# QUANTUM MECHANICS

STEPHEN P. MARTIN
Physics Department
Northern Illinois University
DeKalb IL 60115
`spmartin@niu.edu`

January 19, 2024

# Contents

# 1 Introduction: the quantum revolution

## 1.1 Classical instability of charged matter

Quantum mechanics is often portrayed as nonintuitive, weird, or even paradoxical. Let us start with a retort: it is actually classical physics that is clearly incompatible with basic features of the world we live in. We begin with the observation that if classical mechanics governed the universe, then matter made from charged particles would necessarily be unstable against immediate and catastrophic collapse.

Consider, for example, a classical model of the hydrogen atom, consisting of a very heavy point-like[†] proton and electron, separated by a distance $r$. The electric potential energy is

$$V(r) = -\frac{\mathcal{Q}_e^2}{4\pi\epsilon_0\, r} = -\frac{e^2}{r}, \tag{1.1.1}$$

where in SI metric system units, $\mathcal{Q}_e = -1.60218 \times 10^{-19}\,\mathrm{C}$ is the electronic charge and $\epsilon_0 = 8.85419 \times 10^{-12}\,\mathrm{C}^2/\mathrm{N}\cdot\mathrm{m}^2$ is the permittivity of free space. The positive quantity $e$, which will appear often in this book, is equal to the proton's charge in the Gaussian cgs metric system units. Its square is given numerically by

$$e^2 = 1.43996 \times 10^{-9}\,\mathrm{eV{\cdot}m} = 2.30708 \times 10^{-28}\,\mathrm{J{\cdot}m} = 2.30708 \times 10^{-19}\,\mathrm{erg{\cdot}cm}. \tag{1.1.2}$$

Classical physics seems to have a serious problem: the potential $V(r)$ is unbounded from below as $r$ approaches 0, so that the classical atom should release an arbitrarily large amount of energy as it shrinks to zero size.

This seems dangerous (or possibly useful, depending on your imagination!), and it is certainly not what we observe. One might suppose that safety could still be achieved if the electron is somehow forced to travel in a fixed orbit about the much heavier proton. However, this cannot work in a classical theory, because of energy conservation. Maxwell's equations imply that the classical electron will continuously lose energy in the form of electromagnetic radiation, due to its centripetal acceleration. For a circular orbit, the acceleration is

$$a = \frac{v^2}{r} = \frac{F}{m_e} = \frac{e^2}{m_e r^2}, \tag{1.1.3}$$

---

[†] Protons and nuclei are not exactly point-like, being composed of quarks and gluons. However, the important thing for the following discussion is just that they are tiny ($\sim 10^{-15}$ meters) compared to atoms ($\sim$ few $\times 10^{-10}$ meters). This had become apparent by 1911, before the development of the quantum theory, from the results of Ernest Rutherford's experiments with Hans Geiger and Ernest Marsden in which alpha particles were observed scattering at large angles off of gold nuclei.

where $m_e = 9.109390 \times 10^{-31}$ kg $= 0.510999$ MeV$/c^2$ is the electron's mass, with $c = 2.99792458 \times 10^8$ m/sec, the speed of light in vacuum. The Larmor formula for the radiated power of an accelerating charge,

$$P = -\frac{dE}{dt} = \frac{\mathcal{Q}_e^2 a^2}{6\pi\epsilon_0 c^3} = \frac{2e^2 a^2}{3c^3}, \tag{1.1.4}$$

therefore tells us that the classical electron must lose energy at a rate proportional to $1/r^4$, which in turn will decrease its orbit size at an ever-increasing rate.

Just for fun, let us estimate the tragic fate of the electronic orbit in this classical model, making some simplifying assumptions. If the orbit stays nearly circular, and non-relativistic, as it decays, then the energy will be

$$E = \frac{1}{2}m_e v^2 - \frac{e^2}{r} = -\frac{e^2}{2r}. \tag{1.1.5}$$

Combining eqs. (1.1.4) and (1.1.5) gives

$$3r^2 \frac{dr}{dt} = -K, \tag{1.1.6}$$

where $K \equiv 4e^4/m_e^2 c^3$ is a constant. If $r = r_0$ at $t = 0$, this integrates to $r^3 - r_0^3 = -Kt$, so

$$r = r_0 \left(1 - Kt/r_0^3\right)^{1/3}. \tag{1.1.7}$$

This shows that the decay of the classical electron's orbit is even worse than asymptotic; it collapses all the way to $r = 0$ in a finite time $r_0^3/K$, which turns out to be very short (see homework exercise to find out just how short). Larger atoms, and crystal structures of electrons and nuclei, would have similar instabilities if classical physics governed them.

Quantum mechanics addresses this catastrophe, and allows matter composed of charged particle constituents to be stable, by changing the rules. In the quantum theory, there is no counterpart to the decaying classical orbit with unbounded negative energy. In particular, there are no physical quantum states of the hydrogen atom with arbitrarily low energy, corresponding to a classical electron localized arbitrarily near $r = 0$. Instead, as we will see, there is a single state with the lowest possible energy (about 13.6 eV below a state of ionization), which therefore is stable since it cannot decay by emitting electromagnetic radiation. In this way, quantum mechanics saves the universe. More generally, in quantum mechanics the energies of bound states turn out to be quantized (discrete). There are also unbound (ionized) states with a continuum of allowed energies, but those energies are bounded from below.

## 1.2 Black-body radiation and Planck's constant

In the late 19th century, it became apparent that classical physics theory makes a similarly catastrophic prediction for black-body radiation, which became the subject of detailed experiments. Consider a cavity whose walls are as close as possible to idealized black (a perfect absorber and emitter of electromagnetic radiation), heated as an oven so that it is kept in thermal equilibrium at temperature $T$ with the electromagnetic radiation inside it. The spectrum of the electromagnetic radiation is characterized by the energy density $\rho_E$, per unit volume $V$ and per unit frequency $\nu$, in terms of which the total energy inside the cavity is

$$\text{Total energy} \;=\; V\int_0^\infty d\nu\, \rho_E(\nu, T). \tag{1.2.1}$$

The function $\rho_E(\nu, T)$ can be determined experimentally by making a small hole in the walls of the cavity and measuring the radiation that escapes, analyzed for different frequencies using diffraction gratings, for example.

To obtain a theoretical prediction for $\rho_E(\nu, T)$, we first need to quantify the density of electromagnetic modes per unit frequency. One can assume for simplicity that the cavity is a cubic box of side $L$, and that the allowed electromagnetic radiation modes satisfy periodic boundary conditions. The allowed modes then have wavelengths $\lambda = L/|\vec{n}|$, where the components of $\vec{n}$ are a triplet of integers $(n_x, n_y, n_z)$. The frequencies are therefore

$$\nu \;=\; c|\vec{n}|/L. \tag{1.2.2}$$

Because the allowed $n_x$, $n_y$, and $n_z$ are integers, and each electromagnetic mode can have two transverse polarizations, there are two modes per unit volume in the space of vectors $\vec{n}$. Therefore, the number of allowed modes with frequency between $\nu$ and $\nu + d\nu$ can be evaluated by going to spherical coordinates in $\vec{n}$ space, using

$$d^3\vec{n} \;=\; 4\pi n^2 dn \;=\; \frac{4\pi L^3}{c^3}\nu^2 d\nu, \tag{1.2.3}$$

with $n = |\vec{n}|$. This implies that the total energy per unit frequency is

$$2\,\overline{E}\,\frac{4\pi L^3}{c^3}\nu^2 d\nu, \tag{1.2.4}$$

where the first factor of 2 accounts for the two polarizations, and $\overline{E}$ is the average energy of a mode with frequency $\nu$ when in thermal equilibrium with the walls at temperature $T$. Setting eq. (1.2.4) equal to $L^3\rho_E\, d\nu$ in accord with eq. (1.2.1), we have

$$\rho_E(\nu, T) \;=\; \overline{E}\,\frac{8\pi}{c^3}\nu^2. \tag{1.2.5}$$

We next need to evaluate $\overline{E}$ as a function of temperature.

A fundamental prediction of statistical mechanics is that for an ensemble of identical systems in thermal equilibrium, the probability for one of the systems to have energy $E$ is proportional to the **Boltzmann factor**, named after Ludwig Boltzmann,

$$P(E) \;\; \propto \;\; e^{-E/k_BT}, \tag{1.2.6}$$

where $T$ is the temperature in Kelvin, and Boltzmann's constant

$$k_B \;=\; 1.380649 \times 10^{-23}\,\text{J/K} \;=\; 1.380649 \times 10^{-16}\,\text{erg/K} \;=\; 8.617333 \times 10^{-5}\,\text{eV/K} \tag{1.2.7}$$

is the conversion factor between units of temperature and energy. If the allowed energies of electromagnetic modes are continuous and unrestricted, we therefore have the simple classical prediction for the average energy,

$$\overline{E} \;=\; \left(\int_0^\infty dE\, e^{-E/k_BT}\, E\right) \Big/ \int_0^\infty dE\, e^{-E/k_BT} \;=\; k_BT, \tag{1.2.8}$$

independent of $\nu$. Plugging this into eq. (1.2.5) gives the result

$$\rho_E(\nu, T) \;\;=\;\; 8\pi k_B T \nu^2/c^3. \tag{1.2.9}$$

This is the Rayleigh–Jeans prediction, developed by John W. Strutt, 3rd Baron Rayleigh, and James Jeans. Even without consulting experimental data, this formula is clearly problematic, since it claims that the energy density grows quadratically with frequency. After integrating $\int_0^\infty d\nu\, \rho_E(\nu, T)$, we would find an infinite total energy density per unit volume in black-body radiation. This impossible prediction of classical physics is called the ultraviolet catastrophe.

In order to explain the existing data and avoid the ultraviolet catastrophe, Max Planck proposed in 1900 that the energies of black-body modes with frequency $\nu$ actually can only occur in integer multiples of $h\nu$, where, using modern data,

$$h \;=\; 4.13567 \times 10^{-15}\,\text{eV·s} \;=\; 6.62607 \times 10^{-34}\,\text{J·s} \;=\; 6.62607 \times 10^{-27}\,\text{erg·s} \tag{1.2.10}$$

is known as **Planck's constant**. In that case, the integrals in eq. (1.2.8) are replaced by sums over only the discrete allowed energies,

$$\overline{E} \;=\; \Big(\sum_{n=0}^\infty nh\nu\, e^{-nh\nu/k_BT}\Big) \Big/ \sum_{n=0}^\infty e^{-nh\nu/k_BT} \;=\; \frac{h\nu}{e^{h\nu/k_BT} - 1}. \tag{1.2.11}$$

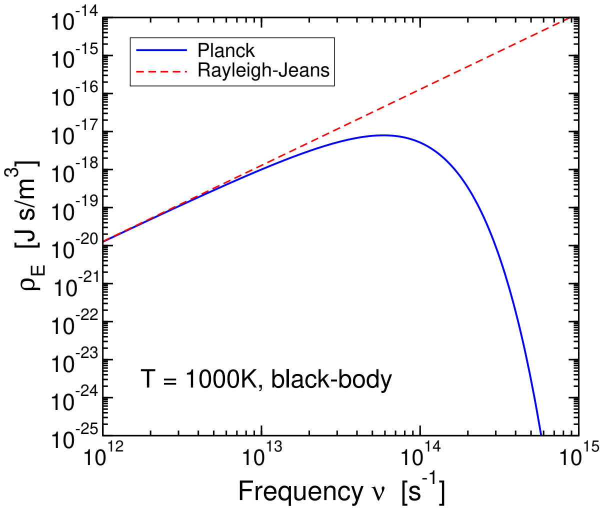

Figure 1.2.1: Comparison of the classical Rayleigh–Jeans (dashed line) and Planck (solid line) predictions for the electromagnetic radiation energy per unit volume and unit frequency, $\rho_E$, in a black-body cavity, as a function of the frequency $\nu$, for temperature $T = 1000$K.

For very small $h\nu/k_BT$, this agrees with eq. (1.2.8). However, unlike that formula, it depends on the frequency, and is exponentially suppressed in the ultraviolet limit of large $h\nu/k_BT$. Putting eq. (1.2.11) into eq. (1.2.5) yields[†]

$$\rho_E(\nu, T) \;=\; \frac{8\pi h}{c^3}\frac{\nu^3}{e^{h\nu/k_BT} - 1} \tag{1.2.12}$$

A comparison of Planck's formula eq. (1.2.12) with the Rayleigh–Jeans formula eq. (1.2.9) is shown in Figure 1.2.1 for black-body radiation with $T = 1000$K. In the far infrared, the two results agree, but they differ significantly in the near infrared, and the Planck prediction for $\rho_E$ decreases with frequency above $\nu = 5.9 \times 10^{13}$ Hz. In the visible range $4 \times 10^{14}$ to $8 \times 10^{14}$ Hz, the Planck formula is many orders of magnitude smaller, in agreement with observation. Furthermore, the Planck prediction for the total energy density per unit volume in the cavity,

$$\int_0^\infty d\nu\, \rho_E(\nu, T) = \frac{8\pi^5(k_BT)^4}{15h^3c^3}, \tag{1.2.13}$$

is finite, resolving the ultraviolet catastrophe. This established that the electromagnetic radiation modes inside the cavity are quantized in energy units equal to Planck's constant multiplied by the frequency.

## 1.3 Photo-electric effect and particle-like features of light

The photo-electric effect is the ejection of electrons from the surface of a material when light shines on it. This is observed to occur if the angular frequency $\omega$ of the light is sufficiently high. For $\omega$ less than a certain threshold value characteristic of the materials used in the experiment,

---

[†] Historically, Planck's derivation of this formula was different, and less compelling, but the result was correct. The origin of eq. (1.2.11) from deeper principles is derived in section 3.5, see eq. (3.5.45).

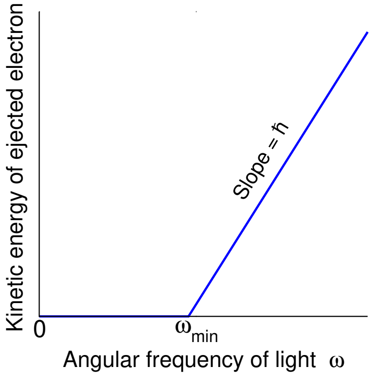

Figure 1.3.1: Sketch of results for the photoelectric effect. Shining electromagnetic radiation on the surface of a metal produces a current of ejected electrons, but to a good approximation this occurs only if the angular frequency of the light exceeds a minimum value $\omega_{\text{min}}$. The maximum kinetic energy of the ejected electrons then rises like $\hbar(\omega - \omega_{\text{min}})$.

essentially no electrons are ejected, even as the intensity of the light is increased. The threshold angular frequency typically corresponds to light in the visible and ultraviolet ranges. Above the threshold photon angular frequency, the maximum kinetic energy of ejected electrons is as sketched in Figure 1.3.1.

This threshold behavior was unexpected when it was first observed, because in classical electrodynamics, the frequencies and energies of electromagnetic waves are continuous and independent of each other. One might have supposed that sufficiently increasing the intensity of the light would result in electrons being ejected for any $\omega$, no matter how small. To explain the observations, Albert Einstein proposed in 1905 that light of a given angular frequency always occurs in chunks, or quanta, which are now called photons. For each photon, the energy is related to the angular frequency by the same formula as proposed by Planck, which can be alternatively written as

$$E = \hbar\omega, \tag{1.3.1}$$

where the **reduced Planck's constant** (or just "h bar", when speaking) is

$$\hbar \;=\; 6.58212 \times 10^{-16}\,\text{eV·s} \;=\; 1.05457 \times 10^{-34}\,\text{J·s} \;=\; 1.05457 \times 10^{-27}\,\text{erg·s} \tag{1.3.2}$$

and is defined to be related to the ordinary Planck's constant by

$$h = 2\pi\hbar. \tag{1.3.3}$$

Einstein's quantization condition conceptually generalized Planck's proposal, which was only intended to apply to electromagnetic modes absorbed and emitted by the black-body cavity.

It follows from special relativity and wave kinematics that the momentum $p$ and the wavelength $\lambda$ of each photon are related by

$$p = E/c = \hbar\omega/c = 2\pi\hbar/\lambda. \tag{1.3.4}$$

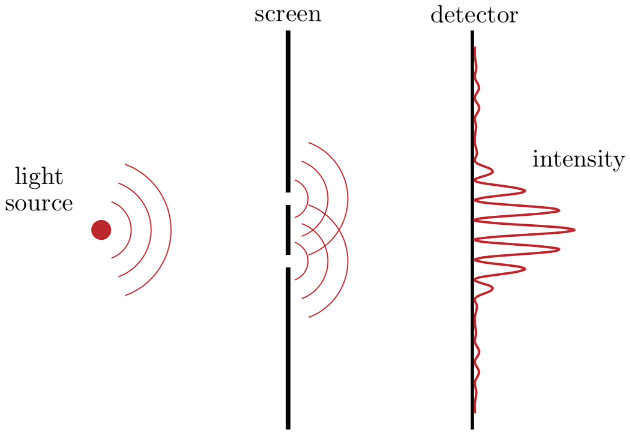

Figure 1.3.2: The wave nature of electromagnetic radiation causes interference and diffraction phenomena when light from a coherent source passes through holes or slits in a screen.

In the photo-electric effect, the discrete particle-like nature of light explains the existence of $\omega_{\rm min}$, because $\hbar\omega_{\rm min}$ is the minimum energy jump needed for the electron to escape the metal and be collected on another plate when it is struck by a single photon.[†] Experiments reported by Robert Millikan in 1916 later verified, using sodium and lithium as the targets, that the maximum kinetic energy of the ejected electron indeed behaves like

$$E = \hbar(\omega - \omega_{\rm min}). \tag{1.3.5}$$

The value of $\omega_{\rm min}$ depends on the collector plate material as well as the material being illuminated, but the most important result of the experiment is the slope of the line in Figure 1.3.1, which is $\hbar$. Millikan's measured value for $\hbar$ was consistent with Planck's result from black-body radiation, but was significantly more accurate and precise.

Although light does behave like a particle in the photo-electric effect, it still has interference properties governed by the wavelength. These effects appear in interference and diffraction experiments like the double-slit experiment illustrated in Figure 1.3.2. A coherent light source impacting on a screen with holes or slits yields a pattern of intensity maxima and minima, where the interference of amplitudes is constructive and destructive, respectively. However, the interpretation of this effect is slightly different in quantum mechanics than in the corresponding classical theory. Classically, the detected intensity is proportional to the magnitude of the time-averaged Poynting vector,

$$\text{Classical intensity} \quad \propto \quad |\vec{E} \times \vec{B}|, \tag{1.3.6}$$

where $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields that each obey superposition and interference due to the linearity of Maxwell's equations. In the quantum theory, the real quantity $\vec{E} \times \vec{B}$

---

[†] Actually, two photons can team up to eject the electron, as has been observed in experiments with high-power lasers, but the rate for this is very small except when the intensity is extremely large.

can be replaced by the square of a complex wavefunction $\psi$,

$$\text{Quantum intensity} \quad \propto \quad |\psi|^2. \tag{1.3.7}$$

Here $\psi(x, y, z, t)$ is the probability amplitude to find a photon at position $(x, y, z)$ at time $t$. The intensity is proportional to a density of probability to detect a photon at a given position and time, given by the squared magnitude of $\psi$.

## 1.4 Electron diffraction and wave-like features of matter

Inspired by the dual wave-particle nature of light, Louis de Broglie in his 1924 PhD thesis suggested that matter particles, such as electrons, will also behave like waves. He proposed that what is now called the **de Broglie wavelength** and wavenumber of the probability amplitude wavefunction are related to the momentum in the same way as for light, given by eq. (1.3.4), or

$$\lambda = 2\pi/k = 2\pi\hbar/p \tag{1.4.1}$$

for a particle with mass $m$ and energy $E = \sqrt{p^2c^2 + m^2c^4}$. De Broglie's wave hypothesis for matter was verified by the discovery of electron diffraction in a series of experiments by Clinton Davisson and Lester Germer from 1923-1928.

In the Davisson–Germer experiments, electrons with controlled energies (of order a few hundred eV) were made to hit a target made of a crystal of nickel,[†] and a detector was used to observe the intensity of electrons scattered at various fixed angles, as shown schematically in the top panel of Figure 1.4.1. The results as a function of varying incident electron momentum (proportional to the inverse de Broglie wavelength) are shown in the second panel. They feature maxima and minima of constructive and destructive interference, depending on the differences in path lengths of the electrons scattering from the regularly spaced crystal sites, with a distance between nuclei of order $3.5 \times 10^{-10}$ meters. This interference effect was similar to the previously known phenomenon of Bragg peaks in x-ray scattering. The electrons' wavelength corresponded to the prediction of the de Broglie relation to momentum in eq. (1.4.1). This provided the first direct evidence that matter particles are also described by a wavefunction which can be interpreted as a probability amplitude, with the crucial feature that this amplitude is subject to superposition and interference. It also provided another example of the central role of Planck's constant in the quantum theory, through the connection between momentum, wavelength, and $\hbar$ in eq. (1.4.1).

---

[†] In fact, their original motivation was to study the surface of nickel, not to verify de Broglie's idea.

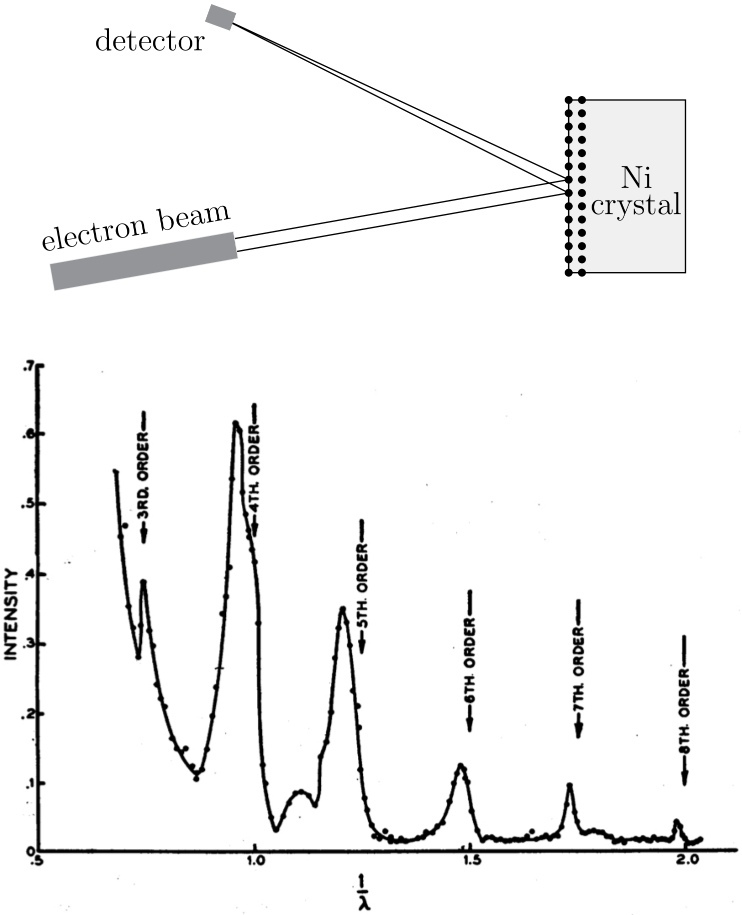

Figure 1.4.1: Schematic setup for the Davisson–Germer electron diffraction experiments (top), and some of their data for intensity of electrons scattered at fixed angle, as a function of the reciprocal of the electron's de Broglie wavelength (bottom). Source: C.J. Davisson and L.H. Germer, "Reflection and refraction of electrons by a crystal of nickel", Proceedings of the National Academy of Sciences of the United States of America, vol. 14, no. 8, p. 619 (1928).

## 1.5 Spin and the Stern–Gerlach experiment

In classical mechanics, angular momentum takes on continuous values, but in quantum mechanics it always occurs in multiples of a fundamental unit. For angular momentum associated with the motion of particles, called **orbital angular momentum**, the fundamental quantum unit is $\hbar$, as we will prove in section 5.3. However, quantum mechanics also allows for **intrinsic angular momentum**, or **spin**, which has no classical counterpart. The spin of a particle is quantized in units of $\hbar/2$, as we will prove in section 8.1, and has a fixed magnitude that depends only on the identity of the particle. Electrons, protons, and neutrons all carry this type of intrinsic angular momentum. Since their spin angular momentum is half of the quantized unit for orbital angular momentum, they are called spin-1/2 particles. Electrons, muons, tau leptons,

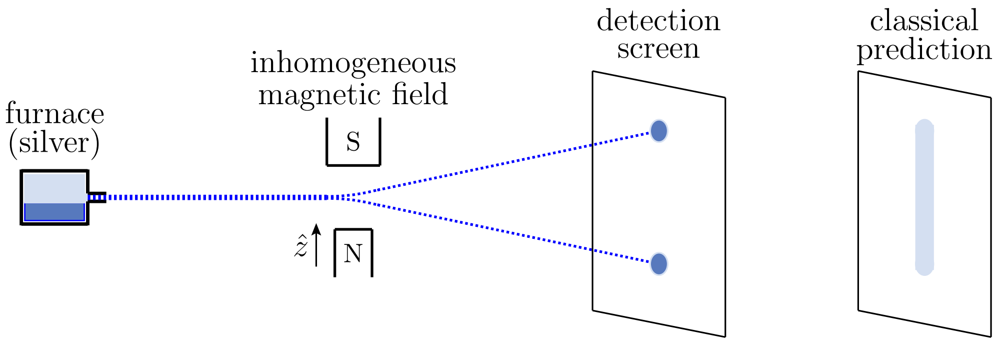

Figure 1.5.1: The Stern–Gerlach experiment. Silver atoms are heated in a furnace, and passed through an inhomogeneous magnetic field, which exerts a force on them proportional to the $\hat{z}$ component of the atomic magnetic dipole moment. The observed result is that the silver beam splits into two "spatially quantized" components on the detection screen. In contrast, the (incorrect) classical prediction is a continuum of deflection magnitudes.

neutrinos, and quarks are examples of particles that are currently believed to be fundamental (not composite collections of other particles) and have spin 1/2. But even composite particles like atoms, atomic nuclei, and mesons and baryons bound together by the strong nuclear force, always have spin that comes in integer multiples of $\hbar/2$.

The first experimental hints of the existence of spin came from the otherwise mysterious doubling of certain spectral lines from atomic transitions. Wolfgang Pauli suggested in 1924 that this was due to the presence of some extra quantum number that could only take on two values. The following year, Samuel Goudsmit and George Uhlenbeck proposed the intrinsic angular momentum interpretation that we now know as spin. This idea took some time to gain acceptance, and the doubters famously included Pauli. Part of the problem was that, at first, the spin was erroneously thought to be due to some kind of internal rotational motion of the structure of the electron, but nobody could get this idea to work in detail. Spin is now understood to be completely distinct from mechanical angular momentum.

In 1922, Otto Stern and Walter Gerlach reported an experiment which eventually provided profound insights into spin and the emerging quantum theory. They heated silver atoms in a furnace to vaporize them. The atoms escaped through a narrow collimating structure to form a beam, which then moved through a region where they were deflected by an inhomogeneous magnetic field, and finally were collected on a measurement screen, as shown in Figure 1.5.1.

The total magnetic dipole moment for the silver atom is a vector $\vec{\mu}$ whose magnitude $\mu$ is a constant, very nearly the same as that of a single electron,[†] which in turn is proportional to the

---

[†] The explanation for this is as follows. Silver atoms have 47 electrons, each of them carrying a magnetic dipole moment along that electron's spin direction. However, 46 of the electrons pair up in such a way that their spins are opposite, and cancel. Furthermore, there is no net orbital angular momentum of the electrons, and the contribution of the heavy nucleus to the atomic magnetic moment is relatively small.

electron's spin, although Stern and Gerlach did not know this at the time. The furnace thoroughly randomizes the directions of the magnetic moments. As a result of the inhomogeneous magnetic field $\vec{B}$, there is a classical force on the atoms,

$$\vec{F} \;=\; \vec{\nabla}(\vec{\mu} \cdot \vec{B}), \tag{1.5.1}$$

causing them to deflect. For simplicity, suppose the magnetic field in the deflection region has a cylindrically symmetric form[‡]

$$\vec{B}(x, y, z) \;=\; B_0 \left[\hat{x}\frac{x}{2a} + \hat{y}\frac{y}{2a} + \hat{z}\left(1 - \frac{z}{a}\right)\right] \tag{1.5.2}$$

where the coordinate system has origin at the center of the magnetic field region, with the unit vector $\hat{z}$ (up in Figure 1.5.1) perpendicular to the beam, and $a$ is a length that is large compared to the size of the region where the magnetic field acts on the atoms in the beam. The atomic dipole moment experiences a torque, obeying an equation of motion of the form

$$\frac{d\vec{\mu}}{dt} \;\propto\; \vec{\mu} \times \vec{B}. \tag{1.5.3}$$

The dominant component $B_0\hat{z}$ therefore causes $\vec{\mu}$ to rotate rapidly about the $\hat{z}$ direction, keeping the magnitude of $\mu_z$ nearly constant but causing the oscillating $\mu_x$ and $\mu_y$ components to average to 0 over the time scale in which the atom is moving through the macroscopic magnetic field region. This implies that $\mu_x$ and $\mu_y$ can be neglected when computing the deflection of the atom. Since only $\mu_z$ contributes, eq. (1.5.1) becomes

$$\vec{F} \;=\; \mu_z \vec{\nabla} B_z = \hat{z}\mu_z \frac{\partial B_z}{\partial z} = -\hat{z}(B_0/a)\mu_z. \tag{1.5.4}$$

Because the gradient of the vertical magnetic field $-B_0/a$ is known and fixed, measuring the deflection of a given atom is equivalent to measuring the value of the $\hat{z}$ component of its magnetic dipole moment.

Note that in the limit of a homogeneous field ($a \to \infty$) there would be no deflection at all. Thus, the role of the large homogeneous part of the magnetic field $\hat{z}B_0$ is to determine which component of $\vec{\mu}$ will be measured, by washing out the effects of the other components, while the smaller inhomogeneous part $-\hat{z}zB_0/a$ provides the force needed to produce the deflection and actually make the measurement.

Since the magnetic dipole moments of silver atoms emerging from the furnace are random in direction, classical physics reasoning suggests that the measured values of $\mu_z$ should have equal

---

[‡] The $\hat{x}$ and $\hat{y}$ components of $\vec{B}$ play no essential role here, except being necessary for consistency in order to satisfy the magnetostatic field equations $\vec{\nabla} \cdot \vec{B} = 0$ and $\vec{\nabla} \times \vec{B} = 0$. More generally, the field need not be cylindrically symmetric, but at least one of $B_x$ or $B_y$ must be non-zero. The $\vec{B}$ given here has the form that would result from a fixed external dipole magnet placed a distance $3a$ below the beam, to linear order in $1/a$.

likelihood to be anything between $-|\mu|$ and $|\mu|$. This in turn would imply that their deflections as observed on the screen should form a continuum between two extremes. Instead, they form two spots with equal numbers of atoms deposited, with a gap in between. The important and surprising conclusion is that the result of measuring $\mu_z$ for silver atoms can only give two discrete, quantized values, with equal probabilities.

The amount of deflection can be related to the spin carried by each atom, which in the case of silver comes mostly from a single unpaired electron. The proportionality between magnetic moment and spin for an electron is very close to

$$\vec{\mu} = \frac{e}{m_e c}\vec{S}. \tag{1.5.5}$$

The numerical magnitudes of the deflections observed in the Stern–Gerlach experiment then turn out to imply that the measurement of $S_z$ for an electron can only give the values

$$\text{measured}\, S_z \;\; = \;\; \pm\hbar/2. \tag{1.5.6}$$

There is nothing special about the $\hat{z}$ direction, so this also applies to $\hat{n} \cdot \vec{S}$ for any unit vector $\hat{n}$.

The Stern–Gerlach experiment has been conducted for other types of atoms and nuclei,[§] in which the angular momentum (and its relationship to the magnetic moment) can be different, resulting in more than two spots on the screen. This can be used to reveal the possible angular momentum properties ("quantum numbers") of the atom in question. A Stern–Gerlach apparatus can even be used to isolate samples with particular desired angular momentum quantum numbers. The results of such experiments are always consistent with quantization of the components of any angular momentum vector in integer multiples of $\hbar/2$. The quantum theory must account for this property, and we will see how in Chapter 8.

One can have more general Stern–Gerlach analyzers (called SG$\hat{n}$ in the following) consisting of the inhomogeneous magnetic field element oriented in any chosen unit vector $\hat{n}$ direction. The result of analyzing silver atoms fresh from the furnace with SG$\hat{n}$ is that half of them will be found to have $\hat{n}\cdot\vec{S} = +\hbar/2$ and the other half will have $-\hbar/2$, for any $\hat{n}$. Stern–Gerlach analyzers play a convenient dual role in further efforts to understand quantum mechanics. First, they are measuring devices, if the output beams are sent directly to a detection screen. Second, because the outgoing beams are separated (sometimes called "spatial quantization"), SG$\hat{n}$ provides a way of preparing a sample of atoms in which the spin component along $\hat{n}$ is known to be either $+\hbar/2$ or $-\hbar/2$. Instead of impacting a detection screen, one or both of the output beams can be

---

[§] However, the Stern–Gerlach setup does not work directly for free electrons, because the necessarily non-zero $B_x$ and/or $B_y$ cause a Lorentz force $-e\vec{v} \times \vec{B}$ on the electron. This produces a large deflection due to the electron's small mass, washing out the magnetic moment effect. The Stern–Gerlach setup relies on the deflected particle either being electrically neutral, or heavy, or both as in the case of silver.

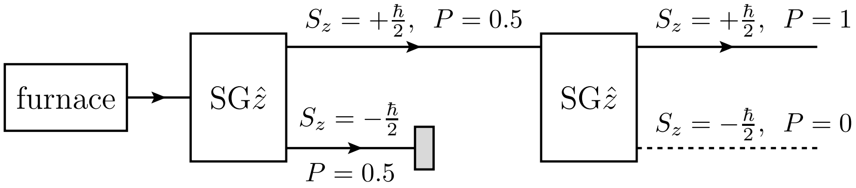

Figure 1.5.2: An experiment with two sequential Stern–Gerlach analyzers. The first analyzer prepares a sample of atoms with $S_z = +\hbar/2$, which are then fed into the second analyzer, which again measures $S_z = +\hbar/2$, with 100% probability in the idealized case.

sent off to some other component of the experiment, which could be another analyzer. Idealized versions of the Stern–Gerlach deflection analyzers are commonly used as modular components in experiments, real or imagined, to probe the implications of quantum theory.

For example, consider the sequence of two ideal Stern–Gerlach analyzers shown in Figure 1.5.2. The atoms in the experiment start in a furnace, which is assumed to produce completely randomized spins. After passing through a first analyzer SG$\hat{z}$, the atoms that had the result $S_z = +\hbar/2$ are sent into a second analyzer SG$\hat{z}$, while those that had $S_z = -\hbar/2$ are thrown away. In this case, the prediction of quantum mechanics for the output of the second analyzer is unlikely to surprise anyone. All of the output atoms on the far right again have $S_z = +\hbar/2$; the second analyzer simply confirms the measurement made by the first.

A more interesting setup is shown in Figure 1.5.3, which differs only by placing a SG$\hat{x}$ analyzer between the two SG$\hat{z}$ analyzers. As before, the experiment uses the first analyzer to select a pure sample of atoms with $S_z = +\hbar/2$. However, now the second analyzer separates the sample by measuring $S_x$. Since the $\hat{z}$ direction has no way of preferring one of $\pm\hat{x}$ over the other, it is no surprise that the output of SG$\hat{x}$ is 50% for each of $S_x = +\hbar/2$ and $-\hbar/2$.

The experiment in Figure 1.5.3 then throws away the atoms with $S_x = -\hbar/2$, and feeds those with $S_x = +\hbar/2$ into a third analyzer SG$\hat{z}$. The final results for $S_z$ can then be determined with a detection screen (not shown). One might perhaps suppose that we should find that the final atoms will all have $S_z = +\hbar/2$, since the first analyzer already selected only atoms with that property. However, this is wrong. In reality, an equal number are measured to have $S_z = +\hbar/2$ and $-\hbar/2$. Inserting the SG$\hat{x}$ analyzer in the middle of the chain affects the atoms in such a way as to restore the possibility of obtaining $S_z = -\hbar/2$.

Another way of thinking about the experiment shows that the restoration of the $S_z = -\hbar/2$ outcomes is a logical necessity, assuming only that the spin is the only thing that makes a difference. (Another way of stating this crucial assumption is that there are no "hidden variables" associated with the atom that are involved in the measurements in some mysterious

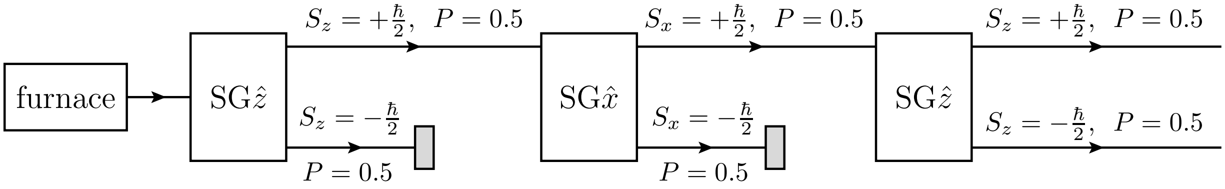

Figure 1.5.3: An experiment with three sequential Stern–Gerlach analyzers. The first analyzer prepares a sample of atoms with $S_z = +\hbar/2$, which are then fed into the second analyzer, which measures $S_x = +\hbar/2$ and $-\hbar/2$, each with probability 0.5. The atoms with $S_x = +\hbar/2$ are then sent to the third analyzer, which measures $S_z = +\hbar/2$ and $-\hbar/2$, each with probability 0.5. The act of measuring $S_x$ restores the possibility of measuring $S_z = -\hbar/2$ at the end, even though the first analyzer had been used to select only atoms with $S_z = +\hbar/2$.

way that we do not know how to take into account.) To see this, cover up everything in Figure 1.5.3 except the last analyzer. Feeding into it are atoms with spin known to be always aligned in the $+\hat{x}$ direction. Assuming no hidden variables containing other information, the $+\hat{x}$ direction has no reason to prefer $+\hat{z}$ over $-\hat{z}$, or vice versa, so no matter what may have occurred earlier, the only possibility is that the final probabilities for $S_z = +\hbar/2$ and $-\hbar/2$ are equal.

It is important that this result does *not* have anything to do with the fact that we threw away the atoms that were measured to have $S_x = -\hbar/2$ coming out of the second analyzer. We could feed those atoms into the final analyzer as well, and they would also be measured to have 50% probability for each of $S_z = +\hbar/2$ and $-\hbar/2$, by the same argument. Evidently, it is the mere act of measuring $S_x$ that causes the restoration of the $S_z = -\hbar/2$ outcomes. When we give the postulates of quantum mechanics in Chapter 3.1, the act of measurement will play a special role in one of them (Postulate 5), consistent with the preceding discussion. This also carries with it the implication that quantum mechanics, as defined by these postulates, is an inherently probabilistic, rather than deterministic, theory.

Readers familiar with the behavior of polarizing filters for light may recognize that the preceding example is closely analogous to a similar experiment that is common and easy to do in optics. Two linear polarization filters arranged with axes of polarization at right angles will not allow any light to pass through, but a third polarizer inserted between the two, with axis at a 45° angle with respect to the axes of each of the others, will restore the transmission of a fraction of the light. This analogy carries over into the quantum regime. In fact, the most sensitive experiments testing quantum mechanics are often done with the polarization of photons replacing the role of the spin in Stern–Gerlach type experiments. One famous example will be discussed in detail in section 21.4.

# 2 Math tools: states, operators, and representations

From the results of many experiments, physicists have arrived at a set of postulates that govern quantum theory. These will be given in the next chapter, but first we must introduce the mathematical language necessary to frame these ideas. This chapter therefore contains many definitions of key concepts, and some useful theorems. In doing so, we employ Paul A.M. Dirac's **bra-ket** notation, which is the modern standard in quantum mechanics.

## 2.1 Complex linear vector spaces

A **complex linear vector space** is a set of vectors, also known as **kets**, denoted for example as $|v\rangle$, $|w\rangle, \ldots$, such that the following properties hold:

- Additive closure: if $|v\rangle$ and $|w\rangle$ are kets, then so is their sum, or **superposition**, $|v\rangle + |w\rangle$.
- Multiplicative closure: if $|v\rangle$ is a ket, then so is the product $c|v\rangle$, where $c$ is any complex number.
- There exists a **null ket**, $|\text{Null}\rangle$, with the properties that $|v\rangle + |\text{Null}\rangle = |v\rangle$ for every ket $|v\rangle$, and $c\,|\text{Null}\rangle = |\text{Null}\rangle$ for every complex number $c$. Multiplying any ket by the complex number 0 results in the null ket: $0\,|v\rangle = |\text{Null}\rangle$.
- Addition of kets and multiplication by complex numbers satisfy the usual commutative, associative, and distributive properties. This means that we have $|v\rangle + |w\rangle = |w\rangle + |v\rangle$, and $(|v\rangle + |w\rangle) + |x\rangle = |v\rangle + (|w\rangle + |x\rangle)$, and $c_1(c_2\,|v\rangle) = (c_1 c_2)\,|v\rangle$, and $(c_1 + c_2)\,|v\rangle = c_1\,|v\rangle + c_2\,|v\rangle$, and $c(|v\rangle + |w\rangle) = c\,|v\rangle + c\,|w\rangle$.

In quantum mechanics, the physical state of a system is completely described by a non-null ket, known as the **state ket** or **state vector**, which we will often denote by $|\psi\rangle$. However, one of the rules is that the ket $c\,|\psi\rangle$ represents the same physical state as $|\psi\rangle$, provided that $c$ is a non-zero complex number. This is true even though $c|\psi\rangle$ and $|\psi\rangle$ are mathematically distinct members of the vector space; they are identified with each other physically.

The null ket, although it is part of the vector space, cannot describe any physical state. There is also a convenient and obvious notational shortcut: if we want to write down that some expression is equal to the null ket, we will just write "$= 0$" instead of "$= |\text{Null}\rangle$". Therefore, the null ket will not appear explicitly any more, but one should always consider the possibility that a ket arising in some calculation might actually be the null ket. A warning: it is common to use the notation $|0\rangle$ for some specific ket that represents a physical state and must not be confused with the null ket. For example, depending on the system under consideration, $|0\rangle$ might represent the lowest energy state of a system, also known as the **ground state**.

Let us list some examples of complex linear vector spaces, which you can check satisfy the properties in the definition.

Example 1: The set of complex numbers $z$.

Example 2: The set of all ordered triples $(z_1, z_2, z_3)$, where $z_1$, $z_2$, and $z_3$ are complex numbers.

Example 3: The set of complex $2 \times 3$ matrices $\begin{pmatrix} z_1 & z_2 & z_3 \\ z_4 & z_5 & z_6 \end{pmatrix}$.

Example 4: The set of all complex linear combinations $c_1 |{\uparrow}\rangle + c_2 |{\downarrow}\rangle$ of two basic kets $|{\uparrow}\rangle$ and $|{\downarrow}\rangle$. (This turns out to be the state space for a single spin-1/2 quantum system.)

Example 5: The set of all complex linear combinations of an infinite number of basic kets $|0\rangle$, $|1\rangle$, …, $|n\rangle$, …, in one-to-one correspondence with the non-negative integers. (This turns out to be a natural notation for the states of fixed energy for a harmonic oscillator in one dimension.)

Example 6: The set of all complex functions of a real variable, $f(x)$, defined on the domain $-\infty < x < \infty$. One can choose to add extra conditions on the functions in a variety of ways, for example requiring that they be continuous, differentiable, or vanishing at certain points or on specified intervals, or satisfy certain integrability conditions.

Intuitively, these vector spaces have different sizes. To make this precise, we define the notions of linear independence and dependence, dimension, and basis. A set of kets $|\psi_j\rangle$ are **linearly independent** (or just **independent**) if $\sum_j c_j |\psi_j\rangle = 0$ can only be satisfied by taking all $c_j = 0$. In other words, $|\psi_j\rangle$ are independent if we cannot write any of them as a complex linear combination of the others. Otherwise, the kets are **linearly dependent**.

A vector space is said to have **dimension** $N$ if one can choose a set of $N$ (but not more) linearly independent vectors. The $N$ linearly independent vectors are then said to form a **basis** for the vector space. The choice of basis is certainly not unique for $N > 1$; it is a common problem that one wants to change the choice of basis, either to make some calculation easier or to make some result simpler to interpret.

You can now check that for our six examples, the dimensions are:

Example 1 has dimension $N = 1$.
Example 2 has dimension $N = 3$.
Example 3 has dimension $N = 6$.
Example 4 has dimension $N = 2$.
Example 5 has dimension $N = \infty$ (countable, discrete basis).
Example 6 has dimension $N = \infty$ (uncountable, continuous basis).

However, the cardinality (countable vs. uncountable) of an infinite-dimensional vector space is mathematically a bit tricky, in ways that we shall find it convenient to mostly ignore. In physics, we are interested in kets that correspond to members of a vector space with the additional structure of a Hilbert space with an inner product, as discussed in the next section. The requirement that all physical state vectors must have a finite inner product implies that the basis is always countable. Despite this, we will see that it is a very convenient fiction to include continuous and uncountable sets of vectors that are not members of the physical state space but are nevertheless extremely useful, both as idealizations and in practical calculations. The most common examples are the kets that describe idealized states in which either a particle's position or its momentum (but not both!) is perfectly known. These continuous sets of vectors can be chosen to obey the other properties of a basis, in which case we will often simply call them basis vectors, as a slight abuse of terminology.

Given a specific basis set $\{|\beta_j\rangle\}$, any vector $|v\rangle$ can be expressed as

$$|v\rangle \;=\; \sum_{j=1}^{N} v_j\, |\beta_j\rangle \tag{2.1.1}$$

where the $N$ complex numbers $v_j$ are called the **components** of $|v\rangle$ in that basis. Using the preceding definitions, one can show that, for a given $|v\rangle$ and a given choice of basis $\{|\beta_j\rangle\}$, the components $v_j$ are unique. In writing eq. (2.1.1) as a sum, we have implicitly assumed that the basis vectors are discrete and countable. If the basis kets are instead continuous,[†] then the sum must be replaced by an integral. For example, if the basis set is $\{|\beta_q\rangle\}$ where $q$ is a continuous real variable with domain $a < q < b$, then we can write any ket $|v\rangle$ as

$$|v\rangle = \int_a^b dq\, v(q)\, |\beta_q\rangle\,. \tag{2.1.2}$$

Here $v(q)$ are the components, which in this case form a function of $q$. The symbol $q$ might represent a coordinate on ordinary space, but it could also be a momentum, or an energy, or some other continuous quantity of interest.

A **subspace** of a vector space is a subset of the elements that also form a vector space by themselves. Below, we will often work with subspaces that consist of states with some feature in common, such as a fixed energy or angular momentum.

## 2.2 Inner products, Hilbert spaces, and orthonormal bases

In quantum mechanics, the vector space of states has the additional structure of a **Hilbert space**, which implies that the complex linear vector space is endowed with an inner product. Given a ket $|v\rangle$ and a second ket $|w\rangle$, the **inner product** (also known as a **scalar product**) returns a complex number, denoted $\langle w|v\rangle$, which must satisfy the following rules:

---

[†] Note we are already engaging in the slight abuse of terminology mentioned in the previous paragraph!

- The order matters, in such a way that exchanging the two kets gives the complex conjugate,
$$\langle v|w\rangle \;=\; (\langle w|v\rangle)^* . \tag{2.2.1}$$
It immediately follows that $\langle v|v\rangle$ is real.

- If $|v\rangle$ is not the null ket, then $\langle v|v\rangle$ is positive.

- If either $|v\rangle$ or $|w\rangle$ is the null ket, then $\langle v|w\rangle = \langle w|v\rangle = 0$.

- Linearity is satisfied. The inner product of $c_1 |v\rangle + c_2 |w\rangle$ and $|x\rangle$, in that order, is
$$\langle x| \big(c_1 |v\rangle + c_2 |w\rangle\big) \;=\; c_1 \langle x|v\rangle + c_2 \langle x|w\rangle . \tag{2.2.2}$$

It follows from eqs. (2.2.1) and (2.2.2) that the inner product of the same two kets, but in the opposite order, is

$$\big(c_1^* \langle v| + c_2^* \langle w|\big) |x\rangle \;=\; c_1^* \langle v|x\rangle + c_2^* \langle w|x\rangle . \tag{2.2.3}$$

The inner product should be thought of as similar to the dot product in the familiar three-dimensional real vector space. However, because quantum mechanics uses complex linear vector spaces, the inner product treats the two input vectors asymmetrically, and interchanging them is the same as taking the complex conjugate.

The **norm** of a ket $|v\rangle$ is defined by $\sqrt{\langle v|v\rangle}$. It follows from the preceding that the norm of a non-null ket is real and positive,[†] and it is 0 if $|v\rangle$ is the null ket. It is often written as $||v|| \equiv \sqrt{\langle v|v\rangle}$, when that notation is unambiguous.

Two useful inequalities that govern the inner product follow.

