the two degenerate levels. We have

$$\det \begin{vmatrix} \frac{|V_{13}|^2}{E_1^{(0)} - E_3^{(0)}} - E^{(2)} & \frac{V_{13}V_{32}}{E_1^{(0)} - E_3^{(0)}} \\ \frac{V_{23}V_{31}}{E_2^{(0)} - E_3^{(0)}} & \frac{|V_{23}|^2}{E_2^{(0)} - E_3^{(0)}} - E^{(2)} \end{vmatrix} \tag{8.479}$$
$$= \det \begin{vmatrix} \frac{|a|^2}{E_1 - E_2} - E^{(2)} & \frac{ab^*}{E_1 - E_2} \\ \frac{ba^*}{E_1 - E_2} & \frac{|b|^2}{E_1 - E_2} - E^{(2)} \end{vmatrix} = 0$$

$$(E^{(2)})^2 - \frac{|a|^2 + |b|^2}{E_1 - E_2} E^{(2)} + \frac{|a|^2 |b|^2}{(E_1 - E_2)^2} - \frac{|a|^2 |b|^2}{(E_1 - E_2)^2}$$
$$= E^{(2)} \left(E^{(2)} - \frac{|a|^2 + |b|^2}{E_1 - E_2} \right) = 0 \tag{8.480}$$

corresponding to

$$E^{(2)} = 0$$
$$E^{(2)} = \frac{|a|^2 + |b|^2}{E_1 - E_2}$$

so that to second-order we have

$$\begin{aligned} &E_1 \\ &E_1 + \frac{|a|^2 + |b|^2}{E_1 - E_2} \\ &E_2 - \frac{|a|^2 + |b|^2}{E_1 - E_2} \end{aligned}$$

which agrees with the exact result.

8.9 Problems

8.9.1 Box with a *Sagging Bottom*

Consider a particle in a 1–dimensional box with a *sagging bottom* given by

$$V(x) = \begin{cases} -V_0 sin(\pi x/L) & \text{for } 0 \le x \le L \\ \infty & \text{for } x < 0 \text{ and } x > L \end{cases}$$

(a) For small V_0 this potential can be considered as a small perturbation of an infinite box with a flat bottom, for which we have already solved the Schrodinger equation. What is the perturbation potential?

(b) Calculate the energy shift due to the sagging for the particle in the n^{th} stationary state to first order in the perturbation.

8.9.2 Perturbing the Infinite Square Well

Calculate the first order energy shift for the first three states of the infinite square well in one dimension due to the perturbation

$$V(x) = V_0 \frac{x}{a}$$

as shown in Figure 8.7 below.

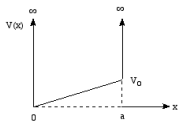


Figure 8.7: Ramp perturbation

8.9.3 Weird Perturbation of an Oscillator

A particle of mass m moves in one dimension subject to a harmonic oscillator potential $\frac{1}{2}m\omega^2x^2$. The particle is perturbed by an additional weak anharmonic force described by the potential $\Delta V = \lambda \sin kx$, $\lambda << 1$. Find the corrected ground state.

8.9.4 Perturbing the Infinite Square Well Again

A particle of mass m moves in a one dimensional potential box

$$V(x) = \begin{cases} \infty & \text{for } |x| > 3a \\ 0 & \text{for } a < x < 3a \\ 0 & \text{for } -3a < x < -a \\ V_0 & \text{for } |x| < a \end{cases}$$

as shown in Figure 8.8 below.

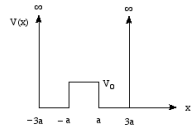


Figure 8.8: Square bump perturbation

Use first order perturbation theory to calculate the new energy of the ground state.

8.9.5 Perturbing the 2-dimensional Infinite Square Well

Consider a particle in a 2-dimensional infinite square well given by

$$V(x,y) = \begin{cases} 0 & \text{for } 0 \le x \le a \text{ and } 0 \le y \le a \\ \infty & \text{otherwise} \end{cases}$$

(a) What are the energy eigenvalues and eigenkets for the three lowest levels?

(b) We now add a perturbation given by

$$V_1(x,y) = \begin{cases} \lambda xy & \text{for } 0 \le x \le a \text{ and } 0 \le y \le a \\ 0 & \text{otherwise} \end{cases}$$

Determine the first order energy shifts for the three lowest levels for $\lambda \ll 1$.

(c) Draw an energy diagram with and without the perturbation for the three energy states, Make sure to specify which unperturbed state is connected to which perturbed state.

8.9.6 Not So Simple Pendulum

A mass m is attached by a massless rod of length L to a pivot P and swings in a vertical plane under the influence of gravity as shown in Figure 8.9 below.

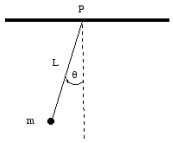


Figure 8.9: A quantum pendulum

(a) In the small angle approximation find the quantum mechanical energy levels of the system.

(b) Find the lowest order correction to the ground state energy resulting from the inaccuracy of the small angle approximation.

8.9.7 1-Dimensional Anharmonic Oscillator

Consider a particle of mass m in a 1–dimensional anharmonic oscillator potential with potential energy

$$V(x) = \frac{1}{2}m\omega^2 x^2 + \alpha x^3 + \beta x^4$$

(a) Calculate the 1^{st}–order correction to the energy of the n^{th} perturbed state. Write down the energy correct to 1^{st}–order.

(b) Evaluate all the required matrix elements of x^3 and x^4 needed to determine the wave function of the n^{th} state perturbed to 1^{st}–order.

8.9.8 A Relativistic Correction for Harmonic Oscillator

A particle of mass m moves in a 1–dimensional oscillator potential

$$V(x) = \frac{1}{2}m\omega^2 x^2$$

In the nonrelativistic limit, where the kinetic energy and the momentum are related by

$$T = \frac{p^2}{2m}$$

the ground state energy is well known to be $E_0 = \hbar\omega/2$.

Relativistically, the kinetic energy and the momentum are related by

$$T = E - mc^2 = \sqrt{m^2c^4 + p^2c^2} - mc^2$$

(a) Determine the lowest order correction to the kinetic energy (a p^4 term).

(b) Consider the correction to the kinetic energy as a perturbation and compute the relativistic correction to the ground state energy.

8.9.9 Degenerate perturbation theory on a spin = 1 system

Consider the spin Hamiltonian for a system of spin = 1

$$\hat{H} = A\hat{S}_z^2 + B(\hat{S}_x^2 - \hat{S}_y^2) \quad , \quad B << A$$

This corresponds to a spin = 1 ion located in a crystal with rhombic symmetry.

(a) Solve the unperturbed problem for $\hat{H}_0 = A\hat{S}_z^2$.

(b) Find the perturbed energy levels to first order.

(c) Solve the problem exactly by diagonalizing the Hamiltonian matrix in some basis. Compare to perturbation results.

8.9.10 Perturbation Theory in Two-Dimensional Hilbert Space

Consider a spin−1/2 particle in the presence of a static magnetic field along the z and x directions,

$$\vec{B} = B_z\hat{e}_z + B_x\hat{e}_x$$

(a) Show that the Hamiltonian is

$$\hat{H} = \hbar\omega_0\hat{\sigma}_z + \frac{\hbar\Omega}{2}\hat{\sigma}_x$$

where $\hbar\omega_0 = \mu_B B_z$ and $\hbar\Omega_0 = 2\mu_B B_x$.

(b) If $B_x = 0$, the eigenvectors are $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$ with eigenvalues $\pm\hbar\omega_0$respectively. Now turn on a weak x field with $B_x \ll B_z$. Use perturbation theory to find the new eigenvectors and eigenvalues to lowest order in B_x/B_z.

(c) If $B_z = 0$, the eigenvectors are $|\uparrow_x\rangle$ and $|\downarrow_x\rangle$ with eigenvalues $\pm\hbar\Omega_0$ respectively. Now turn on a weak z field with $B_z \ll B_x$. Use perturbation theory to find the new eigenvectors and eigenvalues to lowest order in B_z/B_x.

(d) This problem can actually be solved exactly. Find the eigenvectors and eigenvalues for arbitrary values of B_z and B_x. Show that these agree with your results in parts (b) and (c) by taking appropriate limits.

(e) Plot the energy eigenvalues as a function of B_z for fixed B_x. Label the eigenvectors on the curves when $B_z = 0$ and when $B_z \to \pm\infty$.

8.9.11 Finite Spatial Extent of the Nucleus

In most discussions of atoms, the nucleus is treated as a positively charged point particle. In fact, the nucleus does possess a finite size with a radius given approximately by the empirical formula

$$R \approx r_0 A^{1/3}$$

where $r_0 = 1.2 \times 10^{-13}\, cm$ (i.e., 1.2 Fermi) and A is the atomic weight or number (essentially the number of protons and neutrons in the nucleus). A reasonable assumption is to take the total nuclear charge $+Ze$ as being uniformly distributed over the entire nuclear volume (assumed to be a sphere).

(a) Derive the following expression for the electrostatic potential energy of an electron in the field of the *finite* nucleus:

$$V(r) = \begin{cases} -\frac{Ze^2}{r} & \text{for } r > R \\ \frac{Ze^2}{R}\left(\frac{r^2}{2R^2} - \frac{3}{2}\right) & \text{for } r < R \end{cases}$$

Draw a graph comparing this potential energy and the point nucleus potential energy.

(b) Since you know the solution of the point nucleus problem, choose this as the unperturbed Hamiltonian $\hat{H}_0$ and construct a perturbation Hamiltonian $\hat{H}_1$ such that the total Hamiltonian contains the $V(r)$ derived above. Write an expression for $\hat{H}_1$.

(c) Calculate(remember that $R \ll a_0 =$ Bohr radius) the $1^{st}-$order perturbed energy for the $1s$ $(n\ell m) = (100)$ state obtaining an expression in terms of Z and fundamental constants. How big is this result compared to the ground state energy of hydrogen? How does it compare to hyperfine splitting?

8.9.12 Spin-Oscillator Coupling

Consider a Hamiltonian describing a spin$-1/2$ particle in a harmonic well as given below:

$$\hat{H}_0 = \frac{\hbar\omega}{2}\hat{\sigma}_z + \hbar\omega\left(\hat{a}^+\hat{a} + 1/2\right))$$

(a) Show that

$$\{|n\rangle \otimes |\downarrow\rangle = |n,\downarrow\rangle \ , \ |n\rangle \otimes |\uparrow\rangle = |n,\uparrow\rangle\}$$

are energy eigenstates with eigenvalues $E_{n,\downarrow} = n\hbar\omega$ and $E_{n,\uparrow} = (n+1)\hbar\omega$, respectively.

(b) The states associated with the ground-state energy and the first excited energy level are

$$\{|0,\downarrow\rangle \ , \ |1,\downarrow\rangle \ , \ |0,\uparrow\rangle\}$$

What is(are) the ground state(s)? What is(are) the first excited state(s)? Note: two states are degenerate.

(c) Now consider adding an interaction between the harmonic motion and the spin, described by the Hamiltonian

$$\hat{H}_1 = \frac{\hbar\Omega}{2}\left(\hat{a}\hat{\sigma}_+ + \hat{a}^+\hat{\sigma}_-\right)$$

so that the total Hamiltonian is now $\hat{H} = \hat{H}_0 + \hat{H}_1$. Write a matrix representation of $\hat{H}$ in the subspace of the ground and first excited states in the ordered basis given in part (b).

(d) Find the first order correction to the ground state and excited state energy eigenvalues for the subspace above.

8.9.13 Motion in spin-dependent traps

Consider an electron moving in one dimension, in a spin-dependent trap as shown in Figure 8.10 below:

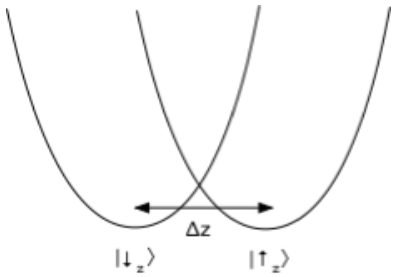


Figure 8.10: A spin-dependent trap

If the electron is in a spin-up state (with respect to the $z-$axis), it is trapped in the right harmonic oscillator well and if it is in a spin-down state (with respect to the $z-$axis), it is trapped in the left harmonic oscillator well. The Hamiltonian that governs its dynamics can be written as:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega_{osc}^2(\hat{z} - \Delta z/2)^2 \otimes |\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}m\omega_{osc}^2(\hat{z} + \Delta z/2)^2 \otimes |\downarrow_z\rangle\langle\downarrow_z|$$

(a) What are the energy levels and stationary states of the system? What are the degeneracies of these states? Sketch an energy level diagram for the first three levels and label the degeneracies.

(b) A small, constant *transverse field* B_x is now added with $|\mu_B B_x| << \hbar\omega_{osc}$. Qualitatively sketch how the energy plot in part (a) is modified.