**Theorem 2.2.1 (Schwarz inequality)** *For any two kets $|v\rangle$ and $|w\rangle$,*

$$|\langle v|w\rangle|^2 \;\leq\; \langle v|v\rangle \langle w|w\rangle , \tag{2.2.4}$$

*Also, equality holds if and only if $|w\rangle$ and $|v\rangle$ are proportional to each other or one of them is the null ket.*

Proof: If either $|v\rangle$ or $|w\rangle$ is the null ket, then eq. (2.2.4) is trivially satisfied with equality. Therefore, we can assume for the remainder of the proof that neither of them is null. Consider the ket $|z\rangle = c_1 |v\rangle - c_2 |w\rangle$, where $c_1$ and $c_2$ are complex numbers. Since $|z\rangle$ is a ket by the additive closure property, $\langle z|z\rangle$ must be non-negative, which gives

$$\langle z|z\rangle \;=\; |c_1|^2 \langle v|v\rangle + |c_2|^2 \langle w|w\rangle - c_1 c_2^* \langle w|v\rangle - c_1^* c_2 \langle v|w\rangle \;\geq\; 0. \tag{2.2.5}$$

---

[†] In quantum field theories with gauge invariance, it is sometimes useful, as a book-keeping trick, to modify the rules by allowing some kets to satisfy $\langle v|v\rangle \leq 0$. However, these negative-norm and null kets represent fictitious (unphysical) states, which must decouple from the true physical states. We will not encounter this issue in the present book.

Choosing $c_1 = \langle w|w\rangle$ and $c_2 = \langle w|v\rangle$, and using $\langle v|w\rangle^* = \langle w|v\rangle$, eq. (2.2.5) becomes

$$(\langle w|w\rangle)^2 \langle v|v\rangle - \langle w|w\rangle \, |\, \langle v|w\rangle\, |^2 \;\;\geq\;\; 0 \tag{2.2.6}$$

Now, since $|w\rangle$ is not null, we can divide by $\langle w|w\rangle$ to get eq. (2.2.4). If the equality condition holds, then it follows that $|z\rangle$ is the null ket, which implies that $|v\rangle$ and $|w\rangle$ are proportional.

**Theorem 2.2.2 (Triangle inequality)** *The norms of the kets $|v\rangle$ and $|w\rangle$ and their superposition $|v + w\rangle = |v\rangle + |w\rangle$ must obey*

$$\|v + w\| \;\;\leq\;\; \|v\| + \|w\|, \tag{2.2.7}$$

*and equality holds if and only if $|w\rangle$ and $|v\rangle$ are proportional to each other or one of them is the null ket.*

The proof can be obtained from the Schwarz inequality, and is left as an exercise. The triangle inequality is similar to the statement in ordinary plane geometry that the sum of the lengths of two sides of a triangle must exceed that of the third side.

Given a non-null ket $|v\rangle$, one can define a new ket by dividing it by its norm. The result

$$|v\rangle \,/ \sqrt{\langle v|v\rangle} \tag{2.2.8}$$

then has norm 1, and in quantum mechanics it is physically equivalent to the original ket, in the sense that it represents the same physical state. We say that the ket has been normalized to unity. This still leaves the freedom to multiply the ket by a complex phase, because

$$e^{i\theta}\, |v\rangle \tag{2.2.9}$$

has the same norm as $|v\rangle$, for any real number $\theta$. This freedom will appear very often as an ambiguity in the determination of a ket that has to satisfy some other specified properties. The resolution of such ambiguities is arbitrary, and equivalent to a choice of convention.

Although the phase of a single ket is not a physically measurable quantity, the relative phases between different kets can be physically meaningful if one considers their sum, also known as their superposition. This is because if we multiply two kets by different phases,

$$|v\rangle \to |v'\rangle = e^{i\theta_v}\, |v\rangle\,, \qquad\qquad |w\rangle \to |w'\rangle = e^{i\theta_w}\, |w\rangle\,, \tag{2.2.10}$$

then their sum is replaced by

$$|v\rangle + |w\rangle \;\;\to\;\; e^{i\theta_v}\left(|v\rangle + e^{i(\theta_w - \theta_v)}\, |w\rangle\right), \tag{2.2.11}$$

which, by removing the overall phase $e^{i\theta_v}$, is then physically equivalent to

$$|v\rangle + e^{i(\theta_w - \theta_v)}\,|w\rangle\,. \tag{2.2.12}$$

In general, this is not proportional to, and therefore not physically equivalent to, $|v\rangle + |w\rangle$, unless $|v\rangle$ and $|w\rangle$ are proportional, or $\theta_w$ and $\theta_v$ differ by an integer multiple of $2\pi$.

Two kets $|v\rangle$ and $|w\rangle$ are **orthogonal** (also known as **perpendicular**) if $\langle w|v\rangle = 0$. An **orthonormal basis**, which we will call an **orthobasis** for short, is a basis of kets $\{|\varphi_j\rangle\}$ in a Hilbert space that satisfies the additional property

$$\langle \varphi_j|\varphi_k\rangle \;=\; \delta_{j,k}. \tag{2.2.13}$$

Here $\delta_{j,k}$ is the **Kronecker delta symbol**,

$$\delta_{j,k} \;=\; \begin{cases} 1 & \text{(for } j = k\text{)}, \\ 0 & \text{(for } j \neq k\text{)}, \end{cases} \tag{2.2.14}$$

and is defined only when $j$ and $k$ are labels that take on discrete values.

As noted in the previous section, we will also often want to deal with basis kets labeled by one or more continuous parameters. In that case, we need a different orthonormality condition, called **Dirac orthonormality**, in which the Kronecker delta symbol is replaced by a **Dirac delta function**.[‡] The Dirac orthonormality condition for kets $|\varphi_q\rangle$ labeled by a single continuous real variable $q$ is thus

$$\langle \varphi_q|\varphi_{q'}\rangle \;=\; \delta(q - q'). \tag{2.2.15}$$

Here $q$ could be, for example, a position coordinate of a particle. The delta function $\delta(x)$ has the properties that

$$\delta(x) \;=\; \begin{cases} \infty & \text{(for } x = 0\text{)}, \\ 0 & \text{(for } x \neq 0\text{)}, \end{cases} \tag{2.2.16}$$

and

$$\int_{-\infty}^{\infty} dx\, \delta(x) = 1, \tag{2.2.17}$$

or, more generally, for sufficiently well-behaved functions $f(x)$,

$$\int_{-\infty}^{\infty} dx\, \delta(x - c) f(x) \;=\; f(c). \tag{2.2.18}$$

---

[‡] Technically, the Dirac delta function is not a function, but a distribution in mathematical language. But, this is not a math textbook, so we will not belabor the distinction too much.

For most purposes, eq. (2.2.18) can be taken as the practical definition of the delta function.

An important technical note: as we have already warned in the previous section, kets satisfying the Dirac orthonormality condition eq. (2.2.15) cannot be physical states, strictly speaking. This is because they do not have finite norm, due to $\delta(0) = \infty$. Such kets with continuous labels are still very useful in practical calculations, where they often represent idealizations with perfectly known position, or momentum, or some other continuous quantity. These are sometimes known as **generalized kets** or **non-normalizable kets**, and they act as a basis for the Hilbert space of physical states, even though they are not themselves part of the Hilbert space. They are so useful as idealizations that mere non-normalizability is not a good enough reason for us to eliminate them from the discussion.

It can also be useful to think of $\delta(x)$ as the "limit" (in a sense that we will not bother to try to make mathematically precise) of a sequence of increasingly narrow and sharply peaked functions that have unit area, in various different ways. For example, the sequence of functions can be taken to be just rectangular functions with width $\Delta$ and height $1/\Delta$,

$$\delta(x) = \lim_{\Delta\to 0} \begin{cases} 1/\Delta & \text{for } |x| < \Delta/2, \\ 0 & \text{for } |x| > \Delta/2. \end{cases} \tag{2.2.19}$$

Another useful representation of $\delta(x)$ is

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}. \tag{2.2.20}$$

To make better intuitive sense of this technically ill-defined expression, we can "cut off" the integration over $k$. One way to do this is to insert a convergence factor $e^{-k^2\Delta^2/2}$ into the integrand, and then take $\Delta \to 0$,

$$\delta(x) = \lim_{\Delta\to 0} \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-k^2\Delta^2/2} e^{ikx} = \lim_{\Delta\to 0} \frac{\exp[-x^2/2\Delta^2]}{\sqrt{2\pi}\Delta}. \tag{2.2.21}$$

This interprets $\delta(x)$ as the $\Delta \to 0$ limit of Gaussian functions with height $1/\sqrt{2\pi}\Delta$ and full width at half maximum (FWHM) approximately equal to $2.35\Delta$. Another useful way to make sense of eq. (2.2.20) is to limit the $k$ integration to a large but finite range $-1/\Delta < k < 1/\Delta$,

$$\delta(x) = \lim_{\Delta\to 0} \frac{1}{2\pi}\int_{-1/\Delta}^{1/\Delta} dk\, e^{ikx} = \lim_{\Delta\to 0} \frac{\sin(x/\Delta)}{\pi x}. \tag{2.2.22}$$

Each of the expressions in eqs. (2.2.19), (2.2.21), and (2.2.22) has the crucial properties of unit integrated area for all $\Delta$, and convergence to 0 at $x \neq 0$ in the limit $\Delta \to 0$. They are illustrated in Figure 2.2.1. Each of them, as well as the formal integral representation of eq. (2.2.20), can be useful, depending on the situation.

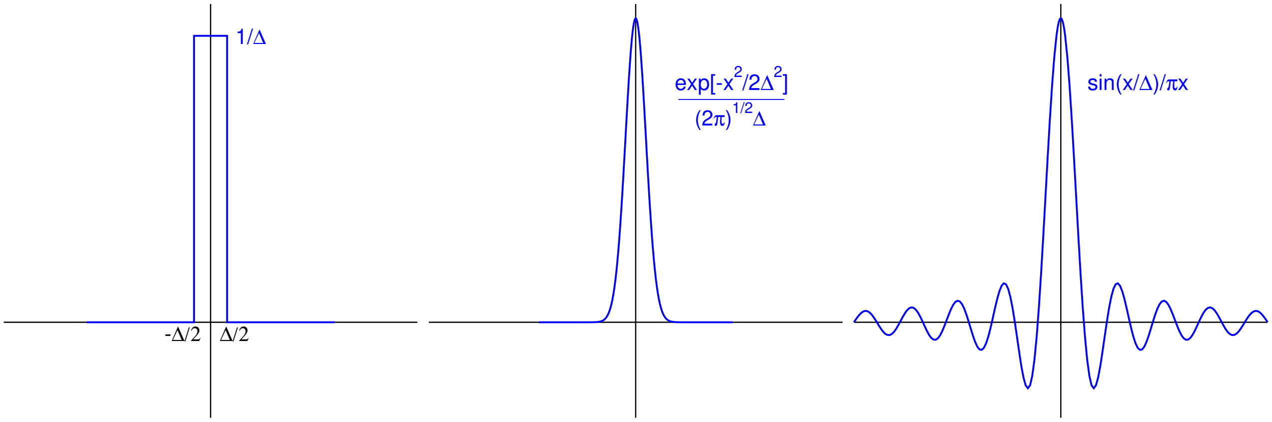

Figure 2.2.1: The Dirac delta function $\delta(x)$ can be thought of as the $\Delta \rightarrow 0$ "limit" of a sequence of increasingly narrow functions peaked near $x = 0$ with width of support proportional to $\Delta$ and unit integrated area, as in the rectangular functions in eq. (2.2.19) [left panel], the Gaussian functions in eq. (2.2.21) [center panel], or the normalized sine functions in eq. (2.2.22) [right panel].

The representation of the delta function in eq. (2.2.20) is related to the theory of Fourier transforms. To see how this works, we can use it to write, for any sufficiently well-behaved function $f(x)$,

$$f(x) \;=\; \int_{-\infty}^{\infty} dx'\, f(x')\, \delta(x - x') \;=\; \int_{-\infty}^{\infty} dx'\, f(x') \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ik(x-x')} \right). \tag{2.2.23}$$

By rearranging the order of integrations, this becomes

$$f(x) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ikx} \int_{-\infty}^{\infty} dx'\, e^{-ikx'} f(x'). \tag{2.2.24}$$

Now, we define the Fourier transform of $f(x)$ by the second integral,

$$F(k) \;\equiv\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx'\, e^{-ikx'} f(x'), \tag{2.2.25}$$

for $-\infty < k < \infty$. Then eq. (2.2.24) becomes

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, e^{ikx} F(k), \tag{2.2.26}$$

which is the inverse Fourier transform relation.

For a finite-dimensional[§] Hilbert space, there is a systematic way to find an orthobasis:

---

[§] In the infinite-dimensional case, the Gram–Schmidt algorithm cannot end in a finite number of steps, but in cases relevant to quantum mechanics a suitable orthobasis is often easy to identify anyway. For example, the Hilbert space may naturally split into an infinite number of mutually orthogonal finite-dimensional subspaces.

**Theorem 2.2.3** *Given a finite-dimensional Hilbert space with an inner product and a known basis* $\{|\beta_j\rangle\}$ *that is not necessarily orthonormal, one can always construct an orthobasis* $\{|\varphi_j\rangle\}$ *by a systematic procedure known as the* **Gram–Schmidt process***.*

We prove this constructively, by giving the steps of the Gram–Schmidt process:
Step 1. Choose any ket in the original basis; call it $|\beta_1\rangle$. The first member of the orthobasis will be the re-scaled version of this ket with unit norm, $|\varphi_1\rangle = |\beta_1\rangle / \sqrt{\langle\beta_1|\beta_1\rangle}$.
Step 2. Choose a second ket $|\beta_2\rangle$ from the original basis set. From it, define a new vector $|\tilde{\varphi}_2\rangle = |\beta_2\rangle - |\varphi_1\rangle \langle\varphi_1|\beta_2\rangle$, which we can describe in words as subtracting off the projection along the vector $|\varphi_1\rangle$. Thus $|\tilde{\varphi}_2\rangle$ is orthogonal to $|\varphi_1\rangle$ by construction. Now take $|\varphi_2\rangle = |\tilde{\varphi}_2\rangle / \sqrt{\langle\tilde{\varphi}_2|\tilde{\varphi}_2\rangle}$, which has norm 1. This is the second member of the orthobasis we are constructing.
Step 3. Choose a third ket $|\beta_3\rangle$ from the original basis set. From it, define a new vector $|\tilde{\varphi}_3\rangle = |\beta_3\rangle - |\varphi_1\rangle \langle\varphi_1|\beta_3\rangle - |\varphi_2\rangle \langle\varphi_2|\beta_3\rangle$. In words, we are subtracting off the projections along both of the vectors $|\varphi_1\rangle$ and $|\varphi_2\rangle$. This means that $|\tilde{\varphi}_3\rangle$ is orthogonal to both $|\varphi_1\rangle$ and $|\varphi_2\rangle$ by construction, so again we just need to re-scale it to have unit norm. Therefore, define $|\varphi_3\rangle = |\tilde{\varphi}_3\rangle / \sqrt{\langle\tilde{\varphi}_3|\tilde{\varphi}_3\rangle}$ as the third member of the orthobasis.

...

Step $n$. Take the $n$th ket $|\beta_n\rangle$ from the original basis set, and subtract off the projections along all of the previously found orthobasis kets, by defining

$$|\tilde{\varphi}_n\rangle \;=\; |\beta_n\rangle - \sum_{j=1}^{n-1} |\varphi_j\rangle \langle\varphi_j|\beta_n\rangle . \tag{2.2.27}$$

By construction, this is orthogonal to all of the previously found kets $|\varphi_j\rangle$. It is not null, because of the linear independence of the original basis. So, to include it in the orthobasis, we only need to re-scale it to also have unit norm,

$$|\varphi_n\rangle \;=\; |\tilde{\varphi}_n\rangle / \sqrt{\langle\tilde{\varphi}_n|\tilde{\varphi}_n\rangle}. \tag{2.2.28}$$

Continuing in this way, when we are finished with Step $N$, where $N$ is the dimension of the Hilbert space, we will have constructed the full orthobasis $\{|\varphi_n\rangle\}$. This concludes the proof.

Suppose that we have chosen an orthobasis $\{|\varphi_j\rangle\}$. To obtain the components $v_j$ of an arbitrary ket

$$|v\rangle \;=\; \sum_j v_j \, |\varphi_j\rangle , \tag{2.2.29}$$

we can take the inner product with $|\varphi_k\rangle$, resulting in

$$\langle\varphi_k|v\rangle \;=\; \sum_j v_j \, \langle\varphi_k|\varphi_j\rangle \;=\; v_k, \tag{2.2.30}$$

where the last equality relies on the orthonormality property eq. (2.2.13). Thus, the components of the ket are simply given by its inner products with the orthobasis vectors.

## 2.3 Dual vector spaces

Given a Hilbert space of kets with an inner product, it is useful to construct another complex linear vector space, called the **dual vector space**. Dual vectors are defined mathematically as linear maps from the space of kets to the complex numbers. For each ket $|v\rangle$, there is a corresponding dual vector denoted $\langle v|$, and we write the association as

$$\begin{array}{ccc} \text{vector} & & \text{dual vector} \\ |v\rangle & \rightarrow & \langle v|\,. \end{array} \tag{2.3.1}$$

Specifically, the dual vector $\langle v|$ is defined to map each ket $|w\rangle$ to the complex number given by the inner product of the ket $|w\rangle$ with the corresponding ket $|v\rangle$,

$$\langle v|\big(|w\rangle\big) \;\equiv\; \langle v|w\rangle\,. \tag{2.3.2}$$

A dual vector defined in this way is also called a **bra**, a silly bit of terminology devised long ago so that the inner product is a "bra-ket", or bracket. From the properties of the inner product, one can easily show that the bra associated with a linear combination of kets is

$$\begin{array}{ccc} \text{vector} & & \text{dual vector} \\ c_1\,|v\rangle + c_2\,|w\rangle & \rightarrow & c_1^*\,\langle v| + c_2^*\,\langle w|\,. \end{array} \tag{2.3.3}$$

A key thing that one must remember is to take the complex conjugates of the coefficients.

The bra $\langle v|$ is also known as the **Hermitian adjoint** of the corresponding ket $|v\rangle$, and vice versa. To understand the relation better, it is useful to consider the following linear algebra analogy: kets are like complex $N$-dimensional column vectors, and bras are like complex $N$-dimensional row vectors, with

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}, \qquad\qquad \vec{v}^\dagger = \begin{pmatrix} v_1^* & v_2^* & \cdots & v_N^* \end{pmatrix}. \tag{2.3.4}$$

The fact that a dual vector maps vectors to complex numbers is just expressed as

$$\vec{v}^\dagger \vec{w} \;=\; \sum_k v_k^* w_k. \tag{2.3.5}$$

As we will discuss more fully in section 2.5, this is not just an analogy; if one has chosen an orthobasis $\{|\varphi_k\rangle\}$, then $w_k = \langle\varphi_k|w\rangle$ are the components of the ket $|w\rangle$, while $v_k^* = \langle v|\varphi_k\rangle$ are the components of the bra $\langle v|$. However, the bra-ket notation has the great virtue of being independent of any choice of orthobasis.

According to the dual vector definition, for each ket $|w\rangle$, there is always a unique corresponding bra $\langle w|$. If the Hilbert space has a finite dimension, then there is also a unique ket for every bra, and the correspondence is one-to-one. However, in the infinite-dimensional case it is possible to construct bras that have no associated ket within the Hilbert space, although this fact is of little practical consequence. For example, consider the generalized kets like those that that satisfy the Dirac orthonormality condition eq. (2.2.15); as we have noted, these have infinite norm and are therefore not part of the physical Hilbert space. The corresponding bras are nevertheless well-defined maps from the physical Hilbert space to the complex numbers, and so are perfectly respectable members of the dual vector space.

You may find it useful to think of kets (vectors) as representing possible states of a system, while bras (dual vectors) represent possible questions that one may ask about the state. For example, we can interrogate the actual state of a system $|\psi\rangle$ about its overlap with another possible state $|\chi\rangle$. Then the question that we are asking is associated with the dual vector $\langle\chi|$. As we will discuss further at the end of section 3.2, the answer that one receives is that the probability of finding the system in the state $|\chi\rangle$ is nothing other than $|\langle\chi|\psi\rangle|^2$, assuming that both $|\psi\rangle$ and $|\chi\rangle$ were normalized to 1.

## 2.4 Operators

An **operator** $A$ is a map from the space of kets to itself. This means that the result of acting with $A$ on any ket $|v\rangle$ must be another ket in the Hilbert space, which we can call either $|Av\rangle$, or equivalently, $A\,|v\rangle$. As a notational convention, we will usually use capital letters to represent operators.

In quantum mechanics, we are almost always interested in **linear operators**, which obey

$$A\big(c_1\,|v\rangle + c_2\,|w\rangle\big) \;=\; c_1 A\,|v\rangle + c_2 A\,|w\rangle\,. \tag{2.4.1}$$

Because there is a bra associated to each ket, a linear operator also maps the dual space to itself. Specifically, for each bra $\langle w|$, the bra $\langle w|\,A$ resulting from the operation of $A$ is defined by the relation

$$\big(\langle w|\,A\big)\,|v\rangle \;=\; \langle w|\,\big(A\,|v\rangle\big). \tag{2.4.2}$$

Adopting this definition, it follows that the operation of $A$ on the dual space also obeys linearity,

$$\big(c_1\,\langle v| + c_2\,\langle w|\big)A \;=\; c_1\,\langle v|\,A + c_2\,\langle w|\,A. \tag{2.4.3}$$

Because of the equality of the two sides of eq. (2.4.2), we can define the **matrix element** of the operator $A$ between $|v\rangle$ and $\langle w|$ as their common value $\langle w|A|v\rangle$, without parentheses.

Addition and subtraction of linear operators and multiplication by complex numbers are defined in the obvious ways, such that

$$\big(c_1 A + c_2 B\big)\,|v\rangle \;=\; A\big(c_1\,|v\rangle\big) + B\big(c_2\,|v\rangle\big) \tag{2.4.4}$$

The product of two operators $A$ and $B$ is defined by

$$(AB)\,|v\rangle \;\equiv\; A\big(B\,|v\rangle\big) \;=\; AB\,|v\rangle\,. \tag{2.4.5}$$

As before, the parentheses make no difference and can be omitted, as indicated in the last equality. Operators also obey **associativity**; for any three operators $A$, $B$, and $C$,

$$(AB)C \;=\; A(BC). \tag{2.4.6}$$

However, the order of operators matters, so that $BA$ and $AB$ are different, in general. One therefore defines the **commutator** of $A$ and $B$ as

$$[A, B] = AB - BA \tag{2.4.7}$$

and the **anticommutator** by

$$\{A, B\} = AB + BA. \tag{2.4.8}$$

The simplest example of a linear operator is the **identity operator** $I$, defined by

$$I\,|v\rangle \;=\; |v\rangle\,, \qquad\qquad \langle v|\,I \;=\; \langle v| \tag{2.4.9}$$

for every $|v\rangle$. The **inverse of an operator** $A$, if it exists, is denoted $A^{-1}$, and is defined by

$$A^{-1}A \;=\; I, \tag{2.4.10}$$

from which it follows that

$$AA^{-1} \;=\; I \tag{2.4.11}$$

as well. However, it is important to recognize that not all operators $A$ have an inverse.

There is an enormously useful way of writing the identity operator, given an orthobasis. Equations (2.2.29) and (2.2.30) can be combined to write

$$|v\rangle \;=\; \sum_j |\varphi_j\rangle\langle\varphi_j|v\rangle\,. \tag{2.4.12}$$

Since this is true for all $|v\rangle$, one has simply

$$I = \sum_j |\varphi_j\rangle\langle\varphi_j|\,. \tag{2.4.13}$$

This extraordinarily important identity is known as the **completeness relation**, or the **closure relation**. We will use it very often.

If $|v\rangle$ and $|w\rangle$ are any two kets, then

$$A \;=\; |v\rangle\langle w| \tag{2.4.14}$$

is a linear operator, sometimes called the **outer product** of $|v\rangle$ and $\langle w|$. It is defined by its action on any other ket $|x\rangle$,

$$A\,|x\rangle \;=\; \big(|v\rangle\langle w|\big)\,|x\rangle \;=\; |v\rangle\,\big(\langle w|x\rangle\big). \tag{2.4.15}$$

As a useful special case, we can form the **projection operator** $P_v$ onto a ket $|v\rangle$, defined by

$$P_v\,|w\rangle \;=\; |v\rangle\,\frac{\langle v|w\rangle}{\langle v|v\rangle} \tag{2.4.16}$$

for every ket $|w\rangle$ that it acts on. This is equivalently written as

$$P_v \;=\; \frac{|v\rangle\langle v|}{\langle v|v\rangle}. \tag{2.4.17}$$

If $|v\rangle$ has norm 1, then one can simply write

$$P_v \;=\; |v\rangle\langle v|\,. \tag{2.4.18}$$

Projection operators have the property

$$P_v^2 \;=\; P_v. \tag{2.4.19}$$

Intuitively, the projection operator $P_v$ acts on a ket by throwing away the part orthogonal to $|v\rangle$, and keeping the rest intact; doing this twice has the same effect as doing it once. As an aside, this is a good example of an operator that has no inverse (with the trivial exception of the case that the state space is one-dimensional). The reason is that when $P_v$ acts on any ket $|w\rangle$ that is orthogonal to $|v\rangle$, it yields 0 (the null ket), and then there is no way to resurrect $|w\rangle$ by acting with another linear operator, the purported inverse.

The completeness relation eq. (2.4.13) can now be equivalently expressed as the statement that the identity operator is equal to

$$I \;=\; \sum_j P_{\varphi_j}, \tag{2.4.20}$$

which is the sum over the projection operators for all of the members of an orthobasis, spanning the whole Hilbert space. Similarly, one can project onto a subspace of the Hilbert space, by summing over projection operators for only a subset of the orthobasis vectors.

For any linear operator $A$, the **Hermitian adjoint** (or just adjoint) $A^\dagger$ can be defined by the way that it acts on dual vectors,

$$\langle v|\, A^\dagger \;\;=\;\; \langle Av|\,, \tag{2.4.21}$$

where $\langle Av|$ is the bra corresponding to the ket $|Av\rangle = A\,|v\rangle$. Consider a generic matrix element involving $A^\dagger$,

$$\langle w|A^\dagger|v\rangle \;=\; \langle Aw|v\rangle \;=\; (\langle v|Aw\rangle)^*\,, \tag{2.4.22}$$

where the second equality has made use of eq. (2.2.1). This we can restate as the useful relation

$$\langle w|A^\dagger|v\rangle \;\;=\;\; \big(\langle v|A|w\rangle\big)^*. \tag{2.4.23}$$

It is left as an exercise to use the definition to show the following facts. The adjoint of the operator $A = |w\rangle\langle v|$ is given by

$$\big(|w\rangle\langle v|\big)^\dagger \;\;=\;\; |v\rangle\,\langle w|\,. \tag{2.4.24}$$

The adjoint of the operator $cI$, where $c$ is a complex number, is $(cI)^\dagger = c^* I$. The adjoint of a product of operators is

$$(AB)^\dagger = B^\dagger A^\dagger, \tag{2.4.25}$$

where the order matters.

It is useful to generalize the concept of taking the adjoint to whole expressions and equalities involving kets, bras, and operators. Given any expression, define the adjoint of it according to the following rules:

- Substitute $A \to A^\dagger$ for all operators.
- Substitute $c \to c^*$ for all complex numbers.
- Substitute $|v\rangle \leftrightarrow \langle v|$ for all kets and bras.
- Reverse the order of kets, bras, and operators within each term.

(Of course, the complex number factors of any term can be written in any desired order.) Then one can show that the adjoint of any valid equality will also be a valid equality. Equations (2.2.1), (2.4.23), (2.4.24) and (2.4.25) are examples of this. As a more random example, if one has an equation

$$ABC\,|v\rangle\,\langle v|w\rangle \;+\; B\,|x\rangle \;\;=\;\; c\,|z\rangle\,, \tag{2.4.26}$$

where $A$, $B$, and $C$ are operators and $c$ is a complex number, then one must also have

$$\langle w|v\rangle\,\langle v|\,C^\dagger B^\dagger A^\dagger \;+\; \langle x|\,B^\dagger \;\;=\;\; c^*\,\langle z|\,. \tag{2.4.27}$$

An operator $A$ is called **Hermitian** (or **self-adjoint**) if it is the same as its adjoint, $A^\dagger = A$. Hermitian operators are particularly important in quantum mechanics, where they are associated with physically measurable quantities. The product of two Hermitian operators is Hermitian if, and only if, they commute. An **anti-Hermitian** operator is one that satisfies $A^\dagger = -A$. Note that any operator can be written as the sum of a Hermitian part and an anti-Hermitian part,

$$A \;\;=\;\; \left(\frac{A + A^\dagger}{2}\right) + \left(\frac{A - A^\dagger}{2}\right). \tag{2.4.28}$$

It is also not hard to show that for two Hermitian operators $A$ and $B$, the commutator is anti-Hermitian, so that $i[A, B]$ is Hermitian.

An operator $U$ is **unitary** if its adjoint is equal to its inverse, so $U^\dagger = U^{-1}$. The product of two unitary operators is always unitary. In quantum mechanics, unitary operators are associated with a change of orthobasis, as we will discuss in the next section. They often appear in the context of defining or exploiting the symmetries of the physical system. The time evolution of a system will also be associated with a unitary operator.

It is possible to define operators that are functions of other operators. The most common example of this that we will encounter in quantum mechanics is the exponentiation of an operator. If $A$ is an operator, then the operator $\exp(A)$ can be defined in two different but equivalent ways, which are useful in different circumstances. Both treat $A$ just as if it were an ordinary number, exploiting the facts that it obviously commutes with itself, and no other operators are involved. First, one can define it as the usual limit,

$$e^A \;\;=\;\; \lim_{N\to\infty}\left(I + \frac{A}{N}\right)^N. \tag{2.4.29}$$

Alternatively, one can define it as a power series expansion:

$$e^A \;\;=\;\; \sum_{n=0}^{\infty} \frac{A^n}{n!}\,, \tag{2.4.30}$$

where of course $A^0 = I$. Some useful results involving exponentials of operators follow.

**Theorem 2.4.1** *For any two operators* $A$ *and* $B$ *such that* $[A,[A,B]]=0$*,*

$$\big[e^A, B\big] \;=\; e^A[A,B]. \tag{2.4.31}$$

The proof can be obtained by first deriving $\big[A^n, B\big] = nA^{n-1}[A,B]$, and then using the power series representation of $e^A$. We also have the following special case of the Baker–Campbell–Hausdorff formula:

**Theorem 2.4.2 (Baker–Campbell–Hausdorff, special case)** *For any two operators* $A$ *and* $B$ *such that* $[A,B]$ *commutes with both* $A$ *and* $B$*, then*

$$e^A e^B = e^{A+B} e^{\frac{1}{2}[A,B]}. \tag{2.4.32}$$

The proof is left as an exercise. These theorems hold in particular if the commutator of $A$ and $B$ is proportional to the identity operator.

From the series definition eq. (2.4.30), taking $A = iB$ where $B$ is Hermitian, it is also not too difficult to show:

**Theorem 2.4.3** *If* $B$ *is a Hermitian operator, then the operator* $\exp(iB)$ *is unitary.*

Again the proof is left as an exercise.

## 2.5 Matrix representations of operators

Suppose we have selected an orthobasis $\{|\varphi_j\rangle\}$, and consider two kets $|v\rangle$ , $|w\rangle$ given as

$$|v\rangle \;=\; \sum_j v_j\,|\varphi_j\rangle\,, \qquad\qquad |w\rangle \;=\; \sum_j w_j\,|\varphi_j\rangle\,, \tag{2.5.1}$$

where $v_j$ and $w_j$ are their components in the chosen orthobasis. The inner product can be written as

$$\langle w|v\rangle \;=\; \sum_{j,k} w_j^* v_k\,\langle\varphi_j|\varphi_k\rangle \;=\; \sum_j w_j^* v_j \;=\; \begin{pmatrix} w_1^* & \cdots & w_N^* \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}, \tag{2.5.2}$$

where the second equality was obtained using the orthonormality property eq. (2.2.13). In the case of a Hilbert space with finite dimension $N$, we can therefore consider the bra $\langle w|$ as associated with a row vector of complex numbers, and the ket $|v\rangle$ with a column vector,

$$\langle w| \;\leftrightarrow\; \begin{pmatrix} w_1^* & \cdots & w_N^* \end{pmatrix}, \qquad\qquad |v\rangle \;\leftrightarrow\; \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}. \tag{2.5.3}$$

The double arrow notation is used to indicate that the bra and ket are being represented by the corresponding row and column vectors. However, it is important to remember that while this representation depends on the choice of basis, the bra and ket themselves exist as abstract objects independent of the choice of basis.

Given an operator $A$, one can construct a matrix representation for it corresponding to the orthobasis $\{|\varphi_j\rangle\}$. We start with

$$A\,|v\rangle \;=\; A\sum_k v_k\,|\varphi_k\rangle = \sum_k v_k A\,|\varphi_k\rangle\,. \tag{2.5.4}$$

Now taking the inner product with a basis ket $|\varphi_j\rangle$ gives

$$\langle\varphi_j|A|v\rangle \;=\; \sum_k \langle\varphi_j|A|\varphi_k\rangle\; v_k. \tag{2.5.5}$$

We therefore define the $N \times N$ complex matrix with elements

$$A_{jk} \;=\; \langle\varphi_j|A|\varphi_k\rangle\,, \tag{2.5.6}$$

so that the components of

$$|v'\rangle \;=\; A\,|v\rangle \tag{2.5.7}$$

are

$$v'_j \;=\; \langle\varphi_j|v'\rangle \;=\; \sum_k A_{jk}v_k, \tag{2.5.8}$$

or, in matrix form,

$$\begin{pmatrix} v'_1 \\ \vdots \\ v'_N \end{pmatrix} \;=\; \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}. \tag{2.5.9}$$

Thus, we complete the correspondence of eq. (2.5.3) by writing, for any operator $A$, the matrix representation

$$A \;\leftrightarrow\; \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix}, \tag{2.5.10}$$

with the orthobasis matrix elements given by eq. (2.5.6).

It is often convenient to dispense with formalities by replacing the $\leftrightarrow$ symbol in equations like (2.5.3) and (2.5.10) with an $=$ symbol. However, again we emphasize that the kets $|v\rangle$, bras $\langle w|$, and operators $A$ are objects that are conceptually independent of the choice of orthobasis, while their components $v_j$, $w_j^*$, and $A_{jk}$ do depend on the choice of orthobasis, and will be different if a different orthobasis is used.

The matrix representation for a product of operators $AB$ is given by

$$(AB)_{jk} = \sum_n A_{jn} B_{nk}. \tag{2.5.11}$$

The proof of this is an almost trivial consequence of the completeness relation eq. (2.4.13), because it can also be read as $\langle\varphi_j|AB|\varphi_k\rangle = \sum_n \langle\varphi_j|A|\varphi_n\rangle \langle\varphi_n|B|\varphi_k\rangle$. It is also not difficult to show that the orthobasis matrix elements of an operator $A$ and its Hermitian adjoint $A^\dagger$ are related by $(A^\dagger)_{jk} = (A_{kj})^*$.

Many operator relations and properties are easiest to see by using the matrix representation associated with an appropriately chosen orthobasis, using the tools of linear algebra. For instance, a Hermitian (or unitary) operator $A$ is represented by a Hermitian (or unitary) matrix with elements $A_{jk}$. The matrix representation of $A^{-1}$, if it exists, is equal to the inverse of the matrix representation of $A$. It is a standard result in linear algebra that $A^{-1}$ exists if and only if the determinant of the matrix $A$ is non-zero,

$$\text{Det}(A) \neq 0. \tag{2.5.12}$$

In that case, the inverse matrix for $A$ is

$$A^{-1} = \frac{1}{\text{Det}(A)} \text{cof}[A^T] \tag{2.5.13}$$

where $\text{cof}[A^T]$ is the so-called co-factor matrix of the transpose of the matrix $A$. For a general $n \times n$ matrix $B$, the co-factor matrix is defined by

$$(\text{cof}[B])_{jk} = (-1)^{j+k} M_{jk} \tag{2.5.14}$$

where the minor $M_{jk}$ is equal to the determinant of the $(n-1) \times (n-1)$ submatrix obtained from $B$ by deleting the $j$th row and the $k$th column.

A unitary operator, as defined in the previous section, satisfies

$$U^\dagger U = U U^\dagger = I. \tag{2.5.15}$$

The components of the matrix representation of $U$ in an orthobasis $\{|\varphi_k\rangle\}$,

$$U = \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1N} \\ U_{21} & U_{22} & \cdots & U_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \cdots & U_{NN} \end{pmatrix}, \tag{2.5.16}$$

obey the rule

$$U_{jk} \;=\; j\text{th component of } |\varphi'_k\rangle, \tag{2.5.17}$$

where

$$|\varphi'_k\rangle \;=\; U\,|\varphi_k\rangle\,. \tag{2.5.18}$$

Now, using eq. (2.5.15), it is not too difficult to show that the set $\{|\varphi'_k\rangle\}$ defined in this way is *another* orthobasis. From eqs. (2.5.16) and (2.5.17), we see that the components of the elements of the new orthobasis $\{|\varphi'_k\rangle\}$ are equal to the the columns of the matrix representation of $U$ in the original orthobasis. In this sense, a unitary operator $U$ performs a change of orthobasis, which can be thought of as a complex rotation in the Hilbert space. The operator $U^\dagger$ performs the inverse change of basis,

$$|\varphi_k\rangle = U^\dagger\,|\varphi'_k\rangle\,, \tag{2.5.19}$$

and the bras for the two orthobases are related by

$$\langle\varphi'_k| \;=\; \langle\varphi_k|\,U^\dagger, \qquad\qquad \langle\varphi_k| \;=\; \langle\varphi'_k|\,U. \tag{2.5.20}$$

Conversely, given any two orthonormal bases $\{|\varphi_k\rangle\}$ and $\{|\varphi'_k\rangle\}$, the inner products are the elements of a unitary matrix,

$$\langle\varphi_j|\varphi'_k\rangle \;=\; \langle\varphi_j|U|\varphi_k\rangle \;=\; U_{jk}. \tag{2.5.21}$$

To verify directly that $U$ as defined by eq. (2.5.21) is indeed unitary, one can use the completeness relation eq. (2.4.13) to show

$$\sum_n U^*_{nj} U_{nk} = \delta_{jk}, \tag{2.5.22}$$

or, rewriting,

$$\sum_n \left(U^\dagger\right)_{jn} U_{nk} = \delta_{jk}. \tag{2.5.23}$$

This expresses the unitarity of the matrix representation of the operator $U$, and is equivalent to the operator equation (2.5.15). Also,

$$|\text{Det}(U)| = 1, \tag{2.5.24}$$

follows from the general linear algebra facts that $\text{Det}(AB) = (\text{Det}A)(\text{DetB})$ and $\text{Det}(A^\dagger) = (\text{Det}A)^*$, for matrices $A, B$.

An important feature of such a unitary change of basis is that it preserves inner products. To see this, consider the transformation[†]

$$|v\rangle \quad \rightarrow \quad |v'\rangle = U\,|v\rangle\,, \tag{2.5.25}$$

$$|w\rangle \quad \rightarrow \quad |w'\rangle = U\,|w\rangle\,, \tag{2.5.26}$$

so that we have performed the same complex rotation on both $|v\rangle$ and $|w\rangle$. Then

$$\langle w|v\rangle \quad \rightarrow \quad \langle w'|v'\rangle = \langle w|U^\dagger U|v\rangle = \langle w|v\rangle\,, \tag{2.5.27}$$

so that inner products are unaffected by a change of orthobasis.

We can also see how the unitary change of basis is realized on the matrix elements of any operator $A$. For any two kets $|v\rangle$ and $|w\rangle$,

$$\langle w'|A|v'\rangle \;=\; \langle w|U^\dagger A U|v\rangle\,. \tag{2.5.28}$$

If we now define a transformation of the operator $A$ according to

$$A \quad \rightarrow \quad A' \;=\; UAU^\dagger, \tag{2.5.29}$$

to go along with eqs. (2.5.25) and (2.5.26), then

$$\langle w'|A'|v'\rangle \;=\; (\langle w|\,U^\dagger)(UAU^\dagger)(U\,|v\rangle) \;=\; \langle w|\,(U^\dagger U)A(U^\dagger U)\,|v\rangle \;=\; \langle w|A|v\rangle\,. \tag{2.5.30}$$

In words, the matrix element for the transformed states of the transformed operator $A'$ is the same as the original matrix element of $A$.

Functions of operators are often easiest to deal with using their matrix representations. If the operator $A$ has a matrix representation that is diagonal in some appropriate basis,

$$A \quad \leftrightarrow \quad \mathrm{diag}(a_1, a_2, \ldots, a_N), \tag{2.5.31}$$

then immediately from the series definition of eq. (2.4.30), with $A$ replaced by $iA$, one finds

$$\exp(iA) \quad \leftrightarrow \quad \mathrm{diag}(e^{ia_1}, e^{ia_2}, \ldots, e^{ia_N}). \tag{2.5.32}$$

Also, if in some other basis the powers of $A$ obey some recurrence relation, then the series for $\exp(iA)$ can often be resummed in the matrix representation, even if it is not diagonal. As an example, suppose $A$ has matrix representation

$$A \quad \leftrightarrow \quad a\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.5.33}$$

---

[†] Here, the arrow "→" can be read as "transforms to".

then we recognize that $A^2 = a^2 I$, and so the unitary operator $e^{iA}$ has matrix representation, obtained by separating the even powers of $A$ from the odd powers of $A$,

$$\sum_{n=0}^{\infty} \frac{(ia)^{2n}}{(2n)!} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sum_{n=0}^{\infty} \frac{(ia)^{2n+1}}{(2n+1)!} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.5.34}$$

or, summing each series,

$$\exp(iA) \;\leftrightarrow\; \begin{pmatrix} \cos(a) & i\sin(a) \\ i\sin(a) & \cos(a) \end{pmatrix}. \tag{2.5.35}$$

## 2.6 Eigenvalues and eigenvectors

In quantum mechanics, we will often need to solve eigenvalue problems. If, for some operator $A$, we can find a complex number $|\alpha\rangle$ and a non-null ket $|v\rangle$ such that

$$A\,|v\rangle \;=\; \alpha\,|v\rangle\,, \tag{2.6.1}$$

then $\alpha$ is called the **eigenvalue** and $|v\rangle$ is the corresponding **eigenvector** or **eigenket**, associated to the **eigenstate** of $A$. When eq. (2.6.1) holds, then it follows immediately that

$$\langle v|\,A^\dagger \;=\; \alpha^*\,\langle v|\,, \tag{2.6.2}$$

or, in words, $\langle v|$ is an eigenbra of $A^\dagger$, with eigenvalue $\alpha^*$.

The **eigenvalue problem** is to find all solutions for both $\alpha$ and $|v\rangle$ as a pair, given $A$. Rewriting eq. (2.6.1) as

$$(A - \alpha I)\,|v\rangle \;=\; 0, \tag{2.6.3}$$

we see that for a solution with a particular $\alpha$ to exist, the operator $(A - \alpha I)^{-1}$ must not exist; otherwise, we could act with it on both sides to discover that $|v\rangle$ could only be the null ket, in contradiction of the assumption.

If the Hilbert space is finite dimensional, we can go to the matrix representation to find that a solution for $\alpha$ must satisfy

$$\text{Det}(A - \alpha I) \;=\; 0. \tag{2.6.4}$$

This is called the characteristic equation, and the left side is a polynomial of degree $N$, the same as the dimension of the Hilbert space. The fundamental theorem of algebra states that this always has exactly $N$ complex solutions, $\alpha_1, \ldots, \alpha_N$, but some of them may be repeated. The integer number of times a particular eigenvalue $\alpha_n$ is repeated in the list is called its **degeneracy**, and we will denote it $g_{\alpha_n}$.

After the eigenvalues $\alpha_n$ have been found, for each of them we can then solve

$$(A - \alpha_n I) |v_n\rangle \;=\; 0 \tag{2.6.5}$$

for the corresponding eigenkets $|v_n\rangle$. If a particular eigenvalue $\alpha_n$ has degeneracy $g_{\alpha_n}$, then there will be a set of $g_{\alpha_n}$ linearly independent eigenkets $|v_n\rangle$ with that eigenvalue. Sometimes the solution of eq. (2.6.5) can be done by inspection, sometimes it will require linear algebra, and in still other cases it will involve solving some differential equation(s). It is often not feasible to solve it exactly, which makes life interesting and necessitates approximation methods. In any case, note that if $|v_n\rangle$ satisfies the equation, then so does $c\,|v_n\rangle$ for every complex constant $c$. That constant can always be chosen, for each eigenket, to normalize it as desired. Even after doing so, the complex phase of each eigenket remains an arbitrary choice.

The most trivial example is the case that $A = I$, the identity operator. Since $I\,|v\rangle = |v\rangle$, every non-null ket $|v\rangle$ is an eigenket, and the corresponding eigenvalue is 1.

Another simple case is the projection operator for a ket $|w\rangle$, as introduced in eq. (2.4.17):

$$P_w \;=\; \frac{|w\rangle\langle w|}{\langle w|w\rangle}. \tag{2.6.6}$$

Then the equation $P_w\,|v\rangle = \alpha\,|v\rangle$ becomes

$$\frac{\langle w|v\rangle}{\langle w|w\rangle}\,|w\rangle \;=\; \alpha\,|v\rangle\,. \tag{2.6.7}$$

There are two ways that this eigenvalue equation can be satisfied. If $\langle w|v\rangle \neq 0$, then $|v\rangle$ must be proportional to $|w\rangle$. Indeed, any $|v\rangle = c\,|w\rangle$, where $c$ is any non-zero complex number, is a solution, and the corresponding eigenvalue is $\alpha = 1$. The second way to satisfy the equation is if $\langle w|v\rangle = 0$, which then implies $\alpha = 0$. So, any ket $|v\rangle$ orthogonal to $|w\rangle$ is an eigenket, with $\alpha = 0$ as the eigenvalue. These are the only solutions to the eigenvalue problem for $P_w$.

As another example, take a Hilbert space that is spanned by an orthobasis of three kets, which we will call $|a\rangle$, $|b\rangle$, and $|c\rangle$. Consider an operator $R_\theta$, which depends on a continuous parameter $\theta$, and is defined by

$$R_\theta\,|a\rangle \;=\; \cos(\theta)\;|a\rangle + \sin(\theta)\;|b\rangle\,, \tag{2.6.8}$$

$$R_\theta\,|b\rangle \;=\; -\sin(\theta)\;|a\rangle + \cos(\theta)\;|b\rangle\,, \tag{2.6.9}$$

$$R_\theta\,|c\rangle \;=\; |c\rangle\,. \tag{2.6.10}$$

Because we have specified how $R_\theta$ acts on each member of the orthobasis, it is completely defined. We can now adopt a representation associated with this orthobasis, so that

$$|a\rangle \leftrightarrow \begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad |b\rangle \leftrightarrow \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad |c\rangle \leftrightarrow \begin{pmatrix}0\\0\\1\end{pmatrix}, \tag{2.6.11}$$

and

$$\langle a| \leftrightarrow \begin{pmatrix}1 & 0 & 0\end{pmatrix}, \qquad \langle b| \leftrightarrow \begin{pmatrix}0 & 1 & 0\end{pmatrix}, \qquad \langle c| \leftrightarrow \begin{pmatrix}0 & 0 & 1\end{pmatrix}. \tag{2.6.12}$$

The matrix representation of $R_\theta$ in this orthobasis is,[†] by applying eqs. (2.5.6) and (2.5.10),

$$R_\theta \quad \leftrightarrow \quad \begin{pmatrix}\cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1\end{pmatrix}. \tag{2.6.13}$$

This is a unitary matrix, and $R_\theta^{-1} = R_\theta^\dagger = R_{-\theta} = R_\theta^T$, where the last equality follows because it is a real matrix. The change of basis produced by $R_\theta$ is a rotation in the $|a\rangle, |b\rangle$ subspace.

To find the eigenvalues and eigenkets, first note that the characteristic equation is

$$\text{Det}(R_\theta - \alpha I) \;=\; (1 - 2\cos(\theta)\alpha + \alpha^2)(1-\alpha) \;=\; 0, \tag{2.6.14}$$

which has eigenvalue solutions

$$\alpha = e^{i\theta},\; e^{-i\theta},\; 1. \tag{2.6.15}$$

These are non-degenerate, and it is not hard to solve $(R_\theta - I\alpha)|V\rangle = 0$ in matrix form for the corresponding eigenvectors. The results are

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 \\ -i \\ 0\end{pmatrix} \quad\leftrightarrow\quad \frac{1}{\sqrt{2}}(|a\rangle - i\,|b\rangle) \qquad \text{for } \alpha = e^{i\theta}, \tag{2.6.16}$$

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 \\ i \\ 0\end{pmatrix} \quad\leftrightarrow\quad \frac{1}{\sqrt{2}}(|a\rangle + i\,|b\rangle) \qquad \text{for } \alpha = e^{-i\theta}, \tag{2.6.17}$$

$$\begin{pmatrix}0 \\ 0 \\ 1\end{pmatrix} \quad\leftrightarrow\quad |c\rangle \qquad \text{for } \alpha = 1. \tag{2.6.18}$$

We have chosen the multiplicative constants in front of these so that each eigenket has norm 1. There remains a freedom to choose each of the phases of the eigenkets; this can only be resolved by arbitrary choice.

We will now discuss several theorems regarding eigenvalues and eigenkets that are important for quantum mechanics.

**Theorem 2.6.1 (Superposition principle)** *If a linear operator $A$ has some eigenkets $|v_n\rangle$ with a common eigenvalue $\alpha$, then any complex linear combination of them, $\sum_n c_n |v_n\rangle$, is also an eigenket of $A$, with the same eigenvalue $\alpha$, unless it is the null ket.*

---

[†] Beware of a common mistake: it is tempting to scan the form of eqs. (2.6.8)–(2.6.10) and incorrectly write down the *transpose* of the matrix representation eq. (2.6.13). For similar examples, see eqs. (2.6.34) and (2.6.57).