(c) Now calculate the perturbed energy levels for this system.

(d) What are the new eigenstates in the ground-state doublet? For Δz macroscopic, these are sometimes called Schrodinger cat states. Explain why.

8.9.14 Perturbed Oscillator

A particle of mass m is moving in the 3–dimensional harmonic oscillator potential

$$V(x,y,z) = \frac{1}{2}m\omega^2(x^2+y^2+z^2)$$

A weak perturbation is applied in the form of the function

$$\Delta V(x,y,z) = kxyz + \frac{k^2}{\hbar\omega}x^2y^2z^2$$

where k is a small constant. Calculate the shift in the ground state energy to *second order* in k. This is *not the same* as second-order perturbation theory!

8.9.15 Another Perturbed Oscillator

Consider the system described by the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{m\omega^2}{2\alpha}(1-e^{-\alpha x^2})$$

Assume that $\alpha << m\omega/\hbar$

(1) Calculate an approximate value for the ground state energy using first-order perturbation theory by perturbing the harmonic oscillator Hamiltonian

$$H = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2$$

(2) Calculate an approximate value for the ground state energy using the variational method with a trial function $\psi = e^{-\beta x^2/2}$.

8.9.16 Helium from Hydrogen - 2 Methods

(a) Using a simple hydrogenic wave function for each electron, calculate by perturbation theory the energy in the ground state of the He atom associated with the electron-electron Coulomb interaction. Use this result to estimate the ionization energy of Helium.

(b) Calculate the ionization energy by using the variational method with an effective charge λ in the hydrogenic wave function as the variational parameter.

(c) Compare (a) and (b) with the experimental ionization energy

$$E_{ion} = 1.807E_0 \quad , \quad E_0 = \frac{\alpha^2 mc^2}{2} \quad , \quad \alpha = \text{ fine structure constant}$$

You will need

$$\psi_{1s}(r) = \sqrt{\frac{\lambda^3}{\pi}}\exp(-\lambda r) \quad , \quad a_0 = \frac{\hbar^2}{me^2} \quad , \quad \int\int d^3r_1 d^3r_2 \frac{e^{-\beta(r_1+r_2)}}{|\vec{r}_1-\vec{r}_2|} = \frac{20\pi^2}{\beta^5}$$

That last integral is very hard to evaluate from first principles.

8.9.17 Hydrogen atom + xy perturbation

An electron moves in a Coulomb field centered at the origin of coordinates. The first excited state ($n = 2$) is 4–fold degenerate. Consider what happens in the presence of a non-central perturbation

$$V_{pert} = f(r)xy$$

where $f(r)$ is some function only of r, which falls off rapidly as $r \to \infty$. To first order, this perturbation splits the 4–fold degenerate level into several distinct levels (some might still be degenerate).

(a) How many levels are there?

(b) What is the degeneracy of each?

(c) Given the energy shift, call it ΔE, for one of the levels, what are the values of the shifts for all the others?

8.9.18 Rigid rotator in a magnetic field

Suppose that the Hamiltonian of a rigid rotator in a magnetic field is of the form

$$\hat{H} = A\vec{L}^2 + B\hat{L}_z + C\hat{L}_y$$

Assuming that $A, B \gg C$, use perturbation theory to lowest nonvanishing order to get approximate energy eigenvalues.

8.9.19 Another rigid rotator in an electric field

Consider a rigid body with moment of inertia I, which is constrained to rotate in the xy–plane, and whose Hamiltonian is

$$\hat{H} = \frac{1}{2I}\hat{L}_z^2$$

Find the eigenfunctions and eigenvalues (zeroth order solution). Now assume the rotator has a fixed dipole moment $\vec{p}$ in the plane. An electric field $\vec{\mathcal{E}}$ is applied in the plane. Find the change in the energy levels to first and second order in the field.

8.9.20 A Perturbation with 2 Spins

Let $\vec{S}_1$ and $\vec{S}_2$ be the spin operators of two spin–1/2 particles. Then $\vec{S} = \vec{S}_1 + \vec{S}_2$ is the spin operator for this two-particle system.

(a) Consider the Hamiltonian

$$\hat{H}_0 = \alpha(\hat{S}_x^2 + \hat{S}_y^2 - \hat{S}_z^2)/\hbar^2$$

Determine its eigenvalues and eigenvectors.

(b) Consider the perturbation $\hat{H}_1 = \lambda(\hat{S}_{1x} - \hat{S}_{2x})$. Calculate the new energies in first-order perturbation theory.

8.9.21 Another Perturbation with 2 Spins

Consider a system with the unperturbed Hamiltonian $\hat{H}_0 = -A(\hat{S}_{1z} + \hat{S}_{2z})$ with a perturbing Hamiltonian of the form $\hat{H}_1 = B(\hat{S}_{1x}\hat{S}_{2x} - \hat{S}_{1y}\hat{S}_{2y})$.

(a) Calculate the eigenvalues and eigenvectors of $?\hat{H}_0$

(b) Calculate the exact eigenvalues of $\hat{H}_0 + \hat{H}_1$

(c) By means of perturbation theory, calculate the first- and the second-order shifts of the ground state energy of $\hat{H}_0$, as a consequence of the perturbation $\hat{H}_1$. Compare these results with those of (b).

8.9.22 Spherical cavity with electric and magnetic fields

Consider a spinless particle of mass m and charge e confined in spherical cavity of radius R, that is, the potential energy is zero for $r < R$ and infinite for $r > R$.

(a) What is the ground state energy of this system?

(b) Suppose that a weak uniform magnetic field of strength B is switched on. Calculate the shift in the ground state energy.

(c) Suppose that, instead a weak uniform electric field of strength $\mathcal{E}$ is switched on. Will the ground state energy increase or decrease? Write down, but do not attempt to evaluate, a formula for the shift in the ground state energy due to the electric field.

(d) If, instead, a very strong magnetic field of strength B is turned on, approximately what would be the ground state energy?

8.9.23 Hydrogen in electric and magnetic fields

Consider the $n = 2$ levels of a hydrogen-like atom. Neglect spins. Calculate to lowest order the energy splittings in the presence of both electric and magnetic fields $\vec{B} = B\hat{e}_z$ and $\vec{\mathcal{E}} = \mathcal{E}\hat{e}_x$.

8.9.24 $n = 3$ Stark effect in Hydrogen

Work out the Stark effect to lowest nonvanishing order for the $n = 3$ level of the hydrogen atom. Obtain the energy shifts and the zeroth order eigenkets.

8.9.25 Perturbation of the $n = 3$ level in Hydrogen - Spin-Orbit and Magnetic Field corrections

In this problem we want to calculate the 1st-order correction to the n=3 unperturbed energy of the hydrogen atom due to spin-orbit interaction and magnetic field interaction for arbitrary strength of the magnetic field. We have

$\hat{H} = \hat{H}_0 + \hat{H}_{so} + \hat{H}_m$ where

$$\hat{H}_0 = \frac{\vec{p}_{op}^2}{2m} + V(r) \quad , \quad V(r) = -e^2\left(\frac{1}{r}\right)$$
$$\hat{H}_{so} = \left[\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV(r)}{dr}\right]\vec{S}_{op}\cdot\vec{L}_{op}$$
$$\hat{H}_m = \frac{\mu_B}{\hbar}(\vec{L}_{op} + 2\vec{S}_{op})\cdot\vec{B}$$

We have two possible choices for basis functions, namely,

$$|n\ell s m_\ell m_s\rangle \quad or \quad |n\ell s j m_j\rangle$$

The former are easy to write down as direct-product states

$$|n\ell s m_\ell m_s\rangle = R_{n\ell}(r)Y_\ell^{m_\ell}(\theta,\varphi)\,|s,m_s\rangle$$

while the latter must be constructed from these direct-product states using addition of angular momentum methods. The perturbation matrix is not diagonal in either basis. The number of basis states is given by

$$\sum_{\ell=0}^{n-1=2}(2\ell+1)\times 2 = 10 + 6 + 2 = 18$$

All the 18 states are degenerate in zero-order. This means that we deal with an 18×18 matrix (mostly zeroes) in degenerate perturbation theory.

Using the direct-product states

(a) Calculate the nonzero matrix elements of the perturbation and arrange them in block-diagonal form.

(b) Diagonalize the blocks and determine the eigenvalues as functions of B.

(c) Look at the $B\to 0$ limit. Identify the spin-orbit levels. Characterize them by $(\ell s j)$.

(d) Look at the large B limit. Identify the Paschen-Bach levels.

(e) For small B show the Zeeman splittings and identify the Lande $g-$factors.

(f) Plot the eigenvalues versus B.

8.9.26 Stark Shift in Hydrogen with Fine Structure

Excluding nuclear spin, the $n = 2$ spectrum of Hydrogen has the configuration: where $\Delta E_{FS}/\hbar = 10\,GHz$ (the fine structure splitting) and $\Delta E_{Lamb}/\hbar = 1\,GHz$ (the Lamb shift - an effect of quantum fluctuations of the electromagnetic field). These shifts were neglected in the text discussion of the Stark effect. This is valid if $ea_0E_z >> \Delta E$. Let $x = ea_0E_z$.

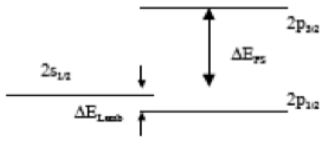


Figure 8.11: $n = 2$ Spectrum in Hydrogen

(a) Suppose $x < \Delta E_{Lamb}$, but $x << \Delta E_{FS}$. . Then we need only consider the $(2s_{1/2}, 2p_{1/2})$ subspace in a near degenerate case. Find the new energy eigenvectors and eigenvalues to first order. Are they degenerate? For what value of the field (in volts/cm) is the level separation doubled over the zero field Lamb shift? HINT: Use the representation of the fine structure eigenstates in the uncoupled representation.

(b) Now suppose $x > \Delta E_{FS}$. We must include all states in the near degenerate case. Calculate and plot numerically the eigenvalues as a function of x, in the range from $0\,GHz < x < 10\,GHz$.

Comment on the behavior of these curves. Do they have the expected asymptotic behavior? Find analytically the eigenvectors in the limit $x/\Delta E_{FS} \to \infty$. Show that these are the expected perturbed states.

8.9.27 2-Particle Ground State Energy

Estimate the ground state energy of a system of two interacting particles of mass m_1 and m_2 with the interaction energy

$$U(\vec{r}_1 - \vec{r}_2) = C\left(|\vec{r}_1 - \vec{r}_2|^4\right)$$

using the variational method.

8.9.28 1s2s Helium Energies

Use first-order perturbation theory to estimate the energy difference between the singlet and triple states of the (1s2s) configuration of helium. The 2s single particle state in helium is

$$\psi_{2s}(\vec{r}) = \frac{1}{\sqrt{4\pi}}\left(\frac{1}{a_0}\right)^{3/2}\left(2 - \frac{2r}{a_0}\right)e^{-r/a_0}$$

8.9.29 Hyperfine Interaction in the Hydrogen Atom

Consider the interaction

$$H_{hf} = \frac{\mu_B \mu_N}{a_B^3}\frac{\vec{S}_1 \cdot \vec{S}_2}{\hbar^2}$$

where μ_B , μ_N are the Bohr magneton and the nuclear magneton, a_B is the Bohr radius, and $\vec{S}_1$, $\vec{S}_2$ are the proton and electron spin operators.

(a) Show that H_{hf} splits the ground state into two levels:

$$E_t = -1\ Ry + \frac{A}{4} \quad , \quad E_s = -1\ Ry - \frac{3A}{4}$$

and that the corresponding states are triplets and singlets, respectively.

(b) Look up the constants, and obtain the frequency and wavelength of the radiation that is emitted or absorbed as the atom jumps between the states. The use of hyperfine splitting is a common way to detect hydrogen, particularly intergalactic hydrogen.

8.9.30 Dipole Matrix Elements

Complete with care; this is real physics. The charge dipole operator for the electron in a hydrogen atom is given by

$$\vec{d}(\vec{r}) = -e\vec{r}$$

Its expectation value in any state vanishes (you should be able to see why easily), but its matrix elements between different states are important for many applications (transition amplitudes especially).

(a) Calculate the matrix elements of each of the components between the 1s ground state and each of the 2p states(there are three of them). By making use of the Wigner-Eckart theorem (which you naturally do without thinking when doing the integral) the various quantities are reduced to a single *irreducible* matrix element and a very manageable set of Clebsch-Gordon coefficients.

(b) By using actual H-atom wavefunctions (normalized) obtain the magnitude of quantities as well as the angular dependence (which at certain points at least are encoded in terms of the (ℓ, m) indices).

(c) Reconstruct the vector matrix elements

$$\langle 1s|\,\vec{d}\,|2p_j\rangle$$

and discuss the angular dependence you find.

8.9.31 Variational Method 1

Let us consider the following very simple problem to see how good the variational method works.