The proof is an immediate consequence of the definitions, including the assumed linearity of $A$.

**Theorem 2.6.2 (Common eigenvectors for commuting operators)** *If operators $A$ and $B$ commute, and $|v\rangle$ is an eigenket of $A$ with eigenvalue $\alpha$, then* (i) *$B|v\rangle$ is also an eigenket of $A$ with the same eigenvalue $\alpha$, and* (ii) *if $\alpha$ is non-degenerate, then $|v\rangle$ is also an eigenket of $B$.*

Proof: To establish (i) takes only one line,

$$A(B|v\rangle) = AB|v\rangle = BA|v\rangle = B\alpha|v\rangle = \alpha(B|v\rangle). \tag{2.6.19}$$

The second equality is where we used that $A$ and $B$ commute. To prove (ii), note that the assumption that $\alpha$ is non-degenerate simply means that the corresponding eigenket is unique up to a multiplicative constant. So, from (i), $B|v\rangle = \beta|v\rangle$ for some constant $\beta$, which is therefore the eigenvalue of $B$ for $|v\rangle$.

**Theorem 2.6.3** *All eigenvalues $\alpha$ of a Hermitian operator $A$ are real.*

Proof: Let $A|v\rangle = \alpha|v\rangle$. Taking the inner product with $|v\rangle$, we have $\langle v|A|v\rangle = \alpha\langle v|v\rangle$. However, we also have $\langle v|A|v\rangle = \langle v|A^\dagger|v\rangle = (\langle v|A|v\rangle)^* = \alpha^*\langle v|v\rangle$, where the first equality made use of the assumption that $A$ is Hermitian, and the second employs eq. (2.4.23). Combining these expressions gives

$$(\alpha - \alpha^*)\langle v|v\rangle = 0, \tag{2.6.20}$$

which establishes that $\alpha = \alpha^*$, because $|v\rangle$, being an eigenket, is not null.

In general, a **selection rule** is a statement that some matrix element vanishes under certain specified conditions. The following theorem is an example.

**Theorem 2.6.4 (Matrix element selection rule)** *If $A$ is a Hermitian operator and $B$ is an operator that commutes with $A$, and $A$ has eigenkets $|v_1\rangle$ and $|v_2\rangle$ with respective eigenvalues $\alpha_1 \neq \alpha_2$, then $\langle v_1|B|v_2\rangle = 0$.*

Proof: Since $A$ and $B$ commute, $AB$ and $BA$ are the same operator, so

$$\langle v_1|AB|v_2\rangle - \langle v_1|BA|v_2\rangle = 0. \tag{2.6.21}$$

Now we can use $A|v_2\rangle = \alpha_2|v_2\rangle$ on the second term, and use eq. (2.6.2), which says $\langle v_1|A = \alpha_1^*\langle v_1|$, on the first term, to get

$$(\alpha_1 - \alpha_2)\langle v_1|B|v_2\rangle = 0, \tag{2.6.22}$$

where we have also used Theorem 2.6.3 which implies $\alpha_1^* = \alpha_1$, since $A$ is Hermitian. Now, since $\alpha_1 \neq \alpha_2$ by assumption, the matrix element $\langle v_1|B|v_2\rangle$ must vanish.

**Theorem 2.6.5 (Inner product selection rule)** *If $A$ is a Hermitian operator with eigenkets $|v_1\rangle$ and $|v_2\rangle$ with different eigenvalues $\alpha_1 \neq \alpha_2$, then the eigenkets are orthogonal, $\langle v_1|v_2\rangle = 0$.*

The proof follows immediately as a special case of Theorem 2.6.4, by taking $B = I$.

The next result establishes a useful connection between a Hermitian operator and a corresponding orthobasis for the Hilbert space:

**Theorem 2.6.6 (Orthobasis of eigenkets of a Hermitian operator)** *If $A$ is a Hermitian operator on a Hilbert space with finite dimension $N$, then one can find a set of its eigenkets that form an orthobasis. The resulting orthobasis elements corresponding to non-degenerate eigenvalues of $A$ are unique (up to multiplication by complex phases).*

Proof: The characteristic equation for the eigenvalue problem for $A$ (using its matrix representation in any orthobasis) has $N$ solutions, according to the fundamental theorem of algebra. The eigenkets corresponding to non-degenerate eigenvalues are orthogonal (according to Theorem 2.6.5) and non-null (from the definition of an eigenket), and so can be normalized to be orthonormal. For each eigenvalue $\alpha$ with degeneracy $g_\alpha > 1$, one can use the Gram–Schmidt process to construct (non-uniquely) a set of orthonormal kets which have the same eigenvalue $\alpha$ due to the linearity of the eigenvalue problem, and which are all orthogonal to the kets corresponding to the other eigenvalues (again using Theorem 2.6.5). The union of all of the kets found in this way are orthonormal with each other, and there are $N$ of them, so they form an orthobasis for the whole Hilbert space.

Let us make some important comments on Theorem 2.6.6. First, it is often convenient to adopt a notation such that the orthobasis eigenkets of $A$ are named by using the corresponding eigenvalue as a label. However, if the eigenvalue $\alpha$ is degenerate, then we need to introduce another label $u_\alpha$, which we refer to as a **degeneracy label**, to distinguish the orthobasis kets that have the same $\alpha$. Thus, the eigenkets can be called $|\alpha, u_\alpha\rangle$, and the eigenvalue equation is written as

$$(A - \alpha I)\,|\alpha, u_\alpha\rangle \;=\; 0, \tag{2.6.23}$$

while the orthonormality condition reads

$$\langle \alpha', u'_{\alpha'}|\alpha, u_\alpha\rangle \;=\; \delta_{\alpha,\alpha'}\,\delta_{u_\alpha,u'_{\alpha'}}. \tag{2.6.24}$$

In the orthobasis of eigenkets of $A$, the matrix representation of $A$ will be diagonal, as follows immediately from eq. (2.6.23), so that

$$A \;\leftrightarrow\; A_{\rm diag} = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_N \end{pmatrix} \tag{2.6.25}$$

in that basis.

Another useful comment on Theorem 2.6.6 is that if we form the matrix whose columns are the components of the orthobasis eigenkets of $A$, then the result is a unitary matrix $U$. This follows from the general discussion surrounding eqs. (2.5.16)–(2.5.23), if in that discussion we let $\{|\varphi_k\rangle\}$ be the "original" basis we are working in, and let the set $\{|\varphi'_k\rangle\} = \{|\alpha, u_\alpha\rangle\}$ be the orthobasis of eigenkets of $A$. Thus, we define the components of the matrix $U$ according to the rule that, in the original basis,

$$\begin{aligned} U_{jk} &= \langle\varphi_j|U|\varphi_k\rangle \;=\; \langle\varphi_j|\varphi'_k\rangle \\ &= j\text{th component of } (k\text{th orthonormal eigenvector of } A). \end{aligned} \tag{2.6.26}$$

The eigenvalue equation (2.6.23) can then be written as, by taking the inner product with $\langle\varphi_n|$,

$$\sum_j A_{nj}U_{jk} = \alpha_k U_{nk} = \sum_l U_{nl}(\delta_{lk}\alpha_l), \tag{2.6.27}$$

or, in index-free matrix notation,

$$AU = UA_{\text{diag}}. \tag{2.6.28}$$

Now, multiplying from the left by $U^\dagger$, we obtain

**Theorem 2.6.7 (Diagonalization of Hermitian operators)** *The matrix representation of a Hermitian operator $A$ in an arbitrary orthobasis $\{|\varphi_k\rangle\}$ is diagonalized by a unitary matrix $U$, according to*

$$U^\dagger A U \;=\; A_{\text{diag}}, \tag{2.6.29}$$

*where $U$ is specified in eq. (2.6.26), and $A_{\text{diag}}$ is the matrix representation of the operator $A$ in the new orthobasis $\{|\varphi'_k\rangle\}$ consisting of its eigenvectors.*

The **trace of an operator** in a finite-dimensional Hilbert space can be defined to be equal to the trace of its matrix representation. For an operator $A$ and any orthobasis $\{|\varphi_k\rangle\}$,

$$\text{Tr}[A] \;=\; \sum_k \langle\varphi_k|A|\varphi_k\rangle \;=\; \sum_k A_{kk}. \tag{2.6.30}$$

It is left as a short exercise to show that the trace of an operator is independent of the choice of orthobasis, and in particular is equal to the sum of its eigenvalues.

Let us do an extended example to illustrate some of the preceding results. Consider a Hilbert space of dimension 3 with an operator $A$ defined by its action on an orthobasis $\{|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle\}$,

$$A\,|\varphi_1\rangle \;=\; 3\,|\varphi_1\rangle\,, \tag{2.6.31}$$

$$A\,|\varphi_2\rangle \;=\; 4\,|\varphi_2\rangle - i\,|\varphi_3\rangle\,, \tag{2.6.32}$$

$$A\,|\varphi_3\rangle \;=\; i\,|\varphi_2\rangle + 4\,|\varphi_3\rangle\,. \tag{2.6.33}$$

The matrix representation of $A$ in this original orthobasis is, using eqs. (2.5.6) and (2.5.10),

$$A \;\leftrightarrow\; \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & i \\ 0 & -i & 4 \end{pmatrix}. \tag{2.6.34}$$

Since the matrix representation is Hermitian, $A$ is a Hermitian operator. The characteristic equation for the eigenvalues $\alpha$ of $A$ is $\mathrm{Det}(A - \alpha I) = (3 - \alpha)(15 - 8\alpha + \alpha^2) = 0$, which yields

$$\alpha = 3, 3, 5. \tag{2.6.35}$$

These eigenvalues are real, as guaranteed by Theorem 2.6.3. For the non-degenerate eigenvalue $\alpha = 5$, the solution of the matrix equation for the corresponding eigenvector gives, after normalization and choice of an arbitrary phase, and conversion to ket language,

$$|5\rangle \;=\; \frac{1}{\sqrt{2}}\left(|\varphi_2\rangle - i\,|\varphi_3\rangle\right). \tag{2.6.36}$$

Note that we have named the eigenket after its eigenvalue. For the eigenvalue $\alpha = 3$, the degeneracy is 2, so there are two linearly independent eigenkets, which can be chosen to be

$$|3,1\rangle \;=\; \frac{1}{\sqrt{2}}\left(|\varphi_2\rangle + i\,|\varphi_3\rangle\right), \tag{2.6.37}$$

$$|3,2\rangle \;=\; |\varphi_1\rangle. \tag{2.6.38}$$

The second entry (1 or 2) on each $\alpha = 3$ ket is a degeneracy label. These kets are automatically orthogonal to $|5\rangle$, as promised by Theorem 2.6.5, but we had to make a choice of linear combination to make them orthonormal, since arbitrary linear combinations of $|3,1\rangle$ and $|3,2\rangle$ would not be orthogonal to each other and would not have norm 1. In this example, the linear combination that makes an orthobasis is not hard to find, but given any other (non-orthonormal) basis for the degenerate eigenvalue subspace, the Gram–Schmidt procedure of Theorem 2.2.3 can be used to construct an orthobasis.

Our orthobasis of eigenkets of $A$ is related to the original orthobasis by a unitary transformation. Let us define an operator $U$ by

$$U\,|\varphi_1\rangle \;=\; |5\rangle, \tag{2.6.39}$$

$$U\,|\varphi_2\rangle \;=\; |3,1\rangle, \tag{2.6.40}$$

$$U\,|\varphi_3\rangle \;=\; |3,2\rangle. \tag{2.6.41}$$

Now using eqs. (2.6.36)–(2.6.38), and applying eqs. (2.5.6) and (2.5.10) again, one finds the matrix representation in the $\{|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle\}$ basis,

$$U \;\leftrightarrow\; \begin{pmatrix} 0 & 0 & 1 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -i/\sqrt{2} & i/\sqrt{2} & 0 \end{pmatrix}, \tag{2.6.42}$$

which is indeed a unitary matrix.

Acting on eqs. (2.6.39)–(2.6.41) with $U^\dagger$, and using $U^\dagger U = I$, we get

$$\begin{aligned} U^\dagger |5\rangle &= |\varphi_1\rangle = |3,2\rangle, & (2.6.43)\\ U^\dagger |3,1\rangle &= |\varphi_2\rangle = (|5\rangle + |3,1\rangle)/\sqrt{2}, & (2.6.44)\\ U^\dagger |3,2\rangle &= |\varphi_3\rangle = i(|5\rangle - |3,1\rangle)/\sqrt{2}. & (2.6.45)\end{aligned}$$

The second equality in each of these equations follows by inverting eqs. (2.6.36)–(2.6.38) to solve for $|\varphi_1\rangle$, $|\varphi_2\rangle$, and $|\varphi_3\rangle$. The results provide the matrix representation of $U^\dagger$ in the orthobasis $\{|5\rangle, |3,1\rangle, |3,2\rangle\}$, by using eqs. (2.5.6) and (2.5.10),

$$U^\dagger \leftrightarrow \begin{pmatrix} 0 & 1/\sqrt{2} & i/\sqrt{2} \\ 0 & 1/\sqrt{2} & -i/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}. \tag{2.6.46}$$

As a check, this matrix is indeed equal to the transpose conjugate of eq. (2.6.42). As promised by eq. (2.6.29),

$$U^\dagger A U = A_{\text{diag}} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \tag{2.6.47}$$

where $A$, $U$, and $U^\dagger$ on the left side are taken to be the matrices given by eqs. (2.6.34), (2.6.42), and (2.6.46), respectively. Note that $A_{\text{diag}}$ on the right side is the matrix representation of the operator $A$ in the orthobasis $\{|5\rangle, |3,1\rangle, |3,2\rangle\}$.

The choice of orthobasis made in eqs. (2.6.36)–(2.6.38) is not the unique one composed of eigenkets of $A$, due to the superposition principle of Theorem 2.6.1. The most general orthonormal pair of basis kets with $A$ eigenvalue $\alpha = 3$ is obtained using the freedom to do unitary transformations within the degenerate eigenvalue subspace, and so is parameterized by two complex numbers $c, s$, subject to the constraint $|c|^2 + |s|^2 = 1$,

$$\begin{aligned} |3,1'\rangle &= c\,|3,1\rangle + s\,|3,2\rangle, & (2.6.48)\\ |3,2'\rangle &= -s^*\,|3,1\rangle + c^*\,|3,2\rangle. & (2.6.49)\end{aligned}$$

These kets have new degeneracy labels ($1'$, $2'$) to distinguish this orthobasis from the unprimed one with $c = 1$, $s = 0$. The change in orthobasis is brought about by another unitary operator $V$, defined by

$$\begin{aligned} V|5\rangle &= |5\rangle, & (2.6.50)\\ V|3,1\rangle &= |3,1'\rangle, & (2.6.51)\\ V|3,2\rangle &= |3,2'\rangle, & (2.6.52)\end{aligned}$$

or, in a matrix representation in the basis $\{|5\rangle, |3,1\rangle, |3,2\rangle\}$, applying eqs. (2.5.6) and (2.5.10) yet again,

$$V \quad \leftrightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & -s^* \\ 0 & s & c^* \end{pmatrix}, \qquad (|5\rangle, |3,1\rangle, |3,2\rangle \text{ basis}). \tag{2.6.53}$$

By combining the information in eqs. (2.6.36)–(2.6.38) and (2.6.48)–(2.6.52), one can also write

$$V|\varphi_1\rangle = c^*|\varphi_1\rangle - \frac{s^*}{\sqrt{2}}|\varphi_2\rangle - i\frac{s^*}{\sqrt{2}}|\varphi_3\rangle, \tag{2.6.54}$$

$$V|\varphi_2\rangle = \frac{s}{\sqrt{2}}|\varphi_1\rangle + \frac{1}{2}(1+c)|\varphi_2\rangle + \frac{i}{2}(c-1)|\varphi_3\rangle, \tag{2.6.55}$$

$$V|\varphi_3\rangle = -i\frac{s}{\sqrt{2}}|\varphi_1\rangle + \frac{i}{2}(1-c)|\varphi_2\rangle + \frac{1}{2}(1+c)|\varphi_3\rangle, \tag{2.6.56}$$

so that the matrix representation of $V$ in the original orthobasis is

$$V \quad \leftrightarrow \quad \begin{pmatrix} c^* & s/\sqrt{2} & -is/\sqrt{2} \\ -s^*/\sqrt{2} & (1+c)/2 & i(1-c)/2 \\ -is^*/\sqrt{2} & i(c-1)/2 & (1+c)/2 \end{pmatrix} \qquad (|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle \text{ basis}). \tag{2.6.57}$$

The comparison of eqs. (2.6.53) and (2.6.57) illustrates the point that even though $V$ is unambiguous as a Hilbert space operator, its matrix representation looks very different depending on the orthobasis used.

We conclude this section by stating two useful theorems involving unitary operators, omitting the proofs (which are not too difficult).

**Theorem 2.6.8** *The eigenvalues of a unitary operator $U$ have magnitude 1.*

Note that eq. (2.6.15) illustrates this.

**Theorem 2.6.9** *If an operator $A$ has eigenvectors $|\alpha_n\rangle$ with eigenvalues $\alpha_n$, and $U$ is a unitary operator, then $UAU^\dagger$ has eigenvectors $U|\alpha_n\rangle$ with the same eigenvalues $\alpha_n$.*

This implies the important result that the eigenvalues of an operator do not depend on the choice of basis we use to calculate them.

## 2.7 Observables

An operator $A$ is an **observable** if it is Hermitian and its eigenkets form an orthobasis $|\alpha, u_\alpha\rangle$, which satisfies the orthonormality relation eq. (2.6.24), and the completeness relation

$$\sum_{\alpha} \sum_{u_\alpha=1}^{g_\alpha} |\alpha, u_\alpha\rangle\langle\alpha, u_\alpha| = I. \tag{2.7.1}$$

Here, $\alpha$ are the eigenvalues with degeneracies $g_\alpha$, and $u_\alpha$ are the degeneracy labels. For a Hilbert space of finite dimension, the completeness part of this definition is redundant, following automatically from Theorem 2.6.6. However it is not automatic for a Hermitian operator in an infinite-dimensional Hilbert space; that is one reason why it must be included in the definition here. The other reason is that we will later find it convenient to slightly generalize our definition of an observable to include ordered sets of Hermitian operators, so that its eigenvalues can be not just numbers, but ordered sets of numbers. (An example is the vector position operator in three-dimensional space, which is really an ordered triple of three Hermitian operators $X$, $Y$, and $Z$.) In such cases, eq. (2.7.1) is again certainly not automatic. From a physical point of view, the crucial requirement we are ensuring with eq. (2.7.1) is that if $A$ is really supposed to be an observable, then every state in the Hilbert space can always be expressed as a linear combination of its eigenstates.

The way that eq. (2.7.1) is written assumes that the eigenvalues of $A$ are countable and discrete. If instead they are uncountable and continuous, then the orthonormality and completeness conditions are of the Dirac type [compare to eq. (2.2.15)],

$$\begin{aligned} \langle \alpha', u'_{\alpha'} | \alpha, u_\alpha \rangle &= \delta(\alpha - \alpha')\, \delta_{u_\alpha, u'_{\alpha'}}, \qquad (2.7.2) \\ \int d\alpha \sum_{u_\alpha = 1}^{g_\alpha} |\alpha, u_\alpha\rangle\langle \alpha, u_\alpha| &= I, \qquad (2.7.3) \end{aligned}$$

where the integral is over the range or ranges of $\alpha$ that occur as eigenvalues. Similarly, it is also possible for the degeneracy label $u_\alpha$ to be continuous, in which case the Kronecker $\delta_{u_\alpha, u'_{\alpha'}}$ is replaced by a Dirac delta function $\delta(u_\alpha - u'_{\alpha'})$ and $\sum_{u_\alpha=1}^{g_\alpha}$ is replaced by $\int du_\alpha$.

It is even possible for the eigenvalue $\alpha$ or the degeneracy label $u_\alpha$ to have a spectrum that includes both discrete (countable) and continuous (uncountable) components. In that case, the orthonormality relations will include both Kronecker and Dirac deltas, and the completeness relation will include both a sum and an integral. We will see an example of this when we study the bound and unbound states of the hydrogen atom, in eqs. (10.2.11)-(10.2.13) and (10.2.22).

A particularly useful consequence of completeness follows from acting with $A$ on both sides of eq. (2.7.1). Then, evaluating $A = \alpha$ when acting on its eigenkets, we obtain the **spectral decomposition** of an observable operator $A$,

$$A = \sum_\alpha \sum_{u_\alpha=1}^{g_\alpha} \alpha\, |\alpha, u_\alpha\rangle\langle \alpha, u_\alpha|\,, \qquad (2.7.4)$$

with an obvious counterpart for continuous eigenvalues by replacing summation with integration. Solving problems in quantum mechanics is often the art of turning expressions involving operators into numbers. The spectral decomposition idea allows us to do this in a systematic way. When we see an observable operator $A$ appearing in a matrix element, a standard trick is to insert a completeness sum directly before or after it, so as to replace $A$ by a sum (or integral) over its eigenvalues, as in eq. (2.7.4). Of course, a prerequisite for using this trick is to solve the eigenvalue problem for $A$.

At the end of section 2.4, we mentioned that it is possible to define functions of operators, and gave the most common example, that of an exponential of an operator. If we have solved the eigenvalue problem for an observable $A$, completeness now gives us yet another way to define a completely general function of it, $F(A)$. For the case of discrete eigenvalues $\alpha$, the spectral decomposition of $F(A)$ is

$$F(A) \;=\; \sum_{\alpha} \sum_{u_\alpha=1}^{g_\alpha} F(\alpha)\; |\alpha, u_\alpha\rangle\langle\alpha, u_\alpha|\,, \tag{2.7.5}$$

again with an obvious counterpart for the continuous case involving integration rather than summation.

We next state a very useful theorem about commuting observables, which we will prove only in the case of a Hilbert space of finite dimension.

**Theorem 2.7.1 (Compatible Observables)** *If $A$ and $B$ are observables, and $[A, B] = 0$, then there exists an orthobasis of kets that are eigenkets of both $A$ and $B$. In other words, for commuting observables $A$ and $B$, there is an orthobasis in which their matrix representations are both diagonal.*

Proof: From the definition of an observable, we already know that we can find an orthobasis of kets $|\alpha, u_\alpha\rangle$ that are eigenkets of $A$. Furthermore, since $[A, B] = 0$, we know from Theorem 2.6.4 that whenever $\alpha \neq \alpha'$,

$$\langle\alpha', u'_{\alpha'}|B|\alpha, u_\alpha\rangle = 0. \tag{2.7.6}$$

Therefore, the matrix representation of $B$ has a block-diagonal form,

$$\begin{pmatrix} B_{\alpha_1} & 0 & 0 & \cdots & 0 \\ 0 & B_{\alpha_2} & 0 & \cdots & 0 \\ 0 & 0 & B_{\alpha_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & B_{\alpha_M} \end{pmatrix}, \tag{2.7.7}$$

where $M \leq N$ is the number of distinct eigenvalues $\alpha_j$ of $A$, and the blocks $B_{\alpha_j}$ along the diagonal represent the possibly non-zero entries, which are confined to the subspaces of the

Hilbert space that have the same eigenvalues $\alpha_j$. Each $B_{\alpha_j}$ is a Hermitian $g_{\alpha_j} \times g_{\alpha_j}$ matrix, where $g_{\alpha_j}$ is the degeneracy[†] of $\alpha_j$, and has entries

$$\langle \alpha_j, u'_{\alpha_j} | B | \alpha_j, u_{\alpha_j} \rangle , \qquad (u_{\alpha_j}, u'_{\alpha_j} = 1, \ldots, g_{\alpha_j}). \tag{2.7.8}$$

Therefore, according to Theorem 2.6.7, each of the matrices $B_{\alpha_j}$ can be diagonalized by a change of basis corresponding to a unitary operator. Each of these unitary operators acts non-trivially only within the corresponding $B_{\alpha_j}$ block, and is the identity operator outside of that block. After doing these basis transformations, we will be left in an orthobasis in which the operators $A$ and $B$ are represented by diagonal matrices of the form

$$A = \text{diag}(\underbrace{\alpha_1, \ldots, \alpha_1}_{g_{\alpha_1}\ \text{times}}, \underbrace{\alpha_2, \ldots, \alpha_2}_{g_{\alpha_2}\ \text{times}}, \cdots, \underbrace{\alpha_M, \ldots, \alpha_M}_{g_{\alpha_M}\ \text{times}}) \tag{2.7.9}$$

and

$$B = \text{diag}(\beta_{\alpha_1,1}, \ldots, \beta_{\alpha_1,g_{\alpha_1}}, \beta_{\alpha_2,1}, \ldots, \beta_{\alpha_2,g_{\alpha_2}}, \cdots, \beta_{\alpha_M,1}, \ldots, \beta_{\alpha_M,g_{\alpha_M}}), \tag{2.7.10}$$

where the $\beta_{\alpha_j,n}$ with $n = 1, \ldots, g_{\alpha_j}$ are the eigenvalues of the $B_{\alpha_j}$ sub-matrix. This is the desired orthobasis that achieves the requirements of the theorem.

Even with two observables $A$ and $B$, there can be unresolved degeneracies, by which we mean that there can be more than one orthobasis ket with the same eigenvalues $\alpha$ and $\beta$. This raises the question of how to tell apart the corresponding degenerate states. From a physics perspective, if two states are genuinely distinct, then there must be some observable that distinguishes them. This leads to the important concept of a **Complete Set of Commuting Observables**, called a **CSCO** from here on. For any Hilbert space of states, a CSCO is a set of observables $A, B, C, \ldots$ that all commute with each other, and whose common eigenkets form an orthobasis with no degeneracies. Thus, we can write

$$\begin{aligned} A\,|\alpha,\beta,\gamma,\ldots\rangle &= \alpha\,|\alpha,\beta,\gamma,\ldots\rangle , & (2.7.11)\\ B\,|\alpha,\beta,\gamma,\ldots\rangle &= \beta\,|\alpha,\beta,\gamma,\ldots\rangle , & (2.7.12)\\ C\,|\alpha,\beta,\gamma,\ldots\rangle &= \gamma\,|\alpha,\beta,\gamma,\ldots\rangle , & (2.7.13) \end{aligned}$$

etc., where each of the eigenkets $|\alpha,\beta,\gamma,\ldots\rangle$ is *uniquely* determined by specifying its CSCO eigenvalues. The eigenvalues of a CSCO are sometimes known as **good quantum numbers**. For each one of the eigenvalue labels $\alpha, \beta, \gamma, \ldots$, the others can be viewed as degeneracy labels.

---

[†] Note that $j = 1, \ldots, M$, and $\sum_{j=1}^{M} g_{\alpha_j} = N$ is the dimension of the Hilbert space.

If we have a set of commuting observables, but the degeneracy in the eigenkets is not fully eliminated, it means that we have not actually found a CSCO yet, and at least one more observable (commuting with all of the others) can be found to add to the list. Conversely, if the degeneracy has been fully eliminated by a list of commuting observables, then we have a CSCO; there is no need to add more observables to the list, even if it is easy to find more observables that commute with all of the others. As we will see in many examples, the number of observables needed to form a CSCO is always finite and typically not enormous for quantum systems with a fixed number of particles, even if the Hilbert space is infinite dimensional.

For a given quantum system, the choice of which operators to include in a CSCO is not unique. For example, even in the case of a spin-less free particle moving in 3 dimensions, there are an infinite number of different choices we can make for the CSCO, all of which must have 3 members. One CSCO choice turns out to be just the three spatial coordinates that specify the position of the particle. Another CSCO choice consists of the three momentum components of the particle. In some cases (those with rotational symmetry about some choice of origin), still another CSCO choice is the Hamiltonian (total energy) of the particle together with two operators associated with its angular momentum with respect to the origin. In general, the best choice of CSCO depends on what physics questions we would like to answer. Furthermore, it is often useful to be able to translate between the orthobases defined by different CSCOs.

Note that the word "complete" in CSCO does not have the same meaning as in the completeness relation. For a CSCO it is the operators that are complete, while for the completeness relation it is the orthobasis that is complete. Also, the second C in CSCO can equivalently be interpreted as standing for the word "compatible". Observables are called **compatible** if they commute with each other, and are called **incompatible** if they have a non-zero commutator.

## 2.8 Wavefunctions

In the preceding, we have been mainly concerned with matrix representations in a finite dimensional Hilbert space. Let us now consider in more detail the case of an observable with a continuous spectrum of eigenvalues. A quintessential example is the position coordinate operator $X$ for a particle moving in one dimension. The eigenvalues and eigenkets of $X$ will be labeled as $x$ and $|x\rangle$, respectively, so that the eigenvalue equation is

$$X\,|x\rangle \;\;=\;\; x\,|x\rangle\,. \tag{2.8.1}$$

The physical interpretation of this is that if the state is $|x\rangle$, then the position of the particle is known to be $x$, probably because it has just been measured to be there. The allowed values of $x$ form an uncountable, continuous set, perhaps $-\infty < x < \infty$, or more generally $a < x < b$ to take into account that the particle might be confined to some range. In the following, we will assume the latter, in order to be more general, and reserve the option to take $a \to -\infty$ and $b \to \infty$.

Since a particle measured to be at position $x$ is not at any other position $x'$ at that precise moment, the states $|x\rangle$ and $|x'\rangle$ are orthogonal. Furthermore, a measurement of the particle's position must return one of the allowed values. These two statements are made precise as a special case of eqs. (2.7.2) and (2.7.3),

$$\begin{aligned} \langle x|x'\rangle &= \delta(x-x'), && (2.8.2) \\ \int_a^b dx\; |x\rangle\langle x| &= I. && (2.8.3) \end{aligned}$$

To check that this is consistent, and that $\delta(x-x')$ in eq. (2.8.2) should really be the Dirac delta function, consider an arbitrary ket $|\psi\rangle$ in the Hilbert space. Then we define

$$\psi(x) \;\equiv\; \langle x|\psi\rangle \tag{2.8.4}$$

to be the **position wavefunction**. The wavefunction $\psi(x)$ can also be viewed as comprising the components of the vector $|\psi\rangle$ in the **position representation**. Using the completeness relation, we have

$$\psi(x) \;=\; \langle x|\psi\rangle \;=\; \int_a^b dx'\; \langle x|x'\rangle\langle x'|\psi\rangle \;=\; \int_a^b dx'\, \delta(x-x')\, \psi(x'). \tag{2.8.5}$$

The fact that this should hold for every function $\psi(x)$ is just the definition of the Dirac delta function $\delta(x-x')$; see eq. (2.2.18).

Note that the kets $|x\rangle$ do not have finite norm, since $\langle x|x\rangle = \delta(0)$ is not finite. This means that they are really generalized kets in the sense defined earlier [see the paragraph after eq. (2.2.18)], and are not associated with physical states in the Hilbert space, strictly speaking. However, eqs. (2.8.3)–(2.8.5) show that they do satisfy the requirements of orthonormality and completeness in the Dirac sense. Therefore, we will refer to the uncountable continuous set $\{|x\rangle\}$ as the **position eigenstate orthobasis**, despite the fact that they are only idealized limits of physical states.

The inner product of two kets can be written in terms of an integral of the wavefunctions, using the completeness relation, as

$$\langle\phi|\psi\rangle \;=\; \int_a^b dx\; \langle\phi|x\rangle\langle x|\psi\rangle \;=\; \int_a^b dx\, \phi^*(x)\, \psi(x). \tag{2.8.6}$$

In particular, the norm of a state can be written in the manifestly non-negative form

$$\langle\psi|\psi\rangle \;=\; \int_a^b dx\, |\psi(x)|^2. \tag{2.8.7}$$

For valid physical states, both of eqs. (2.8.6) and (2.8.7) should be finite, and the latter must be non-zero. In other words, physical states correspond to wavefunctions that are **normalizable**. This set of functions forms a Hilbert space. However, for physics purposes, there is a further constraint that the wavefunctions must also be continuous. The reason for this is as follows. The norm of a ket is not physically significant, but it is often most convenient to choose the normalization of it so that it has norm 1. In particular, if the wavefunction for a single particle is chosen to have unit norm, $\int_a^b dx\, |\psi(x)|^2 = 1$, then $|\psi(x)|^2$ can be interpreted as the probability density, in the sense that the probability to find the particle between $x$ and $x + dx$ is

$$d\mathcal{P}(x) \;=\; |\psi(x)|^2 dx. \tag{2.8.8}$$

This is known as the **Born rule**[†] for the probabilistic interpretation of the wavefunction. On physical grounds, this probability density must be unambiguous, which means that it must be the same for any limiting procedure to approach a given point $x$, implying continuity.

There is a one-to-one correspondence between kets $|\psi\rangle$ and their wavefunctions $\psi(x) = \langle x|\psi\rangle$. Now consider the ket corresponding to the derivative of the wavefunction, $d\psi/dx$. Let us write

$$|d\psi/dx\rangle \;=\; D\,|\psi\rangle\,, \tag{2.8.9}$$

which defines an operator $D$ on the Hilbert space. We then have

$$\langle x|D|\psi\rangle \;=\; \langle x|d\psi/dx\rangle \;=\; d\psi/dx. \tag{2.8.10}$$

Using completeness gives a condition satisfied by the position-eigenket matrix elements of $D$,

$$d\psi/dx \;=\; \int_a^b dx'\; \langle x|D|x'\rangle\langle x'|\psi\rangle \;=\; \int_a^b dx'\; \langle x|D|x'\rangle\, \psi(x'). \tag{2.8.11}$$

We also have

$$d\psi/dx \;=\; \frac{d}{dx}\left(\int_a^b dx'\, \psi(x')\delta(x-x')\right) \;=\; \int_a^b dx'\, \psi(x')\frac{d}{dx}\delta(x-x'). \tag{2.8.12}$$

Comparing the final results in eqs. (2.8.11) and (2.8.12), we find

$$\langle x|D|x'\rangle \;=\; \frac{d}{dx}\delta(x-x'). \tag{2.8.13}$$

To check this, plug in $\psi(x') = \delta(x' - x'')$, do the integrations, and then relabel $x'' \to x'$.

We now see that $D$ cannot be an observable, because it is certainly not Hermitian, since

$$\langle x'|D^\dagger|x\rangle \;=\; \Big(\langle x|D|x'\rangle\Big)^* \;=\; \frac{d}{dx}\delta(x-x') \;=\; -\frac{d}{dx'}\delta(x'-x) \;=\; -\,\langle x'|D|x\rangle\,. \tag{2.8.14}$$

---

[†] Max Born also came up with the Born approximation for scattering, which we will study in section 20.4. More importantly, he was the grandfather of 1970's and 1980's pop music superstar and actress Olivia Newton-John.

To obtain the third equality, we used the fact that $\delta(x - x')$ is an even function of its argument.

Although $D$ is not Hermitian, the minus sign at the end of eq. (2.8.14) suggests that it is anti-Hermitian, and we can simply multiply by a factor of $i$ or $-i$ to make an observable. We therefore define[‡] the **momentum operator**,

$$P = -i\hbar D, \tag{2.8.15}$$

which could be Hermitian and thus a candidate to be an observable. However, since this is an infinite-dimensional Hilbert space, and eq. (2.8.14) was derived using the technically ill-defined derivatives of the Dirac delta function, it is prudent to be suspicious about the Hermiticity of $P$. So, let us check.

From eq. (2.8.9), we have

$$P\,|\psi\rangle = -i\hbar\,|d\psi/dx\rangle\,. \tag{2.8.16}$$

Therefore, given any two kets $|f\rangle$ and $|g\rangle$, with wavefunctions $\langle x|f\rangle = f(x)$ and $\langle x|g\rangle = g(x)$,

$$\langle g|P|f\rangle = \int_a^b dx\ \langle g|x\rangle\langle x|P|f\rangle = -i\hbar \int_a^b g(x)^* \frac{d}{dx} f(x), \tag{2.8.17}$$

where the first equality uses completeness. Interchanging the roles of $f$ and $g$, we also have

$$\langle f|P|g\rangle = -i\hbar \int_a^b f(x)^* \frac{d}{dx} g(x). \tag{2.8.18}$$

Therefore, the difference between $P^\dagger$ and $P$, in an arbitrary matrix element, is

$$\begin{aligned}
\langle f|P^\dagger|g\rangle - \langle f|P|g\rangle &= \left(\langle g|P|f\rangle\right)^* - \langle f|P|g\rangle && (2.8.19)\\
&= i\hbar \int_a^b dx\, g(x) \frac{d}{dx} f(x)^* + i\hbar \int_a^b dx\, f(x)^* \frac{d}{dx} g(x) && (2.8.20)\\
&= i\hbar \int_a^b dx\, \frac{d}{dx}\left[f(x)^* g(x)\right] && (2.8.21)\\
&= i\hbar[f(b)^* g(b) - f(a)^* g(a)]. && (2.8.22)
\end{aligned}$$

So, for $P$ to be a Hermitian operator, we must have

$$f(b)^* g(b) = f(a)^* g(a). \tag{2.8.23}$$

This can be considered a condition on allowed wavefunctions (and their corresponding kets) in the physical Hilbert space of states if $P$ is to be an observable.

---

[‡] The sign and the $\hbar$ normalization in the definition of $P$ are mysteriously arbitrary at this point, but will be justified in what follows; see the discussions following eq. (2.8.39) and surrounding eqs. (3.4.15) and (3.4.16). For now, note that the presence of $\hbar$ at least has the correct units to make $P$ a momentum.

For example, for a particle confined to a box with impenetrable walls at $x = a$ and $x = b$, eq. (2.8.23) is satisfied because all wavefunctions are required to obey $\psi(a) = \psi(b) = 0$. It is also satisfied if all wavefunctions satisfy the weaker condition $\psi(b) = e^{i\varphi}\psi(a)$, where $\varphi$ is any fixed constant. This can occur if $X$ is interpreted to be some kind of periodic or angular coordinate $\theta$, defined for example on the interval $0 \leq \theta \leq 2\pi$; then if $\varphi = 0$ it amounts to the requirement that wavefunctions are single-valued at the point $\theta = 0$ which is identified with $\theta = 2\pi$.

If the domain of $x$ is infinite, with $a = -\infty$ and $b = \infty$, then things are trickier. In practical solutions for wavefunctions, one sometimes finds that the condition $\psi(x) \to 0$ at large spatial distances is naturally satisfied. States with this property are called **bound states**. If either $|f\rangle$ or $|g\rangle$ is one of these, then eq. (2.8.23) is satisfied. However, there may also be other solutions that do not obey this property, called **unbound states** or **scattering states**. These are very useful idealized solutions to simple problems, but strictly speaking they are generalized ket states rather than physical ones, since they cannot have finite norm. In fact, finiteness of the inner product between true physical states $|f\rangle$ and $|g\rangle$ ensures that the products $f(x)^*g(x)$ must approach 0 for large $|x|$. The resolution is that in a more precise formulation that maintains Hermiticity of $P$, the idealized unbound or scattering states should be replaced by more complicated states whose wavefunctions vanish outside of some very large region. Fortunately, this region can simply be taken so ridiculously large (say, a box with sides several light-years in length, to be safe) as to have a completely negligible effect on any physical results of interest, so that for most purposes one can use the idealized unbound states with impunity.

Next, consider the eigenvalue problem for the momentum operator. The eigenkets with definite and constant momentum $p$, corresponding to a **free particle**, satisfy

$$P\,|p\rangle \;=\; p\,|p\rangle\,. \tag{2.8.24}$$

Similarly to the case of position eigenstates, these are taken to obey Dirac orthonormality and completeness conditions, again as a special case of eqs. (2.7.2) and (2.7.3),

$$\langle p|p'\rangle \;=\; \delta(p-p'), \tag{2.8.25}$$

$$\int_{-\infty}^{\infty} dp\;|p\rangle\langle p| \;=\; I. \tag{2.8.26}$$

Here the allowed range of continuous $p$ is taken to be from $-\infty$ to $\infty$. As should now be familiar, the generalized kets $|p\rangle$ are strictly speaking not associated with physical states, because they do not have finite norm, but are still very useful as idealizations, because they satisfy the properties of an orthobasis in the Dirac sense. For any ket $|\psi\rangle$, we define the **momentum wavefunction**

$$\widetilde{\psi}(p) \;\equiv\; \langle p|\psi\rangle\,, \tag{2.8.27}$$

which can be viewed as the components of the ket $|\psi\rangle$ in the **momentum representation**. Using the completeness relation eq. (2.8.26), the inner product of any two states can be written

$$\langle\phi|\psi\rangle \;=\; \int_{-\infty}^{\infty} dp\, \widetilde{\phi}^*(p)\, \widetilde{\psi}(p). \tag{2.8.28}$$

If one has chosen the ket $|\psi\rangle$ to have norm 1, then

$$\int_{-\infty}^{\infty} dp\, |\widetilde{\psi}(p)|^2 \;=\; 1, \tag{2.8.29}$$

and we can interpret

$$d\mathcal{P}(p) \;=\; |\widetilde{\psi}(p)|^2\, dp \tag{2.8.30}$$

as the probability for the particle described by the state $|\psi\rangle$ to have momentum between $p$ and $p + dp$.

Consider the position wavefunctions of the free-particle momentum eigenstates, defined by

$$\phi_p(x) \;\equiv\; \langle x|p\rangle\,. \tag{2.8.31}$$

Now we have

$$\langle x|P|p\rangle \;=\; p\,\langle x|p\rangle \;=\; p\phi_p(x), \tag{2.8.32}$$

but also, using the completeness relation for position eigenkets,

$$\langle x|P|p\rangle \;=\; \int_{-\infty}^{\infty} dx'\; \langle x|P|x'\rangle\langle x'|p\rangle \;=\; \int_{-\infty}^{\infty} dx' \left(-i\hbar\frac{d}{dx}\delta(x-x')\right)\phi_p(x') \tag{2.8.33}$$

$$=\; -i\hbar\frac{d}{dx}\left(\int_{-\infty}^{\infty} dx'\, \delta(x-x')\phi_p(x')\right) \;=\; -i\hbar\frac{d}{dx}\phi_p(x). \tag{2.8.34}$$

Therefore, the momentum eigenvalue equation in the position representation is a linear first-order differential equation,

$$\frac{d}{dx}\phi_p(x) \;=\; i\frac{p}{\hbar}\phi_p(x), \tag{2.8.35}$$

with the solution

$$\phi_p(x) \;=\; c_p e^{ipx/\hbar}, \tag{2.8.36}$$

where $c_p$ is a non-zero complex number. To fix the magnitude of $c_p$, we use completeness in $x$ to obtain

$$\langle p'|p\rangle \;=\; \int_{-\infty}^{\infty} dx\; \langle p'|x\rangle\langle x|p\rangle \;=\; \int_{-\infty}^{\infty} dx\, \phi_{p'}(x)^*\phi_p(x) \;=\; c_{p'}^* c_p \int_{-\infty}^{\infty} dx\, e^{ix(p-p')/\hbar} \tag{2.8.37}$$

$$=\; |c_p|^2\, 2\pi\hbar\, \delta(p-p'), \tag{2.8.38}$$

where the last equality makes use of eq. (2.2.20). Now, in order to enforce the Dirac orthonormality condition of eq. (2.8.25), we choose $c_p = 1/\sqrt{2\pi\hbar}$, and so

$$\langle x|p\rangle \;=\; \phi_p(x) \;=\; \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}, \tag{2.8.39}$$

up to an unavoidably arbitrary choice of constant phase. From the physical consideration that the wavefunctions $\phi_p(x)$ should not blow up at large $|x|$, $p$ must be a real eigenvalue, corresponding to a Hermitian operator $P$. The position dependence of the momentum eigenstate is seen to have the form of a plane wave, with wavenumber $p/\hbar$, and therefore wavelength $2\pi\hbar/p$. Comparison with de Broglie's result in eq. (1.4.1) confirms that we made the correct choice of magnitude of normalization of the momentum operator $P$ in eq. (2.8.15).

As a further check of eq. (2.8.26), take the matrix element of it between two arbitrary position eigenkets,

$$\langle x'|x\rangle \;=\; \int_{-\infty}^{\infty} dp\; \langle x'|p\rangle\langle p|x\rangle \;=\; \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp\, e^{ip(x'-x)/\hbar} \;=\; \delta(x-x'). \tag{2.8.40}$$

In particular, this confirms that we must integrate over the whole range $-\infty < p < \infty$ in the momentum completeness relation eq. (2.8.26).

We can also use completeness to find formulas that convert between the momentum wavefunction and the position wavefunction. Using completeness in $x$,

$$\widetilde{\psi}(p) \;=\; \langle p|\psi\rangle = \int_{-\infty}^{\infty} dx\; \langle p|x\rangle\langle x|\psi\rangle \;=\; \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar}\psi(x). \tag{2.8.41}$$

Conversely, given $\widetilde{\psi}(p)$, we can reconstruct $\psi(x)$ using completeness in $p$,

$$\psi(x) \;=\; \langle x|\psi\rangle = \int_{-\infty}^{\infty} dp\; \langle x|p\rangle\langle p|\psi\rangle \;=\; \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dp\, e^{ipx/\hbar}\widetilde{\psi}(p). \tag{2.8.42}$$

In words, the position and momentum wavefunctions are Fourier transforms of each other, with some $\hbar$'s included. It is also easy to check that

$$\langle x|X|\psi\rangle \;=\; x\psi(x), \qquad\qquad \langle x|P|\psi\rangle \;=\; -i\hbar\frac{d}{dx}\psi(x), \tag{2.8.43}$$

$$\langle p|P|\psi\rangle \;=\; p\,\widetilde{\psi}(p), \qquad\qquad \langle p|X|\psi\rangle \;=\; i\hbar\frac{d}{dp}\widetilde{\psi}(p). \tag{2.8.44}$$

These are analogous to matrix representations of observables in a finite dimensional Hilbert space. If one works in the position representation, with wavefunctions of $x$, then the observable $X$ is represented by $x$ and the observable $P$ is represented by $-i\hbar d/dx$, and inner products are accomplished by integration over all $x$, with complex conjugation for the wavefunction of the bra vector. If, instead, one works in the momentum representation with wavefunctions of $p$, then $P$

| Representation | $\|\psi\rangle$ | $\langle\psi\|$ | $X$ | $P$ | $\langle\psi_1\|\psi_2\rangle$ |
|---|---|---|---|---|---|
| position | $\psi(x)$ | $\psi(x)^*$ | $x$ | $-i\hbar\frac{d}{dx}$ | $\int dx\,\psi_1(x)^*\psi_2(x)$ |
| momentum | $\widetilde{\psi}(p)$ | $\widetilde{\psi}(p)^*$ | $i\hbar\frac{d}{dp}$ | $p$ | $\int dp\,\widetilde{\psi}_1(p)^*\widetilde{\psi}_2(p)$ |

Table 2.8.1: Summary of the position and momentum representations for a particle moving in one dimension. The position and momentum wavefunctions are related to each other as in eqs. (2.8.41) and (2.8.42).

is represented by $p$ and $X$ is represented by $i\hbar d/dp$, and inner products are done by integration over all $p$, again with complex conjugation for the bra vector. These results are summarized in Table 2.8.1.

For example,

$$\langle\psi_1|X|\psi_2\rangle = \int_{-\infty}^{\infty} dx\,\psi_1(x)^*\,x\,\psi_2(x) = \int_{-\infty}^{\infty} dp\,\widetilde{\psi}_1(p)^*\Big(i\hbar\frac{d}{dp}\Big)\widetilde{\psi}_2(p), \tag{2.8.45}$$

$$\langle\psi_1|P|\psi_2\rangle = \int_{-\infty}^{\infty} dx\,\psi_1(x)^*\Big(-i\hbar\frac{d}{dx}\Big)\psi_2(x) = \int_{-\infty}^{\infty} dp\,\widetilde{\psi}_1(p)^*\,p\,\widetilde{\psi}_2(p). \tag{2.8.46}$$

The position representation provides a convenient way to find the commutator of $X$ and $P$. We have

$$\langle x|XP|\psi\rangle = x\Big(-i\hbar\frac{d}{dx}\Big)\psi(x) = -i\hbar x\frac{d\psi}{dx}, \tag{2.8.47}$$

$$\langle x|PX|\psi\rangle = -i\hbar\frac{d}{dx}\big(x\psi(x)\big) = -i\hbar x\frac{d\psi}{dx} - i\hbar\psi(x), \tag{2.8.48}$$

Therefore, for every ket $|\psi\rangle$,

$$\langle x|[X,P]|\psi\rangle = i\hbar\,\langle x|\psi\rangle\,, \tag{2.8.49}$$

so we can conclude that

$$[X,P] = i\hbar, \tag{2.8.50}$$

where the identity operator on the right side is understood. This position-momentum commutation relation was derived after defining the operator $P$ by its action on the orthobasis of position eigenkets, but one could just as easily have worked in the other direction, taking eq. (2.8.50) to be the fundamental definition, and then deriving the operation of $P$ on the kets.