(a) Consider the 1−dimensional harmonic oscillator. Use a Gaussian trial wave function $\psi_n(x) = e^{-\alpha x^2}$. Show that the variational approach gives the exact ground state energy.

(b) Suppose for the trial function, we took a Lorentzian

$$\psi_\alpha(x) = \frac{1}{x^2 + \alpha}$$

Using the variational method, by what percentage are you off from the exact ground state energy?

(c) Now consider the *double oscillator* with potential

$$V(x) = \frac{1}{2} m\omega^2 (|x| - a)^2$$

as shown below:

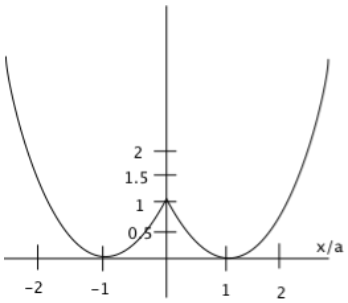


Figure 8.12: Double Oscillator Potential

Argue that a good choice of trial wave functions are:

$$\psi_n^\pm(x) = u_n(x - a) \pm u_n(x + a)$$

where the $u_n(x)$ are the eigenfunctions for a harmonic potential centered at the origin.

(d) Using this show that the variational estimates of the energies are

$$E_n^\pm = \frac{A_n \pm B_n}{1 \pm C_n}$$

where

$$A_n = \int u_n(x - a)\hat{H} u_n(x - a) dx$$
$$B_n = \int u_n(x - a)\hat{H} u_n(x + a) dx$$
$$C_n = \int u_n(x + a)\hat{H} u_n(x - a) dx$$

(e) For a much larger than the ground state width, show that

$$\Delta E_0 = E_0^{(-)} - E_0^{(+)} \approx 2\hbar\omega \sqrt{\frac{2V_0}{\pi\hbar\omega}} e^{-2V_0/\hbar\omega}$$

where $V_0 = m\omega^2 a^2/2$. This is known as the ground *tunneling* splitting. Explain why?

(f) This approximation clearly breaks down as $a \to 0$. Think about the limits and sketch the energy spectrum as a function of a.

8.9.32 Variational Method 2

For a particle in a box that extends from $-a$ to $+a$, try the trial function (within the box)

$$\psi(x) = (x-a)(x+a)$$

and calculate E. There is no parameter to vary, but you still get an upper bound. Compare it to the true energy. Convince yourself that the singularities in ψ'' at $x = \pm a$ do not contribute to the energy.

8.9.33 Variational Method 3

For the attractive delta function potential

$$V(x) = -aV_0\delta(x)$$

use a Gaussian trial function. Calculate the upper bound on E_0 and compare it to the exact answer $-ma^2V_0^2/2h^2$.

8.9.34 Variational Method 4

For an oscillator choose

$$\psi(x) = \begin{cases} (x-a)^2(x+a)^2 & |x| \le a \\ 0 & |x| > a \end{cases}$$

calculate $E(a)$, minimize it and compare to $\hbar\omega/2$.

8.9.35 Variation on a linear potential

Consider the energy levels of the potential $V(x) = g|x|$.

(a) By dimensional analysis, reason out the dependence of a general eigenvalue on the parameters $m = mass$, $\hbar$ and g.

(b) With the simple trial function

$$\psi(x) = c\theta(x+a)\theta(a-x)\left(1 - \frac{|x|}{a}\right)$$

compute (to the bitter end) a variational estimate of the ground state energy. Here both c and a are variational parameters.

(c) Why is the trial function $\psi(x) = c\theta(x+a)\theta(a-x)$ not a good one?

(d) Describe briefly (no equations) how you would go about finding a variational estimate of the energy of the first excited state.

8.9.36 Average Perturbation is Zero

Consider a Hamiltonian

$$H_0 = \frac{p^2}{2\mu} + V(r)$$

H_0 is perturbed by the spin-orbit interaction for a spin= 1/2 particle,

$$H' = \frac{A}{\hbar^2}\vec{S}\cdot\vec{L}$$

Show that the average perturbation of all states corresponding to a given term (which is characterized by a given L and S) is equal to zero.

8.9.37 3-dimensional oscillator and spin interaction

A spin= 1/2 particle of mass m moves in a spherical harmonic oscillator potential

$$U = \frac{1}{2}m\omega^2 r^2$$

and is subject to the interaction

$$V = \lambda\vec{\sigma}\cdot\vec{r}$$

Compute the shift of the ground state energy through second order.

8.9.38 Interacting with the Surface of Liquid Helium

An electron at a distance x from a liquid helium surface feels a potential

$$V(x) = \begin{cases} -K/x & x > 0 \\ \infty & x \le 0 \end{cases}$$

where K is a constant.

In Problem 8.7 we solved for the ground state energy and wave function of this system.

Assume that we now apply an electric field and compute the Stark effect shift in the ground state energy to first order in perturbation theory.

8.9.39 Positronium + Hyperfine Interaction

Positronium is a hydrogen atom but with a positron as the "nucleus" instead of a proton. In the nonrelativistic limit, the energy levels and wave functions are the same as for hydrogen, except for scaling due to the change in the reduced mass.

(a) From your knowledge of the hydrogen atom, write down the normalized wave function for the $1s$ ground state of positronium.

(b) Evaluate the root-mean-square radius for the $1s$ state in units of a_0. Is this an estimate of the physical diameter or radius of positronium?

(c) In the s states of positronium there is a contact *hyperfine* interaction

$$\hat{H}_{\rm int} = -\frac{8\pi}{3}\vec{\mu}_e \cdot \vec{\mu}_p \delta(\vec{r})$$

where $\vec{\mu}_e$ and $\vec{\mu}_p$ are the electron and positron magnetic moments and

$$\vec{\mu} = \frac{ge}{2mc}\hat{S}$$

Using first order perturbation theory compute the energy difference between the singlet and triplet ground states. Determine which lies lowest. Express the energy splitting in GHz. Get a number!

8.9.40 Two coupled spins

Two oppositely charged spin−1/2 particles (spins $\vec{s}_1$=$\hbar\vec{\sigma}_1/2$ and $\vec{s}_2$=$\hbar\vec{\sigma}_2/2$) are coupled in a system with a spin-spin interaction energy ΔE. The system is placed in a uniform magnetic field $\vec{B} = B\hat{z}$. The Hamiltonian for the spin interaction is

$$\hat{H} = \frac{\Delta E}{4}\vec{\sigma}_1 \cdot \vec{\sigma}_2 - (\vec{\mu}_1 + \vec{\mu}_2)\cdot\vec{B}$$

where $\vec{\mu}_j = g_j\mu_0\vec{s}_j/\hbar$ is the magnetic moment of the j^{th} particle.

(a) If we define the 2-particle basis-states in terms of the 1-particle states by

$$|1\rangle = |+\rangle_1|+\rangle_2 \quad , \quad |2\rangle = |+\rangle_1|-\rangle_2 \quad , \quad |3\rangle = |-\rangle_1|+\rangle_2 \quad , \quad |4\rangle = |-\rangle_1|-\rangle_2$$

where

$$\sigma_{ix}|\pm\rangle_i = |\mp\rangle_i \quad , \quad \sigma_{ix}|\pm\rangle_i = \pm i|\mp\rangle_i \quad , \quad \sigma_{iz}|\pm\rangle_i = \pm|\pm\rangle_i$$

and

$$\sigma_{1x}\sigma_{2x}|1\rangle = \sigma_{1x}\sigma_{2x}|+\rangle_1|+\rangle_2 = (\sigma_{1x}|+\rangle_1)(\sigma_{2x}|+\rangle_2) = |-\rangle_1|-\rangle_2 = |4\rangle$$

then derive the results below.

The energy eigenvectors for the 4 states of the system, in terms of the eigenvectors of the z−component of the operators $\vec{\sigma}_{\rm i} = 2\vec{s}_i/\hbar$ are

$$|1'\rangle = |+\rangle_1|+\rangle_2 = |1\rangle \quad , \quad |2'\rangle = d|-\rangle_1|+\rangle_2 + c|+\rangle_1|-\rangle_2 = d|3\rangle + c|2\rangle$$
$$|3'\rangle = c|-\rangle_1|+\rangle_2 - dc|+\rangle_1|-\rangle_2 = c|3\rangle - d|2\rangle \quad , \quad |4'\rangle = |-\rangle_1|-\rangle_2 = |4\rangle$$

where

$$\vec{\sigma}_{zi}\,|\pm\rangle_i = \pm\,|\pm\rangle_i$$

as stated above and

$$d = \frac{1}{\sqrt{2}}\left(1 - \frac{x}{\sqrt{1+x^2}}\right)^{1/2},\ c = \frac{1}{\sqrt{2}}\left(1 + \frac{x}{\sqrt{1+x^2}}\right)^{1/2},\ x = \frac{\mu_0 B(g_2 - g_1)}{\Delta E}$$

(b) Find the energy eigenvalues associated with the 4 states.

(c) Discuss the limiting cases

$$\frac{\mu_0 B}{\Delta E} \gg 1 \quad , \quad \frac{\mu_0 B}{\Delta E} \ll 1$$

Plot the energies as a function of the magnetic field.

8.9.41 Perturbed Linear Potential

A particle moving in one-dimension is bound by the potential

$$V(x) = \begin{cases} ax & x > 0 \\ \infty & x < 0 \end{cases}$$

where $a > 0$ is a constant. Estimate the ground state energy using first-order perturbation theory by the following method: Write $V = V_0 + V_1$ where $V_0(x) = bx^2$, $V_1(x) = ax - bx^2$ (for $x > 0$), where b is a constant and treat V_1 as a perturbation.

8.9.42 The ac-Stark Effect

Suppose an atom is perturbed by a monochromatic electric filed oscillating at frequency ω_L, $\vec{E}(t) = E_z \cos\omega_L t \hat{e}_z$ (such as from a linearly polarized laser), rather than the dc-field studied in the text. We know that such a field can be absorbed and cause transitions between the energy levels: we will systematically study this effect later. The laser will also cause a *shift* of energy levels of the unperturbed states, known alternatively as the *ac-Stark effect*, the *light shift*, and sometimes the *Lamp shift* (don't you love physics humor). In this problem, we will look at this phenomenon in the simplest case that the field is near to resonance between the ground state $|g\rangle$ and some excited state $|e\rangle$, $\omega_L \approx \omega_{eg} = (E_e - E_g)/\hbar$, so that we can ignore all other energy levels in the problem (the *two-level atom* approximation).

(i) **The classical picture**. Consider first the *Lorentz oscillator* model of the atom - a charge on a spring - with natural resonance at ω_0. The Hamiltonian for the system is

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 z^2 - \vec{d}\cdot\vec{E}(t)$$

where $d = -ez$ is the dipole.

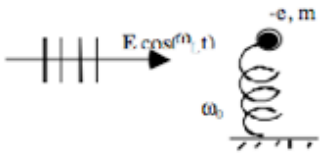


Figure 8.13: Lorentz Oscillator

(a) Ignoring damping of the oscillator, use Newton's Law to show that the induced dipole moment is

$$\vec{d}_{induced}(t) = \alpha \vec{E}(t) = \alpha E_z \cos\omega_L t$$

where

$$\alpha = \frac{e^2/m}{\omega_0^2 - \omega_L^2} \approx \frac{-e^2}{2m\omega_0\Delta}$$

is the polarizability with $\Delta = \omega_L - \omega_0$ the *detuning*.

(b) Use your solution to show that the total energy stored in the system is

$$H = -\frac{1}{2} d_{induced}(t) E(t) = -\frac{1}{2}\alpha E^2(t)$$

or the time average value of H is

$$\bar{H} = -\frac{1}{4}\alpha E_z^2$$

Note the factor of $1/2$ arises because energy is required to *create* the dipole.

(ii) **The quantum picture**. We consider the two-level atom described above. The Hamiltonian for this system can be written in a time independent form (equivalent to the time-averaging done in the classical case).

$$\hat{H} = \hat{H}_{atom} + \hat{H}_{int}$$

where $\hat{H}_{atom} = -\hbar\Delta \left|e\right\rangle \left\langle e\right|$ is the *unperturbed* atomic Hamiltonian and $\hat{H}_{int} = -\frac{\hbar\Omega}{2}\left(\left|e\right\rangle \left\langle g\right| + \left|g\right\rangle \left\langle e\right|\right)$ is the dipole-interaction with $\hbar\Omega = \left\langle e\right| \vec{d} \left|g\right\rangle \cdot \vec{E}$.

(a) Find the *exact* energy eigenvalues and eigenvectors for this simple two dimensional Hilbert space and plot the levels as a function of Δ. These are known as the atomic *dressed states*.

(b) Expand your solution in (a) to lowest nonvanishing order in Ω to find the perturbation to the energy levels. Under what conditions is this expansion valid?