The preceding can all be generalized in a straightforward way to three position and three momentum observables. We define position and momentum operators that are vectors in real space (not the Hilbert vector space, in which they are Hermitian operators), by

$$\vec{R} = \hat{x}X + \hat{y}Y + \hat{z}Z, \tag{2.8.51}$$

$$\vec{P} = \hat{x}P_x + \hat{y}P_y + \hat{z}P_z. \tag{2.8.52}$$

They satisfy commutation relations:

$$[X, P_x] \;=\; i\hbar, \qquad [Y, P_y] \;=\; i\hbar, \qquad [Z, P_z] \;=\; i\hbar, \tag{2.8.53}$$

with all other relevant commutators vanishing. It is convenient to adopt a notation in which the components of $\vec{R}$ and $\vec{P}$ are given an index $a = 1, 2, 3$ corresponding to $x, y, z$ respectively, so that $R_1 = X$, $R_2 = Y$, $R_3 = Z$, and $P_1 = P_x$, $P_2 = P_y$, $P_3 = P_z$. Then the commutation relations take the form

$$[R_a, P_b] \;=\; i\hbar\delta_{ab}, \qquad [R_a, R_b] \;=\; 0, \qquad [P_a, P_b] \;=\; 0. \tag{2.8.54}$$

The operators $\vec{R}$ and $\vec{P}$ have eigenkets $|\vec{r}\rangle$ and $|\vec{p}\rangle$ with[§] eigenvalues $\vec{r} = \hat{x}x + \hat{y}y + \hat{z}z$ and $\vec{p} = \hat{x}p_x + \hat{y}p_y + \hat{z}p_z$,

$$\vec{R}\,|\vec{r}\rangle = \vec{r}\,|\vec{r}\rangle\,, \qquad \vec{P}\,|\vec{p}\rangle = \vec{p}\,|\vec{p}\rangle\,, \tag{2.8.55}$$

which satisfy completeness relations

$$I \;=\; \int d^3\vec{r}\;\; |\vec{r}\rangle\langle\vec{r}|\,, \tag{2.8.56}$$

$$I \;=\; \int d^3\vec{p}\;\; |\vec{p}\rangle\langle\vec{p}|\,, \tag{2.8.57}$$

and have inner products realizing Dirac orthonormality,

$$\langle\vec{r}\,'|\vec{r}\rangle \;=\; \delta^{(3)}(\vec{r} - \vec{r}\,'), \qquad \langle\vec{p}\,'|\vec{p}\rangle \;=\; \delta^{(3)}(\vec{p} - \vec{p}\,'), \tag{2.8.58}$$

where the three-dimensional delta function is given in rectangular coordinates by

$$\delta^{(3)}(\vec{r} - \vec{r}\,') \;=\; \delta(x - x')\delta(y - y')\delta(z - z'), \tag{2.8.59}$$

and in spherical coordinates $(r, \theta, \phi)$ by

$$\delta^{(3)}(\vec{r} - \vec{r}\,') \;=\; \frac{1}{r^2}\delta(r - r')\delta(\cos\theta - \cos\theta')\delta(\phi - \phi'), \tag{2.8.60}$$

with a practical, operational definition that, when integrating over a volume $V$,

$$\int_V d^3\vec{r}\,'\, f(\vec{r}\,')\, \delta^{(3)}(\vec{r} - \vec{r}\,') \;=\; \begin{cases} f(\vec{r}) & \text{if } \vec{r} \text{ is inside the volume } V, \\ 0 & \text{if } \vec{r} \text{ is outside the volume } V. \end{cases} \tag{2.8.61}$$

The wavefunction for a free particle with momentum eigenvalue $\vec{p}$ is a plane wave,

$$\langle\vec{r}|\vec{p}\rangle \;=\; \frac{1}{(2\pi\hbar)^{3/2}} e^{i\vec{p}\cdot\vec{r}/\hbar}. \tag{2.8.62}$$

---

[§] Here we have stretched the previous notion of eigenvalue slightly, as foreshadowed after eq. (2.7.1), because our eigenvalues here are actually not just numbers, but vectors in real space, or equivalently ordered triples of numbers $(x, y, z)$ or $(p_x, p_y, p_z)$. This was a sneaky thing to do, but it is convenient, and works perfectly well provided that the components are compatible observables, as here. This will be discussed further in section 3.2.

The position and momentum wavefunctions for a general state $|\psi\rangle$ are

$$\psi(\vec{r}) = \langle\vec{r}|\psi\rangle, \qquad \widetilde{\psi}(\vec{p}) = \langle\vec{p}|\psi\rangle, \tag{2.8.63}$$

and they are related by Fourier transforms,

$$\widetilde{\psi}(\vec{p}) = \frac{1}{(2\pi\hbar)^{3/2}}\int d^3\vec{r}\, e^{-i\vec{p}\cdot\vec{r}/\hbar}\,\psi(\vec{r}), \tag{2.8.64}$$

$$\psi(\vec{r}) = \frac{1}{(2\pi\hbar)^{3/2}}\int d^3\vec{p}\, e^{i\vec{p}\cdot\vec{r}/\hbar}\,\widetilde{\psi}(\vec{p}), \tag{2.8.65}$$

as can be established using eq. (2.8.62) and the completeness relations (2.8.56) and (2.8.57). The norm of the state can be written in terms of either the position or momentum wavefunction, as

$$\langle\psi|\psi\rangle = \int d^3\vec{r}\ |\psi(\vec{r})|^2 = \int d^3\vec{p}\ |\widetilde{\psi}(\vec{p})|^2. \tag{2.8.66}$$

The interpretation of the integrands is that the probability for the particle described by the state $|\psi\rangle$ to be found in an infinitesimal volume $d^3\vec{r}$ near $\vec{r}$ is given by the Born rule,

$$d\mathcal{P}(\vec{r}) = d^3\vec{r}\,|\psi(\vec{r})|^2, \tag{2.8.67}$$

while the probability for it to have momentum in a volume $d^3\vec{p}$ in momentum space near $\vec{p}$ is

$$d\mathcal{P}(\vec{p}) = d^3\vec{p}\,|\widetilde{\psi}(\vec{p})|^2. \tag{2.8.68}$$

The last claims are based on postulates that will be stated more generally in section 3.1.

The rules for the position and momentum representations for a particle moving in three dimensions are summarized in Table 2.8.2. We also note, for future reference, that in the position representation, the operator $P^2$ is $-\hbar^2\nabla^2$.

| Representation | $\lvert\psi\rangle$ | $\langle\psi\rvert$ | $\overrightarrow{R}$ | $\overrightarrow{P}$ | $\langle\psi_1\lvert\psi_2\rangle$ |
|---|---|---|---|---|---|
| position | $\psi(\vec{r})$ | $\psi(\vec{r})^*$ | $\vec{r}$ | $-i\hbar\overrightarrow{\nabla}$ | $\int d^3\vec{r}\ \psi_1(\vec{r})^*\psi_2(\vec{r})$ |
| momentum | $\widetilde{\psi}(\vec{p})$ | $\widetilde{\psi}(\vec{p})^*$ | $i\hbar\overrightarrow{\nabla}_p$ | $\vec{p}$ | $\int d^3\vec{p}\ \widetilde{\psi}_1(\vec{p})^*\widetilde{\psi}_2(\vec{p})$ |

Table 2.8.2: Summary of the position and momentum representations for a particle moving in three dimensions. The position and momentum wavefunctions are related as in eqs. (2.8.64) and (2.8.65). In rectangular coordinates, $\overrightarrow{\nabla}_p = \hat{x}\partial/\partial p_x + \hat{y}\partial/\partial p_y + \hat{z}\partial/\partial p_z$.

## 2.9 Tensor product Hilbert spaces

We will often want to consider state spaces that are combinations of simpler state spaces. For example, consider a particle moving in three dimensions. In the previous subsection, we worked with a Hilbert space spanned by an orthobasis of kets that are eigenkets of all three position coordinates. An equivalent way to proceed is to first define three separate Hilbert spaces that only describe the $x$, $y$, and $z$ degrees of freedom, and then combine them to form the Hilbert space that describes all three degrees of freedom. As another example, a Hilbert space for a system of two particles can be constructed from the Hilbert spaces of the two particles separately.

The formal construction of Hilbert spaces from simpler components is called a **tensor product** Hilbert space. Consider two distinct Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, spanned respectively by orthobases $\{|\varphi_j\rangle\}$ with dimension $N_1$ and $\{|v_k\rangle\}$ with dimension $N_2$. Then the tensor product Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is defined to be a complex linear vector space with inner product with the following properties:

- There is an orthobasis for $\mathcal{H}$ denoted $|\varphi_j\rangle \otimes |v_k\rangle$ for $j = 1, \ldots, N_1$ and $k = 1, \ldots, N_2$. Thus, $\mathcal{H}$ has dimension $N = N_1 N_2$, and any ket in $\mathcal{H}$ can be written in a unique way as a linear combination of the $N$ kets $|\varphi_j\rangle \otimes |v_k\rangle$.

- Tensor product kets satisfy linearity and distributive properties. This means that if $|\Phi\rangle$ and $|\Psi\rangle$ are any kets in $\mathcal{H}_1$, and $|V\rangle$ and $|W\rangle$ are kets in $\mathcal{H}_2$, and $a, b, c, d$ are complex numbers, then

$$\begin{aligned}(a\,|\Phi\rangle + b\,|\Psi\rangle) \otimes (c\,|V\rangle + d\,|W\rangle) \;&=\; ac\,|\Phi\rangle \otimes |V\rangle + ad\,|\Phi\rangle \otimes |W\rangle \\ &\quad + bc\,|\Psi\rangle \otimes |V\rangle + bd\,|\Psi\rangle \otimes |W\rangle\,. \end{aligned} \tag{2.9.1}$$

- The dual tensor product Hilbert space has an orthobasis $\langle\varphi_j| \otimes \langle v_k|$.

- The inner product of orthobasis kets in $\mathcal{H}$ is inherited from the inner products of the $\mathcal{H}_1$ and $\mathcal{H}_2$ orthobasis kets, according to

$$\big(\langle\varphi_j| \otimes \langle v_k|\big)\big(|\varphi_l\rangle \otimes |v_m\rangle\big) \;=\; \delta_{j,l}\,\delta_{k,m}. \tag{2.9.2}$$

- Given an operator $A$ that acts on $\mathcal{H}_1$ and an operator $B$ that acts on $\mathcal{H}_2$, the tensor product operator $A \otimes B$ is defined to act on $\mathcal{H}$ according to

$$(A \otimes B)(|\Psi\rangle \otimes |V\rangle) \;=\; (A\,|\Psi\rangle) \otimes (B\,|V\rangle), \tag{2.9.3}$$

  for any kets $|\Psi\rangle$ in $\mathcal{H}_1$ and $|V\rangle$ in $\mathcal{H}_2$. In the very common case that $B$ is the identity operator, we simply write $A$ instead of $A \otimes I$. Thus, if it is understood that the operator $A$ acts non-trivially only on $\mathcal{H}_1$, then we write $A(|\Psi\rangle \otimes |V\rangle) = (A\,|\Psi\rangle) \otimes |V\rangle$. Similarly, if $A$ is the identity operator, then we simply write $B$ instead of $I \otimes B$.

Even though an orthobasis of kets for $\mathcal{H}$ can be chosen so that their elements are tensor products $|\varphi_j\rangle \otimes |v_k\rangle$, the same is not true for all kets in $\mathcal{H}$. Only special kets, called **separable**, can be written as a tensor product of kets,

$$|\Phi\rangle \otimes |W\rangle\,. \tag{2.9.4}$$

A ket in a tensor product Hilbert space that cannot be written in this form, for any choice of $|\Phi\rangle$ and $|W\rangle$, is said to be an **entangled state**, a concept first highlighted by Erwin Schrödinger in 1935. For example,

$$|\Psi_1\rangle \otimes |V_1\rangle + |\Psi_2\rangle \otimes |V_2\rangle \tag{2.9.5}$$

is certainly[†] an element of $\mathcal{H}$, but (assuming all the kets involved are not null) it cannot be written as a tensor product of kets $|\Phi\rangle \otimes |W\rangle$ unless either $|\Psi_1\rangle = c\,|\Psi_2\rangle$ or $|V_1\rangle = c\,|V_2\rangle$ for some complex constant $c$.

One of the major uses of the tensor product formalism is as a divide-and-conquer strategy for solving problems. This includes situations where we can make use of an already-solved problem for one or more components of the tensor product. As a simple example, the Hilbert space for a free particle moving in three dimensions can be written as a tensor product of Hilbert spaces that would describe a particle moving in the three rectangular coordinates separately,

$$\mathcal{H}_{\text{three-dimensional}} \;=\; \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_z. \tag{2.9.6}$$

An orthobasis for this Hilbert space is

$$|\vec{r}\,\rangle \;=\; |x\rangle \otimes |y\rangle \otimes |z\rangle\,, \tag{2.9.7}$$

where $|x\rangle$ describes a state in which a particle is known to have $X$ eigenvalue $x$. In this example, as in many other cases, the tensor product notation $|x\rangle \otimes |y\rangle \otimes |z\rangle$ has no real advantage over just writing the simpler

$$|\vec{r}\rangle = |x, y, z\rangle\,, \tag{2.9.8}$$

so that is what we will do from now on. Then,

$$X\,|x, y, z\rangle = x\,|x, y, z\rangle\,, \qquad\qquad Y\,|x, y, z\rangle = y\,|x, y, z\rangle\,, \tag{2.9.9}$$

etc. This is similar to the separation-of-variables strategy for solving partial differential equations problems with several independent variables.

---

[†] Because $\mathcal{H}$ is defined to be a vector space, all linear combinations of its elements must also be included in it.

Of course, one need not choose rectangular coordinates. For example, one could write the same Hilbert space as a different tensor product space appropriate for spherical coordinates,

$$\mathcal{H}_{\text{three-dimensional}} \;=\; \mathcal{H}_r \otimes \mathcal{H}_\theta \otimes \mathcal{H}_\phi, \tag{2.9.10}$$

with an orthobasis

$$|\vec{r}\rangle \;=\; |r, \theta, \phi\rangle \;=\; |r\rangle \otimes |\theta\rangle \otimes |\phi\rangle\,, \tag{2.9.11}$$

and, depending on the problem, this might be much more convenient.

The idea of tensor product Hilbert spaces also applies in situations where a major part of the analysis of a problem really involves only one of the components of the tensor product and the others are just a distraction. For example, we will study the orbital angular momentum eigenvalue problem by isolating only the relevant angular $(\theta, \phi)$ component of the three-dimensional particle Hilbert space, and then apply the results more generally to the whole Hilbert space. Schematically, we use an orthobasis that is a tensor product of the form

$$|\text{radial part}\rangle \otimes |\text{angular momentum part}\rangle\,, \tag{2.9.12}$$

although it is not necessarily notationally convenient to write it that way explicitly. This will be made more precise in section 8.6 after we develop the theory of angular momentum. In fact, the angular momentum component of the tensor product often splits up further into separate components corresponding to different types of angular momenta, for example orbital and intrinsic, and/or the angular momenta for different particles.

Another important use of tensor product spaces is to describe systems of more than one particle. For example, we can construct a Hilbert space that describes two distinct spin-less particles labeled 1 and 2, with an orthobasis of kets

$$|\vec{r}_1, \vec{r}_2\rangle \;=\; |\vec{r}_1\rangle \otimes |\vec{r}_2\rangle\,. \tag{2.9.13}$$

Here, each $|\vec{r}_n\rangle$ is the Hilbert space for just one particle $n = 1, 2$ moving in three dimensions. This construction can naturally be generalized to a tensor product Hilbert space for $N$ spin-less particles with orthobasis kets

$$|\vec{r}_1, \ldots, \vec{r}_N\rangle \;=\; |\vec{r}_1\rangle \otimes \cdots \otimes |\vec{r}_N\rangle\,. \tag{2.9.14}$$

However, when the particles have spin (intrinsic angular momentum) or are identical, things are more complicated, as we will discuss in Chapter 16.

# 3 The core principles of quantum mechanics

## 3.1 The basic postulates

We are now ready to present the six basic postulates that define quantum mechanics. Although they are not postulates in the traditional formal and rigorous mathematical sense, they constitute the key principles that should govern all quantum systems and can be used to make physical predictions. Some sources give versions that differ from the ones presented here, by combining two or all three of Postulates 2, 3, and 4, or by leaving out the controversial Postulate 5, or by including statements about subsystems. Others include a postulate governing identical particles, which we do not include for reasons discussed in Chapter 16.

**Postulate 1: States.** The state of a quantum system is specified by a non-null ket $|\psi\rangle$, a vector in a Hilbert space (a complex linear vector space with an inner product). For a physical state, the norm $\langle\psi|\psi\rangle$ is real and positive and finite, but is not otherwise physically significant; two state kets $|\psi\rangle$ and $c\,|\psi\rangle$ are physically equivalent if $c$ is any non-zero complex number.

**Postulate 2: Observables.** Physically measurable quantities, called observables, correspond to Hermitian operators whose eigenstates can be chosen to be a complete orthonormal basis (orthobasis) of the state space. This means that for an observable $A$ there is a basis $|\alpha, u_\alpha\rangle$ satisfying $A\,|\alpha, u_\alpha\rangle = \alpha\,|\alpha, u_\alpha\rangle$, where $\alpha$ are the eigenvalues, and $u_\alpha$ are the degeneracy labels for each $\alpha$, with

$$\langle\alpha', u'_{\alpha'}|\alpha, u_\alpha\rangle \;=\; \delta_{\alpha,\alpha'}\delta_{u_\alpha,u'_{\alpha'}}, \tag{3.1.1}$$

and the completeness relation

$$I \;=\; \sum_\alpha \sum_{u_\alpha} |\alpha, u_\alpha\rangle\,\langle\alpha, u_\alpha|\,. \tag{3.1.2}$$

The preceding assumes that $\alpha$ and $u_\alpha$ have discrete values. If instead $\alpha$ has continuous values, then Dirac orthonormality and completeness are used: $\delta_{\alpha,\alpha'}$ is replaced by $\delta(\alpha - \alpha')$ and $\sum_\alpha$ is replaced by a definite integral $\int d\alpha$ over the range of allowed $\alpha$. Similarly, if $u_\alpha$ is continuous, then $\delta_{u_\alpha,u'_{\alpha'}}$ is replaced by $\delta(u_\alpha - u'_{\alpha'})$, and $\sum_{u_\alpha}$ is replaced by a definite integral $\int du_\alpha$. In some cases $\alpha$ and/or $u_\alpha$ can have both some discrete and some continuous values, which are then summed and integrated over the possible values.

**Postulate 3: Allowed results of measurements.** The result of the measurement of an observable $A$ is always one of its eigenvalues, $\alpha$. This rule is sensible and consistent because the eigenvalues of a Hermitian operator do not depend on the arbitrary choice of orthobasis used to calculate them, as we observed following Theorem 2.6.9.

**Postulate 4: Probabilities of results of measurements.** If a system is in a state given by a ket $|\psi\rangle$ which has been normalized so that $\langle\psi|\psi\rangle = 1$, and an observable $A$ is measured, then the probability of getting a particular discrete result $\alpha$ is

$$\mathcal{P}(\alpha) \;=\; \langle\psi|P_\alpha|\psi\rangle \;=\; \sum_{u_\alpha} |\langle\alpha, u_\alpha|\psi\rangle|^2\,, \tag{3.1.3}$$

where $|\alpha, u_\alpha\rangle$ are normalized as in eq. (3.1.1), and

$$P_\alpha \;=\; \sum_{u_\alpha} |\alpha, u_\alpha\rangle\langle\alpha, u_\alpha| \tag{3.1.4}$$

is the projection operator to the subspace of states with eigenvalue $\alpha$ for $A$. If instead $\alpha$ is continuous, then the probability of getting a result between $\alpha$ and $\alpha + d\alpha$ is

$$d\mathcal{P}(\alpha) \;=\; d\alpha\; \langle\psi|P_\alpha|\psi\rangle \;=\; d\alpha \sum_{u_\alpha} |\langle\alpha, u_\alpha|\psi\rangle|^2\,. \tag{3.1.5}$$

This postulate generalizes the Born rule for the probabilistic interpretation of the position wavefunction, eq. (2.8.8) and its three-dimensional version (2.8.67), and the corresponding rules for momentum, eqs. (2.8.30) and (2.8.68). Note that eq. (3.1.3) ensures that the probabilities are always positive,

$$\mathcal{P}(\alpha) \;>\; 0, \tag{3.1.6}$$

and the completeness relation (3.1.2) ensures the sum of the probabilities for all possible outcomes for a measurement must be unity,

$$\sum_{\alpha} \mathcal{P}(\alpha) \;=\; 1. \tag{3.1.7}$$

Thus, these important consistency requirements demanded by any sensible theory of probability are built into the postulates of quantum mechanics.

**Postulate 5: Collapse of the state due to measurement.** If a system is in a state $|\psi\rangle$ and an observable $A$ is measured and found to have the value $\alpha$, then the state immediately after the measurement will be an eigenstate of $A$ with eigenvalue $\alpha$, specifically,

$$P_\alpha\,|\psi\rangle\,, \tag{3.1.8}$$

where $P_\alpha$ is the projection operator given by eq. (3.1.4). This ket has a norm that is typically less than 1, but positive; in fact, from eq. (3.1.3) and the projection operator rules $P_\alpha^\dagger = P_\alpha$ and $P_\alpha^2 = P_\alpha$, we see that the norm is simply equal to the probability $\mathcal{P}(\alpha)$. Therefore, one can always divide the ket by the square root of its norm, to obtain

$$\frac{P_\alpha\,|\psi\rangle}{\sqrt{\langle\psi|P_\alpha|\psi\rangle}}, \tag{3.1.9}$$

a physically equivalent ket with norm 1.

**Postulate 6: Time evolution of the quantum state.** Between measurements described by Postulate 5, the time dependence of a state obeys the Schrödinger differential equation,

$$i\hbar \frac{d}{dt} |\psi(t)\rangle \;\; = \;\; H\, |\psi(t)\rangle \,, \tag{3.1.10}$$

where the Hamiltonian $H$ is a Hermitian observable operator whose eigenvalues are the allowed energies of the system.

## 3.2 Valid and invalid questions

Scientific advances often reveal not just how to answer certain difficult questions, but that other seemingly sensible questions do not need to be answered or even considered. From the theory of special relativity, we learn that it makes no sense to ask questions about collisions of particles with relative speed larger than the speed of light in vacuum $c$; these are not valid questions because the very structure of the theory implies that such collisions do not occur. One also learns not to ask questions concerning events that are supposedly simultaneous in two different reference frames that are moving with respect to each other, because this also is not meaningful in special relativity. Even though our experience with non-relativistic systems might make such questions seem legitimate, they are in fact invalid.

Similarly, in quantum mechanics, there are questions that have no good answer because they are not valid to begin with. A prominent example is "what are the position and momentum of this particle at time $t$?". In classical mechanics this question makes perfect sense, and we learn to calculate the answer given some initial conditions and the equations of motion. However, in quantum mechanics, even in the most idealized case, we can only ask for the probability that the measurement of an observable has a specific result from among the allowed list of eigenvalues. In any particular measurement, that observable could be the position of a particle, or it could be the momentum, but it cannot be both.

To see why, suppose we attempt to define a clever new multi-component operator

$$\Omega \;\; = \;\; (\vec{R}, \vec{P}), \tag{3.2.1}$$

which is the ordered pair whose components are the position and momentum vectors of a particle. If one could measure $\Omega$, the result would be the answer to the simultaneous position and momentum of a particle. Since $\vec{R}$ and $\vec{P}$ are each Hermitian, $\Omega$ may indeed be defined as a Hermitian operator. However, it is not an observable, because it fails the part of the definition that calls for the eigenstates to form an orthobasis that satisfies the completeness relation. In fact, $\Omega$ has no eigenstates at all; this is directly related to the fact that $\vec{R}$ and $\vec{P}$ do not commute.

Notice that if they had a vanishing commutator, then Theorem 2.7.1 says that they would have had a whole orthobasis of simultaneous eigenstates, and so would $\Omega$, which could therefore have been an acceptable observable. But, as things are, $\Omega$ is not an observable, and the postulates of quantum mechanics do not allow for its measurement.

More generally, we can ask whether a set of observables can be measured simultaneously (as a multi-component operator), and whether it makes sense to ask what the results of such a measurement are. This depends on the commutation relations of the observable operators.

First suppose that the operators $A, B, C, \ldots$ all commute with each other (are compatible). Then the **simultaneous measurement** can be performed, and can be defined as a sequence of consecutive measurements, performed immediately after one another so that there is no time for the system to evolve between the measurements. It is left as an exercise to show, using Postulates 4 and 5, that for compatible observables the final results for the probabilities of the different outcomes $(\alpha, \beta, \gamma, \ldots)$ and for the corresponding final state do not depend on which order one performs the measurements, as long as they are all performed with no intervening time delay, so that Postulate 6 does not come into play.

Next, suppose that two of the operators are incompatible, with a commutator $[A, B]$ that is an operator with no vanishing eigenvalues. This includes the case that $[A, B]$ is a non-zero constant, notably if $A$ and $B$ are a position operator and the corresponding momentum. In this case, Postulate 5 tells us that after measuring $B$ the system will be left in some state $|\beta\rangle$ that is an eigenstate of $B$, but it is definitely not an eigenstate of $A$. (Otherwise, it would be an eigenstate of $[A, B]$ with eigenvalue 0, which we are assuming does not exist.) Similarly, if $A$ is measured, the system will be left in a state that is definitely not an eigenstate of $B$. Therefore, the order of making the measurements certainly makes a difference, and one cannot define their simultaneous measurement.

A third case is that $[A, B]$ is an operator that has some vanishing but also some non-vanishing eigenvalues. In this case, one might measure $A$, and find a result that leaves the state in an eigenstate of $B$, but this will not always happen. The same is true if $B$ is measured first. An evaluation using Postulates 4 and 5 will be necessary on a case-by-case basis to decide what the outcomes are that might leave $A$ and $B$ simultaneously determined in the final state.

Consider a classical observable, for example $f(a, b, c, \ldots)$ where $a, b, c, \ldots$ are quantities that have quantum observable counterparts $A, B, C, \ldots$. Then, there is always at least one quantum operator $F(A, B, C, \ldots)$ which is also an observable. However, one must be careful in defining it if $A, B, C, \ldots$ do not all commute, due to quantum ordering ambiguities. For example, if $f(x, p) = xp$ where $x$ and $p$ are the position and momentum of a particle moving in one dimension, we could try quantum operator versions $F(X, P) = XP$ or $F(X, P) = PX$ or $F(X, P) =$

$(XP + PX)/2$. The first two of these are not Hermitian. However, the last is Hermitian, and is an observable. More generally, by completely symmetrizing each term of $F(A, B, C, \ldots)$, we can always systematically construct a Hermitian version of it, as can easily be proved using eq. (2.4.25).

For the converse, there can be quantum observables that have no classical counterpart; the quintessential example of this is spin, or intrinsic angular momentum. The magnitude of the spin of a particle is a fixed multiple of $\hbar/2$. In particular, unlike ordinary angular momentum, there are no states in which it can take on classical values arbitrarily large compared to $\hbar$.

A perhaps unexpected example of a valid question in quantum mechanics is: "Given a system in a state $|\psi\rangle$, what is the probability of finding it in another state $|\chi\rangle$?". (Such a question has a valid counterpart in classical physics, but there it is somewhat trivial since all information about the state of a classical system is, in principle, more directly accessible.) The observable operator corresponding to this question is the projection operator

$$P_\chi = |\chi\rangle\langle\chi| . \tag{3.2.2}$$

It is Hermitian, and has eigenvalues 1 (with eigenstate $|\chi\rangle$) and 0 (with eigenstates consisting of all states orthogonal to $|\chi\rangle$). The result of the measurement of $P_\chi$ will therefore always be either 1 ("yes, we are in the state $|\chi\rangle$") or 0 ("no, we are not in the state $|\chi\rangle$"), even though the state $|\psi\rangle$ before the measurement need not have had either of these definite attributes. The probability of obtaining the result 1 is $|\langle\chi|\psi\rangle|^2$, assuming both kets are normalized. If we do measure $P_\chi$ and obtain the result 1, then the state after the measurement will be $P_\chi |\psi\rangle$, which is simply the same as $|\chi\rangle$, up to normalization. If we obtain the result 0, then the state after the measurement will be $(I - P_\chi) |\psi\rangle$. A crucial feature of quantum mechanics, as embodied in the collapse Postulate 5, is that making the measurement changes the state; it is not the same after the measurement as it was before, unless the system was already in an eigenstate of $P_\chi$.

## 3.3 Expectation values and uncertainties

In addition to the quantities directly associated to single measurements of an observable, there are statistical quantities that result from making many measurements. Consider an idealized situation in which we have access to an arbitrarily large number $N$ of copies of a system in the same state $|\psi\rangle$. This mythical group of identical and independent quantum systems is called a **pure ensemble**. The **expectation value** of an operator $A$ in the state $|\psi\rangle$ is defined to be the average value obtained by measuring $A$ in these independent experiments, as $N \to \infty$. According to the frequentist interpretation of probabilities, this is the same as the sum of the possible outcomes for each experiment multiplied by their respective probabilities, which can be evaluated using Postulate 4 as

$$\sum_{\alpha} \alpha \mathcal{P}(\alpha) = \sum_{\alpha} \sum_{u_\alpha} \alpha \, \langle \psi | \alpha, u_\alpha \rangle \langle \alpha, u_\alpha | \psi \rangle = \sum_{\alpha} \sum_{u_\alpha} \langle \psi | A | \alpha, u_\alpha \rangle \langle \alpha, u_\alpha | \psi \rangle = \langle \psi | A | \psi \rangle \, . \quad (3.3.1)$$

Here, the first equality assumed that $|\psi\rangle$ is normalized to unity, the second equality used the fact that $|\alpha, u_\alpha\rangle$ are eigenstates of $A$ with eigenvalue $\alpha$, and the completeness relation was used to get the last equality. In cases where the state $|\psi\rangle$ is understood by context, it is customary to denote the expectation value by

$$\langle A \rangle \;\; \equiv \;\; \langle \psi | A | \psi \rangle \, , \quad (3.3.2)$$

still assuming that $|\psi\rangle$ is normalized to unity. If that is not convenient for some reason, one has the more general relation

$$\langle A \rangle \;\; = \;\; \frac{\langle \psi | A | \psi \rangle}{\langle \psi | \psi \rangle} . \quad (3.3.3)$$

The expectation value of an observable is easily shown to be a real number.

Another quantity defined through measurements on a large pure ensemble of identical states $|\psi\rangle$ is the **uncertainty** of an observable $A$, denoted $\Delta A$. Here again the notation assumes that $|\psi\rangle$ is understood by context. It is defined by

$$(\Delta A)^2 \;\; = \;\; \langle (A - \langle A \rangle)^2 \rangle \; = \; \langle \psi | (A - \langle \psi | A | \psi \rangle)^2 | \psi \rangle \, . \quad (3.3.4)$$

Intuitively, the uncertainty tells us how much the measurement of the observable is expected to fluctuate about its expectation value, over the course of many independent measurements each performed starting in the same state $|\psi\rangle$. It is the same as the concept of standard deviation in statistics. By expanding eq. (3.3.4), one obtains the equivalent form that is most commonly used in practical calculations,

$$(\Delta A)^2 \;\; = \;\; \langle \psi | A^2 | \psi \rangle - (\langle \psi | A | \psi \rangle)^2 \, . \quad (3.3.5)$$

There is a fundamental obstacle to having states with arbitrarily small uncertainties for incompatible observables, imposed by the following result due to Howard P. Robertson and Schrödinger:

**Theorem 3.3.1 (Uncertainty relation)** *In any state $|\psi\rangle$, the uncertainties of two observables $A$ and $B$ obey*

$$(\Delta A)(\Delta B) \;\; \geq \;\; \frac{1}{2} \big| \langle [A, B] \rangle \big| . \quad (3.3.6)$$

Proof: Define observables $\widetilde{A} = A - \langle A\rangle$ and $\widetilde{B} = B - \langle B\rangle$. These are Hermitian, because $\langle A\rangle$ and $\langle B\rangle$ are real numbers. It follows from the definition of uncertainty that

$$(\Delta A)^2\,(\Delta B)^2 \;=\; \langle\psi|\widetilde{A}^2|\psi\rangle\,\langle\psi|\widetilde{B}^2|\psi\rangle \;=\; \langle\widetilde{A}\psi|\widetilde{A}\psi\rangle\,\langle\widetilde{B}\psi|\widetilde{B}\psi\rangle \;\geq\; \big|\langle\widetilde{A}\psi|\widetilde{B}\psi\rangle\big|^2, \tag{3.3.7}$$

where the Schwarz inequality eq. (2.2.4) was used at the end. Therefore, we have

$$(\Delta A)^2\,(\Delta B)^2 \;\geq\; \big|\langle\psi|\widetilde{A}\widetilde{B}|\psi\rangle\big|^2 = \left|\frac{1}{2}\,\langle\psi|[\widetilde{A},\widetilde{B}]|\psi\rangle + \frac{1}{2}\,\langle\psi|\{\widetilde{A},\widetilde{B}\}|\psi\rangle\right|^2. \tag{3.3.8}$$

Using the Hermiticity of $\widetilde{A}$ and $\widetilde{B}$ yet again, $\langle\psi|[\widetilde{A},\widetilde{B}]|\psi\rangle$ is a pure imaginary number and $\langle\psi|\{\widetilde{A},\widetilde{B}\}|\psi\rangle$ is a pure real number. Therefore, the squared magnitude of the sum is equal to the sum of the square magnitudes, and

$$(\Delta A)^2\,(\Delta B)^2 \;\geq\; \frac{1}{4}\big|\langle\psi|[\widetilde{A},\widetilde{B}]|\psi\rangle\big|^2 + \frac{1}{4}\big|\langle\psi|\{\widetilde{A},\widetilde{B}\}|\psi\rangle\big|^2. \tag{3.3.9}$$

Now, because the last term is certainly non-negative, we can drop it without affecting the validity of the inequality. Furthermore, $[\widetilde{A},\widetilde{B}]$ is just equal to $[A,B]$. So, eq. (3.3.6) follows from taking the square root of both sides of eq. (3.3.9).

A famous special case is obtained for a particle moving in one dimension, when $A = X$ and $B = P$. Because $[X,P] = i\hbar$, the uncertainties must satisfy the **Heisenberg position-momentum uncertainty relation**, named after Werner Heisenberg (who originally proposed a weaker version of it) and derived rigorously first by Earle H. Kennard and shortly after by Hermann Weyl,

$$(\Delta X)\,(\Delta P) \;\geq\; \hbar/2, \tag{3.3.10}$$

in any state. It might seem that an even stronger version might be possible, since we simply discarded the non-negative last term in eq. (3.3.9). However, for the case of position and momentum, we will show later, in section 6.1, that there do exist states (those with Gaussian wavefunctions), in which eq. (3.3.10) is saturated, in other words equality holds. The same wavefunctions will reappear in section 7.4. So, eq. (3.3.10) is the strongest possible general version of the position-momentum uncertainty relation.

For a particle moving in three dimensions, one finds in the same way that each of $(\Delta X)(\Delta P_x)$ and $(\Delta Y)(\Delta P_y)$ and $(\Delta Z)(\Delta P_z)$ cannot be less than $\hbar/2$. However, since $X$ and $P_y$ commute, there is no uncertainty relation for the product $(\Delta X)(\Delta P_y)$. This means that, in principle, one could simultaneously specify the exact values of a particle's coordinate along some direction and the momentum component in an orthogonal direction.

## 3.4 How states change

According to Postulate 6, the time evolution of a quantum state obeys a linear first-order differential equation, the Schrödinger equation. Let the initial condition for the state at time $t = t_0$ be $|\psi(t_0)\rangle$. We then define the **time-evolution operator** $U(t, t_0)$ such that the state ket at time $t$ is

$$|\psi(t)\rangle \;=\; U(t, t_0)\, |\psi(t_0)\rangle\,. \tag{3.4.1}$$

It follows from this definition, and Schrödinger's eq. (3.1.10), that the operator $U(t, t_0)$ satisfies

$$i\hbar \frac{d}{dt} U(t, t_0) \;=\; H U(t, t_0). \tag{3.4.2}$$

Our goal is to solve this differential equation for $U(t, t_0)$ as a function of $t$, subject to the boundary condition $U(t_0, t_0) = I$. As we will see, it is a unitary operator if $H(t)$ is Hermitian.

First, consider the case that the Hamiltonian does not have any dependence on $t$. In that case, the solution is simply

$$U(t - t_0) \;\equiv\; U(t, t_0) \;=\; e^{-i(t-t_0)H/\hbar}, \tag{3.4.3}$$

which only depends on the time difference $t - t_0$, not the individual times. To check this claim, note that it clearly satisfies the initial condition at $t = t_0$, and

$$\begin{aligned} i\hbar \frac{d}{dt} |\psi(t)\rangle \;&=\; i\hbar \frac{d}{dt} \left[ e^{-i(t-t_0)H/\hbar} |\psi(t_0)\rangle \right] \;=\; i\hbar \left[ (-iH/\hbar) e^{-i(t-t_0)H/\hbar} \right] |\psi(t_0)\rangle \\ &=\; H |\psi(t)\rangle \end{aligned} \tag{3.4.4}$$

recovers the Schrödinger equation, as required. It is crucial in the preceding derivation that the operator $H$ does not depend on time, so that it can be treated just like a number in the exponential, as it obviously commutes with itself.

The unitarity of $U(t - t_0)$ in eq. (3.4.3) is simple to prove, given that $H$ is Hermitian. From the rules for taking adjoints,

$$U(t - t_0)^\dagger \;=\; \left[ e^{-i(t-t_0)H/\hbar} \right]^\dagger \;=\; e^{i(t-t_0)H^\dagger/\hbar} \;=\; e^{i(t-t_0)H/\hbar} \;=\; U(t - t_0)^{-1} \;=\; U(t_0 - t). \tag{3.4.5}$$

The last equality is a bonus, which shows that evolving a state backward in time is the inverse operation of evolving it forward in time by the same amount, as one might expect. Since $U(t-t_0)$ is a unitary operator, time evolution can be regarded as equivalent to a change of orthobasis.

While eq. (3.4.3) is a neat formal solution of the Schrödinger equation, in practice it leaves more to do, because the exponential of an operator as an infinite series can be non-trivial to evaluate in matrix elements. To make further progress, we can apply the spectral decomposition trick of eq. (2.7.5) to the operator $U(t - t_0)$ as given by eq. (3.4.3). To accomplish this, first consider the eigenvalue problem for the Hamiltonian, which we are still assuming is Hermitian and does not depend on time. This eigenvalue equation is called the **time-independent Schrödinger equation**, and is written as

$$H\,|E, u_E\rangle \;=\; E\,|E, u_E\rangle\,, \tag{3.4.6}$$

where $u_E$ is a degeneracy label, and the kets $|E, u_E\rangle$ do not depend on $t$. Suppose that this equation has been solved for the energy eigenvalues $E$ and corresponding orthobasis eigenstates $|E, u_E\rangle$. Then, using completeness of the energy orthobasis,

$$U(t - t_0) \;=\; \sum_E \sum_{u_E} e^{-i(t-t_0)E/\hbar}\,|E, u_E\rangle\langle E, u_E|\,, \tag{3.4.7}$$

where we have turned the operator $H$ into the number $E$ when acting on each of its eigenstates. Applying this to eq. (3.4.1) gives

$$|\psi(t)\rangle \;=\; \sum_E \sum_{u_E} e^{-i(t-t_0)E/\hbar}\,|E, u_E\rangle\langle E, u_E|\psi(t_0)\rangle\,, \tag{3.4.8}$$

the spectral decomposition of the state with respect to energy.

Note that eq. (3.4.7) is different from the identity operator only because the phases $e^{-i(t-t_0)E/\hbar}$ are different for the eigenstates with different energies. As a special case, acting on a state that is already an eigenstate of energy, time evolution just multiplies by a global phase (that is, a single phase that multiplies the whole state ket). Because such a global phase is not physically significant, the state has not really changed. For this reason, an eigenstate of the Hamiltonian is also known as a **stationary state**.

To illustrate this, consider the time evolution of a stationary state, starting from $t = 0$,

$$|\psi_E(t)\rangle = e^{-itE/\hbar}\,|\psi_E(0)\rangle\,, \tag{3.4.9}$$

and suppose that at time $t$ we measure some observable $A$, which is assumed to have no explicit time dependence built into it.[†] Recall from Postulate 4 that the probability of getting a particular measurement result $\alpha$ is

$$\begin{aligned}\mathcal{P}(\alpha, t) \;&=\; \sum_{u_\alpha} |\,\langle \alpha, u_\alpha|\psi_E(t)\rangle\,|^2 \;=\; \sum_{u_\alpha} |e^{-itE/\hbar}\;\langle \alpha, u_\alpha|\psi_E(0)\rangle\,|^2 \;=\; \sum_{u_\alpha} |\,\langle \alpha, u_\alpha|\psi_E(0)\rangle\,|^2 \\ &=\; \mathcal{P}(\alpha, 0).\end{aligned} \tag{3.4.10}$$

---

[†] By "explicit time dependence", we just mean an explicit appearance of $t$ in the definition of the operator. For example, the position operator $X$ has no explicit time dependence, but $A = X\sin(\omega t)$ does depend explicitly on time, which we write as $\partial A/\partial t = \omega X\cos(\omega t)$.

In a stationary state, the probability of measuring each particular possible result $\alpha$ stays the same for all time, as long as the time evolution is not interrupted by a measurement.

Returning to the case of a general state $|\psi(t)\rangle$ and a general observable $A$, the time dependence of the expectation value obeys

$$\frac{d}{dt}\langle\psi|A|\psi\rangle \;=\; \langle\psi|\,A\left(\frac{d}{dt}\,|\psi\rangle\right) + \left(\frac{d}{dt}\,\langle\psi|\right)A\,|\psi\rangle + \langle\psi|\frac{\partial A}{\partial t}|\psi\rangle\,, \tag{3.4.11}$$

where $\partial A/\partial t$ is the derivative of the explicit time dependence of $A$. Applying Schrödinger's equation and its Hermitian conjugate to the first two terms on the right, respectively, gives

$$\frac{d}{dt}\langle\psi|A|\psi\rangle \;=\; -\frac{i}{\hbar}\,\langle\psi|[A,H]|\psi\rangle + \langle\psi|\frac{\partial A}{\partial t}|\psi\rangle\,, \tag{3.4.12}$$

or, in the more compact notation of expectation values, just

$$\frac{d}{dt}\langle A\rangle \;=\; -\frac{i}{\hbar}\,\langle[A,H]\rangle + \Big\langle\frac{\partial A}{\partial t}\Big\rangle, \tag{3.4.13}$$

in which the state $|\psi\rangle$ should be given by context. This general result is known as **Ehrenfest's Theorem**, after Paul Ehrenfest.

In the special case that $|\psi\rangle$ is a stationary state, $H\,|\psi\rangle = E\,|\psi\rangle$ and $\langle\psi|\,H = E\,\langle\psi|$, so $\langle[A,H]\rangle$ evaluates to $(E-E)\,\langle A\rangle = 0$. In that case, $\frac{d}{dt}\langle A\rangle = \langle\frac{\partial A}{\partial t}\rangle$. The only change in an expectation value in a stationary state is due to the explicit dependence of the observable on time.

As an example, suppose that the Hamiltonian is that of a particle moving in a potential in one dimension, with

$$H = \frac{1}{2m}P^2 + V(X), \tag{3.4.14}$$

and consider $A = P$, the momentum operator. Here, $\partial P/\partial t = 0$, as there is no explicit dependence of the operator $P$ on time. Also, $[P,H] = [P,V(X)] = -i\hbar dV/dX$. Therefore,

$$\frac{d}{dt}\langle P\rangle \;=\; -\Big\langle\frac{dV}{dX}\Big\rangle, \tag{3.4.15}$$

which is the quantum mechanical version of Newton's second law. Similarly, for $A = X$, using $\partial X/\partial t = 0$ and $[X,H] = [X,P^2]/2m = i\hbar P/m$, we get

$$\frac{d}{dt}\langle X\rangle \;=\; \langle P\rangle/m. \tag{3.4.16}$$

Ehrenfest's Theorem says that the momentum and position expectation values in quantum mechanics obey the same equations of motion as the corresponding quantities in classical mechanics. Equations (3.4.15) and (3.4.16) therefore confirm the choices of sign and magnitude in our definition of the momentum operator $P$ in eq. (2.8.15).

We now turn to the more difficult case that the Hamiltonian operator depends on time. First, consider time evolution over an infinitesimal interval from time $t_0$ to time $t_0 + \Delta t$. Then, from the Schrödinger equation,

$$|\psi(t_0 + \Delta t)\rangle \;=\; \Bigl[1 - \frac{i}{\hbar}\Delta t\, H(t_0)\Bigr]\, |\psi(t_0)\rangle\,. \tag{3.4.17}$$

Note that because $\Delta t$ is infinitesimal, it does not matter here whether we use $H(t_0)$, or $H(t)$, or $H$ evaluated at some intermediate time, because the difference will be higher order in $\Delta t$. Up to terms of order $(\Delta t)^2$, we can rewrite this as an exponential,

$$|\psi(t_0 + \Delta t)\rangle \;=\; \exp\Bigl[-\frac{i}{\hbar}\Delta t\, H(t_0)\Bigr]\, |\psi(t_0)\rangle\,. \tag{3.4.18}$$

An advantage of writing it this way is that the exponential is a unitary operator if $H(t_0)$ is Hermitian, so that $|\psi(t_0 + \Delta t)\rangle$ has the same norm as $|\psi(t_0)\rangle$. Now, if we evolve the state further from time $t_0 + \Delta$ to $t_0 + 2\Delta t$ in the same way, we have

$$|\psi(t_0 + 2\Delta t)\rangle \;=\; \exp\Bigl[-\frac{i}{\hbar}\Delta t\, H(t_0 + \Delta t)\Bigr] \exp\Bigl[-\frac{i}{\hbar}\Delta t\, H(t_0)\Bigr]\, |\psi(t_0)\rangle\,. \tag{3.4.19}$$

Here, the exponentials cannot easily be combined into a single exponential, because $H(t_0 + \Delta t)$ and $H(t_0)$ are different operators, and need not commute. Continuing in this way,

$$|\psi(t_0 + N\Delta t)\rangle \;=\; \left(\prod_{n=0}^{N-1} \exp\left[-\frac{i}{\hbar}\Delta t\, H(t_0 + n\Delta t)\right]\right) |\psi(t_0)\rangle\,, \tag{3.4.20}$$

where it is important that the terms in the product, are understood to be arranged from higher to lower $n$, reading from left to right. Since each of the terms in the product is a unitary operator, the whole product is a unitary operator as well. Now we can take $N \to \infty$ with $\Delta t = (t - t_0)/N$, to obtain

$$|\psi(t)\rangle \;=\; U(t, t_0)\, |\psi(t_0)\rangle\,, \tag{3.4.21}$$

where the unitary time-evolution operator is

$$U(t, t_0) \;=\; \lim_{N\to\infty} \prod_{n=0}^{N-1} \exp\Bigl[-\frac{i}{\hbar}\Delta t\, H(t_0 + n(t - t_0)/N)\Bigr]. \tag{3.4.22}$$

From its construction, it satisfies

$$[U(t_2, t_1)]^\dagger = [U(t_2, t_1)]^{-1} = U(t_1, t_2), \tag{3.4.23}$$

$$U(t_3, t_2) U(t_2, t_1) \;=\; U(t_3, t_1). \tag{3.4.24}$$

This time evolution operator can depend on both arguments separately; in general $U(t, t_0) \neq U(t - t_0)$ if the Hamiltonian depends on time.

Formally, one can also write for eq. (3.4.22)

$$U(t,t_0) \;=\; T\exp\Big[-\frac{i}{\hbar}\int_{t_0}^{t} dt'\, H(t')\Big], \tag{3.4.25}$$

where the $T$ is a time-ordering symbol, an instruction to rearrange the Hamiltonians in the expansion of the exponential so that each $H(t_j)$ appears to the right of $H(t_k)$ whenever $t_j < t_k$. More explicitly, it takes the form of a **Dyson series**, named after Freeman J. Dyson,

$$U(t,t_0) \;=\; I + \sum_{N=1}^{\infty}\left(-\frac{i}{\hbar}\right)^{N} \int_{t_0}^{t} dt_N \int_{t_0}^{t_N} dt_{N-1}\cdots\int_{t_0}^{t_2} dt_1\, H(t_N)H(t_{N-1})\cdots H(t_1), \tag{3.4.26}$$

where each $t_k$ integration has lower limit $t_0$ and upper limit $t_{k+1}$, for $k = 1,\ldots,N$, with $t_{N+1}$ interpreted as $t$. Note that the nested upper limits of integration have neatly removed the need for the $1/N!$ usually present in the series expansion of the exponential function. In eq. (3.4.26) the ordering of the Hamiltonians is again important because in general they do not commute at different times. In the special case that they do all commute, one can write

$$U(t,t_0) \;=\; \exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} dt'\, H(t')\right], \tag{3.4.27}$$

which agrees with the result eq. (3.4.3) when $H$ does not depend on time at all. You can also check that eq. (3.4.26) satisfies the differential equation (3.4.2), by plugging it in. Another Dyson series, based on the interaction picture of quantum mechanics and useful for time-dependent perturbation theory, will be discussed in section 17.2.

We have seen that there are two very different ways that a state can change in quantum mechanics: Schrödinger time evolution governed by the Hamiltonian, and collapse of the state ket due to measurement. It is natural to ask whether the latter might be a special case of the former. At least within the standard formulation of quantum mechanics, as given by the postulates listed in section 3.1, the answer is “No!”. Hamiltonian time evolution is accomplished by multiplying the state by a *unitary* operator, while collapse of the state due to measurement is associated with multiplying by a projection operator, which is instead *Hermitian.* Thus, there is a fundamental difference between Hamiltonian time evolution and measurement collapse.

The time evolution due to the Hamiltonian is perfectly causal and deterministic; the state at a given time is uniquely determined by the state at earlier times, just as in classical physics, provided that a measurement of the type described in Postulates 4 and 5 has not taken place in the interim. In contrast, the collapse of the wavefunction associated with measurement is inherently probabilistic, rather than deterministic. This means that, unlike in classical physics, you cannot predict the future with certainty, even in principle. But, even worse: you cannot even predict the past. Given complete knowledge of the present state of a quantum system, the past state of the system before the most recent measurement cannot be known, because Postulate 5 says that the act of measurement changes the state in a way that destroys information. If you measured an observable $A$ and got a result $\alpha$, leaving the system in a state $|\alpha\rangle$, then you know that the state of the system $|\psi\rangle$ before the measurement must have had a non-zero matrix element $\langle\alpha|\psi\rangle$, but that is all. This is clearly very incomplete information.

The insistence on a fundamental distinction between the measuring agent and the quantum system being measured, as required by Postulate 5, was developed and promoted by Niels Bohr and collaborators, and is often referred to as the **Copenhagen interpretation** of quantum mechanics. According to the Copenhagen interpretation, the measuring apparatus apparently can be thought of in classical terms, or at least we do not ask questions about its quantum behavior. This may seem troubling, as one can always imagine treating any particular measuring apparatus as itself a quantum system undergoing unitary time evolution. For this reason, many people have proposed changing the postulates of quantum mechanics to eliminate or replace Postulate 5 dealing with the collapse of the state due to measurement. It is difficult not to be sympathetic to this view. Indeed, we could draw a big sphere of radius several million light years around the Earth, and think of the contents (including us, all other known observers, and all of our measuring devices) as one big quantum system evolving strictly according to unitary time evolution. In any case, there seems to be no reasonable principle that tells us where we should put the boundary between the quantum system and the measuring apparatus.

However, from a practical point of view, Postulate 5 is indispensable, because it provides a straightforward and reliable way of making predictions for the actual experiments that we do in the real world. No matter how philosophically attractive it might be to banish the measurement collapse of the wavefunction, it is not scientifically necessary, with the possible exception of some interesting attempts to treat the quantum dynamics of the universe as a whole. So far, the postulates of quantum mechanics as given in section 3.1 have stood the test of time, successfully providing accurate predictions of every experimental phenomenon with which they have been confronted.

## 3.5 Mixed ensembles and the density matrix operator

The expectation value and uncertainty for an operator were defined for a single quantum state in section 3.3, using the concept of a pure ensemble. However, it is often more realistic to suppose that in a large ensemble of quantum systems of the same type, some fraction of them $p_1$ will be in a state $|\psi_1\rangle$, a fraction $p_2$ will be in a different state $|\psi_2\rangle$, etc. Such a large collection of systems of the same type, but in different states, is called a **mixed ensemble**. If we choose one of the systems at random from a mixed ensemble, there is a probability $p_I$ that it will be in the state $|\psi_I\rangle$, with $p_I > 0$ for each $I$, and

$$\sum_I p_I = 1. \tag{3.5.1}$$

In the following, the kets $|\psi_I\rangle$ are assumed to have unit norm, but there is no reason why they must be linearly independent or orthogonal to each other. There is not even any constraint on the number of distinct states $|\psi_I\rangle$ that are found in the ensemble. The mixed ensemble generalizes the concept of a pure ensemble, for which one of the $p_I$ is equal to 1 and all others are 0.

Suppose we choose one system of the mixed ensemble at random, and measure an observable $A$ with eigenvalues $\alpha$ and orthonormal eigenstates $|\alpha, u_\alpha\rangle$. Then, applying Postulate 4, the probability of getting a particular result $\alpha$ is equal to the sum over $|\psi_I\rangle$ of the product of the probability of choosing a system in that state and the probability that a measurement in that state will give $\alpha$,

$$\mathcal{P}(\alpha)_{\text{mixed}} = \sum_I p_I \sum_{u_\alpha} |\, \langle \alpha, u_\alpha|\psi_I\rangle \,|^2. \tag{3.5.2}$$

We can similarly compute the average result obtained by measuring the observable $A$ many times on systems chosen at random from the mixed ensemble,

$$\overline{A} = \sum_I p_I \, \langle\psi_I|A|\psi_I\rangle \,. \tag{3.5.3}$$

We use an overline notation to denote this **mixed ensemble average**, to distinguish it from the expectation value associated with measuring $A$ in a single state in a pure ensemble.

Note that there are two very different types of probabilities at work in eqs. (3.5.2) and (3.5.3). First, there are the **ensemble probabilities** $p_I$, which simply reflect the fact that the mixed ensemble is populated by different states. These ensemble probabilities would exist even if our systems were classical. Second, we have the probabilities associated with the inherently non-deterministic nature of measurement in quantum systems, which are manifested in $\sum_{u_\alpha} |\, \langle \alpha, u_\alpha|\psi_I\rangle \,|^2$ and in the expectation value $\langle\psi_I|A|\psi_I\rangle$. The results for $\mathcal{P}(\alpha)_{\text{mixed}}$ and $\overline{A}$ incorporate both types of probabilities.

One should not confuse the concepts of a pure ensemble based on a *superposition* of quantum states and a mixed ensemble containing a *population* of the same quantum states. A simple example will illustrate the distinction. Consider a state space with two orthobasis kets $|1\rangle$ and $|2\rangle$. Suppose that initially we have a pure ensemble, with all systems in the superposition state

$$|\psi\rangle = c_1|1\rangle + c_2|2\rangle \,, \tag{3.5.4}$$

where $c_1$ and $c_2$ are complex numbers subject to $|c_1|^2 + |c_2|^2 = 1$. Now we can conduct a measurement to ask if a system is in the state $|1\rangle$. The probability of finding the result 1 (yes) is $|c_1|^2$, and the probability of finding the result 0 (no) is $|c_2|^2$. If we do this measurement on each and every system in the pure ensemble, but disregard the results, we will afterwards have a mixed ensemble, with $p_1 = |c_1|^2$ for the state $|1\rangle$, and $p_2 = |c_2|^2$ for the state $|2\rangle$. This illustrates that one way to prepare a mixed ensemble is to conduct measurements on the members of a pure ensemble.

Continuing with this example, if we now make the same measurement again on the mixed ensemble, we will get the same results as for the pure ensemble; the probability is still $|c_1|^2$ to find the state $|1\rangle$. But now consider instead the probability that measurement of some other observable $A$ will yield the result $\alpha$, with corresponding eigenket $|\alpha\rangle$. For the pure ensemble with state $|\psi\rangle$, this is

$$\mathcal{P}(\alpha)_{\text{pure}} \;=\; |\langle\alpha|\psi\rangle|^2 \;=\; \big|c_1 \langle\alpha|1\rangle + c_2 \langle\alpha|2\rangle\big|^2, \tag{3.5.5}$$

but for the mixed ensemble, we find instead, from eq. (3.5.2),

$$\mathcal{P}(\alpha)_{\text{mixed}} \;=\; |c_1|^2|\langle\alpha|1\rangle|^2 + |c_2|^2|\langle\alpha|2\rangle|^2. \tag{3.5.6}$$

The key difference is that in the pure ensemble result there are interference terms that are absent for the mixed ensemble.

There is an elegant way of encoding all of the physical information about a mixed ensemble, called the **density matrix operator**, or just the **density operator**, due to John von Neumann. It is defined by

$$\rho \;=\; \sum_I p_I \,|\psi_I\rangle \langle\psi_I| \,. \tag{3.5.7}$$

In the special case of a pure ensemble, $\rho$ is simply the projection operator for the state $|\psi\rangle$. More generally, it is a sum of projection operators weighted by the frequencies of occurrence of states within the ensemble. In terms of the density operator, the result of eq. (3.5.2) for the probability of getting the result $\alpha$ for a single measurement of $A$ becomes

$$\mathcal{P}(\alpha)_{\text{ensemble}} \;=\; \sum_{u_\alpha} \langle\alpha, u_\alpha|\, \rho \,|\alpha, u_\alpha\rangle \,. \tag{3.5.8}$$

We can rewrite this by choosing an arbitrary orthobasis of kets $\{|\phi_k\rangle\}$, and then using completeness followed by a rearrangement,

$$\mathcal{P}(\alpha)_{\text{ensemble}} \;=\; \sum_{u_\alpha} \langle\alpha, u_\alpha| \Big(\sum_k |\phi_k\rangle \langle\phi_k|\Big) \rho \,|\alpha, u_\alpha\rangle \;=\; \sum_k \langle\phi_k|\rho P_\alpha|\phi_k\rangle \;=\; \text{Tr}[\rho P_\alpha], \tag{3.5.9}$$

where $P_\alpha$ is the projection operator for the result $\alpha$, defined in eq. (3.1.4), and at the end we used the definition of the trace of an operator in eq. (2.6.30). Recall that the trace is independent of the choice of orthobasis.

The average of the results of many measurements of $A$ in a mixed ensemble, as computed in eq. (3.5.3), can also be re-expressed in terms of the density operator. Again using completeness followed by a rearrangement,

$$\overline{A} \;=\; \sum_I p_I \,\langle\psi_I|\, A\Big(\sum_k |\phi_k\rangle\,\langle\phi_k|\Big)\,|\psi_I\rangle \;=\; \sum_k \langle\phi_k|\rho A|\phi_k\rangle \;=\; \mathrm{Tr}[\rho A]. \tag{3.5.10}$$

As a special case,

$$\mathrm{Tr}[\rho] \;=\; 1, \tag{3.5.11}$$

which simply re-expresses the conservation of probability from eq. (3.5.1). It is also straightforward to show $\mathrm{Tr}[\rho^2] \leq 1$, with equality only in the special case that the ensemble is a pure one, in which case one also has $\rho^2 = \rho$.

From its definition in eq. (3.5.7), $\rho$ is clearly a Hermitian operator. So, according to Theorem 2.6.6, one can find a special orthobasis consisting of its eigenkets, call them $|\varphi_k\rangle$, with eigenvalues $p_k$ that are real because of Theorem 2.6.3. Therefore, no matter what states $|\psi_I\rangle$ were involved in the original preparation of the mixed ensemble, or how many such states there were, we can always rewrite the density operator as

$$\rho \;=\; \sum_k p_k \,|\varphi_k\rangle\langle\varphi_k|\,, \tag{3.5.12}$$

where the index $k$ now takes on a limited number of values up to the dimension of the state space, and the $p_k$ can be interpreted as ensemble probabilities. This illustrates the more general fact that the density operator is not tied to any specific set of states $|\psi_I\rangle$, even if we used one when preparing the mixed ensemble. The density operator can also be used to summarize our (incomplete) information about a single system chosen at random from the mixed ensemble, called a **mixed state**.

A pure ensemble is one extreme special case of a mixed ensemble, in which all systems are in the same state. The opposite extreme is the **completely random ensemble**, which we can define by choosing any orthobasis $|\phi_k\rangle$ with $k = 1, \ldots, n$, where $n$ is the dimension of the state space, and writing

$$\rho \;=\; \frac{1}{n}\sum_{k=1}^{n} |\phi_k\rangle\langle\phi_k|\,. \tag{3.5.13}$$

Here $1/n$ is the ensemble probability for each of the orthobasis states. This density operator is proportional to the identity operator, so it is actually independent of the choice of orthobasis, and is the unique one associated with maximum randomness of states in the ensemble.

The extent to which an ensemble of quantum systems is randomized can more generally be quantified by the **von Neumann entropy**,

$$\sigma \;=\; -\mathrm{Tr}[\rho \ln \rho] \;=\; -\sum_k p_k \ln p_k, \tag{3.5.14}$$

where the final result is in terms of the orthobasis ensemble probabilities in eq. (3.5.12), which are defined to be the eigenvalues of $\rho$. This is equivalent to the **Shannon entropy** introduced by Claude E. Shannon in the study of information and communication theory. It is also equivalent, up to a multiplicative constant factor, to the **Gibbs entropy** defined by Josiah Willard Gibbs in statistical mechanics and thermodynamics,

$$S \;=\; k_B \sigma \;=\; -k_B \mathrm{Tr}[\rho \ln \rho], \tag{3.5.15}$$

where $k_B$ is Boltzmann's constant. In our two extreme cases,

$$\sigma \;=\; 0 \qquad\qquad \text{(pure ensemble)}, \tag{3.5.16}$$

$$\sigma \;=\; -n\left[\frac{1}{n}\ln(1/n)\right] \;=\; \ln(n) \qquad\qquad \text{(completely random ensemble)}, \tag{3.5.17}$$

where $n$ is the number of orthobasis states available to the systems in the ensemble, generally the same as the dimension of the state space. In the case of a mixed state, the entropy is a measure of our ignorance, and it is always between 0 and $\ln(n)$. The result $S = k_B \ln(n)$ obtained for the special case of the completely random ensemble is the **Boltzmann entropy**.

In general, the density operator for a mixed state or a mixed ensemble depends on time. From the general form in eq. (3.5.7), we have

$$\frac{d\rho}{dt} \;=\; \sum_I \left[\left(\frac{d}{dt}\left|\psi_I\right\rangle\right)\left\langle\psi_I\right| + \left|\psi_I\right\rangle\left(\frac{d}{dt}\left\langle\psi_I\right|\right)\right], \tag{3.5.18}$$

and evaluating the time derivatives using the Schrödinger equation, we obtain

$$\frac{d\rho}{dt} \;=\; -\frac{i}{\hbar}[H, \rho]. \tag{3.5.19}$$

Note that this vanishes in the special case of a completely random ensemble. It also vanishes in the case of a pure ensemble if the state is an energy eigenstate, but not if it is a superposition of states with different energies. The historical reason for the name "density operator" is that this equation is analogous to Liouville's Theorem in classical mechanics, which says that for an ensemble of classical systems, the phase-space density $\rho_{\text{classical}}$ (the number of classical ensemble members per unit position and momentum) obeys

$$\frac{\partial \rho_{\text{classical}}}{\partial t} = \{H,\, \rho_{\text{classical}}\}_{\text{PB}}, \tag{3.5.20}$$

which has a Poisson bracket (and a factor of $i\hbar$) on the right-hand side replacing the commutator. This is an example of the classical–quantum correspondence principle to be discussed further in section 4.1.

Equation (3.5.19) gives the instantaneous change in the density operator as the ensemble of systems undergoes time evolution. More generally, in terms of the unitary time evolution operator that we defined in eq. (3.4.1),

$$\rho(t) = U(t, t_0)\, \rho(t_0)\, U(t, t_0)^\dagger, \tag{3.5.21}$$

with a short proof left as an exercise. Since $\rho(t)$ and $\rho(t_0)$ are related by a unitary transformation, they have the same eigenvalues, according to Theorem 2.6.9. It follows, from the expression in eq. (3.5.14) in terms of the eigenvalues $p_k$, that

$$\sigma(t) = \sigma(t_0) \qquad \text{(unitary time evolution)}, \tag{3.5.22}$$

in the absence of external measurements or other influences. So, we have found that the entropy of a closed system does not change with time.

Let us now work out what the density operator will be after a measurement on a mixed state. Suppose we have a density operator $\rho$ as given in eq. (3.5.12) in terms of orthobasis states $|\varphi_k\rangle$ with ensemble probabilities $p_k$, and we make a measurement on a randomly chosen ensemble state of an observable $A$ and obtain the result $\alpha$. For each of the ensemble states $|\varphi_k\rangle$, Postulate 5 tells us that the state after the measurement will be

$$\frac{P_\alpha\, |\varphi_k\rangle}{\sqrt{\langle \varphi_k | P_\alpha | \varphi_k \rangle}}, \tag{3.5.23}$$

where $P_\alpha$ is the projection operator for the result $\alpha$. Therefore, we can write the post-measurement density operator as

$$\rho_\alpha = \sum_k \mathcal{P}(k|\alpha) \left( \frac{P_\alpha\, |\varphi_k\rangle}{\sqrt{\langle \varphi_k | P_\alpha | \varphi_k \rangle}} \right) \left( \frac{\langle \varphi_k |\, P_\alpha}{\sqrt{\langle \varphi_k | P_\alpha | \varphi_k \rangle}} \right), \tag{3.5.24}$$

where $\mathcal{P}(k|\alpha)$ is the conditional probability that the state selected from the ensemble was $|\varphi_k\rangle$, given that the result $\alpha$ was obtained for $A$. To evaluate this, we use **Bayes' Theorem**, the fundamental result in the theory of conditional probabilities, which says

$$\mathcal{P}(k|\alpha) = \frac{\mathcal{P}(\alpha|k)\, p_k}{\mathcal{P}(\alpha)}. \tag{3.5.25}$$

Here, as given by Postulate 4,

$$\mathcal{P}(\alpha|k) \quad = \quad \langle\varphi_k|P_\alpha|\varphi_k\rangle \tag{3.5.26}$$

is the conditional probability that the result of a measurement of $A$ is $\alpha$, assuming that the state was $|\varphi_k\rangle$, while $\mathcal{P}(\alpha) = \text{Tr}[\rho P_\alpha]$ is the probability, obtained in eq. (3.5.9), that a measurement of $A$ in the ensemble resulted in $\alpha$. Putting these results together, eq. (3.5.24) for the density operator after the measurement becomes

$$\rho_\alpha \quad = \quad \frac{P_\alpha \rho P_\alpha}{\text{Tr}[\rho P_\alpha]}. \tag{3.5.27}$$

This is the density operator version of Postulate 5.

We have now succeeded in finding the rules for measurement probabilities and averages, post-measurement collapse, and time evolution for a mixed state in terms of the density matrix operator $\rho$, in eqs. (3.5.9), (3.5.10), (3.5.19), (3.5.21), and (3.5.27). Since each of these results only depends on $\rho$, and not on the individual ensemble states or probabilities, we have justified the assertion that all of the physical properties of the mixed ensemble or mixed state are contained in the density operator.

A common situation is that a measurement of $A$ has been made, but we do not know the result. Perhaps we were not looking, or we lost the data, or the measurement was made by something or someone else who is unwilling or unable to communicate with us. In that case, the density operator describing the mixed state after the measurement will be

$$\rho' \quad = \quad \sum_\alpha \mathcal{P}(\alpha)\, \rho_\alpha \;=\; \sum_\alpha P_\alpha \rho P_\alpha. \tag{3.5.28}$$

We can now show that the entropy always increases when this occurs, except in the trivial special case $\rho' = \rho$. To do so, we will use the following fact from linear algebra about traces of functions of matrices, due to Oskar Klein, the proof of which is omitted.

**Theorem 3.5.1 (Klein's inequality, general)** *Suppose that the function $f(x)$ is differentiable and strictly convex ($f''(x) > 0$) for all $0 < x < \infty$, and that $A$ and $B$ are Hermitian matrices with non-negative eigenvalues. Then*

$$\text{Tr}[f(A) - f(B) + (B - A)f'(B)] \;\geq\; 0, \tag{3.5.29}$$

*with equality if and only if $A = B$.*

We now apply this to the case $f(x) = x \ln x$, and let $A = \rho$ and $B = \rho'$ be any two density operators; we are not yet assuming the special form of eq. (3.5.28). Then, using $\text{Tr}[\rho] = \text{Tr}[\rho'] = 1$, we obtain:

**Theorem 3.5.2 (Klein's inequality for density operators)** *Suppose that $\rho$ and $\rho'$ are any two density operators on a common state space. Then*

$$\mathrm{Tr}[\rho(\ln\rho - \ln\rho')] \;\geq\; 0, \tag{3.5.30}$$

*with equality if and only if $\rho = \rho'$.*

Now consider the case of $\rho'$ given by eq. (3.5.28), which arose from having made a measurement of $A$ on a mixed state described by $\rho$. The entropy after the measurement is

$$\sigma' = -\mathrm{Tr}[\rho' \ln\rho'] = -\sum_\alpha \mathrm{Tr}[P_\alpha\, \rho\, P_\alpha \ln\rho'] = -\sum_\alpha \mathrm{Tr}[\rho\, P_\alpha \ln\rho' P_\alpha], \tag{3.5.31}$$

where the cyclic property of the trace was used at the end. Since $P_\alpha^2 = 1$, eq. (3.5.28) shows that $P_\alpha$ commutes with $\rho'$, which implies that $P_\alpha$ also commutes with $\ln\rho'$. Therefore, $P_\alpha \ln\rho' P_\alpha = P_\alpha^2 \ln\rho' = P_\alpha \ln\rho'$. Then, using completeness in the form $\sum_\alpha P_\alpha = I$, we obtain

$$\sigma' = -\mathrm{Tr}[\rho \ln\rho']. \tag{3.5.32}$$

Using $\sigma = -\mathrm{Tr}[\rho\ln\rho]$ and Klein's inequality (3.5.30), we finally obtain the claimed result,

$$\sigma' \;\geq\; \sigma. \tag{3.5.33}$$

The entropy increases whenever a non-trivial measurement is made but the result is unknown.

Note that this increase of the von Neumann entropy does not apply to a situation in which we made a measurement on a single mixed state and the result is known. In fact, if the known result of the measurement $\alpha$ is a non-degenerate eigenvalue, then the resulting density operator describing the system will be that of a pure state, with vanishing entropy.

As a very important example, consider an ensemble consisting of a bottle of, say, $\sim 10^{24}$ gas molecules, each of which can be in states characterized by energy eigenvalues $E$ and degeneracy labels $u_E$. The molecules interact with each other and with the bottle walls, but weakly enough that they can be considered an ensemble of independent quantum states of the same type. Intuitively, each interaction can be thought of as a sort of external measurement on the gas molecule, but the results of these measurements remain unknown, so that eq. (3.5.33) applies so as to make the entropy as large as it can be, subject to the constraint of energy conservation. Thus, when the molecules reach thermal equilibrium, the ensemble probability to find one of them in a particular orthobasis state $|E, u_E\rangle$ can be determined by the statistical principle that the entropy should be maximized, subject to the constraint that the ensemble average energy has a fixed value $\overline{E}$.

To see the implications of this, we write the density operator in the form

$$\rho = \sum_E \sum_{u_E} p_E \, |E, u_E\rangle\langle E, u_E| \, . \tag{3.5.34}$$

Here, we have already implemented the idea that maximizing the entropy will require the density operator to correspond to complete randomization within each subspace of fixed energy eigenvalue $E$, but the relative probabilities $p_E$ for each energy level remain to be found. Equation (3.5.34) gives

$$\sigma = -\sum_E g_E \, p_E \ln p_E, \tag{3.5.35}$$

where $g_E$ is the degeneracy of the energy eigenvalue $E$. Now, to maximize $\sigma$ subject to the constraints $1 = \text{Tr}[\rho] = \sum_E g_E \, p_E$ and fixed $\overline{E} = \sum_E g_E \, p_E E$, it is simplest to use the method of Lagrange multipliers. The function to be extremized is

$$f(p_E, \alpha, \beta) \;=\; -\sum_E g_E \, p_E \ln p_E \;+\; \alpha\Bigl(1 - \sum_E g_E \, p_E\Bigr) \;+\; \beta\Bigl(\overline{E} - \sum_E g_E \, p_E E\Bigr), \tag{3.5.36}$$

where $\alpha$ and $\beta$ are the Lagrange multipliers for the trace constraint and the energy constraint, respectively, and $\overline{E}$ is fixed. We then obtain, for each $E$,

$$0 \;=\; \frac{\partial f}{\partial p_E} \;=\; -g_E(\ln p_E + 1) - \alpha g_E - \beta g_E E, \tag{3.5.37}$$

which has the solution

$$p_E \;=\; e^{-(\beta E + \alpha + 1)}. \tag{3.5.38}$$

The $e^{-(\alpha+1)}$ factor is independent of $E$, and so can be absorbed into a common normalization factor; the important point is that we have derived that the canonical ensemble probabilities that maximize the entropy must be proportional to the **Boltzmann factor**,

$$p_E \;\propto\; e^{-\beta E}. \tag{3.5.39}$$

The Lagrange multiplier $\beta$ is related to temperature by the definition

$$\beta \;=\; \frac{1}{k_B T}. \tag{3.5.40}$$

Since $\beta$ has units of 1/energy, this definition shows that Boltzmann's constant is really just a conversion factor between energy and temperature. If we agreed to measure temperature in units of energy, then Boltzmann's constant would be 1.

The result of eq. (3.5.34) with $p_E \propto e^{-\beta E}$ is called the **canonical ensemble**, with density operator

$$\rho \;=\; \frac{1}{Z} \sum_E \sum_{u_E} e^{-\beta E} \, |E, u_E\rangle\langle E, u_E| \,, \tag{3.5.41}$$

where the normalization factor $Z$ is called the **partition function**. Thus, the canonical ensemble is completely randomized at each fixed energy level $E$, but with relative probabilities between energy levels determined by the Boltzmann factor. The partition function can be computed using the requirement $\mathrm{Tr}[\rho] = 1$, which gives

$$Z \;=\; \sum_E \sum_{u_E} e^{-\beta E} \;=\; \sum_E g_E \, e^{-\beta E}. \tag{3.5.42}$$

Equation (3.5.41) can be recognized as the spectral decomposition form [see eq. (2.7.5)] of

$$\rho \;=\; \frac{1}{Z} e^{-\beta H}. \tag{3.5.43}$$

This obviously commutes with $H$, so according to eq. (3.5.19), $\rho$ is constant in time. For any observable $A$ defined for each molecule, the canonical ensemble average is

$$\overline{A} \;=\; \frac{1}{Z} \mathrm{Tr}\left[e^{-\beta H} A\right] \;=\; \frac{1}{Z} \sum_E \sum_{u_E} e^{-\beta E} \, \langle E, u_E | A | E, u_E \rangle \,. \tag{3.5.44}$$

In particular, the average energy for states in the canonical ensemble is

$$\overline{E} \;=\; \frac{1}{Z} \sum_E g_E \, E \, e^{-\beta E} \;=\; -\frac{\partial}{\partial \beta} \ln Z. \tag{3.5.45}$$

It is left as an exercise to check that, with the entropy definition $S = k_B \sigma$,

$$\overline{E} - TS \;=\; -\beta \ln Z \;=\; F, \tag{3.5.46}$$

where $F$ is called the **Helmholtz free energy** after Hermann von Helmholtz. Let us stop our discussion of the canonical ensemble here, before this book accidentally turns into one on statistical mechanics.

In all of the preceding, we have assumed for notational simplicity that the states $|\psi_I\rangle$ appearing in the ensemble are discrete and countable. As usual, one can also consider a continuum of states, which entails turning summations into integrals. If the ensemble states are labeled by some continuous parameter $q$, then the density operator is

$$\rho \;=\; \int dq \, p(q) \; |\psi_q\rangle\langle\psi_q| \,, \tag{3.5.47}$$

where $p(q)$ is a probability density subject to the constraint $\int dq \, p(q) = 1$, but otherwise arbitrary.

# 4 Canonical variables and the Hamiltonian

## 4.1 Quantum observables from classical dynamics

The postulates of quantum mechanics refer to observable operators with time evolution governed by a Hamiltonian, but do not specifically identify these objects. This is intentional, in order to be general. In many physical situations, one can determine the appropriate Hamiltonian and observables by considering a classical version, but this is not always true. Indeed, one should think of classical physics as an approximate limit obtained from quantum mechanics, and not the reverse. Still, in favorable circumstances the classical properties of a system can be used to infer the basic observable operators, including the Hamiltonian, and their algebraic commutator properties in the quantum theory.

In the Lagrangian formulation of classical mechanics, one starts with some dynamical variables $q_n$, often called generalized coordinates, which we will label by an index $n$. The Lagrangian is a function of the $q_n$ and their first time derivatives $\dot{q}_n = dq_n/dt$,

$$L(q_n, \dot{q}_n, t). \tag{4.1.1}$$

The classical equations of motion for the system are then

$$\frac{\partial L}{\partial q_n} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_n}, \tag{4.1.2}$$

for each $n$. A short calculation, found in any good classical physics textbook, shows that this follows from a variational principle involving the action obtained by integrating the Lagrangian with respect to time. However, we will postpone our own discussion of that until section 22.4. Our reason for doing so is that rather than accept the variational principle as a postulate of classical mechanics, we will be able to *derive* it as a consequence of the path integral formulation of quantum mechanics.

The Hamiltonian formulation of classical mechanics recasts the Lagrangian formulation by defining a canonical momentum conjugate to each generalized coordinate,

$$p_n = \frac{\partial L}{\partial \dot{q}_n}. \tag{4.1.3}$$

Now one defines the Hamiltonian as

$$H(q_n, p_n, t) = \sum_n p_n \dot{q}_n - L(q_n, \dot{q}_n, t), \tag{4.1.4}$$

where it is important that the $\dot{q}_n$ are to be completely eliminated in favor of the $p_n$ using eq. (4.1.3). This implies that $H$ is a function only of the generalized coordinates and their canonical conjugate momenta, and not their time derivatives. The $q_n$ and $p_n$ are collectively called the **phase-space coordinates**. As shown in the same good textbook on classical mechanics, the Lagrangian equations of motion eq. (4.1.2) are equivalent to the Hamiltonian equations of motion,

$$\dot{q}_n = \frac{\partial H}{\partial p_n}, \qquad \dot{p}_n = -\frac{\partial H}{\partial q_n}. \tag{4.1.5}$$

The generalized coordinates $q_n$ need not be the rectangular coordinates of a particle, but rather could be any quantities that fully specify the classical state of the system at a given time. Likewise, the canonical momenta need not coincide with mechanical momenta (equal to the product of mass and velocity for particles).

Consider any quantity $a(q_n, p_n, t)$, built out of the generalized coordinates and their conjugate momenta. According to the same good classical physics textbook, the time evolution of $a$ is

$$\frac{da}{dt} = \big\{a, H\big\}_{\text{PB}} + \frac{\partial a}{\partial t}, \tag{4.1.6}$$

where the **Poisson bracket** for any two functions on phase space $a$ and $b$ is defined as

$$\big\{a, b\big\}_{\text{PB}} \equiv \sum_n \left( \frac{\partial a}{\partial q_n}\frac{\partial b}{\partial p_n} - \frac{\partial b}{\partial q_n}\frac{\partial a}{\partial p_n} \right). \tag{4.1.7}$$

Dirac noted that the Poisson brackets $\big\{a, b\big\}_{\text{PB}}$ of classical mechanics are closely analogous to the commutators $[A, B]$ for the corresponding observables in the quantum theory. Both are antisymmetric under interchange of the observables, and at least for the position and momentum observables, one has the exact correspondence

$$\begin{array}{ccc} \underline{\text{classical}} & & \underline{\text{quantum}} \\ \big\{q_n, p_k\big\}_{\text{PB}} = \delta_{nk} & \longleftrightarrow & \big[Q_n, P_k\big] = i\hbar\delta_{nk} \end{array} \tag{4.1.8}$$

Commutators obtained in this way are called **canonical commutation relations**. Furthermore, eq. (4.1.6) has a striking resemblance to Ehrenfest's Theorem in quantum mechanics, which we found in eq. (3.4.13). Indeed, one finds from the latter equation that

$$\frac{d}{dt}\langle Q_n\rangle = \Big\langle \frac{\partial H}{\partial P_n} \Big\rangle \qquad \frac{d}{dt}\langle P_n\rangle = -\Big\langle \frac{\partial H}{\partial Q_n} \Big\rangle, \tag{4.1.9}$$

directly analogous to the Hamiltonian equations of motion (4.1.5).

For a single particle of mass $m$ moving in three dimensions in a potential $V$, it is natural to choose the $Q_n$ to be the usual rectangular coordinates $X = R_x$, $Y = R_y$, and $Z = R_z$, and their conjugate **canonical momenta** $P_x$, $P_y$, $P_z$, defined to satisfy the commutation relations already given in eq. (2.8.54),

$$[R_a, P_b] = i\hbar\delta_{ab}, \qquad [R_a, R_b] = 0, \qquad [P_a, P_b] = 0, \tag{4.1.10}$$

for $a, b = x, y, z$. The Hamiltonian operator is then

$$H = \frac{P^2}{2m} + V(\vec{R}), \tag{4.1.11}$$

where $P^2 \equiv \vec{P} \cdot \vec{P}$.

The connection between classical and quantum physics just described is often called the **correspondence principle**, but it has some weaknesses. First, there are some quantum observables that do not have a classical counterpart at all, for example spin, also known as intrinsic angular momentum. Second, there is no guarantee that every generalized coordinate and its canonical momentum will obey a canonical commutation relation. Although it is valid for the rectangular coordinates of a particle, or a collection of particles, in more complicated cases one might encounter higher order corrections in $\hbar$, or ambiguities in connecting the classical observables to the quantum ones. This is why we preferred to derive the position-momentum commutation relations by the method given in section 2.8. More generally, the most logical (but perhaps not the simplest) way to draw the connection is to derive the classical theory as an approximation to the quantum theory, as we will do using the Feynman sum-over-paths approach in section 22.4.

## 4.2 The two-body problem

An important special case is that of two particles that are free except for a potential energy of interaction that depends only on their separation. This occurs, for example, in the hydrogen atom to be treated in Chapter 10, where the two particles are the electron and the (much heavier) proton. Another example is neutron-proton scattering, to be studied in section 20.8, where the masses are almost the same.

In general, the two-body Hamiltonian has the form

$$H = \frac{P_1^2}{2m_1} + \frac{P_2^2}{2m_2} + V(\vec{R}_1 - \vec{R}_2), \tag{4.2.1}$$

where we have allowed for the possibility that the potential energy depends on the vector displacement (not just its magnitude), and the two particles have masses $m_1$ and $m_2$ and canonical position and momentum operators $(\vec{R}_1, \vec{P}_1)$ and $(\vec{R}_2, \vec{P}_2)$. The components of these observables satisfy commutation relations $[R_{1a}, P_{1b}] = i\hbar\delta_{ab}$ and $[R_{2a}, P_{2b}] = i\hbar\delta_{ab}$, for $a, b = x, y, z$, with other combinations vanishing. In particular, each of the observables for particle 1 commutes with those of particle 2. As an orthobasis, one can choose a tensor product of the eigenkets of $\vec{R}_1$ and $\vec{R}_2$,

$$|\vec{r}_1, \vec{r}_2\rangle = |\vec{r}_1\rangle \otimes |\vec{r}_2\rangle\,, \tag{4.2.2}$$

defined to satisfy the eigenvalue equations

$$\vec{R}_1 \left|\vec{r}_1, \vec{r}_2\right\rangle = \vec{r}_1 \left|\vec{r}_1, \vec{r}_2\right\rangle, \qquad \vec{R}_2 \left|\vec{r}_1, \vec{r}_2\right\rangle = \vec{r}_2 \left|\vec{r}_1, \vec{r}_2\right\rangle. \tag{4.2.3}$$

However, the solution of the Hamiltonian eigenvalue problem is complicated by the fact that the potential couples the two particle degrees of freedom together.

Fortunately, as in classical mechanics, such cases can be reduced to a simpler problem that is effectively the same as for two decoupled particles, by making a change of variables to separate the relative motion from that of the center of mass. The relative motion is described by

$$\vec{R} = \vec{R}_1 - \vec{R}_2, \qquad \vec{P} = \frac{m_2 \vec{P}_1 - m_1 \vec{P}_2}{m_1 + m_2}, \tag{4.2.4}$$

and the motion of the center of mass by

$$\vec{R}_{\text{cm}} = \frac{m_1 \vec{R}_1 + m_2 \vec{R}_2}{m_1 + m_2}, \qquad \vec{P}_{\text{tot}} = \vec{P}_1 + \vec{P}_2. \tag{4.2.5}$$

It is a short exercise to show that the pairs $(\vec{R}, \vec{P})$ and $(\vec{R}_{\text{cm}}, \vec{P}_{\text{tot}})$ each satisfy canonical commutation relations, and do not interfere with each other. Furthermore, if one defines the total mass $M$ and the **reduced mass** $\mu$ according to

$$M = m_1 + m_2, \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}, \tag{4.2.6}$$

then the Hamiltonian eq. (4.2.1) can be rewritten in the decoupled form

$$H = H_{\text{cm}} + H_{\text{rel}} \tag{4.2.7}$$

where

$$H_{\text{cm}} = \frac{P_{\text{tot}}^2}{2M}, \qquad H_{\text{rel}} = \frac{P^2}{2\mu} + V(\vec{R}). \tag{4.2.8}$$

The center-of-mass degrees of freedom have the same Hamiltonian as that of a completely free particle with mass $M$, whose eigenvalue problem is easy to solve (plane waves). The dynamics of $H_{\text{rel}}$ are the same as for a single particle with mass equal to the reduced mass $\mu$, moving in the potential $V(\vec{R})$. One can now rewrite the orthobasis eq. (4.2.2) as the tensor product of eigenkets of $\vec{R}$ and $\vec{R}_{\text{cm}}$,

$$\left|\vec{r}, \vec{r}_{\text{cm}}\right\rangle = \left|\vec{r}\right\rangle \otimes \left|\vec{r}_{\text{cm}}\right\rangle, \tag{4.2.9}$$

and look for stationary-state wavefunction solutions of the form

$$\Psi(\vec{r}, \vec{r}_{\text{cm}}) = \langle \vec{r}, \vec{r}_{\text{cm}} | \Psi \rangle = \frac{1}{(2\pi\hbar)^{3/2}} e^{i\vec{k}_{\text{tot}} \cdot \vec{r}_{\text{cm}}} \psi(\vec{r}), \tag{4.2.10}$$

where $\vec{p}_{\rm tot} = \hbar\vec{k}_{\rm tot}$ is the eigenvalue of $\vec{P}_{\rm tot}$, and the relative coordinate wavefunction satisfies

$$\left[-\frac{\hbar^2\nabla^2}{2\mu} + V(\vec{r}) - E\right]\psi(\vec{r}) \;=\; 0, \tag{4.2.11}$$

and the total energy eigenvalue is $E + \hbar^2 k_{\rm tot}^2/2M$. The second contribution to the energy can be treated as a constant by noting that the center of mass of the particles acts like a free particle. We can then solve the eigenvalue problem in eq. (4.2.11) for $E$ and $\psi(\vec{r})$ as if it were a single particle. If one of the particles is much heavier than the other, as in the case of electrons compared to atomic nuclei, then $\mu$ is equal to the mass of the lighter particle, to a good approximation.

If the number of particles $N$ is three or more, and fixed, then one can construct appropriate **Jacobi coordinates**, and their canonical momenta, by iteration. First one chooses two of the particles and defines their relative and center-of-mass coordinates and momenta, just as in the preceding. The two-particle center-of-mass coordinates and momenta are then combined in the same way with those of a third particle, to give another relative coordinate/momentum pair and a three-particle center-of-mass coordinate and momentum. The three-particle center-of-mass coordinate and momentum are combined with those of a fourth particle, etc. In the end one will have a single center-of-mass coordinate and momentum for the whole system with total mass $M = \sum_{i=1}^N m_i$,

$$\vec{R}_{\rm cm} \;=\; \frac{1}{M}\sum_{i=1}^N m_i \vec{R}_i, \qquad\qquad \vec{P}_{\rm tot} \;=\; \sum_{i=1}^N \vec{P}_i, \tag{4.2.12}$$

and $N-1$ translationally invariant coordinate/momentum pairs, each satisfying canonical commutation relations. The kinetic energy terms for these momenta are all decoupled from each other, but with the complication that they have different effective masses even if the particle masses $m_i$ are all the same. If there are no external forces, so that the potential energy is translationally invariant, then $\vec{R}_{\rm cm}$ will not appear in the Hamiltonian at all. The energy eigenstates will therefore have the form of a tensor product, of plane-wave free particle wavefunctions that are eigenstates of $\vec{P}_{\rm tot}$, and eigenstates of the remaining, more complicated, part of the Hamiltonian. In the case of multi-electron atoms, where one of the particles is a nucleus that is much more massive than the electrons, it is much more common to make the simple and good approximation that the nucleus is infinitely massive and fixed at the origin in the center-of-mass frame, and the remaining coordinates and momenta are simply those of the individual electrons.

## 4.3 Charged particle in external electromagnetic fields

We now turn our attention to the dynamics of charged particles in external electromagnetic fields. We will follow the example of most quantum mechanics books by using Gaussian cgs unit normalizations for electrodynamics quantities, rather than the SI units that you may be more familiar with. This means that Maxwell's equations for the electric and magnetic fields are (with the SI versions indicated parenthetically, for comparison)

$$\vec{\nabla}\cdot\vec{E} = 4\pi\rho \qquad (\text{SI:}\quad \rho/\epsilon_0), \tag{4.3.1}$$

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad (\text{SI:}\quad 0), \tag{4.3.2}$$

$$\vec{\nabla}\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t} \qquad (\text{SI:}\quad -\frac{\partial\vec{B}}{\partial t}), \tag{4.3.3}$$

$$\vec{\nabla}\times\vec{B} = \frac{1}{c}\frac{\partial\vec{E}}{\partial t} + \frac{4\pi}{c}\vec{j} \qquad (\text{SI:}\quad \frac{1}{c^2}\frac{\partial\vec{E}}{\partial t} + \mu_0\vec{j}\,), \tag{4.3.4}$$

which imply that the charge density $\rho$ and the current density $\vec{j}$ obey local charge conservation,

$$\vec{\nabla}\cdot\vec{j} + \frac{\partial\rho}{\partial t} = 0. \tag{4.3.5}$$

The electromagnetic energy density and Poynting vector (power per unit area) are

$$u_{\text{EM}} = \frac{1}{8\pi}(E^2 + B^2) \qquad (\text{SI:}\quad \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2), \tag{4.3.6}$$

$$\vec{S} = \frac{c}{4\pi}\vec{E}\times\vec{B} \qquad (\text{SI:}\quad \frac{1}{\mu_0}\vec{E}\times\vec{B}). \tag{4.3.7}$$

The electromagnetic fields are obtained as derivatives of the scalar and vector potentials,

$$\vec{E} = -\vec{\nabla}\Phi - \frac{1}{c}\frac{\partial\vec{A}}{\partial t}, \tag{4.3.8}$$

$$\vec{B} = \vec{\nabla}\times\vec{A}. \tag{4.3.9}$$

It follows that $\vec{E}$ and $\vec{B}$ remain unchanged if one makes a simultaneous change in $\Phi$ and $\vec{A}$, called a **gauge transformation**,

$$\Phi \rightarrow \Phi - \frac{1}{c}\frac{\partial\Lambda}{\partial t}, \qquad \vec{A} \rightarrow \vec{A} + \vec{\nabla}\Lambda. \tag{4.3.10}$$

Here $\Lambda(\vec{r}, t)$ is an arbitrary function of position and time.

In classical electrodynamics, the Lagrangian for a non-relativistic particle with mass $m$ and charge[†] $q$ and position $\vec{r}(t)$, moving in the potentials $\Phi$ and $\vec{A}$, is

$$L = \frac{1}{2}m\left(\frac{d\vec{r}}{dt}\right)^2 + \frac{q}{c}\frac{d\vec{r}}{dt}\cdot\vec{A}(\vec{r}, t) - q\Phi(\vec{r}, t). \tag{4.3.11}$$

---

[†] The convention in this book is that the electric charge for a particle is given by $q = Qe$, where $Q$ is a dimensionless number, while $e$ is the proton's charge, numerically given in Gaussian cgs units by eq. (1.1.2), and therefore always positive. (Some other sources take $e$ to be negative when referring to the electron.) Thus, for the electron, $Q = -1$ and $q = -e$, and for the proton, $Q = 1$ and $q = e$. All known particles have $Q$ equal to integer multiples of 1/3, and the ones unconfined by the strong nuclear force have integer $Q$. For example $Q = 2/3$ for up, charm, and top quarks, and $Q = -1/3$ for down, strange, and bottom quarks, and $Q = -1$ for the electron, muon, and tau lepton. Within the assumed structure of the Standard Model of particle physics, this remarkable charge quantization can be understood as a requirement of anomaly cancellation, a consistency constraint on quantum field theories with gauge interactions. Grand Unified Theories (based on non-Abelian gauge groups like $SU(5)$, $SO(10)$, or $E_6$) go further, elegantly explaining why all particles, known or unknown, must have integer values of $3Q$, but it is not presently known if these theories are correct.

The equation of motion resulting from applying eq. (4.1.2) is the Lorentz force law,

$$m\frac{d^2\vec{r}}{dt^2} \;=\; q\left(\vec{E} + \frac{1}{c}\frac{d\vec{r}}{dt} \times \vec{B}\right). \tag{4.3.12}$$

The motion of the particle is thus determined only by $\vec{E}$ and $\vec{B}$. This illustrates that, even though the Lagrangian is written in terms of the potentials $\Phi(\vec{r}, t)$ and $\vec{A}(\vec{r}, t)$, they also contain unphysical information, since the physics is equally well described by the potentials modified by the gauge transformation in eq. (4.3.10). Unlike $\vec{E}$ and $\vec{B}$, which have unambiguous physical meaning, quantities like $\Phi$ and $\vec{A}$ that have gauge dependence are useful book-keeping tools, but cannot be physical observables.

Applying the procedure in section 4.1 to eq. (4.3.11), the classical canonical momentum conjugate to $\vec{r}$ is

$$\vec{p} \;=\; m\frac{d\vec{r}}{dt} + \frac{q}{c}\vec{A}, \tag{4.3.13}$$

and the classical Hamiltonian is

$$H \;=\; \frac{1}{2m}\left[\vec{p} - \frac{q}{c}\vec{A}(\vec{r}, t)\right]^2 + q\Phi(\vec{r}, t). \tag{4.3.14}$$

Like the Lagrangian, the Hamiltonian and canonical momentum are not written directly in terms of the fields $\vec{E}$ and $\vec{B}$, but rather in terms of the potentials $\Phi$ and $\vec{A}$, even though the latter are gauge-dependent. One must be careful to distinguish the canonical momentum $\vec{p}$ of the particle from the **kinetic momentum** (also known as **mechanical momentum**), defined as the product of mass and velocity,

$$\vec{\pi} \;\equiv\; m\frac{d\vec{r}}{dt} \;=\; \vec{p} - \frac{q}{c}\vec{A}. \tag{4.3.15}$$

The kinetic momentum $\vec{\pi}$ is a gauge-invariant observable, since $d\vec{r}/dt$ does not depend on the choice of gauge. In contrast, the canonical momentum $\vec{p}$ is not a gauge-invariant observable, due to the appearance of $\vec{A}$ in eq. (4.3.13).

In quantum mechanics, we promote the classical position $\vec{r}$ and canonical momentum $\vec{p}$ to operators, and thus the electromagnetic potentials become operators $\Phi(\vec{R}, t)$ and $\vec{A}(\vec{R}, t)$ that are functions of $\vec{R}$. So, naively, the quantum Hamiltonian should be

$$H \;=\; \frac{1}{2m}\left[\vec{P} - \frac{q}{c}\vec{A}(\vec{R}, t)\right]^2 + q\Phi(\vec{R}, t), \tag{4.3.16}$$

where $\vec{R}$ and $\vec{P}$ satisfy the usual canonical commutation relations of eq. (2.8.54). This implies that in the position representation, these canonical operators are represented by

$$\vec{R} \;\leftrightarrow\; \vec{r}, \qquad\qquad \vec{P} \;\leftrightarrow\; -i\hbar\vec{\nabla}, \tag{4.3.17}$$

as in Table 2.8.2. Therefore, one must be careful with the ordering of $\vec{P}$ and $\vec{A}$, so that $[\vec{P} - \frac{q}{c}\vec{A}(\vec{R}, t)]^2$ is interpreted as the symmetrized form $P^2 - \frac{q}{c}(\vec{P} \cdot \vec{A} + \vec{A} \cdot \vec{P}) + \frac{q^2}{c^2}A^2$, in order that $H$ is Hermitian. However, eq. (4.3.16) is still not complete, because it does not include the (purely quantum!) effect of intrinsic angular momentum, or spin.

We will discuss spin more thoroughly in 8.2, but for the present discussion we only need to know that the spin for a particle is an observable vector operator $\vec{S}$. The intrinsic magnetic moment of a particle is always proportional to its spin (since, in the rest frame of the particle, there is no other special direction in which it could point):

$$\vec{\mu} = \gamma\vec{S}. \tag{4.3.18}$$

The constant of proportionality $\gamma$ is a property of the particle type, called the **gyromagnetic ratio**. For the electron, the gyromagnetic ratio is often written as

$$\gamma_e = -\frac{g_e e}{2m_e c}, \tag{4.3.19}$$

where the dimensionless quantity $g_e$ is called the $g$-factor. The Dirac equation of relativistic quantum mechanics predicts $g_e = 2$, but there are small corrections to this coming from the quantum field theory of relativistic quantum electrodynamics (QED). It has been predicted very precisely in perturbation theory in QED, and measured experimentally with comparable accuracy, with the results

$$g_e = 2.00231930436321(46) \quad \text{(QED prediction, 5th-order perturbation theory)}, \tag{4.3.20}$$

$$g_e = 2.00231930436182(52) \quad \text{(experiment)}, \tag{4.3.21}$$

a famous agreement of better than 12 significant digits between theory and experiment. The quantity $(g_e - 2)/2$ is called the **anomalous magnetic moment** of the electron. In this book, we will simply use the approximation $g_e = 2$.

For the proton and the neutron, the gyromagnetic ratios are often written as

$$\gamma_p = \frac{g_p e}{2m_p c}, \tag{4.3.22}$$

$$\gamma_n = \frac{g_n e}{2m_p c}, \tag{4.3.23}$$

which again define dimensionless $g$-factors. Note that the neutron has a magnetic moment, even though it has no net charge, because it is a composite particle with charged constituents (quarks). The conventional definition of $g_n$ for the neutron in eq. (4.3.23) uses the proton's charge and mass. The nucleon $g$-factors are also very accurately known experimentally:

$$g_p = 5.5856946893(16), \tag{4.3.24}$$

$$g_n = -3.82608545(90), \tag{4.3.25}$$

but the theoretical predictions of these quantities are not nearly as accurate. The reason is that, unlike the electron, the proton and neutron are complicated composite particles made up of quarks and gluons (and virtual quark/antiquark pairs) held together by the strong nuclear force, (quantum chromodynamics, or QCD) for which perturbation theory does not converge, and non-perturbative methods are highly advanced but limited by finite computing power.