(c) Confirm your answer to (b) using perturbation theory. Find also the mean induced dipole moment (to lowest order in perturbation theory), and from this show that the atomic polarizability, defined by $\langle \vec{d} \rangle = \alpha \vec{E}$ is given by

$$\alpha = -\frac{|\langle e|\vec{d}|g\rangle|^2}{\hbar\Delta}$$

so that the second order perturbation to the ground state is $E_g^{(2)} = -\alpha E_z^2$ as in part (b).

(d) Show that the ratio of the polarizability calculated classically in (b) and the quantum expression in (c) has the form

$$f = \frac{\alpha_{quantum}}{\alpha_{classical}} = \frac{|\langle e|z|g\rangle|^2}{(\Delta z^2)_{SHO}}$$

where $\left(\Delta z^2\right)_{SHO}$ is the SHO zero point variance. This is also known as the oscillator strength.

We see that in lowest order perturbation theory an atomic resonance looks just like a harmonic oscillator with a correction factor given by the oscillator strength and off-resonance harmonic perturbations cause energy level shifts as well as absorption and emission(Chapter 11).

8.9.43 Light-shift for multilevel atoms

We found the ac-Stark (light shift) for the case of a two-level atom driven by a monchromatic field. In this problem we want to look at this phenomenon in a more general context, including arbitrary polarization of the electric field and atoms with multiple sublevels.

Consider then a general monochromatic electric field $\vec{E}(\vec{x},t) = \Re(\vec{E}(\vec{x})e^{-i\omega_L t})$, driving an atom near resonance on the transition $|g;J_g\rangle \rightarrow |e;J_e\rangle$, where the ground and excited manifolds are each described by some total angular momentum J with degeneracy $2J+1$. The generalization of the ac-Stark shift is now the light-shift operator acting on the $2J_g+1$ dimensional ground manifold:

$$\hat{V}_{LS}(\vec{x}) = -\frac{1}{4}\vec{E}^*(\vec{x})\cdot\hat{\vec{\alpha}}\cdot\vec{E}(\vec{x})$$

Here,

$$\hat{\vec{\alpha}} = -\frac{\hat{\vec{d}}_{ge}\hat{\vec{d}}_{eg}}{\hbar\Delta}$$

is the atomic polarizability tensor operator, where $\hat{\vec{d}}_{eg} = \hat{P}_e\hat{\vec{d}}\hat{P}_g$ is the dipole operator, projected between the ground and excited manifolds; the projector onto the excited manifold is

$$\hat{P}_e = \sum_{M_e=-J_e}^{J_e} |e;J_e,M_e\rangle\langle e;J_e,M_e|$$

and similarly for the ground manifold.

(a) By expanding the dipole operator in the spherical basis$(\pm, 0)$, show that the polarizability operator can be written

$$\hat{\tilde{\alpha}} = \tilde{\alpha}\left(\begin{array}{c}\sum\limits_{q,M_g}\left|C_{M_g}^{M_g+q}\right|^2 \vec{e}_q\,|g,J_g,M_g\rangle\,\langle g,J_g,M_g|\,\vec{e}_q{}^* \\ +\sum\limits_{q\neq q',M_g} C_{M_g+q-q'}^{M_g+q}\,C_{M_g}^{M_g+q}\,\vec{e}_{q'}\,|g,J_g,M_g+q-q'\rangle\,\langle g,J_g,M_g|\,\vec{e}_q{}^*\end{array}\right)$$

where

$$\tilde{\alpha} = -\frac{|\langle e;J_e\,\|d\|\,g;J_g\rangle|^2}{\hbar\Delta}$$

and

$$C_{M_g}^{M_e} = \langle J_e M_e\,|\,1qJ_gM_g\rangle$$

Explain physically, using dipole selection rules, the meaning of the expression for $\hat{\tilde{\alpha}}$.

(b) Consider a polarized plane wave, with complex amplitude of the form $\vec{E}(\vec{x}) = E_1\vec{\varepsilon}_L e^{i\vec{k}\cdot\vec{x}}$ where E_1 is the amplitude and $\vec{\varepsilon}_L$ the polarization (possibly complex). For an atom driven on the transition $|g;J_g=1\rangle \to |e;J_e=2\rangle$ and the cases (i) linear polarization along z, (ii) positive helicity polarization, (iii) linear polarization along x, find the eigenvalues and eigenvectors of the light-shift operator. Express the eigenvalues in units of

$$V_1 = -\frac{1}{4}\tilde{\alpha}|E_1|^2.$$

Please comment on what you find for cases (i) and (iii). Repeat for $|g;J_g=1/2\rangle \to |e;J_e=3/2\rangle$ and comment.

(c) A deeper insight into the light-shift potential can be seen by expressing the polarizability operator in terms of irreducible tensors. Verify that the total light shift is the sum of scalar, vector, and rank-2 irreducible tensor interactions,

$$\hat{V}_{LS} = -\frac{1}{4}\left(|\vec{E}(\vec{x})|^2\hat{\alpha^{(0)}} + (\vec{E}^*(\vec{x})\times\vec{E}(\vec{x})\cdot\hat{\alpha^{(1)}} + \vec{E}^*(\vec{x})\cdot\hat{\alpha^{(2)}}\cdot\vec{E}(\vec{x})\right)$$

where

$$\hat{\alpha^{(0)}} = \frac{\hat{\vec{d}}_{ge}\cdot\hat{\vec{d}}_{eg}}{-3\hbar\Delta} \quad , \quad \hat{\alpha^{(1)}} = \frac{\hat{\vec{d}}_{ge}\times\hat{\vec{d}}_{eg}}{-2\hbar\Delta}$$

and

$$\hat{\alpha^{(2)}}_{ij} = \frac{1}{-\hbar\Delta}\left(\frac{\hat{\vec{d}}^i_{ge}\hat{\vec{d}}^j_{ge}+\hat{\vec{d}}^j_{ge}\hat{\vec{d}}^i_{ge}}{2} - \hat{\alpha^{(0)}}\delta_{ij}\right)$$

(d) For the particular case of $|g; J_g = 1/2\rangle \rightarrow |e; J_e = 3/2\rangle$, show that the rank-2 tensor part vanishes. Show that the light-shift operator can be written in a basis independent form of a scalar interaction (independent of sublevel), plus an effective Zeeman interaction for a fictitious B-field interacting with the spin−1/2 ground state,

$$\hat{V}_{LS} = V_0(\vec{x})\hat{I} + \vec{B}_{fict}(\vec{x}) \cdot \hat{\vec{\sigma}}$$

where

$$V_0(\vec{x}) = \frac{2}{3}U_1|\vec{\varepsilon}_L(\vec{x})|^2 \rightarrow \text{ proportional to field intensity}$$

and

$$\vec{B}_{fict}(\vec{x}) = \frac{1}{3}U_1\left(\frac{\vec{\varepsilon}_L{}^*(\vec{x}) \times \vec{\varepsilon}_L(\vec{x})}{i}\right) \rightarrow \text{ proportional to field ellipticity}$$

and we have written $\vec{E}(\vec{x}) = E_1\vec{\varepsilon}_L(\vec{x})$. Use this form to explain your results from part (b) on the transition $|g; J_g = 1/2\rangle \rightarrow |e; J_e = 3/2\rangle$.