For electrons, protons, and neutrons, the magnitude of $\vec{S}$ is always the same, $\hbar/2$. The magnitude of the electron's intrinsic magnetic moment is therefore within about 0.1% of the **Bohr magneton**,

$$\mu_B \;=\; \frac{e\hbar}{2m_e c} = 5.78838 \times 10^{-5}\frac{\text{eV}}{\text{Tesla}} = 9.27401 \times 10^{-21}\frac{\text{ergs}}{\text{gauss}}, \tag{4.3.26}$$

and the proton and neutron have magnetic moment magnitudes equal to about 2.79 and 1.91 times the **nuclear magneton**,

$$\mu_N \;=\; \frac{e\hbar}{2m_p c} = 3.15245 \times 10^{-8}\frac{\text{eV}}{\text{Tesla}} = 5.05078 \times 10^{-24}\frac{\text{ergs}}{\text{gauss}}. \tag{4.3.27}$$

Because $\mu_B/\mu_N = m_p/m_e \approx 1836$ is a large number, electronic and atomic magnetic moments are typically 3 orders of magnitude larger than nuclear magnetic moments.

The classical energy of interaction of a magnetic moment $\vec{\mu}$ with an external magnetic field is $-\vec{\mu}\cdot\vec{B}$. So, we add this to the quantum Hamiltonian for a non-relativistic particle, to get

$$H \;=\; \frac{1}{2m}\left[\vec{P} - \frac{q}{c}\vec{A}(\vec{R},t)\right]^2 + q\Phi(\vec{R},t) - \gamma\vec{S}\cdot\vec{B}, \tag{4.3.28}$$

where $q$ is the electric charge and $\gamma$ is the appropriate gyromagnetic ratio for the particle. Even this Hamiltonian is not complete, for it does not include relativistic effects suppressed by further powers of $1/c$. These will be discussed when needed for the fine and hyperfine contributions to the hydrogen atom, in sections 15.1 and 15.2.

If the Hamiltonian eq. (4.3.28) is written in terms of the kinetic momentum operator

$$\vec{\Pi} = \vec{P} - \frac{q}{c}\vec{A}, \tag{4.3.29}$$

it will appear simpler, since it then does not depend on the vector potential $\vec{A}$,

$$H \;=\; \frac{1}{2m}\Pi^2 + q\Phi(\vec{R},t) - \gamma\vec{S}\cdot\vec{B}. \tag{4.3.30}$$

However, it is important to recognize that the kinetic momentum does not satisfy canonical commutation relations. While one still has

$$[R_a, \Pi_b] \;=\; i\hbar\delta_{ab}, \qquad\qquad (a, b = x, y, z), \tag{4.3.31}$$

one can show from its definition[‡] that $[\Pi_a, \Pi_b] = i\frac{q\hbar}{c}(\nabla_a A_b - \nabla_b A_a)$. This can be rewritten directly in terms of the magnetic field as

$$[\Pi_x, \Pi_y] = i\frac{q\hbar}{c}B_z, \qquad [\Pi_y, \Pi_z] = i\frac{q\hbar}{c}B_x, \qquad [\Pi_z, \Pi_x] = i\frac{q\hbar}{c}B_y. \tag{4.3.32}$$

This should be contrasted with the canonical commutator $[P_a, P_b] = 0$. If one chooses to write the Hamiltonian in terms of the kinetic momentum, the simplicity comes with a cost; the information about the vector potential and the magnetic field is hidden in the commutation relations for $\vec{\Pi}$. Since $\vec{\Pi}$ is the product of mass and velocity for the particle, we see that in the presence of a magnetic field one cannot simultaneously specify the three components of the velocity of a charged particle, because they are not compatible observables.

The freedom to make gauge transformations as in eq. (4.3.10) always allows us to select **Coulomb gauge** (also known as **transverse gauge**), defined by

$$\vec{\nabla} \cdot \vec{A} \; = \; 0. \tag{4.3.33}$$

Since $\vec{P}$ is given in the position representation by $-i\hbar\vec{\nabla}$ acting on everything to its right, the Coulomb gauge condition implies

$$\vec{P} \cdot \vec{A} \;\; = \;\; \vec{A} \cdot \vec{P}, \tag{4.3.34}$$

with the consequent advantage that (unlike other gauge choices) there is no operator ordering issue with the cross-terms in the Hamiltonian, of the type mentioned after eq. (4.3.17). Thus, in Coulomb gauge, one is free to write

$$H \;\; = \;\; \frac{P^2}{2m} - \frac{q}{mc}\vec{A} \cdot \vec{P} + \frac{q^2}{2mc^2}A^2 + q\Phi - \gamma\vec{S} \cdot \vec{B}. \tag{4.3.35}$$

This form will be useful to us when we discuss absorption and emission of light, in Chapter 19.

---

[‡] Note that in the position representation, $\vec{P} \leftrightarrow -i\hbar\vec{\nabla}$, so that $\vec{\Pi} \leftrightarrow -i\hbar\vec{\nabla} - \frac{q}{c}\vec{A}(\vec{r})$.

# 5 Transformations, symmetries, and conservation laws

## 5.1 Continuous unitary transformations and symmetries

As we saw in section 2.5, a unitary operator can be viewed as implementing a change of orthobasis for the space of states. Certain unitary operators also have the interpretation of physical changes, or **transformations** on a system. Notable examples include translations, rotations, inversion of the coordinate system, and displacements in time.

If a transformation leaves the Hamiltonian unchanged, then we say, as a definition, that the transformation is a **symmetry** of the quantum system. In this section we will show, in a general way, that there is always a conserved quantity corresponding to each continuous symmetry. As special cases, we will see that the conserved quantities associated with time translation, spatial translation, and rotation symmetries are, respectively, energy, momentum, and angular momentum.

Consider a set of transformation operators

$$U(\alpha) \;=\; \exp\left(-i\alpha_a G_a\right), \tag{5.1.1}$$

where the $G_a$ are $N$ Hermitian operators, typically observables without explicit time dependence, called the **generators** of the transformations, and the $\alpha_a$ are $N$ real numbers that parameterize the transformations. Here, and in the following, repeated indices $a = 1, \ldots, N$ are implicitly summed over. Theorem 2.4.3 confirms that since the $G_a$ are Hermitian, $U(\alpha)$ is a unitary operator. Following the discussion surrounding eqs. (2.5.25)-(2.5.30), the transformations for the state ket and all observables $A$ are defined by

$$\begin{aligned} |\psi\rangle \;&\rightarrow\; |\psi'\rangle = U\,|\psi\rangle\,, && (5.1.2)\\ A \;&\rightarrow\; A' = UAU^\dagger, && (5.1.3)\end{aligned}$$

so that matrix elements are invariant under the transformation, because $U^\dagger U = I$. The inverse of the transformation parameterized by $\alpha_a$ is parameterized by $-\alpha_a$,

$$U(\alpha)^\dagger \;=\; U(\alpha)^{-1} \;=\; U(-\alpha). \tag{5.1.4}$$

A complete set of transformations have the closure property, which says that the combination of two transformations $\alpha_a$ and $\beta_a$ should always be another transformation, parameterized by some real numbers $\gamma_a$,

$$U(\beta)U(\alpha) \;=\; U(\gamma). \tag{5.1.5}$$

A continuous set of transformations obeying these properties has the structure of a **Lie group**, named after the mathematician Marius Sophus Lie.

Consider the following combination of unitary transformations that are infinitesimally close to the identity,

$$\begin{aligned} U(\epsilon)U(\delta)U(\epsilon)^\dagger U(\delta)^\dagger &= (I - i\epsilon_a G_a + \cdots)(I - i\delta_b G_b + \cdots)(I + i\epsilon_c G_c + \cdots)(I + i\delta_d G_d + \cdots) \\ &= I - \epsilon_a \delta_b [G_a, G_b] + \cdots , \end{aligned} \tag{5.1.6}$$

where terms of higher order in either $\epsilon_a$ or $\delta_b$ have been dropped. The closure property says that this must also be a unitary transformation infinitesimally close to the identity, so $[G_a, G_b]$ must be a linear combination of generators. Therefore,

$$[G_a, G_b] = i f_{abc} G_c \tag{5.1.7}$$

for some set of numbers $f_{abc}$, called the **structure constants** of the Lie group of transformations. Equation (5.1.7) is called the **Lie algebra** of the group. Since the commutator of any two Hermitian operators $[G_a, G_b]$ is anti-Hermitian, and the $G_c$ are Hermitian, the quantities $f_{abc}$ must all be real. If they are all 0, so that the generators all commute, then the group is said to be Abelian, otherwise it is non-Abelian. The mathematical study and classification of Lie groups is a rich and beautiful subject that we will not delve into further here.

For the remainder of this section, let us consider unitary transformations that are symmetries of a quantum system. According to our definition, this means that $H$ does not change, so eq. (5.1.3) gives

$$U(\alpha) H U(\alpha)^\dagger = H. \tag{5.1.8}$$

Taking the special case that the parameters $\alpha_a = \epsilon_a$ are infinitesimal, we have

$$(I - i\epsilon_a G_a + \cdots)\, H\, (I + i\epsilon_b G_b + \cdots) = H, \tag{5.1.9}$$

which implies $\epsilon_a [H, G_a] = 0$. Since this is supposed to hold for any $\epsilon_a$, we learn that

$$[H, G_a] = 0. \tag{5.1.10}$$

It follows from Theorem 2.7.1 that an orthobasis of energy eigenstates can also be chosen to be eigenstates of any subset of the symmetry generators $G_a$ that are mutually commuting (called a Cartan subalgebra by mathematicians, after Élie Cartan).

Consider an eigenstate $|g_a\rangle$ of one of the symmetry generators $G_a$, labeled by its eigenvalue $g_a$. Since the symmetry generators $G_a$ commute with the Hamiltonian, $|g_a\rangle$ will remain an eigenstate of $G_a$ at later times, because

$$G_a \left(e^{-itH/\hbar} |g_a\rangle\right) = e^{-itH/\hbar} G_a |g_a\rangle = g_a \left(e^{-itH/\hbar} |g_a\rangle\right). \tag{5.1.11}$$

This holds even if the Hamiltonian depends on time, provided that $[H(t), G_a] = 0$ for all $t$, as one can see by replacing the factors $e^{-itH/\hbar}$ by the time-evolution operator $U(t, t_0)$ from eq. (3.4.26).

Furthermore, the symmetry generators $G_a$ are **conserved quantities**, also known as **constants of the motion**. The meaning of this is given by the following result.

**Theorem 5.1.1 (Noether's principle)** *Suppose that an observable $G_a$ has no explicit time dependence, and that it commutes with the Hamiltonian, which may be time-dependent. Then, for an arbitrary state, the probability of each possible outcome $g_a$ for a measurement of $G_a$ is independent of time.*

Proof: The key point is that $[H(t), G_a] = 0$ implies that, for any particular time $t = t_0$, one can find an orthobasis of common eigenkets $|g_a, E, u\rangle$, defined such that $G_a |g_a, E, u\rangle = g_a |g_a, E, u\rangle$ and $H(t_0) |g_a, E, u\rangle = E |g_a, E, u\rangle$. Here $u = u_{g_a,E}$ are possible degeneracy labels for $g_a$ and $E$, with the subscripts dropped for typographical simplicity. These orthobasis kets are fixed, determined by the Hamiltonian at the time $t_0$. Now, Postulate 4 tells us that if the system is in a state $|\psi(t)\rangle$, then the probability of measuring $G_a$ at time $t$ and getting the result $g_a$ is

$$\mathcal{P}(g_a, t) \;=\; \sum_{E,u} \langle g_a, E, u|\psi(t)\rangle \langle \psi(t)|g_a, E, u\rangle . \tag{5.1.12}$$

Here $E$ and $u$ are summed over, as the degeneracy labels for $g_a$. Taking the time derivative, and then using the time-dependent Schrödinger equation $\frac{d}{dt}|\psi\rangle = -\frac{i}{\hbar}H(t)|\psi\rangle$ and its adjoint $\frac{d}{dt}\langle\psi| = \frac{i}{\hbar}\langle\psi| H(t)$, we get

$$\begin{aligned}\frac{d}{dt}\mathcal{P}(g_a, t) \;=\; -\frac{i}{\hbar}\sum_{E,u}\Big(&\langle g_a, E, u|H(t)|\psi(t)\rangle \langle \psi(t)|g_a, E, u\rangle \\ &- \langle g_a, E, u|\psi(t)\rangle \langle \psi(t)|H(t)|g_a, E, u\rangle\Big).\end{aligned} \tag{5.1.13}$$

Since $H(t_0)|g_a, E, u\rangle = E|g_a, E, u\rangle$ and $\langle g_a, E, u| H(t_0) = E\langle g_a, E, u|$, the two terms on the right side of eq. (5.1.13) simply cancel if we evaluate them at $t = t_0$. Thus, we obtain

$$\frac{d}{dt}\mathcal{P}(g_a, t)\Big|_{t=t_0} = 0. \tag{5.1.14}$$

This shows that, for an arbitrary state, $\mathcal{P}(g_a, t)$ has a vanishing time derivative at any given $t = t_0$, so it must be constant in time, as claimed.

The expectation value of an operator in a state is the sum of its measurement outcomes weighted by the probabilities; see eq. (3.3.1). Therefore, Theorem 5.1.1 immediately implies a weaker but still interesting and important result, that the expectation value of a symmetry generator $G_a$ in an arbitrary state does not depend on time:

$$\frac{d}{dt}\langle G_a\rangle = 0. \tag{5.1.15}$$

This also follows directly from Ehrenfest's Theorem eq. (3.4.13) with $G_a$ playing the role of $A$. Theorem 5.1.1 and eq. (5.1.15) are quantum mechanical versions of Emmy Noether's celebrated 1918 theorem in classical mechanics, that every continuous symmetry implies a conservation law; the observables $G_a$ that generate the symmetries of the Hamiltonian are conserved quantities.

The most straightforward special case is that of **time translation** symmetry, for which the generator is the Hamiltonian itself, as expressed by the Schrödinger equation (3.1.10) in Postulate 6. If the Hamiltonian does not depend explicitly on time, then the energy is a conserved quantity, since $[H, H] = 0$ is trivially satisfied. In that case, for any state the probability of measuring the energy to be any particular value is independent of time, and the expectation value of the energy does not change in time.

## 5.2 Translations

For a particle moving in one dimension with position operator $X$ and momentum operator $P$, consider the operator

$$T(a) \;=\; e^{-iaP/\hbar}, \tag{5.2.1}$$

where $a$ is a constant length. Because $P$ is Hermitian, Theorem 2.4.3 tells us that $T(a)$ is a unitary operator,

$$T(a)^\dagger \;=\; T(a)^{-1} \;=\; T(-a). \tag{5.2.2}$$

Using $[X, P] = i\hbar$, one can compute the commutator using Theorem 2.4.1,

$$[X,\, T(a)] \;=\; aT(a). \tag{5.2.3}$$

From this we get, acting on a position eigenstate $|x\rangle$,

$$X\left(T(a)\,|x\rangle\right) \;=\; T(a)(X+a)\,|x\rangle \;=\; (x+a)\left(T(a)\,|x\rangle\right), \tag{5.2.4}$$

This shows that $T(a)\,|x\rangle$ is an eigenstate of $X$ with eigenvalue $x + a$, which means that it must be equal to $|x + a\rangle$ up to a multiplicative constant. If $|x\rangle$ is normalized according to the Dirac condition, then so will be $T(a)\,|x\rangle$, since $T(a)$ is unitary. Therefore, the constant is just a phase, and $T(a)\,|x\rangle = e^{i\theta}\,|x + a\rangle$ for some $\theta$. The only way to resolve the ambiguity represented by $\theta$ is to arbitrarily choose a value for it, and $\theta = 0$ is as good as any, so we define

$$|x+a\rangle \;=\; T(a)\,|x\rangle\,. \tag{5.2.5}$$

In words, $T(a)$ operates by changing a state in which the particle is known to be at $x$ into a state where it is known to be at $x + a$. We therefore call $T(a)$ a **translation operator**.

Together, eqs. (5.2.5) and (5.2.2) imply

$$\langle x|\, T(a) \;=\; \langle x-a|\,, \tag{5.2.6}$$

so that for any state $|\psi\rangle$, the wavefunction obeys

$$\psi(x-a) \;=\; \langle x|T(a)|\psi\rangle\,. \tag{5.2.7}$$

This can be interpreted as shifting the particle's position to the right by $a$ (the "active view"), or equivalently as leaving the particle's position unchanged but shifting the coordinate system to the left by $a$ (the "passive view").

If we do a translation on any normalized state $|\psi\rangle$,

$$|\psi\rangle \;\rightarrow\; T(a)\,|\psi\rangle\,, \tag{5.2.8}$$

then the expectation value of the position is changed according to

$$\langle\psi|X|\psi\rangle \;\rightarrow\; \langle\psi|T(a)^\dagger X T(a)|\psi\rangle \;=\; \langle\psi|(X+a)|\psi\rangle \;=\; \langle\psi|X|\psi\rangle + a. \tag{5.2.9}$$

The effect of $T(a)$ on momentum eigenstates is just to multiply by a phase,

$$T(a)\,|p\rangle = e^{-iap/\hbar}\,|p\rangle\,, \tag{5.2.10}$$

and the expectation value of momentum is unaffected by the transformation,

$$\langle\psi|P|\psi\rangle \;\rightarrow\; \langle\psi|T(a)^\dagger P T(a)|\psi\rangle \;=\; \langle\psi|P|\psi\rangle\,. \tag{5.2.11}$$

Following the general example of a unitary transformation of an operator given in eq. (5.1.3), we can also define the translated version of an arbitrary observable $A$,

$$A' \;=\; T(a) A T(a)^\dagger, \tag{5.2.12}$$

so that if we do a simultaneous transformation of both states and operators,

$$|\psi\rangle \;\rightarrow\; |\psi'\rangle = T(a)\,|\psi\rangle\,, \tag{5.2.13}$$

$$A \;\rightarrow\; A' = T(a) A T(a)^\dagger, \tag{5.2.14}$$

then matrix elements are unaffected. From eq. (5.2.3), the translated position operator is

$$X' \;=\; T(a) X T(a)^\dagger \;=\; X - a, \tag{5.2.15}$$

while the momentum operator does not change,

$$P' \;=\; T(a) P T(a)^\dagger \;=\; P. \tag{5.2.16}$$

From eq. (5.2.5), it follows that translations simply add,

$$T(a)T(b) \;=\; T(a+b). \tag{5.2.17}$$

One consequence of this is that any finite translation can be viewed as the combination of many infinitesimal translations. For an infinitesimal translation, the exponential can be truncated to linear order,

$$T(\epsilon) \;=\; 1 - i\frac{\epsilon}{\hbar}P. \tag{5.2.18}$$

This is described in words by saying that the momentum operator is the generator of translations. Since the momentum operator commutes with itself, in the general language of section 5.1 where $T(a)$ plays the role of $U(\alpha)$ and $P$ plays the role of $G_a$, the Lie algebra for translations is simply

$$[P, P] \;=\; 0. \tag{5.2.19}$$

This is an Abelian algebra; the antisymmetric structure constants defined in general by eq. (5.1.7) trivially vanish in this case, because there is only one generator, $P$.

Now suppose that a quantum mechanical system has translations as a symmetry, by which we mean that the Hamiltonian is invariant. In particular, for infinitesimal translations, we require $T(\epsilon)HT(\epsilon)^\dagger = H$, from which it follows that

$$\left(1 - i\frac{\epsilon}{\hbar}P\right) H \left(1 + i\frac{\epsilon}{\hbar}P\right) - H \;=\; i\frac{\epsilon}{\hbar}[H, P] + \mathcal{O}(\epsilon^2) \tag{5.2.20}$$

must vanish, so

$$[H, P] = 0. \tag{5.2.21}$$

In general, this requires the potential $V$ to have no dependence on $X$. The Hamiltonian could be that of a free particle with $H = P^2/2m$, but it could also have some extra terms that may involve other functions of $P$ or other degrees of freedom (for example, spin), but not $X$. In any case, eq. (5.2.21) implies that there must be an orthobasis of common eigenstates of $P$ and $H$. An eigenstate of $P$ with momentum $p$ remains so at later times, since

$$P\left(e^{-itH/\hbar}|p\rangle\right) \;=\; p\left(e^{-itH/\hbar}|p\rangle\right). \tag{5.2.22}$$

Also, Ehrenfest's Theorem eq. (3.4.13) says

$$\frac{d}{dt}\langle P\rangle \;=\; 0, \tag{5.2.23}$$

and Theorem 5.1.1 says even more, that the probability to measure the momentum within any given range will be constant in time. Note that these statements are true for any state, including an arbitrary superposition of free particle eigenstates with different momenta, provided that $[H, P] = 0$. A non-trivial illustration will be provided at the end of section 6.2.

For a particle moving in three dimensions, the preceding discussion can be adapted in a straightforward way. The translation operator for a vector distance $\vec{a}$ is

$$T(\vec{a}) \,=\, e^{-i\vec{a}\cdot\vec{P}/\hbar}, \tag{5.2.24}$$

and it satisfies

$$\begin{aligned} T(\vec{a})T(\vec{b}) &= T(\vec{a}+\vec{b}), &(5.2.25)\\ T(\vec{a})^{-1} &= T(\vec{a})^{\dagger} \,=\, T(-\vec{a}). &(5.2.26)\end{aligned}$$

The operation on position and momentum eigenkets is

$$\begin{aligned} T(\vec{a})\,|\vec{r}\rangle &= |\vec{r}+\vec{a}\rangle\,, \qquad\qquad \langle\vec{r}|\,T(\vec{a}) \,=\, \langle\vec{r}-\vec{a}|\,, &(5.2.27)\\ T(\vec{a})\,|\vec{p}\rangle &= e^{-i\vec{a}\cdot\vec{p}/\hbar}\,|\vec{p}\rangle\,, \qquad\qquad \langle\vec{p}|\,T(\vec{a}) \,=\, e^{i\vec{a}\cdot\vec{p}/\hbar}\,\langle\vec{p}|\,, &(5.2.28)\end{aligned}$$

and the remaining discussion for the one-dimensional case likewise follows through for the three-dimensional case with $P$ replaced by $\vec{P}$. Although there are now three generators $P_x$, $P_y$, and $P_z$, they all commute with each other, so the structure constants are all 0, and the Lie group is Abelian. Note that in the case of a charged particle moving in an electromagnetic field as discussed in section 4.3, it is the canonical momentum $\vec{P}$ that generates translations, not the kinetic momentum $\vec{\Pi}$.

Consider a quantum system describing two distinguishable particles labeled 1 and 2. Then one can have translation invariance even with a non-zero potential, provided that the Hamiltonian has the form considered in our discussion of the two-body problem in section 4.2,

$$H = \frac{P_1^2}{2m_1} + \frac{P_2^2}{2m_2} + V(\vec{R}_1 - \vec{R}_2). \tag{5.2.29}$$

Now, one can define individual translation operators for particles 1 and 2, by

$$T_1(\vec{a}) \,=\, e^{-i\vec{a}\cdot\vec{P}_1/\hbar}, \qquad\qquad T_2(\vec{a}) \,=\, e^{-i\vec{a}\cdot\vec{P}_2/\hbar}. \tag{5.2.30}$$

These are not symmetries of the Hamiltonian unless the potential $V(\vec{R}_1 - \vec{R}_2)$ is neglected. However, defining the total momentum operator $\vec{P}_{\rm tot} = \vec{P}_1 + \vec{P}_2$ as in section 4.2, one can check that each component of $\vec{P}_{\rm tot}$ commutes with each component of $\vec{R}_1 - \vec{R}_2$, so

$$\left[H,\, \vec{P}_{\rm tot}\right] \,=\, 0. \tag{5.2.31}$$

Therefore, the total translation operator

$$T(\vec{a}) \,=\, \exp\left[-i\vec{a}\cdot\vec{P}_{\rm tot}/\hbar\right] \tag{5.2.32}$$

leaves the full Hamiltonian eq. (5.2.29) invariant,

$$T(\vec{a})HT(\vec{a})^\dagger \;=\; H, \tag{5.2.33}$$

and is therefore a symmetry of the system. [Note that this would not be true if one added individual potentials $V_1(\vec{R}_1)$ or $V_2(\vec{R}_2)$ to the Hamiltonian.] Equation (5.2.33) simply reflects the feature that the total translation moves both particles 1 and 2 by the same amount, and so does not change the separation between them, nor either of their momenta. It follows that one can find an orthonormal basis of simultaneous eigenstates of $H$ and $\vec{P}_{\text{tot}} = \vec{P}_1 + \vec{P}_2$, and that an eigenstate of $\vec{P}_{\text{tot}}$ remains so at later times, and that in an arbitrary state the probability of measuring a given result for $\vec{P}_{\text{tot}}$ does not change in time, and its expectation value is conserved.

## 5.3 Rotations

In this section, we will explore the connection between rotations and the angular momentum operators that generate them. Classically, the angular momentum of a particle about the point chosen as the origin is defined by

$$\vec{l} = \vec{r} \times \vec{p}. \tag{5.3.1}$$

In quantum mechanics, we promote this to a vector operator, and define the **orbital angular momentum operator** for a particle as

$$\vec{L} = \vec{R} \times \vec{P} = \hat{x}L_x + \hat{y}L_y + \hat{z}L_z, \tag{5.3.2}$$

where the components

$$L_x \;=\; YP_z - ZP_y, \qquad L_y \;=\; ZP_x - XP_z, \qquad L_z \;=\; XP_y - YP_x. \tag{5.3.3}$$

are each observables. There is no problem with operator ordering to worry about here, because $[Y, P_z] = [Z, P_y] = [Z, P_x] = [X, P_z] = [X, P_y] = [Y, P_x] = 0$.

The commutator algebra for the angular momentum components $L_x$, $L_y$, and $L_z$ can be computed using the commutators of the position and momentum operators that they are built out of. For example,

$$[L_x, L_y] \;=\; [YP_z, ZP_x] + [ZP_y, XP_z] \;=\; i\hbar(XP_y - YP_x) \;=\; i\hbar L_z. \tag{5.3.4}$$

Similarly,

$$[L_y, L_z] \;=\; i\hbar L_x, \qquad [L_z, L_x] \;=\; i\hbar L_y. \tag{5.3.5}$$

The last three equations can be summarized as[†]

$$[L_a, L_b] = i\hbar\epsilon_{abc}L_c, \qquad (a, b, c = x, y, z). \tag{5.3.6}$$

Here, $\epsilon_{abc}$ is the totally antisymmetric Levi-Civita symbol, with

$$\epsilon_{xyz} = \epsilon_{yzx} = \epsilon_{zxy} = 1, \qquad \epsilon_{xzy} = \epsilon_{yxz} = \epsilon_{zyx} = -1, \tag{5.3.7}$$

and all other components $\epsilon_{abc} = 0$. It obeys the identities

$$\epsilon_{abc}\epsilon_{dec} = \delta_{ad}\delta_{be} - \delta_{ae}\delta_{bd}, \tag{5.3.8}$$

$$\epsilon_{abc}\epsilon_{dbc} = 2\delta_{ad}, \tag{5.3.9}$$

$$\epsilon_{abc}\epsilon_{abc} = 6. \tag{5.3.10}$$

In the general language of section 5.1, the role of the generators $G_a$ is taken by $L_a/\hbar$ for rotations, as we are about to show. Comparing to eq. (5.1.7), we see that the Lie algebra of angular momentum operators is non-Abelian, with structure constants $f_{abc} = \epsilon_{abc}$. Note that there is always an arbitrary normalization in the definition of the generators; it was convenient to include the factor of $1/\hbar$ here so that the structure constants are dimensionless.

Rotations are defined by the property that, as changes in coordinates, they leave invariant the distances of points from the origin. The composition of two rotations is another rotation; in mathematical language, rotations correspond to the Lie group called $SO(3)$. In quantum mechanics, we can think of rotations as a unitary change of basis corresponding to the change in coordinates; a rotation by an angle $\alpha$ about the axis defined by a unit vector $\vec{n}$ is implemented by a unitary operator $U(\vec{\alpha})$, where $\vec{\alpha} = \hat{n}\alpha$. The fact that rotations form a group means that for any $\vec{\alpha}$ and $\vec{\beta}$,

$$U(\vec{\beta})U(\vec{\alpha}) = U(\vec{\gamma}), \tag{5.3.11}$$

for some $\vec{\gamma}$. In particular, arbitrary finite rotations can be constructed from the limit of a large number of infinitesimal rotations.

Let us now show that (for the case of a single spin-less particle) $L_z$ is the generator of rotations about the $z$-axis (analogously to how $P_z$ is the generator of translations along the $z$ direction). For such a rotation by an angle $\alpha$, the coordinates transform as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5.3.12}$$

---

[†] Here, and from now on, we use the **repeated index summation convention**, which says that repeated indices are implicitly summed over, except when they appear on both sides of an equation. Thus, in this case, $c$ is summed over, but $a$ and $b$ are not.

If $\alpha = \epsilon$ is infinitesimal, then this becomes $x \to x - \epsilon y$ and $y \to y + \epsilon x$, and of course $z \to z$. We therefore define an infinitesimal unitary rotation operator by its action on the position eigenkets,

$$U(\epsilon\hat{z})\,|x,\, y,\, z\rangle \;=\; |x - \epsilon y,\, y + \epsilon x,\, z\rangle\,, \tag{5.3.13}$$

which implies

$$\langle x, y, z|\, U(\epsilon\hat{z})^\dagger \;=\; \langle x - \epsilon y,\, y + \epsilon x,\, z| \tag{5.3.14}$$

and $U(\epsilon\hat{z})^\dagger = U(\epsilon\hat{z})^{-1} = U(-\epsilon\hat{z})$. It follows that

$$\begin{aligned}\langle x, y, z|U(\epsilon\hat{z})|\psi\rangle &= \langle x + \epsilon y,\, y - \epsilon x,\, z|\psi\rangle = \psi(x + \epsilon y, y - \epsilon x, z)\\ &= \psi(x, y, z) + \epsilon y\frac{\partial}{\partial x}\psi(x, y, z) - \epsilon x\frac{\partial}{\partial y}\psi(x, y, z),\end{aligned} \tag{5.3.15}$$

where the last equality is the result of the Taylor series expansion to linear order in $\epsilon$. Now, since $\partial/\partial x \leftrightarrow iP_x/\hbar$ and $\partial/\partial y \leftrightarrow iP_y/\hbar$ in the position representation, we have

$$\langle x, y, z|U(\epsilon\hat{z})|\psi\rangle \;=\; \langle x, y, z|\left[I - i\frac{\epsilon}{\hbar}(XP_y - YP_x)\right]|\psi\rangle\,, \tag{5.3.16}$$

for every state $|\psi\rangle$, so comparing with eq. (5.3.3), we get

$$U(\epsilon\hat{z}) \;=\; I - \frac{i}{\hbar}\epsilon L_z. \tag{5.3.17}$$

This establishes that $L_z$ generates rotations about the $z$ axis, and

$$L_z \;\leftrightarrow\; -i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) \tag{5.3.18}$$

in the position wavefunction representation.

There is nothing special about the $z$-axis in the preceding discussion, so repeating the preceding process for infinitesimal rotations about the $x$ and $y$ axes, one obtains

$$U(\epsilon\hat{x}) = I - \frac{i}{\hbar}\epsilon L_x, \qquad U(\epsilon\hat{y}) = I - \frac{i}{\hbar}\epsilon L_y, \tag{5.3.19}$$

with $L_x$ and $L_y$ given by eqs. (5.3.3), and position representations

$$L_x \leftrightarrow -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right), \qquad L_y \leftrightarrow -i\hbar\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right). \tag{5.3.20}$$

Thus $L_x$, $L_y$, and $L_z$ are generators for rotations about the $x$, $y$, and $z$ axes, respectively (assuming that there is no intrinsic angular momentum).

The unitary rotation operator for a non-infinitesimal angle $\alpha$ can be built from the limit of a large number $N$ of infinitesimal operators with $\epsilon = \alpha/N$, acting sequentially, so

$$U(\alpha\hat{z}) = \lim_{N\to\infty}\left(I - \frac{i}{\hbar}\frac{\alpha}{N}L_z\right)^N = \exp\left(-\frac{i}{\hbar}\alpha L_z\right), \tag{5.3.21}$$

where we have used the definition of the exponential of an operator in eq. (2.4.29). For a rotation by an angle $\alpha$ about an arbitrary axis $\hat{n}$, this generalizes to

$$U(\alpha\hat{n}) \;=\; \exp\left(-\frac{i}{\hbar}\vec{\alpha}\cdot\vec{L}\right), \tag{5.3.22}$$

where $\vec{\alpha} = \hat{n}\alpha$.

For the position representation, we could also use cylindrical or spherical coordinates instead of rectangular coordinates, so that the position eigenkets are $|r, \phi, z\rangle$ or $|r, \theta, \phi\rangle$, respectively. For spherical coordinates in particular, this has the advantage that one of the coordinates, $r$, is not involved in the differential operators that represent the angular momenta. In spherical coordinates, eq. (5.3.15) becomes

$$\langle r, \theta, \phi|U(\epsilon\hat{z})|\psi\rangle \;=\; \psi(r, \theta, \phi) + \epsilon\frac{\partial}{\partial\phi}\psi(r, \theta, \phi), \tag{5.3.23}$$

so that as an equivalent to eq. (5.3.18),

$$L_z \;\;\leftrightarrow\;\; -i\hbar\frac{\partial}{\partial\phi} \tag{5.3.24}$$

in the position representation. For the special case of rotations about the $z$ axis, this also holds in cylindrical coordinates.

Let us now solve the eigenvalue problem for the operator $L_z$, using the position representation. Because eq. (5.3.24) only involves the coordinate $\phi$, the following derivation works equally well in spherical or cylindrical coordinates. We start with

$$L_z|l_z\rangle \;=\; l_z|l_z\rangle, \tag{5.3.25}$$

where $l_z$ is the eigenvalue, also used as a label for the eigenstate, and we have suppressed any degeneracy labels. In the position representation, this becomes

$$-i\hbar\frac{\partial}{\partial\phi}\psi(\phi) \;=\; l_z\psi(\phi), \tag{5.3.26}$$

also suppressing the dependence on other coordinates. The solutions are

$$\psi(\phi) \;=\; ce^{il_z\phi/\hbar}, \tag{5.3.27}$$

where $c$ is a non-zero normalization constant. The range of $\phi$ is the continuous interval $0 \leq \phi \leq 2\pi$. Therefore, for $L_z$ to be a Hermitian operator, by following exactly the same derivation that led to eq. (2.8.23) with $P$ replaced by $L_z$, we find that

$$[\psi_1(2\pi)]^*\psi_2(2\pi) \;=\; [\psi_1(0)]^*\psi_2(0) \tag{5.3.28}$$

must hold for any wavefunctions $\psi_1$ and $\psi_2$. A sufficient condition for this to be satisfied is that all wavefunctions obey $\psi(2\pi) = \psi(0)$, and more generally $\psi(\phi + 2\pi) = \psi(\phi)$. It is necessary to impose this periodic boundary condition anyway, in order for the wavefunction to have a unique expansion in terms of position eigenstate kets. It follows that $e^{il_z 2\pi/\hbar} = 1$, which requires that

$$l_z = \hbar m, \tag{5.3.29}$$

where $m$ is an integer. The use of the letter $m$ is traditional in this role, and it is sometimes called the **magnetic quantum number**. We therefore use it as a label for the corresponding eigenstates of $L_z$, and write, from now on,

$$L_z|m\rangle \;=\; \hbar m|m\rangle. \tag{5.3.30}$$

All of the preceding was derived with the assumption that the angular momentum was associated with the motion of a single particle. More generally, a Hilbert space can describe more than one particle. Furthermore, each of the particles may also have an intrinsic angular momentum, or spin, which has no classical counterpart, and is not associated at all with the quantum mechanical position wavefunction of the particle. These distinct types of angular momenta can also be combined to form new angular momenta.

In general, we define an angular momentum operator

$$\vec{J} \;=\; \hat{x}J_x + \hat{y}J_y + \hat{z}J_z \tag{5.3.31}$$

to be one that satisfies a commutator algebra that has the same form as for orbital angular momentum. Specifically,

$$[J_x, J_y] \;=\; i\hbar J_z \qquad [J_y, J_z] \;=\; i\hbar J_x \qquad [J_z, J_x] \;=\; i\hbar J_y, \tag{5.3.32}$$

or equivalently

$$[J_a,\, J_b] \;=\; i\hbar\epsilon_{abc}J_c. \tag{5.3.33}$$

Because the components of an angular momentum operator $\vec{J}$ do not commute with each other, they are not compatible, and one cannot find an orthobasis of $\vec{J}$ eigenstates. The only solution to the eigenvalue equation

$$\vec{J}\,|\vec{j}\rangle \;=\; \vec{j}\,|\vec{j}\rangle \tag{5.3.34}$$

has $\vec{j} = 0$. Thus, an angular momentum vector $\vec{J}$ is not an observable, although each of its components is. If we choose eigenstates of $J_z$, they cannot also be eigenstates of $J_x$ or $J_y$, except in the very special case that all of the eigenvalues are 0. However, you can check that the **angular momentum squared operator**

$$J^2 \;=\; J_x^2 + J_y^2 + J_z^2 \tag{5.3.35}$$

does commute with $J_z$. This means that one can find common eigenkets for $J^2$ and $J_z$, and they can be part of a CSCO. We will work out the the details of the simultaneous eigenvalue problem for $J^2$ and $J_z$ in Chapter 8.

For each particle, the intrinsic angular momentum, or spin $\vec{S}$, is the part of the angular momentum that is inseparable from the identity of the particle, and is distinct from the orbital angular momentum. The operator $\vec{S}$ is a special case of $\vec{J}$ defined by the property that it has fixed magnitude, in the sense that all particles of a given type have the same eigenvalue of $S^2$, which can never change. For any single particle, the total angular momentum operator is simply the sum of the orbital and spin parts, $\vec{J} = \vec{L} + \vec{S}$. (Note that we use the same symbol $\vec{J}$ as for a generic angular momentum operator defined in the previous paragraph.)

For the two-body problem of section 4.2, you can check that the total orbital angular momentum operator (not including the spins) can be written in two equivalent ways as

$$\vec{L}_{\rm tot} \;=\; \vec{L}_1 + \vec{L}_2 \;=\; \vec{L} + \vec{L}_{\rm cm}, \tag{5.3.36}$$

where $\vec{L}_1 = \vec{R}_1 \times \vec{P}_1$ and $\vec{L}_2 = \vec{R}_2 \times \vec{P}_2$ for the individual particles, and

$$\vec{L} = \vec{R} \times \vec{P}, \qquad\qquad \vec{L}_{\rm cm} = \vec{R}_{\rm cm} \times \vec{P}_{\rm tot}, \tag{5.3.37}$$

define the relative and center-of-mass contributions to the angular momenta. It is often sensible to restrict to the subspace of the Hilbert space consisting of eigenstates of $\vec{P}_{\rm tot}$ with eigenvalue 0, corresponding to vanishing total momentum in the center-of-mass frame. If we do so, then $\vec{L}_{\rm cm}$ vanishes identically, and the relative angular momentum $\vec{L}$ is equal to the total angular momentum $\vec{L}_{\rm tot}$.

More generally, for a given choice of origin, every three-dimensional physical system has a **total angular momentum** operator $\vec{J}$, which adds both orbital and spin angular momentum contributions for all of the particles that are present. By definition, $\vec{J}$ is the operator that acts on the full Hilbert space of states to generate rotations about an axis $\hat{n}$ by an angle $\alpha$, with

$$U(\vec{\alpha}) \;=\; \exp\Bigl(-i\alpha\hat{n}\cdot\vec{J}/\hbar\Bigr), \tag{5.3.38}$$

where $\vec{\alpha} = \hat{n}\alpha$. To rotate a state $|\psi\rangle$, the unitary transformation is

$$|\psi\rangle \;\to\; |\psi'\rangle \;=\; U(\vec{\alpha})\,|\psi\rangle\,. \tag{5.3.39}$$

We also define rotated operators

$$A \;\rightarrow\; A' \;=\; U(\vec{\alpha}) A U(\vec{\alpha})^\dagger, \tag{5.3.40}$$

so that, due to the unitarity of $U(\vec{\alpha})$, matrix elements of the rotated operators between rotated states are the same as the original matrix elements, $\langle\chi'|A'|\psi'\rangle \;=\; \langle\chi|A|\psi\rangle$.

Suppose that the Hamiltonian for a system has the symmetry of invariance under rotations generated by an angular momentum component $\hat{n}\cdot\vec{J}$, so that rotations about the unit vector axis $\hat{n}$ are a symmetry of the system, and

$$[H,\, \hat{n}\cdot\vec{J}\,] \;=\; 0. \tag{5.3.41}$$

(It is traditional to choose the coordinate system so that $\hat{n} = \hat{z}$, unless there is a good reason not to, but let us be more general.) One can then choose an orthobasis of simultaneous eigenstates of $H$ and $\hat{n}\cdot\vec{J}$, and as shown on general grounds in section 5.1, $\hat{n}\cdot\vec{J}$ is a conserved quantity. This means that eigenstates of $\hat{n}\cdot\vec{J}$ will remain so under time evolution, and for an arbitrary state the probability to obtain a given outcome for a measurement of $\hat{n}\cdot\vec{J}$ is constant, and

$$\frac{d}{dt}\langle\hat{n}\cdot\vec{J}\,\rangle \;=\; 0. \tag{5.3.42}$$

If the Hamiltonian is invariant under all rotations, it is convenient to choose a CSCO to include

$$H, J^2, J_z, \tag{5.3.43}$$

since these observables commute with each other.

## 5.4 Parity

Another kind of coordinate transformation is **parity** (also known as **space inversion**), defined as the operation of replacing each rectangular position coordinate by minus itself. Unlike translations and rotations, parity is a discrete group known as $Z_2$; acting twice with parity gives back the identity operation, and there are no infinitesimal parity transformations.

Let us start by defining the **parity operator** $\Pi$ for a particle moving in one dimension by giving its action on the position-eigenstate orthobasis kets,

$$\Pi\,|x\rangle \;=\; |{-x}\rangle\,. \tag{5.4.1}$$

(Note that $|{-x}\rangle$ is the ket that describes a particle known to be at the point $-x$, and is not the same thing as $-|x\rangle$, which still describes a particle known to be at the point $x$.) Since $\Pi\,(\Pi\,|x\rangle) = |x\rangle$, we have $\Pi\,\Pi = I$, so that so parity is its own inverse,

$$\Pi^{-1} \;=\; \Pi. \tag{5.4.2}$$

Also, taking the Hermitian adjoint of eq. (5.4.1) gives $\langle x|\,\Pi^\dagger = \langle -x|$, so

$$\langle x'|\Pi^\dagger|x\rangle \;=\; \langle -x'|x\rangle \;=\; \delta(x+x') \;=\; \langle x'|-x\rangle \;=\; \langle x'|\Pi|x\rangle\,. \tag{5.4.3}$$

Since this is true for every $|x\rangle$ and $|x'\rangle$ in the position orthobasis, it must be that

$$\Pi^\dagger \;=\; \Pi. \tag{5.4.4}$$

Comparing eqs. (5.4.2) and (5.4.4), we see that the parity operator is both unitary (a map from an orthobasis to another orthobasis) and Hermitian (an observable). Since all Hermitian operators have real eigenvalues, and all unitary operators have eigenvalues with magnitude 1, the only possible eigenvalues of $\Pi$ are 1 and $-1$.

Given a wavefunction for an arbitrary state $\langle x|\psi\rangle = \psi(x)$, we have

$$\langle x|\Pi|\psi\rangle \;=\; \langle -x|\psi\rangle \;=\; \psi(-x). \tag{5.4.5}$$

Applying this to the case of a momentum eigenstate,

$$\langle x|\Pi|p\rangle \;=\; \langle -x|p\rangle \;=\; \frac{1}{\sqrt{2\pi\hbar}}e^{-ipx/\hbar} \;=\; \langle x|-p\rangle\,, \tag{5.4.6}$$

from which we learn that

$$\Pi|p\rangle \;=\; |-p\rangle. \tag{5.4.7}$$

Similarly, it is easy to show that

$$\Pi X\Pi \;=\; -X, \qquad\qquad \Pi P\Pi \;=\; -P, \tag{5.4.8}$$

so that the position and momentum operators are both said to be odd under parity. More generally, one defines the **parity of an operator** $A$ as

$$\pi_A \;=\; \pm 1, \qquad \text{if} \qquad \Pi A\Pi = \pm A. \tag{5.4.9}$$

However, not all operators have definite parity in this sense.

If $A$ is even under parity, it follows that

$$\Pi A - A\Pi = 0, \tag{5.4.10}$$

so that $\Pi$ and $A$ are compatible operators, and can have common eigenstates. For example, the Hamiltonian $H = P^2/2m + V(X)$ is parity-even if, and only if, $V(X)$ is an even function of $X$.

In the special case of a free particle with $V = 0$, the corresponding simultaneous eigenstates of $H$ and $\Pi$ are

$$|E, +1\rangle \;=\; \frac{1}{\sqrt{2}}\left(|p\rangle \,+\, |{-p}\rangle\right), \tag{5.4.11}$$

$$|E, -1\rangle \;=\; \frac{1}{\sqrt{2}}\left(|p\rangle \,-\, |{-p}\rangle\right), \tag{5.4.12}$$

where $p = \sqrt{2mE}/\hbar$. However, since $P$ does not commute with $\Pi$, one cannot find simultaneous eigenstates of them, and indeed the parity eigenstates $|E, \pm 1\rangle$ are not eigenstates of momentum.

As in the case of translations, parity generalizes straightforwardly to three dimensions. By definition,

$$\Pi|\vec{r}\,\rangle \;=\; |{-\vec{r}}\,\rangle\,, \tag{5.4.13}$$

from which it follows that

$$\Pi|\vec{p}\,\rangle \;=\; |{-\vec{p}}\,\rangle \tag{5.4.14}$$

and position and momentum operators carry odd parity,

$$\Pi\vec{R}\,\Pi \;=\; -\vec{R}, \qquad\qquad \Pi\vec{P}\,\Pi \;=\; -\vec{P}. \tag{5.4.15}$$

Angular momentum operators, as defined in the previous section, always have even parity,

$$\Pi\vec{J}\,\Pi \;=\; \vec{J}. \tag{5.4.16}$$

In the case of orbital angular momentum, $\Pi\vec{L}\,\Pi = \vec{L}$ follows directly from the definition in eq. (5.3.3). For more general angular momentum operators (including spin), the even parity can be inferred from the general commutator algebra structure in eq. (5.3.33). If the potential energy depends only on the radial coordinate, so that the Hamiltonian has the form $H = P^2/2m + V(R)$, then this $H$ commutes not only with $L^2$ and $L_z$, but also with $\Pi$. Such a system therefore admits simultaneous eigenstates of all four operators. In that case, parity is a symmetry of the system, and eigenstates of parity will remain so under time evolution.

A powerful application of parity is to the identification of selection rules for matrix elements. Suppose that an operator $A$ has definite parity $\pi_A$ as defined by eq. (5.4.9), and that the states $|\psi\rangle$ and $|\phi\rangle$ are parity eigenstates with eigenvalues $\pi_\psi$ and $\pi_\phi$, respectively. Now we note that

$$\langle\phi|A|\psi\rangle \;=\; \pi_A\,\langle\phi|\Pi A \Pi|\psi\rangle \;=\; \pi_A\pi_\phi\pi_\psi\,\langle\phi|A|\psi\rangle\,. \tag{5.4.17}$$

The product $\pi_A\pi_\phi\pi_\psi$ is either $+1$ or $-1$. In the former case, eq. (5.4.17) tells us nothing, but in the latter case, we obtain:

**Theorem 5.4.1 (Parity selection rule)** *If states* $|\psi\rangle$ *and* $|\phi\rangle$ *and an operator* $A$ *all have definite parities, and* $\pi_A \pi_\phi \pi_\psi = -1$*, then the matrix element* $\langle\phi|A|\psi\rangle$ *must vanish.*

This result has many practical applications, including elegantly explaining the absence of certain atomic transitions. Before investing your valuable time in calculating a quantity, it is always a good idea to consider first whether it must vanish due to a matrix element selection rule.

## 5.5 Gauge transformations

Physical systems that involve electromagnetic interactions with charged particles can be formulated in terms of potentials, subject to gauge transformations of the form

$$\Phi \rightarrow \Phi - \frac{1}{c}\frac{\partial}{\partial t}\Lambda, \qquad \vec{A} \rightarrow \vec{A} + \vec{\nabla}\Lambda. \tag{5.5.1}$$

As discussed in section 4.3, the physical situation described by a pair of potentials $(\Phi, \vec{A})$ is equally well described by any pair $(\Phi', \vec{A}')$ related to them by a gauge transformation parameterized by some arbitrary function $\Lambda(\vec{r}, t)$.

Before proceeding, we pause to emphasize that gauge transformations have a completely different character than the other transformations discussed in the preceding sections. If we do a translation or rotation on a system, we have transformed the system into a physically distinct one. A particle that has been translated by 1 centimeter is in a different place, distinguishable by experiment. To say that a system has translation symmetry means that we can physically move the whole system in an experimentally measurable way without changing the Hamiltonian that describes its time evolution. In contrast, the whole point of gauge transformations is that although they change the Hamiltonian, they do so without changing the actual physical situation at all. There is no way an experiment can tell whether you have chosen Coulomb gauge or not! Although it is common to refer to "gauge symmetries", the freedom to do a gauge transformation really is not a symmetry in the same way that translations or rotations can be, but rather is a manifestation of the fact that our description in terms of potentials has an arbitrariness, in the form of redundancies that can be eliminated by fixing the gauge.

Consider a classical charged particle in an electromagnetic field. Clearly, if we do a gauge transformation the position and the velocity of the charged particle are unaffected, since they can be measured experimentally. However, the classical canonical momentum $\vec{p}$ defined in eq. (4.3.13) does change with a gauge transformation, because it involves not just the velocity but also the gauge-dependent potential $\vec{A}$.

In the quantum description, the state ket used to describe the system is similarly gauge-dependent. Naively, this might seem like a problem, but it is not, because the state ket by itself is not a physically measurable observable. For example, we already have pointed out that multiplying a ket by a constant complex phase does not change probabilities. In the following, we will show that the gauge transformation of the ket describing a single particle of charge $q$ is realized as a unitary transformation

$$U_\Lambda = \exp\left[i\frac{q}{\hbar c}\Lambda(\vec{R},t)\right], \tag{5.5.2}$$

with

$$|\psi\rangle \rightarrow |\psi'\rangle = U_\Lambda |\psi\rangle, \tag{5.5.3}$$
$$\Phi(\vec{R},t) \rightarrow \Phi(\vec{R},t)' = \Phi(\vec{R},t) - \frac{1}{c}\frac{\partial}{\partial t}\Lambda(\vec{R},t), \tag{5.5.4}$$
$$\vec{A}(\vec{R},t) \rightarrow \vec{A}(\vec{R},t)' = \vec{A}(\vec{R},t) + \vec{\nabla}\Lambda(\vec{R},t). \tag{5.5.5}$$

Thus, a gauge transformation in quantum mechanics is defined as the simultaneous changes in eqs. (5.5.3)–(5.5.5) with the same $\Lambda$. Primes are used to indicate the state ket and operators after the gauge transformation. For every operator $\Omega$, the definition of the gauge transformation

$$\Omega \rightarrow \Omega' \tag{5.5.6}$$

is obtained by applying eq. (5.5.4) and (5.5.5) to the explicit dependence of $\Omega$ on the potentials. We now need to show that physical predictions are unaffected by such a transformation.

We start by noting that a prerequisite for an operator $\Omega$ to be a **physical gauge-invariant observable** is that its expectation values should be the same before and after the gauge transformation,

$$\langle\psi|\Omega|\psi\rangle = \langle\psi'|\Omega'|\psi'\rangle. \tag{5.5.7}$$

Requiring this to be true for all state kets $|\psi\rangle$, and using eq. (5.5.3), we obtain the defining property of a gauge-invariant observable,

$$\Omega' = U_\Lambda \Omega U_\Lambda^\dagger. \tag{5.5.8}$$

Not all Hermitian operators that would otherwise satisfy the requirements of being an observable will have this property.

For example, consider the canonical operators $\vec{R}$ and $\vec{P}$. Since they do not depend explicitly on the potentials, we have

$$\vec{R}' = \vec{R}, \qquad \vec{P}' = \vec{P}. \tag{5.5.9}$$

Meanwhile, using the definition of $U_\Lambda$ in eq. (5.5.2), we find

$$U_\Lambda \vec{R} U_\Lambda^\dagger = \vec{R}, \tag{5.5.10}$$
$$U_\Lambda \vec{P} U_\Lambda^\dagger = \vec{P} - \frac{q}{c}\vec{\nabla}\Lambda. \tag{5.5.11}$$

The first equation is an immediate consequence of the fact that $U_\Lambda$ involves only $\vec{R}$ and not $\vec{P}$, and the second follows from applying Theorem 2.4.1. Comparison of eqs. (5.5.9)–(5.5.11) with eq. (5.5.8) shows that $\vec{R}$ is a gauge-invariant observable, but the canonical momentum operator $\vec{P}$ is not. In contrast, it follows from eq. (5.5.5) that the gauge transformation of the kinetic momentum $\vec{\Pi} = \vec{P} - \frac{q}{c}\vec{A}$ is $\vec{\Pi}' = \vec{P} - \frac{q}{c}(\vec{A} + \vec{\nabla}\Lambda)$. From eq. (5.5.11), it follows that

$$\vec{\Pi}' \;=\; U_\Lambda \vec{\Pi} U_\Lambda^\dagger. \tag{5.5.12}$$

This establishes that the kinetic momentum $\vec{\Pi}$ is a gauge-invariant observable, according to our defining requirement of eq. (5.5.8).

Since this was perhaps somewhat tricky and unexpected, we reiterate. Even though the canonical momentum operator $\vec{P}$ does not change when we do a gauge transformation, its expectation values do change, so it is not a gauge-invariant observable. And, although the kinetic momentum operator $\vec{\Pi}$ changes when we do a gauge transformation, its expectation values do not, so it is a gauge-invariant observable. These are the quantum versions of the statements made about the gauge dependences of the classical quantities $\vec{p}$ and $\vec{\pi}$ following eq. (4.3.15).

The Hamiltonian operator certainly changes when we do a gauge transformation, since it depends explicitly on the potentials $\Phi$ and $\vec{A}$. Using eqs. (5.5.4) and (5.5.5) in eq. (4.3.28) gives the gauge transformation of $H$,

$$H \;\rightarrow\; H' = \frac{1}{2m}\left[\vec{P} - \frac{q}{c}\vec{A} - \frac{q}{c}\vec{\nabla}\Lambda\right]^2 + q\Phi - \frac{q}{c}\frac{\partial\Lambda}{\partial t} - \gamma\vec{S}\cdot\vec{B}. \tag{5.5.13}$$

Equations (5.5.10) and (5.5.11) allow us to rewrite this in the convenient form

$$H' \;=\; U_\Lambda H U_\Lambda^\dagger + i\hbar\left(\frac{\partial U_\Lambda}{\partial t}\right)U_\Lambda^\dagger. \tag{5.5.14}$$

Now, given the Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi(t)\rangle \;=\; H|\psi(t)\rangle, \tag{5.5.15}$$

it is straightforward to use $|\psi(t)'\rangle = U_\Lambda|\psi(t)\rangle$ and eq. (5.5.14) to obtain

$$i\hbar\frac{d}{dt}|\psi(t)'\rangle \;=\; H'\,|\psi(t)'\rangle. \tag{5.5.16}$$

This shows the key result that the Schrödinger equation for time evolution is also satisfied if $H$ and $|\psi(t)\rangle$ are replaced by their gauge-transformed counterparts.

Another consequence of eq. (5.5.14) is that the Hamiltonian is not, in general, a gauge-invariant observable, due to the presence of the last term. However, the requirement $H' =$

$U_\Lambda H U_\Lambda^\dagger$ is satisfied for the subset of gauge transformations such that $\partial\Lambda/\partial t = 0$. Thus, if we limit the gauge transformations to those that do not depend explicitly on time, then $H$ is a gauge-invariant observable in that restricted sense. Also, if the potentials $\Phi$ and $\vec{A}$ do not depend on time, then according to the results at the end of section 5.1, $H$ is a conserved quantity, the total energy of the particle.

The defining requirement for gauge-invariant observables, eq. (5.5.8), was obtained by requiring that expectation values do not depend on the choice of gauge. Now we will use eq. (5.5.8) to prove the stronger result that all probabilities for measurements of such observables are gauge independent. Consider a gauge-invariant observable $\Omega$, and choose an orthobasis of its eigenkets with degeneracy labels $u_\omega$, so that

$$\Omega|\omega, u_\omega\rangle \;=\; \omega|\omega, u_\omega\rangle. \tag{5.5.17}$$

According to Postulates 3 and 4, the probabilities for allowed results of the measurement of $\Omega$ in a state $|\psi\rangle$ are

$$\mathcal{P}(\omega) = \sum_{u_\omega} |\langle\omega, u_\omega|\psi\rangle|^2. \tag{5.5.18}$$

Now, we observe that

$$\Omega'\, U_\Lambda|\omega, u_\omega\rangle \;=\; U_\Lambda \Omega U_\Lambda^\dagger\, U_\Lambda|\omega, u_\omega\rangle \;=\; U_\Lambda\Omega|\omega, u_\omega\rangle \;=\; \omega U_\Lambda|\omega, u_\omega\rangle, \tag{5.5.19}$$

which shows that the states $U_\Lambda|\omega, u_\omega\rangle$ are eigenkets of $\Omega'$, with the same eigenvalues $\omega$ and the same degeneracies. Therefore, we can compute the probability to obtain the result $\omega$ from a measurement of $\Omega'$ in the gauge-transformed description as

$$\mathcal{P}(\omega)' = \sum_{u_\omega} |\langle\omega, u_\omega|U_\Lambda^\dagger|\psi'\rangle|^2 = \sum_{u_\omega} |\langle\omega, u_\omega|U_\Lambda^\dagger U_\Lambda|\psi\rangle|^2 = \sum_{u_\omega} |\langle\omega, u_\omega|\psi\rangle|^2 = \mathcal{P}(\omega). \tag{5.5.20}$$

Thus, we have succeeded in our goal of showing that the predictions for measurements of gauge-invariant observables are not changed by the gauge transformation. This is in accord with the general principle that gauge transformations affect our equations in intermediate steps, but do not change physical reality.

For simplicity, in the preceding we have treated the case of a single particle with charge $q$. In the case of more than one particle, with charges $q_n$ and masses $m_n$, the same discussion goes through, with Hamiltonian

$$H \;=\; \sum_n \left( \frac{1}{2m_n}\left[\vec{P}_n - \frac{q_n}{c}\vec{A}(\vec{R}_n, t)\right]^2 + q_n\Phi(\vec{R}_n, t) - \gamma_n \vec{S}_n \cdot \vec{B}(\vec{R}_n, t) \right), \tag{5.5.21}$$

subject to gauge transformations as given by eqs. (5.5.3)–(5.5.5), but with

$$U_\Lambda \;=\; \exp\left[\frac{i}{\hbar c}\sum_n q_n \Lambda(\vec{R}_n, t)\right]. \tag{5.5.22}$$

The gauge transformation of the wavefunction in the position representation, $\psi(\vec{r}_1, \ldots, \vec{r}_n) = \langle \vec{r}_1, \ldots, \vec{r}_n | \psi\rangle$, is therefore given by

$$\psi(\vec{r}_1, \ldots, \vec{r}_n) \;\rightarrow\; \exp\left[\frac{i}{\hbar c}\sum_n q_n \Lambda(\vec{r}_n, t)\right] \psi(\vec{r}_1, \ldots, \vec{r}_n). \tag{5.5.23}$$

This consists of multiplication by a phase that may depend on time and the particle positions, but not in an arbitrary way when there is more than one particle, since there is only one function $\Lambda(\vec{r}, t)$. Thus, gauge invariance can be viewed as the statement that multiplying the wavefunction for a system of charged particles by a phase of this special form can be compensated by a redefinition of the electromagnetic potentials. In practice, one may avoid this redundancy in the description of the physical system by fixing the gauge. This just means that one chooses a specific form of the potentials, or at least imposes some condition on them that will not be satisfied if one does an arbitrary gauge transformation.

## 5.6 Currents and local conservation of probability

From Postulate 4, and the completeness of position eigenstates, we know that the probability to find a particle within an infinitesimal volume $d^3\vec{r}$ is given by the Born rule,

$$d\mathcal{P} \;=\; \rho(\vec{r}, t)\, d^3\vec{r}, \tag{5.6.1}$$

where the probability density per unit volume is

$$\rho(\vec{r}, t) \;=\; |\langle \vec{r} | \psi(t)\rangle|^2 \;=\; |\psi(\vec{r}, t)|^2. \tag{5.6.2}$$

Assuming the total probability of finding the particle somewhere is fixed and equal to 1, there must be a law of conservation of probability. The local form of this law is a differential equation

$$\frac{\partial \rho}{\partial t} \;=\; -\vec{\nabla} \cdot \vec{J}, \tag{5.6.3}$$

where $\vec{J}$ is a probability current density. The left side of this equation is the rate at which probability density is accumulating at a point, which the right side tells us is the negative of a source for the vector field $\vec{J}$.

To prove eq. (5.6.3), and identify the current density, start with the Schrödinger equation in the position representation, and multiply by $-i\psi^*/\hbar$,

$$\psi^* \frac{\partial}{\partial t}\psi \;=\; \frac{i\hbar}{2m}\psi^* \nabla^2 \psi - \frac{i}{\hbar} V \psi^* \psi. \tag{5.6.4}$$

Adding this to its complex conjugate, we get

$$\frac{\partial \rho}{\partial t} \;=\; \frac{i\hbar}{2m}\left(\psi^*\nabla^2\psi - \psi\nabla^2\psi^*\right), \tag{5.6.5}$$

in which the contribution proportional to $V(\vec{r})$ has canceled due to the fact that the potential is real. Now if we define the probability current density by

$$\vec{J} \;\equiv\; \frac{i\hbar}{2m}\left(\psi\vec{\nabla}\psi^* - \psi^*\vec{\nabla}\psi\right), \tag{5.6.6}$$

then

$$\vec{\nabla}\cdot\vec{J} \;=\; \frac{i\hbar}{2m}\left(\psi\nabla^2\psi^* - \psi^*\nabla^2\psi\right), \tag{5.6.7}$$

where two terms proportional to $\vec{\nabla}\psi^*\cdot\vec{\nabla}\psi$ have canceled. Comparison of eqs. (5.6.5) and (5.6.7) establishes the local conservation of probability, eq. (5.6.3).

In the case of a charged particle, an important modification is needed if the vector potential $\vec{A}$ is non-zero. Starting with the Hamiltonian in eq. (4.3.28), one can show by steps similar to those followed in eqs. (5.6.4)–(5.6.7) that the current density satisfying $\vec{\nabla}\cdot\vec{J} = -\partial\rho/\partial t$ is

$$\vec{J} \;=\; \frac{1}{m}\mathrm{Re}\left[\psi^*\left(-i\hbar\vec{\nabla} - \frac{q}{c}\vec{A}\right)\psi\right], \tag{5.6.8}$$

generalizing eq. (5.6.6). Here, $-i\hbar\vec{\nabla} - \frac{q}{c}\vec{A}$ is the position representation of the kinetic momentum operator, $\vec{\Pi}$, which was introduced in eq. (4.3.29) of section 4.3. This expression for $\vec{J}$ is invariant under gauge transformations, and is related to the electric current density $\vec{j}$ by $\vec{j} = q\vec{J}$, where $q$ is the electric charge of the particle.

As an example, consider a free particle in a plane-wave simultaneous eigenstate of momentum and energy with eigenvalues $\vec{p} = \hbar\vec{k}$ and $E = \hbar^2k^2/2m$,

$$\psi(\vec{r}, t) \;=\; Ce^{i\vec{k}\cdot\vec{r} - iEt/\hbar}. \tag{5.6.9}$$

In terms of the complex normalization constant $C$, the probability and current densities are just constants in both time and position,

$$\rho \;=\; |C|^2, \tag{5.6.10}$$

$$\vec{J} \;=\; \frac{i\hbar}{2m}(-i\vec{k} - i\vec{k})|C|^2 \;=\; \frac{\vec{p}}{m}\rho. \tag{5.6.11}$$

If the domain of the particle is all space, then the wavefunction is not normalizable to unity for any finite $C$, but at least the ratio of the current density to the probability density is well-defined and equal to the velocity eigenvalue of the particle. Although the probability density is constant, it is constantly flowing in the direction of $\vec{v} = \vec{p}/m$. So, in the case of a free particle,

$$\frac{\partial \rho}{\partial t} \;=\; 0, \qquad -\vec{\nabla}\cdot\vec{J} \;=\; 0, \tag{5.6.12}$$

satisfying the local conservation of probability in the most trivial possible way.

In the case of one-dimensional problems, the probability density and current are

$$\rho \;=\; |\psi(x,t)|^2, \qquad\qquad J \;=\; \frac{i\hbar}{2m}\left(\psi\frac{\partial\psi^*}{\partial x} - \psi^*\frac{\partial\psi}{\partial x}\right), \tag{5.6.13}$$

and the statement of local conservation of probability is

$$\frac{\partial\rho}{\partial t} \;=\; -\frac{\partial J}{\partial x}. \tag{5.6.14}$$

The current density vector is just a number in the one-dimensional case, and is positive for probability density flowing to the right, and negative for flow to the left.

In some cases, one can think of probability as *not* being conserved. Suppose, for example, that we are using quantum mechanics to model the behavior of an unstable particle. If our Hilbert space only describes the unstable particle, and not the other particles that it decays into, then we should expect that the total integrated probability to find it somewhere should decrease with time. To describe situations like this, we can break the law for a good cause, by taking the Hamiltonian to be non-Hermitian.

To see how this works, suppose that we defy the authority of Postulate 6 by taking $H$ to have complex eigenvalues, with corresponding normalized eigenkets $|\phi_n\rangle$, so that

$$H\,|\phi_n\rangle = (E_n - i\Gamma_n/2)\,|\phi_n\rangle\,, \tag{5.6.15}$$

where $E_n$ and $\Gamma_n$ are real numbers. Of course, this is only possible if $H$ is not Hermitian, due to Theorem 2.6.3. If the state of the system at time $t = 0$ is one of them, $|\psi(0)\rangle = |\phi_n\rangle$, with unit norm, then the time evolution predicted by the Schrödinger equation will be non-unitary,

$$|\psi(t)\rangle \;=\; e^{-i(E_n - i\Gamma_n/2)t/\hbar}\,|\phi_n\rangle\,. \tag{5.6.16}$$

The norm of this ket as a function of time is therefore

$$\langle\psi(t)|\psi(t)\rangle \;=\; e^{-\Gamma_n t/\hbar}. \tag{5.6.17}$$

We can then interpret the norm of the ket as the probability that the particle exists at time $t > 0$, given that it existed at time $t = 0$. The mean lifetime of our unstable state $|\phi_n\rangle$ is thus $\tau = \hbar/\Gamma_n$, where $\Gamma_n/2$ is the negative of the imaginary part of the Hamiltonian eigenvalue.

The unstable states could correspond to particles that undergo spontaneous decay, such as the neutron, the muon, or many atomic nuclei. The formalism could also be applied to a set of atomic states, if we imagine that our quantum treatment does not include the complete set of states to which they could decay, including photons released in the process. It could even apply in a description of a single absolutely stable particle like the electron, if it can be removed from the system under discussion by a process like electron-capture inverse beta decay, which in some nuclei occurs as $e^- p \to n\nu_e$. This would correspond to an effective potential $V(\vec{r})$ with negative imaginary parts localized at the nuclei. By repeating the steps of eqs. (5.6.4)-(5.6.7), one can see that, in the case of a non-Hermitian potential, the equation that governs the local probability density and current is

$$\frac{\partial \rho}{\partial t} = -\vec{\nabla} \cdot \vec{J} + \frac{2}{\hbar}\text{Im}[V]\rho. \tag{5.6.18}$$

In all such cases of particles decaying to other particles, or interacting in such a way as to change their numbers, the use of a non-Hermitian Hamiltonian is really a sign of an incomplete description. There is always a more fundamental description in which the *full* Hamiltonian will be Hermitian. Quantum field theories provide the natural way to incorporate processes that change the numbers of particles, in such a way that the postulates of quantum mechanics hold, including unitary time evolution as predicted by Postulate 6.

# 6 Particle moving in one dimension

## 6.1 Gaussian wavefunctions

Consider a particle moving in one dimension in the domain $-\infty < x < \infty$. Suppose that at some moment in time, the position wavefunction has a Gaussian form,

$$\langle x|\psi\rangle \;=\; \psi(x) \;=\; N \exp\left[-(x-a)^2/4\sigma^2\right]. \tag{6.1.1}$$

Here $a$ and $\sigma$ are constants with units of [length], corresponding respectively to the center and the width of the probability density $|\psi(x)|^2$. More precisely, the full width at half maximum (FWHM) of the Gaussian peak of the probability density is $2\sqrt{2\ln 2}\sigma \approx 2.35\sigma$. The factor $N$ is a normalization. If we require the ket $|\psi\rangle$ to have unit norm, we need

$$1 \;=\; |N|^2 \int_{-\infty}^{\infty} dx\, \exp\left[-(x-a)^2/2\sigma^2\right] \;=\; |N|^2\sqrt{2\pi}\sigma, \tag{6.1.2}$$

so we can choose $N = e^{i\theta}/(2\pi\sigma^2)^{1/4}$, where $\theta$ is any phase. If $\theta$ is a constant, then it is a global phase and has no physical significance at all, so let us choose the next simplest possibility, that $\theta$ depends linearly on $x$. The Gaussian wavefunction that we choose to study is thus

$$\psi(x) \;=\; \frac{1}{(2\pi\sigma^2)^{1/4}}\, e^{ikx} \exp\left[-(x-a)^2/4\sigma^2\right] \tag{6.1.3}$$

for some constant $k$, whose interpretation will become clear soon. We will not need to assume anything in particular about the Hamiltonian of the system in this section.

According to Postulate 4, the probability of finding the particle between $x$ and $x + dx$ is

$$d\mathcal{P}(x) \;=\; |\langle x|\psi\rangle|^2\, dx \;=\; \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(x-a)^2/2\sigma^2\right]\, dx. \tag{6.1.4}$$

Therefore, the probability of finding the particle in a range $b < x < c$ is

$$\mathcal{P}(b < x < c) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \int_b^c dx\, \exp\left[-(x-a)^2/2\sigma^2\right]. \tag{6.1.5}$$

We can also find the expectation value of $X$ in the state $|\psi\rangle$,

$$\begin{aligned}
\langle X\rangle \;&=\; \langle\psi|X|\psi\rangle \;=\; \int_{-\infty}^{\infty} dx\, \langle\psi|X|x\rangle\langle x|\psi\rangle \;=\; \int_{-\infty}^{\infty} dx\, x|\psi(x)|^2 \\
&=\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} dx\, x \exp\left[-(x-a)^2/2\sigma^2\right] \;=\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} du\, (u+a) \exp\left[-u^2/2\sigma^2\right] \\
&=\; a.
\end{aligned} \tag{6.1.6}$$

The second equality uses the completeness relation, the third uses $X|x\rangle = x|x\rangle$, and the fourth uses the common trick of "completing the square", which means that we define a new integration variable $u$ so that the exponential in the integrand depends only on $u^2$ (with no linear term in $u$). Here, the change of variable was $x = u + a$. Similarly, we have

$$\langle X^2 \rangle \;=\; \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} du\, (u+a)^2 \exp\left[-u^2/2\sigma^2\right] \;=\; a^2 + \sigma^2. \tag{6.1.7}$$

Therefore, the uncertainty of $X$ for this state is

$$\Delta X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2} = \sigma. \tag{6.1.8}$$

So far, the constant $k$ has not made any difference at all.

The momentum wavefunction for the same state is

$$\widetilde{\psi}(p) \;=\; \langle p|\psi\rangle \;=\; \int_{-\infty}^{\infty} dx\; \langle p|x\rangle\langle x|\psi\rangle \;=\; \frac{1}{\sqrt{2\pi\hbar}(2\pi\sigma^2)^{1/4}} \int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} e^{ikx} e^{-(x-a)^2/4\sigma^2}, \tag{6.1.9}$$

where we have used the complex conjugate of eq. (2.8.39). We now use the completing-the-square trick again; this time the appropriate change of variables is $x = u + b$, where we choose $b = a + i2\sigma^2(k - p/\hbar)$, so

$$\int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} e^{ikx} e^{-(x-a)^2/4\sigma^2} \;=\; e^{(b^2-a^2)/4\sigma^2} \int_{-\infty}^{\infty} du\, e^{-u^2/4\sigma^2} \;=\; 2\sqrt{\pi}\sigma\, e^{(b^2-a^2)/4\sigma^2}. \tag{6.1.10}$$

It follows that

$$\widetilde{\psi}(p) \;=\; \left(\frac{2\sigma^2}{\pi\hbar^2}\right)^{1/4} e^{ia(k-p/\hbar)} e^{-\sigma^2(p/\hbar-k)^2}. \tag{6.1.11}$$

The last factor shows that a state with a Gaussian position wavefunction also has a Gaussian momentum wavefunction, centered at $p = \hbar k$. The momentum wavefunction also contains a phase that depends on $p$; this factor encodes the information about the center of the position wavefunction Gaussian peak, $a$.

It is now clear that our state depends on three physically significant parameters: the average momentum $\hbar k$, the position center $a$, and the position width $\sigma$, which is also the uncertainty in $X$. There is a duality between the position and momentum wavefunctions, for if one of them has a Gaussian magnitude with a linear phase, then so does the other, with parameters that are related by comparing eqs. (6.1.3) and (6.1.11). In particular, the widths of the position and momentum Gaussian wavefunctions are inversely proportional.

Using eq. (6.1.11), one can now obtain

$$\langle P \rangle \;=\; \langle\psi|P|\psi\rangle = \int_{-\infty}^{\infty} dp\; \langle\psi|P|p\rangle\langle p|\psi\rangle = \int_{-\infty}^{\infty} dp\, p|\widetilde{\psi}(p)|^2 = \hbar k. \tag{6.1.12}$$

Similarly,

$$\langle P^2 \rangle \;=\; \int_{-\infty}^{\infty} dp\, p^2 |\widetilde{\psi}(p)|^2 \;=\; \hbar^2(k^2 + 1/4\sigma^2). \tag{6.1.13}$$

It follows that the momentum uncertainty in the state with a Gaussian position wavefunction is

$$\Delta P \;=\; \sqrt{\langle P^2\rangle - \langle P\rangle^2} \;=\; \hbar/2\sigma, \tag{6.1.14}$$

and so the product of the position and momentum uncertainties is

$$(\Delta X)(\Delta P) = \hbar/2. \tag{6.1.15}$$

This result does not depend on the center of the Gaussian $a$, or its width $\sigma$, or its average momentum $\hbar k$. Thus, Gaussian wavefunctions always realize the minimum possible product of uncertainties consistent with the general position-momentum uncertainty relation, eq. (3.3.10).

## 6.2 Time evolution of free-particle state in one dimension

For a free particle with mass $m$ moving in one dimension, the Schrödinger equation is

$$i\hbar\frac{d}{dt}\left|\psi\right\rangle = H\left|\psi\right\rangle = \frac{P^2}{2m}\left|\psi\right\rangle, \tag{6.2.1}$$

where $P$ is the momentum operator. Since $[H, P] = 0$, we know that there is an orthobasis of stationary states that are also eigenstates of $P$ with eigenvalue $p$. Acting on such states, $E = P^2/2m = p^2/2m$. Therefore, for a given $E$ there are exactly two solutions, $p = \pm\sqrt{2mE}$, and the plane-wave stationary states can be labeled

$$|E,R\rangle \;=\; |p\!=\!\sqrt{2mE}\rangle, \qquad\qquad |E,L\rangle \;=\; |p\!=\!-\sqrt{2mE}\rangle, \tag{6.2.2}$$

where $R, L$ is a degeneracy label that tells us whether the particle is moving right or left.

The time-dependent wavefunction for a stationary state with momentum $p$ is, combining eq. (2.8.39) with the time-evolution phase factor,

$$\psi_p(x,t) \;=\; \frac{1}{\sqrt{2\pi\hbar}}e^{i(kx-\omega t)}, \tag{6.2.3}$$

where $k = p/\hbar$ and $\omega = E/\hbar$. The position of constant phase is $x = \omega t/k = Et/p = pt/2m$, so the phase velocity of one of these waves is

$$v_{\text{phase}} \;=\; \omega/k = p/2m, \tag{6.2.4}$$

which is half of the classical velocity $p/m$.

To understand the classical speed of propagation, one must consider the group velocity for wave-packet superpositions of states with a continuous distribution of $k$. It is a general feature of wave kinematics in the presence of dispersion that the group velocity for wavepackets is not $v_{\text{phase}}$, but instead

$$v_{\text{group}} \;=\; \partial\omega/\partial k. \tag{6.2.5}$$

In the present case, with $\omega = E/\hbar$, we have

$$v_{\text{group}} = \frac{\partial}{\partial k}\left(\frac{\hbar k^2}{2m}\right) = \hbar k/m = p/m, \tag{6.2.6}$$

which is the classical value. For wave-packets in quantum mechanics with momentum sharply peaked near $p$, this group velocity corresponds to the motion of the expectation value of the position, as can be seen in general from eq. (3.4.16). Before the end of this section, we will verify this for the particular case of a Gaussian superposition of plane waves.

Using eq. (3.4.7), the unitary time evolution operator is

$$U(t) = \int_{-\infty}^{\infty} dp \; |p\rangle\langle p| \, e^{-itp^2/2m\hbar}. \tag{6.2.7}$$

The matrix element of this operator between different position eigenstates is therefore

$$\langle x|U(t)|x'\rangle = \int_{-\infty}^{\infty} dp \; \langle x|p\rangle\langle p|x'\rangle \, e^{-itp^2/2m\hbar} = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp \, e^{ip(x-x')/\hbar} e^{-itp^2/2m\hbar}. \tag{6.2.8}$$

This integral is again done by the completing-the-square trick, this time by changing the integration variable to $p' = p - m(x - x')/t$, with the result

$$\langle x|U(t)|x'\rangle = \left(\frac{m}{2\pi\hbar i t}\right)^{1/2} e^{im(x-x')^2/2\hbar t}. \tag{6.2.9}$$

By now applying the completeness relation, the time dependence of the wavefunction for a free particle reduces to an integral involving the wavefunction at time $t = 0$,

$$\begin{aligned}\psi(x,t) &= \langle x|\psi(t)\rangle = \langle x|U(t)|\psi(0)\rangle = \int_{-\infty}^{\infty} dx' \; \langle x|U(t)|x'\rangle\langle x'|\psi(0)\rangle \\ &= \left(\frac{m}{2\pi\hbar i t}\right)^{1/2} \int_{-\infty}^{\infty} dx' \, e^{im(x-x')^2/2\hbar t}\psi(x',0).\end{aligned} \tag{6.2.10}$$

In principle, the time evolution of an arbitrary free-particle state has therefore been solved.

As a good example for which the integral can be done analytically, consider the motion and spreading with time of a Gaussian wavefunction. We start at time $t = 0$ with a state $|\psi(0)\rangle$ that has

$$\psi(x',0) = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{ip_0 x'/\hbar} e^{-x'^2/4\sigma^2}, \tag{6.2.11}$$

which as we saw in the previous section can be interpreted as having center at $x' = 0$, width $\sigma$, and average momentum $p_0$. At time $t$, eq. (6.2.10) gives

$$\psi(x,t) = \left(\frac{m}{2\pi\hbar i t}\right)^{1/2} \frac{1}{(2\pi\sigma^2)^{1/4}} \int_{-\infty}^{\infty} dx' \, \exp\left[-\frac{x'^2}{4\sigma^2} + i\frac{m(x-x')^2}{2\hbar t} + i\frac{p_0 x'}{\hbar}\right]. \tag{6.2.12}$$

By the usual completing-the-square integration variable change trick, this becomes, after some algebraic manipulation,

$$\psi(x,t) \;=\; \frac{1}{\sqrt{\sqrt{2\pi}(\sigma + i\hbar t/2m\sigma)}} \exp\left[-\frac{(x - p_0 t/m)^2}{4\sigma^2 + 2i\hbar t/m}\right] \exp\left[i\frac{p_0}{\hbar}\left(x - \frac{p_0 t}{2m}\right)\right], \tag{6.2.13}$$

or, after further rearrangement,

$$\psi(x,t) \;=\; \frac{e^{i\theta(x,t)}}{\sqrt{\sqrt{2\pi}(\sigma + i\hbar t/2m\sigma)}} \exp\left[-\frac{(x - p_0 t/m)^2}{4\sigma^2 + \hbar^2 t^2/m^2\sigma^2}\right], \tag{6.2.14}$$

where

$$\theta(x,t) \;=\; \frac{p_0}{\hbar}\left(x - \frac{p_0 t}{2m}\right) + \frac{\hbar t(x - p_0 t/m)^2}{2m(4\sigma^2 + \hbar^2 t^2/m^2)}. \tag{6.2.15}$$

The probability density at time $t$ is therefore

$$|\psi(x,t)|^2 \;=\; \frac{1}{\sqrt{2\pi(\sigma^2 + \hbar^2 t^2/4m^2\sigma^2)}} \exp\left[-\frac{(x - p_0 t/m)^2}{2(\sigma^2 + \hbar^2 t^2/4m^2\sigma^2)}\right]. \tag{6.2.16}$$

This is a Gaussian function of $x$ whose peak moves to the right at exactly the constant speed $p_0/m$ of naive classical expectation, but whose width grows with time $t$. Using the results of the previous section, we obtain

$$\langle X \rangle \;=\; p_0 t/m, \tag{6.2.17}$$

$$\Delta X \;=\; \sqrt{\sigma^2 + \hbar^2 t^2/4m^2\sigma^2}. \tag{6.2.18}$$

The speed at which the expectation value $\langle X \rangle$ moves is also the group velocity $v_{\text{group}} = \partial\omega/\partial k$. At large $t$, $\Delta X \approx \hbar t/2m\sigma$ grows linearly with time, and ironically is larger for smaller $\sigma$; the more we try to confine the wavefunction initially, the more spread out the particle's probability density support will be at late times. The explanation is that, due to the uncertainty relation eq. (3.3.10), a highly constrained particle position has a larger amplitude for momenta that deviate from the central value.

One can also compute the momentum wavefunction as a function of time, by applying eq. (2.8.41) to eq. (6.2.13). After another integration and more algebraic juggling, one finds

$$\widetilde{\psi}(p,t) \;=\; \left(\frac{2\sigma^2}{\pi\hbar^2}\right)^{1/4} \exp\left[-\sigma^2(p - p_0)^2/\hbar^2\right] e^{ip^2 t/2\hbar m}. \tag{6.2.19}$$

This is remarkably simple; the time dependence is entirely in the complex phase, so

$$\left|\widetilde{\psi}(p,t)\right|^2 \;=\; \sqrt{\frac{2}{\pi}\frac{\sigma}{\hbar}} \exp\left[-2\sigma^2(p - p_0)^2/\hbar^2\right] \tag{6.2.20}$$

does not depend on time at all, despite the fact that the width of the support of the position wavefunction grows with time. This is an illustration of Theorem 5.1.1; because the momentum is a conserved quantity for a free particle Hamiltonian, the probability density to measure the momentum between $p$ and $p + dp$ is a constant in time.

## 6.3 Properties of stationary states in one-dimensional potentials

Consider a particle moving in one dimension in a potential that is given classically by $V(x)$, so that the Hamiltonian operator is

$$H = \frac{P^2}{2m} + V(X). \tag{6.3.1}$$

In general, we want to find stationary states,

$$H \left|\psi_E\right\rangle = E \left|\psi_E\right\rangle . \tag{6.3.2}$$

To accomplish this, consider the position representation, in which $X \to x$ and $P \to -i\hbar d/dx$, by multiplying on the left by $\langle x|$. Then the wavefunction $\psi_E(x) = \langle x|\psi_E\rangle$ obeys the eigenvalue differential equation

$$\frac{d^2}{dx^2}\psi_E(x) = \frac{2m}{\hbar^2}\left[V(x) - E\right]\psi_E(x). \tag{6.3.3}$$

This is the time-independent Schrödinger equation for a spin-less particle in one dimension.

Before doing some special cases, we make some general statements about this problem. First, it may be that the potential $V(x)$ has some special points where it may not be smooth, or may even diverge. We would like to know what can be said about the behavior of the wavefunction $\psi(x)$ at such special points. Consider a particular special point $x_0$. We can integrate eq. (6.3.3) with respect to $x$ over a small neighborhood of that point (dropping the subscript $E$ on the wavefunction) to get

$$\int_{x_0-\epsilon}^{x_0+\epsilon} dx\, \frac{d}{dx}\left(\frac{d\psi}{dx}\right) = \psi'(x_0+\epsilon) - \psi'(x_0-\epsilon) = \frac{2m}{\hbar^2}\int_{x_0-\epsilon}^{x_0+\epsilon} dx\, \left[V(x) - E\right]\psi(x), \tag{6.3.4}$$

where the first equality follows from the fundamental theorem of calculus.

Suppose that $V(x)$ is bounded near $x = x_0$. Then the right side of eq. (6.3.4) tends to 0 as $\epsilon \to 0$, so we can conclude that the first derivative of $\psi(x)$ must be continuous at $x_0$. Note that this holds even if $V(x)$ is discontinuous, as long as it stays finite in the neighborhood of $x = x_0$. It follows that $\psi(x)$ itself is also continuous at $x = x_0$.

Another possibility is that $V(x) = C\delta(x - x_0) + \cdots$, where $C$ is a constant, and the ellipses represent a contribution that is possibly discontinuous but bounded near $x = x_0$. In that case, eq. (6.3.4) reveals that

$$\lim_{\epsilon\to 0}\left[\psi'(x_0+\epsilon)-\psi'(x_0-\epsilon)\right] \;=\; \frac{2m}{\hbar^2}C\psi(x_0), \tag{6.3.5}$$

so that the derivative of the wavefunction at $x_0$ is discontinuous by an amount that we now know. Equation (6.3.5) is consistent with $\psi'(x)$ being bounded in a neighborhood of $x = x_0$, even if it is not continuous. Then, integrating $\int_{x_0-\epsilon}^{x_0+\epsilon} dx\,\psi'(x)$, we obtain the continuity of the wavefunction,

$$\lim_{\epsilon\to 0}\left[\psi(x_0+\epsilon)-\psi(x_0-\epsilon)\right] \;=\; 0. \tag{6.3.6}$$

More generally, **continuity of the wavefunction** is a requirement that we always impose on physically sensible states. The idea is that because $|\psi(x)|^2$ represents the probability density, its value at $x = x_0$ must not depend on how one approaches the point $x_0$.

Yet another possibility is that $V(x) = \infty$ for a whole range $x < x_0$ but it is finite for $x > x_0$. In that case, the particle is forbidden to enter the region of infinite potential, and so $\psi(x) = 0$ for $x \leq x_0$. Now, continuity of the probability density tells us that $\psi(x_0) = 0$, which acts as a boundary condition for the solution in the range $x \geq x_0$. Of course, the same holds if the inequalities are reversed: if $V(x) = \infty$ for $x > x_0$ but it is finite for $x < x_0$, then $\psi(x) = 0$ for $x \geq x_0$.

A bound state is a stationary state whose wavefunction vanishes at large distances. A useful result is that bound states in one-dimensional potentials in quantum mechanics never have degenerate energy eigenvalues. To prove it, suppose there are two states $|\psi_1\rangle$ and $|\psi_2\rangle$ with the same energy eigenvalue $E$. In the position representation, the wavefunctions satisfy

$$-\frac{\hbar^2}{2m}\psi_1'' + V\psi_1 \;=\; E\psi_1, \qquad\qquad -\frac{\hbar^2}{2m}\psi_2'' + V\psi_2 \;=\; E\psi_2. \tag{6.3.7}$$

Multiplying the first equation by $\psi_2$ and the second by $\psi_1$ and taking the difference gives $\psi_2\psi_1'' - \psi_1\psi_2'' = 0$, or

$$\frac{d}{dx}(\psi_2\psi_1' - \psi_1\psi_2') = 0, \tag{6.3.8}$$

so that integrating with respect to $x$ gives

$$\psi_2\psi_1' - \psi_1\psi_2' = C \tag{6.3.9}$$

where $C$ is a constant of integration. Now, by definition the bound states have $\psi_1 = \psi_2 = 0$ when $x = \pm\infty$, so evaluating eq. (6.3.9) at $x = \infty$, we learn that $C$ must be 0. Therefore $\psi_1'/\psi_1 = \psi_2'/\psi_2$, or

$$\frac{d}{dx}(\ln\psi_1) = \frac{d}{dx}(\ln\psi_2), \tag{6.3.10}$$

which integrates to

$$\ln(\psi_1) = \ln(\psi_2) + K \tag{6.3.11}$$

where $K$ is another constant of integration. So, $\psi_1 = e^K \psi_2$, and since $e^K$ is a non-zero constant, $\psi_1$ and $\psi_2$ are proportional and therefore represent the same state. Note that in this proof it is crucial that we are working in one dimension. In two or three dimensions, bound states certainly can have degenerate energies, as we will see by finding examples in Chapter 9.

## 6.4 Particle in a one-dimensional box

Consider a particle of mass $m$ confined to a box of length $L$, so that its potential energy as a function of $x$ is

$$V(x) = \begin{cases} 0 & \text{for } |x| < L/2, \\ \infty & \text{for } |x| > L/2. \end{cases} \tag{6.4.1}$$

Since this is an even function of $x$, we know from the discussion in section 5.4 that there should be simultaneous eigenstates of energy and parity. Let us find them.

In the region $|x| < L/2$, eq. (6.3.3) becomes

$$\psi'' \;=\; -\frac{2mE}{\hbar^2}\psi. \tag{6.4.2}$$

The general solution to this differential equation is

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{6.4.3}$$

where $A$ and $B$ are constants and

$$k = \sqrt{2mE}/\hbar. \tag{6.4.4}$$

Because the particle is completely confined to the box region, one must have $\psi(x) = 0$ for $|x| \geq L/2$, and so by continuity of the wavefunction, $\psi(L/2) = 0$ and $\psi(-L/2) = 0$, or

$$Ae^{ikL/2} + Be^{-ikL/2} \;=\; 0, \tag{6.4.5}$$

$$Ae^{-ikL/2} + Be^{ikL/2} \;=\; 0. \tag{6.4.6}$$

This system of equations will have a non-trivial solution for $A$ and $B$ if and only if the matrix

$$M \;=\; \begin{pmatrix} e^{ikL/2} & e^{-ikL/2} \\ e^{-ikL/2} & e^{ikL/2} \end{pmatrix} \tag{6.4.7}$$

is not invertible, which means $\text{Det}(M) = e^{ikL} - e^{-ikL} = 2i\sin(kL) = 0$. The solutions to the particle-in-a-box eigenvalue problem therefore must have $k = n\pi/L$, where $n$ is an integer. It follows that $B = -e^{ikL}A = (-1)^{n+1}A$.

For odd $n$, we have $B = A$, and so for $-L/2 \leq x \leq L/2$,

$$\psi_n(x) \;=\; 2A\cos(n\pi x/L). \tag{6.4.8}$$

To normalize the states to unity, we require

$$1 = \int_{-\infty}^{\infty} dx\,|\psi_n(x)|^2 \;=\; 4|A|^2 \int_{-L/2}^{L/2} dx\,\cos^2(n\pi x/L) = 2|A|^2 L, \tag{6.4.9}$$

and so we can choose $A = 1/\sqrt{2L}$, yielding the even-parity solutions

$$\psi_n(x) = \sqrt{\frac{2}{L}}\cos(n\pi x/L) \qquad (n = 1, 3, 5, \ldots). \tag{6.4.10}$$

Note that we do not need to include negative $n$, because those just have the same wavefunction up to a physically irrelevant sign, and are therefore the same states. Similarly, for even $n$, one finds the odd-parity solutions

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin(n\pi x/L) \qquad (n = 2, 4, 6, \ldots). \tag{6.4.11}$$

Here, not only do we not need negative $n$, but also the case $n = 0$ is excluded, because it would result in $\psi_0(x) = 0$, which would correspond to the null ket, which is not a physical state. For both odd and even $n$, eq. (6.4.4) gives the allowed energy levels:

$$E_n = \frac{\hbar^2\pi^2 n^2}{2mL^2} \qquad (n = 1, 2, 3, \ldots). \tag{6.4.12}$$

These are non-degenerate and discrete, as is true for any bound-state solutions in one dimension.

The ground state energy for the particle in a box, $E_1 = \hbar^2\pi^2/2mL^2$, becomes very large as the confining box is taken smaller ($L \rightarrow 0$). This can be seen to be in accord with the uncertainty relation eq. (3.3.10), as follows. First, note that $\langle P \rangle = 0$ for each of the stationary states; this can be obtained either by direct computation in terms of the wavefunction, or seen as a consequence of the parity selection rule discussed at the end of section 5.4. It follows that

$$\langle H \rangle \;=\; \frac{1}{2m}\langle P^2 \rangle \;=\; \frac{1}{2m}(\Delta P)^2. \tag{6.4.13}$$

Now the uncertainty relation eq. (3.3.10) says that $(\Delta P)^2 \geq (\hbar/2\Delta X)^2$, and it is also clear that $\langle X \rangle = 0$, again either by direct computation or as a consequence of the parity selection rule. Therefore, $(\Delta X)^2 = \langle X^2 \rangle$, and we have a bound

$$\langle H \rangle \;\geq\; \frac{\hbar^2}{8m^2\langle X^2 \rangle}. \tag{6.4.14}$$

Since the particle is confined to a box of length $L$, a crude, conservative estimate valid for any state is that $\langle X^2 \rangle \leq (L/2)^2$. Using this to compare our estimate of the lower bound on $\langle H \rangle$

from the uncertainty relation to eq. (6.4.12) with $n = 1$, we see that the latter is a factor of $\pi^2$ larger. For the ground state, a much better estimate of $\langle X^2 \rangle$ comes from actually computing it, with the result

$$\langle X^2 \rangle \;=\; L^2 \left( \frac{1}{12} - \frac{1}{2\pi^2} \right), \tag{6.4.15}$$

so

$$\langle H \rangle \;\geq\; \frac{\hbar^2 \pi^2}{2mL^2} \left( \frac{3}{\pi^2 - 6} \right). \tag{6.4.16}$$

Comparing again with eq. (6.4.12) with $n = 1$, we see that the ground state energy is a factor of $\pi^2/3 - 2$, or about 1.29, larger than the estimate eq. (6.4.16) of its lower bound following from the uncertainty relation.

## 6.5 Bound states for the one-dimensional square well

As a generalization of the previous section, consider a particle of mass $m$ in a finite square-well potential with linear width $L$ and energy depth $V_0$:

$$V(x) = \begin{cases} V_0 & \text{for } x \leq -L/2 \qquad \text{(region I)}, \\ 0 & \text{for } -L/2 < x < L/2 \quad \text{(region II)}, \\ V_0 & \text{for } x \geq L/2 \qquad \text{(region III)}. \end{cases} \tag{6.5.1}$$

Before proceeding, we note that the special case $V_0 \rightarrow \infty$ should give the results of the previous section. Again, we expect to find energy eigenstate solutions with definite parity, because the potential is invariant under $x \rightarrow -x$.

The strategy for finding the stationary states is to first solve the differential equation (6.3.3) separately in each of the three regions I, II, and III, and then stitch these solutions together using eqs. (6.3.5) and (6.3.6) as boundary conditions at the points $x = \pm L/2$.

In region II, the differential equation is exactly the same as for the particle in a box,

$$\psi'' \;=\; -\frac{2mE}{\hbar^2}\psi, \tag{6.5.2}$$

and so the general solution is an arbitrary linear combination of $e^{ikx}$ and $e^{-ikx}$, where $k = \sqrt{2mE}/\hbar$, as before. So, we can write

$$\psi_{\text{II}}(x) = A\cos(kx) + B\sin(kx). \tag{6.5.3}$$

In regions I and III, the differential equation is

$$\psi'' \;=\; \frac{2m(V_0 - E)}{\hbar^2}\psi. \tag{6.5.4}$$

Assume that $E < V_0$, so that we will have a bound state. Then, defining

$$\kappa = \sqrt{2m(V_0 - E)}/\hbar, \tag{6.5.5}$$

the general solutions in regions I and III are

$$\psi_{\rm I}(x) = Ce^{\kappa x} + De^{-\kappa x}, \tag{6.5.6}$$

$$\psi_{\rm III}(x) = Fe^{\kappa x} + Ge^{-\kappa x}. \tag{6.5.7}$$

The coefficient $D$ must be 0, because otherwise $\psi_I(x)$ would blow up for $x \rightarrow -\infty$, giving a non-normalizable unphysical solution. Similarly, from requiring $\psi_{\rm III}(x)$ to be well-behaved at $x \rightarrow \infty$, we get $F = 0$. It remains to solve for $A$, $B$, $C$, and $G$, using the boundary conditions at the points $x = -L/2$ and $L/2$.

Let us first look for wavefunction solutions that are even under the parity transformation $x \rightarrow -x$. It follows that $B = 0$ and $G = C$, and they must have the form

$$\psi(x) = \begin{cases} Ce^{\kappa x} & \text{for } x \leq -L/2, \\ A\cos(kx) & \text{for } -L/2 \leq x \leq L/2, \\ Ce^{-\kappa x} & \text{for } x \geq L/2. \end{cases} \tag{6.5.8}$$

Now we can apply the requirements that the wavefunction and its first derivatives are both continuous at $x = L/2$, as proved on general grounds in eqs. (6.3.5) and (6.3.6). In the present case, these conditions amount to

$$A\cos(kL/2) = Ce^{-\kappa L/2}, \tag{6.5.9}$$

$$-Ak\sin(kL/2) = -C\kappa e^{-\kappa L/2}. \tag{6.5.10}$$

By taking the ratio of these equations, one obtains $k\tan(kL/2) = \kappa$. It is convenient to define dimensionless quantities $X = kL/2$ and $Y = \kappa L/2$, so that

$$X\tan X = Y, \qquad \text{(even parity)}, \tag{6.5.11}$$

$$X^2 + Y^2 = mV_0L^2/2\hbar^2, \tag{6.5.12}$$

where eq. (6.5.12) follows from eq. (6.5.5) and $E = \hbar^2k^2/2m$.

It is not possible to solve the simultaneous transcendental equations (6.5.11) and (6.5.12) analytically, but one can use graphical methods to understand the solutions and then obtain numerical results. In Figure 6.5.1 we graph in the $X, Y$ plane the curves $Y = X\tan X$ (darker solid curves) and the circle eq. (6.5.12) (dashed), for some sample values (1.4, 4, and 8) of the dimensionless radius $R = \sqrt{mV_0L^2/2\hbar^2}$. For a given value of $R$, the intersections provide the numerical eigenvalue solutions for $X$ and $Y$, and thus for $k$ and $\kappa$. We only need to consider positive $X$ and $Y$, because $\kappa$ is positive, and $k$ can be taken positive without loss of generality.