8.9.44 A Variational Calculation

Consider the one-dimensional box potential given by

$$V(x) = \begin{cases} 0 & \text{for } |x| < a \\ \infty & \text{for } |x| > a \end{cases}$$

Use the variational principle with the trial function

$$\psi(x) = |a|^\lambda - |x|^\lambda$$

where λ is a variational parameter. to estimate the ground state energy. Compare the result with the exact answer.

8.9.45 Hyperfine Interaction Redux

An important effect in the study of atomic spectra is the so-called *hyperfine interaction* – the magnetic coupling between the electron spin and the nuclear spin. Consider Hydrogen. The hyperfine interaction Hamiltonian has the form

$$\hat{H}_{HF} = g_s g_i \mu_B \mu_N \frac{1}{r^3}\hat{s} \cdot \hat{i}$$

where $\hat{s}$ is the electron's spin−1/2 angular momentum and $\hat{i}$ is the proton's spin−1/2 angular momentum and the appropriate g-factors and magnetons are given.

(a) In the absence of the hyperfine interaction, but including the electron and proton spin in the description, what is the degeneracy of the ground state? Write all the quantum numbers associated with the degenerate sublevels.

(b) Now include the hyperfine interaction. Let $\hat{f} = \hat{i} + \hat{s}$ be the total spin angular momentum. Show that the ground state manifold is described with the good quantum numbers $|n = 1, \ell = 0, s = 1/2, i = 1/2, f, m_f\rangle$. What are the possible values of f and m_f?

(c) The perturbed $1s$ ground state now has hyperfine splitting. The energy level diagram is sketched below.



Figure 8.14: Hyperfine Splitting

Label all the quantum numbers for the four sublevels shown in the figure.

(d) Show that the energy level splitting is

$$\Delta E_{HF} = g_s g_i \mu_B \mu_N \langle \frac{1}{r^3} \rangle_{1s}$$

Show numerically that this splitting gives rise to the famous $21\,cm$ radio frequency radiation used in astrophysical observations.

8.9.46 Find a Variational Trial Function

We would like to find the ground-state wave function of a particle in the potential $V = 50(e^{-x} - 1)^2$ with $m = 1$ and $\hbar = 1$. In this case, the true ground state energy is known to be $E_0 = 39/8 = 4.875$. Plot the form of the potential. Note that the potential is more or less quadratic at the minimum, yet it is skewed. Find a variational wave function that comes within 5% of the true energy. OPTIONAL: How might you find the exact analytical solution?

8.9.47 Hydrogen Corrections on 2s and 2p Levels

Work out the first-order shifts in energies of $2s$ and $2p$ states of the hydrogen atom due to relativistic corrections, the spin-orbit interaction and the so-called Darwin term,

$$-\frac{p^4}{8m_e^3c^2} + g\frac{1}{4m_e^2c^2}\frac{1}{r}\frac{dV_c}{dr}(\vec{L}\cdot\vec{S}) + \frac{\hbar^2}{8m_e^2c^2}\nabla^2 V_c \;\;,\;\; V_c = -\frac{Ze^2}{r}$$

where you should be able to show that $\nabla^2 V_c = 4\pi\delta(\vec{r})$. At the end of the calculation, take $g = 2$ and evaluate the energy shifts numerically.

8.9.48 Hyperfine Interaction Again

Show that the interaction between two magnetic moments is given by the Hamiltonian

$$H = -\frac{2}{3}\mu_0(\vec{\mu}_1 \cdot \vec{\mu}_2)\delta(\vec{x} - \vec{y}) - \frac{\mu_0}{4\pi}\frac{1}{r^3}\left(3\frac{r_i r_j}{r^2} - \delta_{ij}\right)\mu_1^i \mu_2^j$$

where $r_i = x_i - y_i$. (NOTE: Einstein summation convention used above). Use first-order perturbation to calculate the splitting between $F = 0, 1$ levels of the hydrogen atoms and the corresponding wavelength of the photon emission. How does the splitting compare to the temperature of the cosmic microwave background?

8.9.49 A Perturbation Example

Suppose we have two spin−1/2 degrees of freedom, A and B. Let the initial Hamiltonian for this joint system be given by

$$H_0 = -\gamma B_z \left(S_z^A \otimes I^B + I^A \otimes S_z^B\right)$$

where I^A and I^B are identity operators, S_z^A is the observable for the $z-$component of the spin for the system A, and S_z^B is the observable for the $z-$component of the spin for the system B. Here the notation is meant to emphasize that both spins experience the same magnetic field $\vec{B} = B_z\hat{z}$ and have the same gyromagnetic ratio γ.

(a) Determine the energy eigenvalues and eigenstates for H_0

(b) Suppose we now add a perturbation term $H_{total} = H_0 + W$, where

$$W = \lambda \vec{S}^A \cdot \vec{S}^B = \lambda\left(S_x^A \otimes S_x^B + S_y^A \otimes S_y^B + S_z^A \otimes S_z^B\right)$$

Compute the first-order corrections to the energy eigenvalues.

8.9.50 More Perturbation Practice

Consider two spi−1/2 degrees of freedom, whose joint pure states can be represented by state vectors in the tensor-product Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ are each two-dimensional. Suppose that the initial Hamiltonian for the spins is

$$H_0 = \left(-\gamma_A B_z S_z^A\right) \otimes I^B + I^A \otimes \left(-\gamma_B B_z S_z^B\right)$$

(a) Compute the eigenstates and eigenenergies of H_0, assuming $\gamma_A \neq \gamma_B$ and that the gyromagnetic ratios are non-zero. If it is obvious to you what the eigenstates are, you can just guess them and compute the eigenenergies.

(b) Compute the first-order corrections to the eigenstates under the perturbation

$$W = \alpha S_x^A \otimes S_x^B$$

where α is a small parameter with appropriate units.