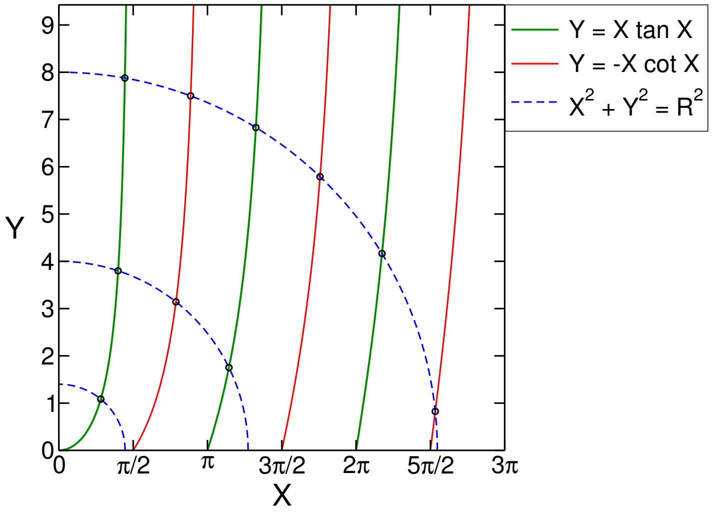

Figure 6.5.1: Graphical solutions for the bound energy eigenstates of a particle of mass $m$ in a one-dimensional square well potential with depth $V_0$ and width $L$. The dashed circles are $X^2 + Y^2 = R^2 = mV_0L^2/2\hbar^2$, for sample values $R = 1.4, 4, 8$. Even-parity solutions are the intersections with $Y = X \tan X$ (darker, green, solid curves) and odd-parity solutions are the intersections with $Y = -X \cot X$ (lighter, red, solid curves). The energy eigenvalues for each solution are $E = 2\hbar^2X^2/mL^2$. The cases $R = 1.4$, 4, and 8 are seen to have 1, 3, and 6 bound state solutions, respectively.

The graphical approach makes clear that there is always at least one bound-state solution, no matter how small $V_0$ (and thus $R$) is. The ground state, which we will call $|\psi_1\rangle$ with energy eigenvalue $E_1$, is the solution with smallest $X$, which always has even parity and occurs for $0 < X < \pi/2$, or $0 < k < \pi/L$, and so

$$0 < E_1 < \hbar^2\pi^2/2mL^2. \tag{6.5.13}$$

The lower bound corresponds to small $R$, which means the limit of small $V_0 \ll \hbar^2/mL^2$. Conversely, the upper bound is relevant for large $R$ and thus $V_0 \gg \hbar^2/mL^2$, the limit of the particle completely confined to a box as treated in the previous section. As $V_0$ is increased (for fixed $m, L$), the radius of the dashed circle grows, so the number of solutions increases, but for any finite $V_0$ the number of bound state solutions is always finite.

Consider the lone bound state solution in the limit of small $V_0$, which means $X$ and $Y$ are also very small. By expanding $X \tan X$ to quadratic order in $X^2$, and solving eqs. (6.5.11) and (6.5.12) in that limit, we obtain

$$E_1 = \frac{\hbar^2k^2}{2m} \approx V_0\left(1 - \frac{mV_0L^2}{2\hbar^2} + \cdots\right). \tag{6.5.14}$$

Because this energy is only slightly lower than $V_0$, the state is very weakly bound, but it always exists, no matter how small $V_0$ is.

Now consider the stationary states that are odd under parity ($x \to -x$), with wavefunctions that are therefore of the form

$$\psi(x) = \begin{cases} -Ce^{\kappa x} & \text{for } x \leq -L/2, \\ B\sin(kx) & \text{for } -L/2 \leq x \leq L/2, \\ Ce^{-\kappa x} & \text{for } x \geq L/2. \end{cases} \tag{6.5.15}$$

The continuity of the wavefunction and its derivative give

$$B\sin(kL/2) = Ce^{-\kappa L/2}, \tag{6.5.16}$$

$$Bk\cos(kL/2) = -C\kappa e^{-\kappa L/2}. \tag{6.5.17}$$

Defining $X$ and $Y$ in exactly the same way as before, we now have

$$-X\cot X = Y, \qquad \text{(odd parity)} \tag{6.5.18}$$

in place of eq. (6.5.18), with the same equation for $X^2+Y^2 = R^2$. The graph of $Y = -X\cot X$ is shown in Figure 6.5.1 as the lighter solid lines. This time, we see that if $V_0$ is small enough, there will not be any odd-parity bound-state solution. In order for there to exist at least one bound-state solution with an odd wavefunction, the dashed circle must have a large enough radius to intersect with the curve $Y = -X\cot X$ for positive $X$ and $Y$, specifically, $\sqrt{\frac{mV_0L^2}{2\hbar^2}} > \pi/2$, so that $V_0 > \pi^2\hbar^2/2mL^2$. In this case, the first excited state will have $\pi/2 < X < \pi$, so $\hbar^2\pi^2/2mL^2 < E_2 < 2\hbar^2\pi^2/2mL^2$.

Combining the information for even and odd parity states (which have odd and even $n$, respectively) from Figure 6.5.1 we can see that the energy levels alternate between even and odd parity, and the bound state $|\psi_n\rangle$ exists if and only if the potential well is deep enough, which requires that $R^2$ is sufficiently large,

$$\frac{mV_0L^2}{2\hbar^2} > \frac{\pi^2}{4}(n-1)^2. \tag{6.5.19}$$

If $|\psi_n\rangle$ does exist as a bound state, then it has $\pi(n-1)/2 < X_n < \pi n/2$, which implies

$$\frac{\hbar^2\pi^2(n-1)^2}{2mL^2} < E_n < \frac{\hbar^2\pi^2n^2}{2mL^2} \qquad (n = 1, 2, 3, \ldots). \tag{6.5.20}$$

Here $E_n$ will approach its upper bound in the limit of large $V_0$, in agreement with the result found in section 6.4.

The square well potential also has a continuum of unbound states with energies $E \geq V_0$, with wavefunctions whose magnitudes approach a constant for large distances $|x|$. These are also known as scattering states. We will discuss the problem of scattering from the one-dimensional square well in section 6.6.

## 6.6 Scattering problems in one dimension

**Scattering theory** deals with unbound quantum states with continuous energy eigenvalues. The particles described by these states can be thought of as originating far away, in a measured or controlled configuration, typically a superposition of free plane waves moving in a common

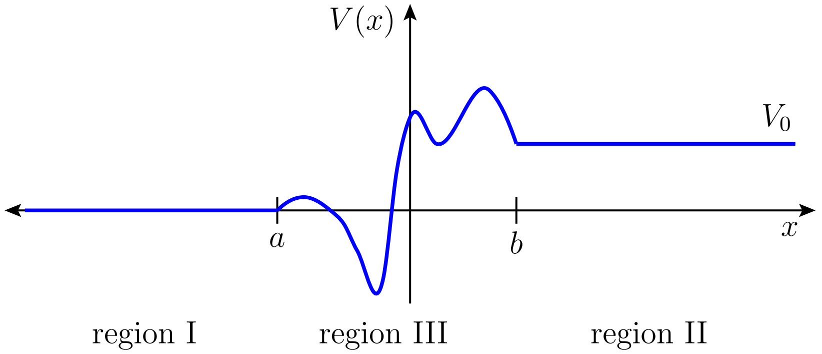

Figure 6.6.1: A generic one-dimensional scattering potential of the type in eq. (6.6.1). The potential $V_0$ on the far right can be either positive or negative.

direction. In a localized region of space, the particles then interact with a potential, or with another group of particles, after which they escape to large distances again where they can be measured. In practice, one can learn about the potential, or interactions between particles, by studying the asymptotic forms of the unbound states. In this section we will consider potential scattering problems in the simplified realm of one dimension, and turn to the problem of scattering in three dimensions in Chapter 20.

Consider two asymptotic regions I and II with constant potentials, and an intermediate region III where the potential can be arbitrary, so

$$V(x) = \begin{cases} 0 & \text{region I} \quad (x < a), \\ U(x) & \text{region III} \quad (a < x < b), \\ V_0 & \text{region II} \quad (x > b), \end{cases} \tag{6.6.1}$$

as illustrated in Figure 6.6.1. The constant potential in region I, which will contain the incident particles, is taken to be 0, by subtraction from $V(x)$ if necessary. This entails no loss of generality, because as usual the effect of a constant contribution to the energy can be absorbed into a global phase that is the same for all states. The potential $V_0$ in region II can be either positive or negative. We will be interested only in unbound states with energy $E > 0$, although there may also be bound states if the potential $U(x)$ in region III goes negative for some $x$.

The stationary states for this potential have wavefunctions that might be complicated in region III, but they are definitely simple in regions I and II. Assuming that $E > V_0$, one class of such stationary states is

$$\psi(x,t) = \left(Ae^{ikx} + Be^{-ikx}\right) e^{-iEt/\hbar}, \qquad \text{(region I)}, \tag{6.6.2}$$

$$\psi(x,t) = Ce^{ik'x}e^{-iEt/\hbar}, \qquad \text{(region II)}. \tag{6.6.3}$$

Here the component $A$ corresponds to an initial beam of particles moving to the right in region I from $x = -\infty$, and $B$ corresponds to a reflected component going back to $x = -\infty$. In region

II, there is only a transmitted component moving to the right; as a boundary condition we are imposing that there is no left-moving component that would correspond to incident particles arriving from $x = +\infty$.

Now, for a given energy $E$ and normalization $A$ treated as known inputs characteristic of the incident beam of particles, we can solve for $k$, $k'$ and then, in principle, for $B$ and $C$. The easy part is that from the Schrödinger equation,

$$\frac{\hbar^2 k^2}{2m} = E, \qquad \frac{\hbar^2 k'^2}{2m} = E - V_0. \tag{6.6.4}$$

To solve for $B$ and $C$ requires the scattering potential $U(x)$ to be specified. Once $B$ and $C$ are known, eq. (5.6.13) can be used to find the probability and current densities in each region. In region I, the probability density is

$$\rho = |A|^2 + |B|^2 + 2\text{Re}[A^* B]\cos(2kx) + 2\text{Im}[A^* B]\sin(2kx). \tag{6.6.5}$$

The last two terms will each give 0 after averaging over a range of $x$ that is large compared to $1/k$. The current density in region I is

$$J = \frac{\hbar k}{m}(|A|^2 - |B|^2), \tag{6.6.6}$$

where the cross-terms between $A$ and $B$ canceled completely, even without doing any averaging. Meanwhile, in region II, the probability density and current are

$$\rho = |C|^2, \tag{6.6.7}$$

$$J = \frac{\hbar k'}{m}|C|^2. \tag{6.6.8}$$

The interpretation of the probability densities of eqs. (6.6.5) and (6.6.7) is

$$\text{density of particles in} \begin{pmatrix} \text{incident} \\ \text{reflected} \\ \text{transmitted} \end{pmatrix} \text{beam} = \begin{pmatrix} |A|^2 \\ |B|^2 \\ |C|^2 \end{pmatrix}, \tag{6.6.9}$$

while the interpretation of the current densities of eqs. (6.6.6) and (6.6.8) is

$$\text{flux} = \text{particles/time in} \begin{pmatrix} \text{incident} \\ \text{reflected} \\ \text{transmitted} \end{pmatrix} \text{beam} = \begin{pmatrix} \hbar k|A|^2/m \\ -\hbar k|B|^2/m \\ \hbar k'|C|^2/m \end{pmatrix} = \begin{pmatrix} J_A \\ J_B \\ J_C \end{pmatrix}. \tag{6.6.10}$$

The effect of the scattering potential on the incident mono-energetic beam can therefore be given in terms of reflection and transmission ratios, reminiscent of quantities in classical optics,

$$R = \frac{|J_B|}{|J_A|} = \frac{|B|^2}{|A|^2}, \qquad T = \frac{|J_C|}{|J_A|} = \frac{k'|C|^2}{k|A|^2}. \tag{6.6.11}$$

In specific problems, we can always set $A = 1$ at the beginning, since only the ratios $B/A$ and $C/A$ are needed, and $B$ and $C$ will always be proportional to $A$ due to the linearity of the wavefunction in the Schrödinger equation.

We cannot find $R$ or $T$ without specifying the scattering potential $U(x)$. However, in general, probability conservation implies that they are related by

$$R + T = 1. \tag{6.6.12}$$

To prove this, we note that for a stationary state,

$$\frac{\partial J}{\partial x} \;=\; -\frac{\partial \rho}{\partial t} \;=\; -\frac{\partial}{\partial t}|\psi(x,t)|^2 \;=\; -\frac{\partial}{\partial t}|\psi(x,t_0)e^{-i(t-t_0)E/\hbar}|^2 \;=\; 0, \tag{6.6.13}$$

so, by the fundamental theorem of calculus,

$$0 \;=\; \int_{-\infty}^{\infty} dx\, \frac{\partial J}{\partial x} \;=\; J(\infty) - J(-\infty). \tag{6.6.14}$$

This can be rewritten as

$$J_C = |J_A| - |J_B|, \tag{6.6.15}$$

from which eq. (6.6.12) follows immediately.

To find $R$ and $T$ in specific examples, it is necessary to solve the Schrödinger equation in region III and to apply boundary conditions in the form of continuity of the wavefunction, and its derivative if the potential is finite, at the boundaries between adjacent regions. If the potential at a boundary has a delta function, then eq. (6.3.5) can be used to obtain the boundary condition for the first derivative of the wavefunction.

As a first example, consider a step-function potential

$$V(x) \;=\; \begin{cases} 0 & \text{region I } (x < 0), \\ V & \text{region II } (x > 0), \end{cases} \tag{6.6.16}$$

for which region III does not exist, and regions I and II meet at $x = 0$. Continuity of the wavefunction eq. (6.6.3) and its first derivative at $x = 0$ give

$$\begin{aligned} 1 + B \;&=\; C, & (6.6.17)\\ ik + (-ik)B \;&=\; ik'C. & (6.6.18) \end{aligned}$$

Here we have taken the opportunity to set $A = 1$, since we will be interested in the ratios $R$ and $T$. The solutions for $B$ and $C$ are

$$B \;=\; \frac{k-k'}{k+k'}, \qquad\qquad C \;=\; \frac{2k}{k+k'}. \tag{6.6.19}$$

The reflection and transmission coefficients are therefore

$$R = \left|\frac{1-k'/k}{1+k'/k}\right|^2, \qquad T = \frac{4k'/k}{(1+k'/k)^2}, \tag{6.6.20}$$

where

$$k'/k = \sqrt{1-V/E}. \tag{6.6.21}$$

Note that we need $E > V$ in order for $k'$ and $T$ to be real. Otherwise, the wavefunction for $x > 0$ is instead an exponential of the form $\psi_{\rm II}(x) = Ce^{-\kappa x}$ with $\hbar^2\kappa^2/2m = V - E$, and one finds $R = 1$. Since the flux decreases exponentially in region II rather than maintaining constant magnitude, $T = 0$ in that case; the particles are all reflected, although they still have a non-zero probability to be found at any point $x > 0$, proportional to $e^{-2\kappa x}$. Region II in this case is said to be a classically forbidden region.

As a second example, consider the symmetric rectangular barrier potential

$$V(x) = \begin{cases} 0 & \text{region I} \;\; (x < -a/2), \\ V & \text{region III} \;\; (|x| < a/2), \\ 0 & \text{region II} \;\; (x > a/2), \end{cases} \tag{6.6.22}$$

as illustrated in Figure 6.6.2.

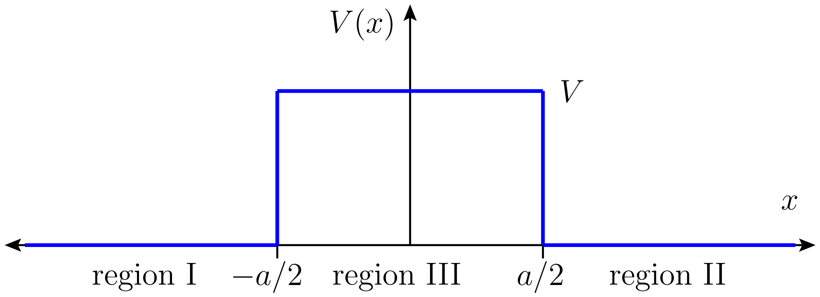

Figure 6.6.2: A one-dimensional scattering potential of the type in eq. (6.6.22). The potential can be either a barrier (if $V > 0$, as shown) or a well (if $V < 0$).

Let us first consider the case that $E > V > 0$, so that classically, transmission past the barrier is possible. Then we can write the wavefunctions in the three regions as

$$\psi_{\rm I}(x) = e^{ikx} + Be^{-ikx}, \tag{6.6.23}$$

$$\psi_{\rm III}(x) = De^{ik'x} + Fe^{-ik'x}, \tag{6.6.24}$$

$$\psi_{\rm II}(x) = Ce^{ikx}, \tag{6.6.25}$$

where $k' = \sqrt{2m(E-V)}/\hbar$ and $k = \sqrt{2mE}/\hbar$ are real numbers, and again we choose $A = 1$. There are four remaining unknowns, $B, C, D, F$, and only two of these enter into the measurable quantities $R = |B|^2$ and $T = |C|^2$. Continuity of the wavefunction and its first derivative at $x = -a/2$, using $\psi_{\rm I}$ and $\psi_{\rm III}$, give us

$$\begin{aligned} e^{-ika/2} + Be^{ika/2} &= De^{-ik'a/2} + Fe^{ik'a/2}, \qquad (6.6.26) \\ ike^{-ika/2} - ikBe^{ika/2} &= ik'De^{-ik'a/2} - ik'Fe^{ik'a/2}, \qquad (6.6.27) \end{aligned}$$

Also, at $x = a/2$ we find from $\psi_{\rm II}$ and $\psi_{\rm III}$,

$$\begin{aligned} Ce^{ika/2} &= De^{ik'a/2} + Fe^{-ik'a/2}, \qquad (6.6.28) \\ ikCe^{ika/2} &= ik'De^{ik'a/2} - ik'Fe^{-ik'a/2}. \qquad (6.6.29) \end{aligned}$$

It is convenient to first solve eqs. (6.6.28) and (6.6.29) for $D$ and $F$ in terms of $C$, and plug the results into eqs. (6.6.26) and (6.6.27) which then involve only $B$ and $C$ as unknowns. The resulting eq. (6.6.26) and eq. (6.6.27) then combine to give

$$B = i\frac{k'^2 - k^2}{2kk'}\sin(k'a)\, C. \qquad (6.6.30)$$

Now, since we also know $R + T = |B|^2 + |C|^2 = 1$, we can solve to get

$$T = |C|^2 = \left[1 + \left(\frac{k'^2 - k^2}{2kk'}\right)^2 \sin^2(k'a)\right]^{-1} \qquad (6.6.31)$$

for the transmission coefficient. Then, $R$ is just $1 - T$.

To express the result directly in terms of the incident energy and the height of the potential barrier, we can now plug in $k' = k\sqrt{1 - V/E}$, to get

$$T = \left[1 + \frac{V^2}{4E(E-V)}\sin^2\left(\frac{a}{\hbar}\sqrt{2m(E-V)}\right)\right]^{-1}, \qquad (E \geq V). \qquad (6.6.32)$$

As a check, if $V = 0$, then there is no barrier, and $T = 1$ and $R = 0$. In the high-energy limit, $E \gg V$ gives $T \approx 1$, and the barrier is almost transparent, as the potential is too weak to have much effect on the incident particles. It is perhaps more surprising that even if $V$ and $E$ are comparable, we also get $T = 1$ whenever $\sin(k'a) = 0$, which occurs if $k' = n\pi/a$ for any integer $n$. This shows that the barrier is transparent to the incident mono-energetic particle beam if its width $a$ is equal to an integer times half of the de Broglie wavelength $2\pi/k'$ inside the barrier. The discrete energies at which this occurs, called **scattering resonances**, are

$$E = E_n + V, \qquad E_n = \frac{\hbar^2\pi^2 n^2}{2ma^2}, \qquad (n = 1, 2, 3, \ldots). \qquad (6.6.33)$$

These $E_n$ coincide with the energy eigenvalues for a particle in a box of width $a$, but with walls of infinite potential height, as found in section 6.4. Finally, note that if $E \approx V$, then we have

$\sin^2\left(a\sqrt{2m(E-V)}/\hbar\right) \approx 2a^2m(E-V)/\hbar^2$, so

$$T \approx \left(1 + \frac{mVa^2}{2\hbar^2}\right)^{-1}, \qquad (E \approx V). \tag{6.6.34}$$

This is the transmission coefficient for the case that classical transmission past the barrier is just barely possible; for $\hbar \to \infty$ it approaches 1.

Now we turn to the case that $E < V$, so that classically the particle would not be expected to make it past the barrier. Then, within the barrier region III, we have instead of eq. (6.6.24),

$$\psi_{\text{III}}(x) \;=\; De^{-\kappa x} + Fe^{\kappa x}, \tag{6.6.35}$$

where

$$\kappa = \sqrt{2m(V-E)}/\hbar. \tag{6.6.36}$$

All of the subsequent algebra is the same, but with $k' \to i\kappa$, so that the sine function is replaced by a hyperbolic sine, and

$$T = \left[1 + \frac{V^2}{4E(V-E)}\sinh^2\left(\frac{a}{\hbar}\sqrt{2m(V-E)}\right)\right]^{-1}, \qquad (E \leq V). \tag{6.6.37}$$

In this case, $T < 1$ always, but it is never 0. This is an example of **quantum tunneling**. Classically, there would be no transmission at all, but the Schrödinger equation gives a non-zero probability for the particle to enter the classically forbidden region and emerge on the right side of the barrier. For $E \ll V$, one finds that $T \to 0$; the transmission due to tunneling becomes exponentially small in the limit of an incident energy much smaller than the barrier potential height. The results for the transmission coefficient $T$ are shown as a function of the incident particle energy $E$ in Figure 6.6.3, for two different values of the barrier height $V$.

So far we have assumed that the potential barrier height in Figure 6.6.2 is positive, with $V > 0$. Now let us consider the case of scattering from a potential well, so that $V < 0$ in eq. (6.6.22). In that case, all of the same algebra goes through as before, with $V = -|V|$. So, we have

$$T \;=\; \left[1 + \frac{V^2}{4E(E+|V|)}\sin^2\left(\frac{a}{\hbar}\sqrt{2m(E+|V|)}\right)\right]^{-1}, \qquad (E > 0 > V). \tag{6.6.38}$$

Classically, there would never be reflection, as the particle incident from the left would have enough momentum and energy to follow a trajectory that takes it ineluctably to $x = +\infty$. The prediction of quantum mechanics in the low-energy limit is very different, as eq. (6.6.38) gives $T \to 0$ for $E \to 0$. In the high-energy limit, one finds $T \to 1$ for $E \to \infty$, in agreement with the

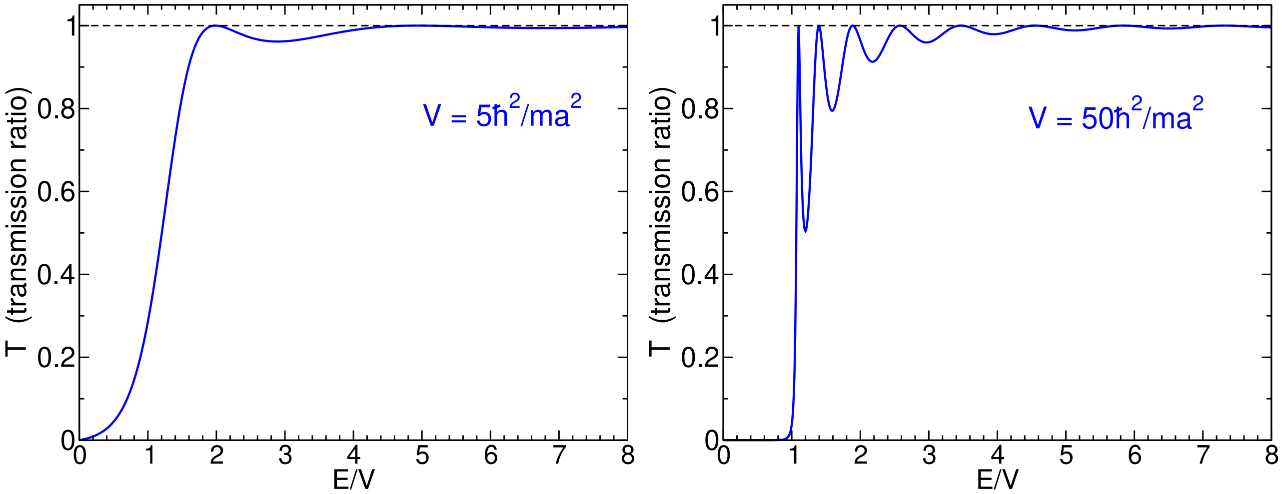

Figure 6.6.3: Transmission ratio $T$ for scattering from a one-dimensional rectangular barrier of width $a$ and height $V = 5\hbar^2/ma^2$ (left) or $V = 50\hbar^2/ma^2$ (right), as a function of the energy $E$ of the incident particles of mass $m$. For $E/V < 1$, the transmission is classically forbidden, and the non-zero $T$ is an example of quantum tunneling. For $E/V > 1$, the barrier is perfectly transparent if its width is equal to an integer multiple of half of the particle's de Broglie wavelength inside the barrier. Transparency also occurs in the limit $E \gg V$.

classical expectation. Just as we found for $V > 0$, there are scattering resonances that occur when $E = E_n - |V|$ where $E_n = \hbar^2\pi^2 n^2/2ma^2$ are the binding energies of bound states in a box (with walls at infinite potential) of width $a$. The results for the transmission coefficient $T$ are shown as a function of the incident particle energy $E$ in Figure 6.6.4, for two different values of the well depth $|V|$. For larger negative $V$, the resonance energies are more sharply defined.

More generally, including in three-dimensional problems, an attractive potential can become almost transparent to scattering for certain resonance energies, which are determined by the geometry of the potential and the de Broglie wavelengths of the incident particles. This phenomenon is known as the **Ramsauer–Townsend effect**, as it was first observed independently by Carl Ramsauer and John S. Townsend in 1921, before its subsequent explanation by quantum mechanics, in the scattering of electrons from noble gas atoms Ar, Kr, and Xe. The qualitative explanation for this is that close to the positively charged nucleus, the scattering electrons see a spherically symmetric attractive potential well that is cut off at larger distances by the screening of the nuclear charge due to the atomic electrons. This geometry leads to a strong suppression of the scattering cross-section, corresponding to near transparency for incident electron energies $E \approx 0.7$ eV. The preceding results for the first peak in $T$ in the one-dimensional well scattering problem with large negative $V$ are a rough qualitative model for this phenomenon.

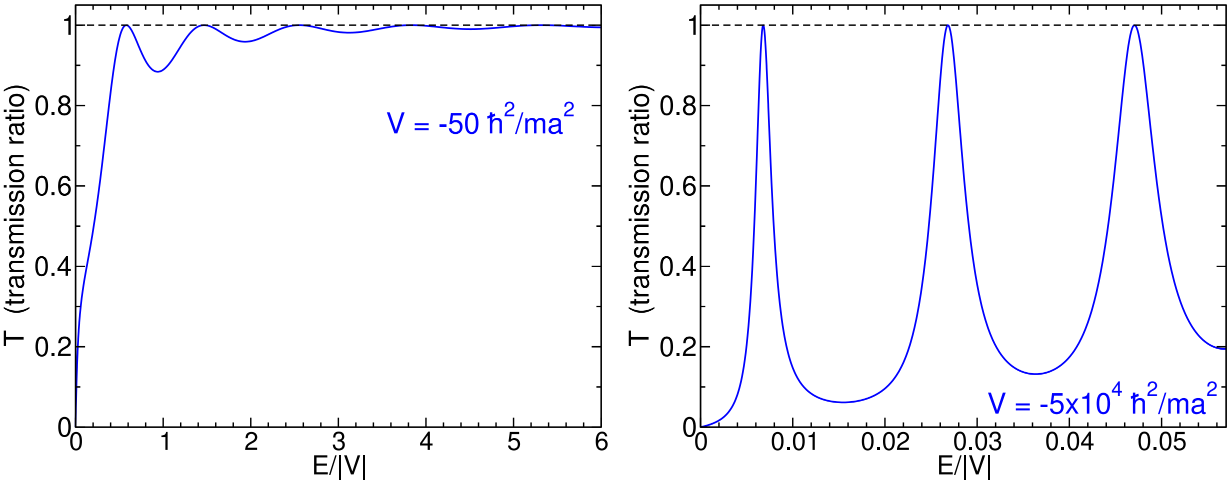

Figure 6.6.4: Transmission ratio $T$ for scattering from a one-dimensional well of width $a$ and $V = -50\hbar^2/ma^2$ (left) or $V = -5 \times 10^4\,\hbar^2/ma^2$ (right). In the very low energy limit, the particles are entirely reflected. Transparency occurs in the limit of large $E$, and also when $E$ is equal to a resonance energy, given by the eigenvalues $E_n = \frac{\hbar^2\pi^2n^2}{2ma^2}$ of a particle confined to a box of width $a$.

# 7 The harmonic oscillator

## 7.1 The reasonable effectiveness of the harmonic oscillator

A harmonic oscillator is any system in which the restoring force is proportional to the displacement from equilibrium. Consider the case of a particle of mass $m$ moving in one dimension. In terms of energy, the classical harmonic oscillator is defined by

$$\text{kinetic energy} \;=\; \frac{p^2}{2m}, \qquad \text{potential energy} \;=\; V(x) \;=\; \frac{1}{2}m\omega^2 x^2, \tag{7.1.1}$$

where the angular frequency $\omega$ has units of 1/time and parameterizes the strength of the restoring force, according to

$$\text{force} = -\frac{\partial V}{\partial x} = -m\omega^2 x = -kx, \tag{7.1.2}$$

with $\omega = \sqrt{k/m}$. The classical Hamiltonian is the sum of kinetic and potential energies,

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2, \tag{7.1.3}$$

which leads to the phase-space equations of motion

$$\dot{x} \;=\; \frac{\partial H}{\partial p} = p/m, \qquad \dot{p} \;=\; -\frac{\partial H}{\partial x} = -m\omega^2 x, \tag{7.1.4}$$

with the general solution

$$x(t) \;=\; x_0 \cos(\omega t + \phi_0), \qquad p(t) \;=\; -m\omega x_0 \sin(\omega t + \phi_0), \tag{7.1.5}$$

where $x_0$ and $\phi_0$ are constants determined by the initial conditions.

A wise theorist, Sidney Coleman, was fond of remarking that the job of a physicist consists of "treating the harmonic oscillator in ever-increasing levels of abstraction".

This may be a slight exaggeration, but there are at least two reasons why the quantum harmonic oscillator is particularly worthy of study.

First, the harmonic oscillator is a problem that can actually be solved exactly in closed form in quantum mechanics. Most problems in quantum mechanics do not have this property.

Second, it arises quite often as a good approximation to more complicated problems. To understand this, consider a more general potential $V(x)$, which is assumed to have a local minimum at some point $x = x_0$, and to be smooth in a neighborhood of that point. Then we can expand it in a Taylor series for small $x - x_0$,

$$V(x) \;=\; V(x_0) + (x - x_0)\frac{dV}{dx}\bigg|_{x=x_0} + \frac{1}{2}(x - x_0)^2 \frac{d^2V}{dx^2}\bigg|_{x=x_0} + \frac{1}{6}(x - x_0)^3 \frac{d^3V}{dx^3}\bigg|_{x=x_0} + \cdots . \tag{7.1.6}$$

The first term is a constant, and so can be absorbed into a redefinition of the zero point of energy; as a constant part of the Hamiltonian, it gives the same overall phase to all states, and therefore can be ignored. The second term is zero, since by definition $dV/dx$ vanishes at a minimum of the potential. Thus, if the terms in the expansion are getting smaller, the leading behavior is captured by the quadratic term. After redefining coordinates according to $x \to x + x_0$, we can hope to approximate

$$V(x) \;\approx\; \frac{x^2}{2}\left(\frac{d^2V}{dx^2}\bigg|_{x=0}\right) \tag{7.1.7}$$

where the constant quantity in parentheses is defined to be $m\omega^2$.

The same argument generalizes to a particle moving in a three dimensional potential. Let us suppose that the $x, y, z$ coordinates have already been redefined so that the minimum of the potential is at the origin $(x, y, z) = (0, 0, 0)$. Then we have

$$V(x, y, z) \;=\; V(0, 0, 0) + \frac{1}{2}x_a x_b \mathcal{V}_{ab} + \cdots\,, \tag{7.1.8}$$

where by convention repeated indices $a, b, \ldots$ are implicitly summed over $1, 2, 3$, with $x_1 = x$, $x_2 = y$, $x_3 = z$, and

$$\mathcal{V}_{ab} \;=\; \frac{\partial^2 V}{\partial x_a \partial x_b}\bigg|_{x_c=0} \tag{7.1.9}$$

is a real symmetric matrix. A linear term involving $\frac{\partial V}{\partial x_a}\big|_{x_b=0}$ vanishes because we are at the minimum of the potential. A theorem in linear algebra says that a real symmetric matrix can always be diagonalized by some orthogonal matrix $O$, according to

$$\mathcal{V} \;=\; O^T \tilde{\mathcal{V}}\, O, \tag{7.1.10}$$

where $\tilde{\mathcal{V}} \;=\; \mathrm{diag}(\tilde{\mathcal{V}}_1, \tilde{\mathcal{V}}_2, \tilde{\mathcal{V}}_3)$, so that

$$\mathcal{V}_{ab} \;=\; O_{ca}\tilde{\mathcal{V}}_c O_{cb}, \tag{7.1.11}$$

and the orthogonality of $O$ is written as

$$O_{ac}O_{bc} \;=\; O_{ca}O_{cb} \;=\; \delta_{ab}. \tag{7.1.12}$$

Dropping the constant $V(0, 0, 0)$, and assuming the higher-order terms in the expansion can be neglected, the quantum mechanical Hamiltonian can be approximated as

$$H \;=\; \frac{1}{2m}(P_x^2 + P_y^2 + P_z^2) + \frac{1}{2}O_{ca}\tilde{\mathcal{V}}_c O_{cb} X_a X_b. \tag{7.1.13}$$

Now we can do a change of variables, by defining

$$\tilde{X}_c \;=\; O_{cb}X_b, \qquad\qquad \tilde{P}_c \;=\; O_{cb}P_b. \tag{7.1.14}$$

A key feature of this change of variables is that $\tilde{X}_a$ and $\tilde{P}_a$ have canonical commutators, since

$$[\tilde{X}_c, \tilde{P}_d] \;=\; O_{cb}O_{da}[X_b, P_a] \;=\; i\hbar O_{ca}O_{da} \;=\; i\hbar\delta_{cd}, \tag{7.1.15}$$

where the second equality used the canonical commutation relation for the original variables, $[X_b, P_a] = i\hbar\delta_{ab}$, and eq. (7.1.12) was used to get the last equality. Also,

$$\tilde{P}_a\tilde{P}_a \;=\; O_{ca}P_aO_{cb}P_b \;=\; \delta_{ab}P_aP_b \;=\; P_aP_a \;=\; P_x^2 + P_y^2 + P_z^2. \tag{7.1.16}$$

Therefore, the Hamiltonian in terms of the new canonical variables is

$$H \;=\; \sum_{a=1}^{3}\left(\frac{\tilde{P}_a^2}{2m} + \frac{1}{2}m\omega_a^2\tilde{X}_a^2\right), \tag{7.1.17}$$

where $m\omega_a^2 = \tilde{\mathcal{V}}_a$. This shows that $H$ is the sum of three independent harmonic oscillator Hamiltonians, with possibly different natural frequencies. The general case is called the anisotropic three-dimensional harmonic oscillator. The special case $\omega_x = \omega_y = \omega_z$ is called the isotropic three-dimensional harmonic oscillator; we will study it in more detail later in section 9.5, using its spherical coordinate wavefunction.

The preceding illustrates why many systems can be treated as if they were, effectively, systems of harmonic oscillators, with various higher-order effects that one may hope to either neglect or treat as perturbations. Of course, there are some important problems that are not approximated well by the harmonic oscillator. The free particle has no restoring force at all. The potential of a particle in a box or a square well is not close to its Taylor series expansion. The hydrogen atom has a potential with minimum $-\infty$ at $r = 0$, and again the Taylor series expansion fails. Fortunately, these three cases are also exactly solvable, and one can say that together with the harmonic oscillator they are the four important examples of potential problems in non-relativistic quantum mechanics that should be familiar to a well-educated physicist.

The three-dimensional harmonic oscillator energy eigenvalue problem can be solved as the tensor product of three one-dimensional harmonic oscillators. In the following two sections, we will use two very different methods to derive the solutions of the one-dimensional harmonic oscillator with Hamiltonian

$$H \;=\; \frac{P^2}{2m} + \frac{1}{2}m\omega^2X^2. \tag{7.1.18}$$

But first, we make a general observation: the eigenstates of this $H$ must all have positive energy. This should be intuitively plausible, since both contributions to the classical energy are manifestly positive. To prove it formally in the quantum theory, note that in a stationary state $|\psi\rangle$ with energy $E$,

$$E = \langle\psi|H|\psi\rangle = \frac{1}{2m}\langle\psi|P^2|\psi\rangle + \frac{1}{2}m\omega^2\langle\psi|X^2|\psi\rangle \tag{7.1.19}$$

$$= \frac{1}{2m}\langle P\psi|P\psi\rangle + \frac{1}{2}m\omega^2\langle X\psi|X\psi\rangle > 0. \tag{7.1.20}$$

Here we have applied the fact that $P$ and $X$ are Hermitian, so that for example $\langle\psi|\,P = \langle P^\dagger\psi| = \langle P\psi|$, and then used the positivity of the inner product for non-null kets.

## 7.2 Position and momentum representations: the differential equations approach

In this section, we will solve for the energy eigenstates of the harmonic oscillator in the position representation using the differential equations approach. It must be admitted that this is more difficult and less elegant than the algebraic (energy representation) approach given in the following section. The position representation differential equation method for the harmonic oscillator is nevertheless important to learn, because it can be generalized to more complicated Hamiltonians, for which algebraic approaches are often not available.

The position-representation version of the time-independent Schrödinger equation for the harmonic oscillator $H\,|\psi\rangle = E\,|\psi\rangle$ is

$$\langle x|\left(\frac{P^2}{2m} + \frac{1}{2}m\omega^2X^2\right)|\psi\rangle = E\,\langle x|\psi\rangle\,, \tag{7.2.1}$$

or, in terms of the wavefunction $\psi(x) = \langle x|\psi\rangle$, after using $X \to x$ and $P \to -i\hbar d/dx$,

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}\left(E - \frac{1}{2}m\omega^2x^2\right)\psi = 0. \tag{7.2.2}$$

It is convenient to introduce dimensionless variables corresponding to the position and energy,

$$y = x/b, \qquad \mathcal{E} = E/\hbar\omega, \tag{7.2.3}$$

where

$$b = \sqrt{\hbar/m\omega} \tag{7.2.4}$$

is a constant length scale. In terms of $y$ and $\mathcal{E}$, eq. (7.2.2) becomes

$$\frac{d^2\psi}{dy^2} + (2\mathcal{E} - y^2)\psi = 0, \tag{7.2.5}$$

which we will now solve as an eigenvalue problem for both $\mathcal{E}$ and $\psi(y)$. Since $H$ commutes with the parity operator $\Pi$, we know that we can find simultaneous eigenstates of energy and parity.

To help us understand the character of the solutions, first consider the behavior at large $|y|$, where we can neglect the constant $\mathcal{E}$ compared to $y^2$, so that $d^2\psi/dy^2 \approx y^2\psi$. Let us try a solution of the form

$$\psi = Ay^n e^{-\alpha y^2}, \tag{7.2.6}$$

where $n$ and $\alpha$ are constants. Then

$$\frac{d\psi}{dy} = A\left(ny^{n-1} - 2\alpha y^{n+1}\right) e^{-\alpha y^2} \approx -2\alpha y\psi, \tag{7.2.7}$$

where at large $|y|$ the first term in parentheses is neglected compared to the second. Repeating this yields

$$\frac{d^2\psi}{dy^2} \approx (2\alpha y)^2\psi. \tag{7.2.8}$$

Therefore the guess eq. (7.2.6) indeed works for large $y$, with $\alpha = \pm 1/2$, regardless of the value of the constant $n$. We can reject $\alpha = -1/2$ on physical grounds, since the wavefunction would blow up at large $|y|$ and would not be normalizable. A purported wavefunction that grows exponentially with $|y|$ must be unphysical because it would imply that no matter how far from the origin you look, the probability that the particle would be found farther way would be infinitely larger, defying the fact that the potential is attractive.

This motivates trying a solution of the form

$$\psi = u(y)\, e^{-y^2/2}, \qquad \text{with} \qquad u = \sum_{j=0}^{\infty} c_j y^j, \tag{7.2.9}$$

where the $c_j$ are constants to be determined. No negative powers are included in this guess, because we expect that $\psi$ will be well-behaved as $y \to 0$, where the potential smoothly vanishes.

Plugging eq. (7.2.9) into eq. (7.2.5) gives

$$\sum_{j=0}^{\infty} c_j \left[j(j-1)y^{j-2} + (2\mathcal{E} - 1 - 2j)y^j\right] = 0. \tag{7.2.10}$$

Now we use a trick on the first term. Since it vanishes for $j = 0$ and $j = 1$, we can rewrite it as $\sum_{j=2}^{\infty} c_j j(j-1)y^{j-2}$, and then let $j = k+2$ so that it becomes $\sum_{k=0}^{\infty} c_{k+2}(k+2)(k+1)y^k$, and finally rename $k \to j$, so that it becomes $\sum_{j=0}^{\infty} c_j(j+2)(j+1)y^j$. Now it can be nicely combined with the second term, combining like powers of $y$,

$$\sum_{j=0}^{\infty} y^j \left[c_{j+2}(j+1)(j+2) + c_j(2\mathcal{E} - 1 - 2j)\right] = 0. \tag{7.2.11}$$

Because each power $y^j$ in the sum is independent, their coefficients must vanish separately for each $j$, so we have a recurrence relation,

$$c_{j+2} = \frac{2j + 1 - 2\mathcal{E}}{(j + 1)(j + 2)} c_j. \tag{7.2.12}$$

This shows that we only need to know two constants $c_0$ and $c_1$, because if $c_0$ is known, then so are $c_2$, $c_4$, $c_6$, ..., and if $c_1$ is known, then so are $c_3$, $c_5$, $c_7$, etc. So, the solution is

$$\begin{aligned} u(y) \;=\; & c_0 \left[1 + \left(\frac{1-2\mathcal{E}}{2}\right) y^2 + \left(\frac{1-2\mathcal{E}}{2}\right)\left(\frac{5-2\mathcal{E}}{12}\right) y^4 + \cdots\right] \\ & + c_1 \left[y + \left(\frac{3-2\mathcal{E}}{6}\right) y^3 + \left(\frac{3-2\mathcal{E}}{6}\right)\left(\frac{7-2\mathcal{E}}{20}\right) y^5 + \cdots\right], \end{aligned} \tag{7.2.13}$$

which, naively, appears to be an infinite series.

However, there is something horribly wrong with the solution if it is really an infinite series. For sufficiently large $|y|$, the series will be dominated by terms with large powers $j$, where eq. (7.2.12) appears to give

$$\frac{c_{j+2}}{c_j} \approx \frac{2}{j + \mathcal{E} + 5/2}, \tag{7.2.14}$$

up to contributions to the denominator that vanish as $j \to \infty$. If arbitrarily large powers $j$ are present, the function $u(y)$ will necessarily grow too fast as $|y| \to \infty$. To see this, consider for comparison the function $f(y) = y^p e^{y^2}$, with series expansion

$$f(y) = \sum_{k=0}^{\infty} y^{2k+p}/k!. \tag{7.2.15}$$

Now, writing $j = 2k + p$, this series has terms $C_j y^j$ where $C_{j+2}/C_j = 1/(k + 1) = 2/(j - p + 2)$. Comparing to eq. (7.2.14), we see that if $u(y)$ is really a non-terminating series in powers of $y$, then $u(y) \sim y^p e^{y^2}$ for $p = -\mathcal{E} - 1/2$, and the wavefunction behaves for large $|y|$ like

$$\psi(y) \sim (y^p e^{y^2}) e^{-y^2/2} = y^p e^{y^2/2}. \tag{7.2.16}$$

These are recognized as the unphysical solutions with $\alpha = -1/2$ that we had already rejected.

For a sensible physical solution, there is only one way out: the series for $u(y)$ must terminate. To find a basis of such solutions, we can first consider $c_0 \neq 0$ but $c_1 = 0$, so that only even-index coefficients are present. If one of them vanishes, $c_{2k} = 0$ for some $k$, then all higher coefficients $c_j$ with $j > 2k$ will also vanish, according to the recurrence relation. Likewise, we can take $c_1 \neq 0$ but $c_0 = 0$, so that only odd-index coefficients are present. If one of them vanishes, $c_{2k+1} = 0$ for some $k$, then all $c_j$ with $j > 2k + 1$ will also vanish.

In either case, the condition for the series in $u(y)$ to terminate, yielding a physical solution, is that the numerator in the recurrence relation eq. (7.2.12) must vanish for some non-negative integer $j = n$. Therefore, the allowed energy eigenvalues are $\mathcal{E} = n + 1/2$, or

$$E_n = (n + 1/2)\hbar\omega, \qquad (n = 0, 1, 2, \ldots). \tag{7.2.17}$$

It follows that $u(y)$ will be a polynomial of degree $n$ in $y$, and contains only even (odd) powers of $y$ if $n$ is even (odd). For any given $n$, they can be constructed from the recurrence relation, eq. (7.2.12), up to an overall multiplicative constant given by either $c_0$ or $c_1$. The resulting $u(y) = H_n(y)$ for a given $n$ are called **Hermite polynomials**. From eqs. (7.2.5) and (7.2.9) with $\mathcal{E} = n + 1/2$, they satisfy the differential equation

$$\left(\frac{d^2}{dy^2} - 2y\frac{d}{dy} + 2n\right) H_n = 0. \tag{7.2.18}$$

An explicit expression for the Hermite polynomials is

$$H_n(y) = e^{y^2/2}\left(y - \frac{d}{dy}\right)^n e^{-y^2/2}. \tag{7.2.19}$$

This can be verified by plugging it into the differential equation (7.2.18), but an even nicer derivation will be found near the end of the section 7.3.

The first few Hermite polynomials are

$$H_0(y) = 1, \qquad H_1(y) = 2y, \tag{7.2.20}$$

$$H_2(y) = -2 + 4y^2, \qquad H_3(y) = -12y + 8y^3, \tag{7.2.21}$$

$$H_4(y) = 12 - 48y^2 + 16y^4, \qquad H_5(y) = 120y - 160y^3 + 32y^5. \tag{7.2.22}$$

In general, they also obey the identities

$$H_n(-y) = (-1)^n H_n(y), \tag{7.2.23}$$

$$\frac{d}{dy}H_n(y) = 2nH_{n-1}(y), \tag{7.2.24}$$

$$H_{n+1}(y) = 2yH_n(y) - 2nH_{n-1}(y), \tag{7.2.25}$$

and satisfy a generating function relation

$$\sum_{n=0}^{\infty} H_n(y)\frac{t^n}{n!} = \exp(2yt - t^2). \tag{7.2.26}$$

They can be shown to satisfy an orthonormality condition

$$\int_{-\infty}^{\infty} dy\, H_n(y)H_m(y)e^{-y^2} = \delta_{n,m}\sqrt{\pi}\, 2^n\, n!. \tag{7.2.27}$$

The orthonormal energy basis wavefunctions for the harmonic oscillator can now be written

$$\psi_n(x) \;=\; \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\sqrt{\frac{1}{2^n n!}}H_n(y)e^{-y^2/2}, \tag{7.2.28}$$

where $y = x/b = x\sqrt{m\omega/\hbar}$, so that

$$\int_{-\infty}^{\infty} dx\,[\psi_n(x)]^*\psi_m(x) \;=\; \delta_{n,m}. \tag{7.2.29}$$

(In this case, the wavefunctions are all real, so the complex conjugation does nothing.) In particular, the normalized ground state wavefunction is just a pure Gaussian,

$$\psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp(-m\omega x^2/2\hbar). \tag{7.2.30}$$

The wavefunctions $\psi_n$ and the corresponding probability densities $|\psi_n|^2$ are shown for $n = 0, 1, 2, 3, 4,$ and 16 in Figure 7.2.1.

The wavefunctions are related to the energy eigenstate kets $|n\rangle$ of the harmonic oscillator by

$$\langle x|n\rangle = \psi_n(x), \tag{7.2.31}$$

with

$$H|n\rangle \;=\; \hbar\omega(n+1/2)|n\rangle. \tag{7.2.32}$$

This shows that the energy levels of the one-dimensional harmonic oscillator are equally spaced and non-degenerate, with a quantum of energy given by $\Delta E = \hbar\omega$. There is also a **zero-point energy** of the ground state, $E_0 = \hbar\omega/2$, in agreement with our earlier proof that the energies had to be positive. From eq. (7.2.23), the states with even (odd) $n$ have even (odd) parity.

Another interesting relation satisfied by the harmonic oscillator wavefunctions, and thus the Hermite polynomials, is

$$\sum_{n=0}^{\infty}\psi_n(x')\psi_n(x) \;=\; \delta(x-x'). \tag{7.2.33}$$

This can be derived immediately from $\langle x|x'\rangle = \delta(x-x')$ by applying the completeness relation $\sum_{n=0}^{\infty}|n\rangle\langle n| = 1$.

Having found the position wavefunctions for the harmonic oscillator, we now turn to the momentum representation. One way to evaluate the momentum wavefunctions of the energy eigenstates $|n\rangle$ is to use the completeness relation:

$$\widetilde{\psi}_n(p) = \langle p|n\rangle = \int_{-\infty}^{\infty} dx\;\langle p|x\rangle\,\langle x|n\rangle = \int_{-\infty}^{\infty} dx\,\frac{1}{\sqrt{2\pi\hbar}}e^{-ipx/\hbar}\,\psi_n(x), \tag{7.2.34}$$

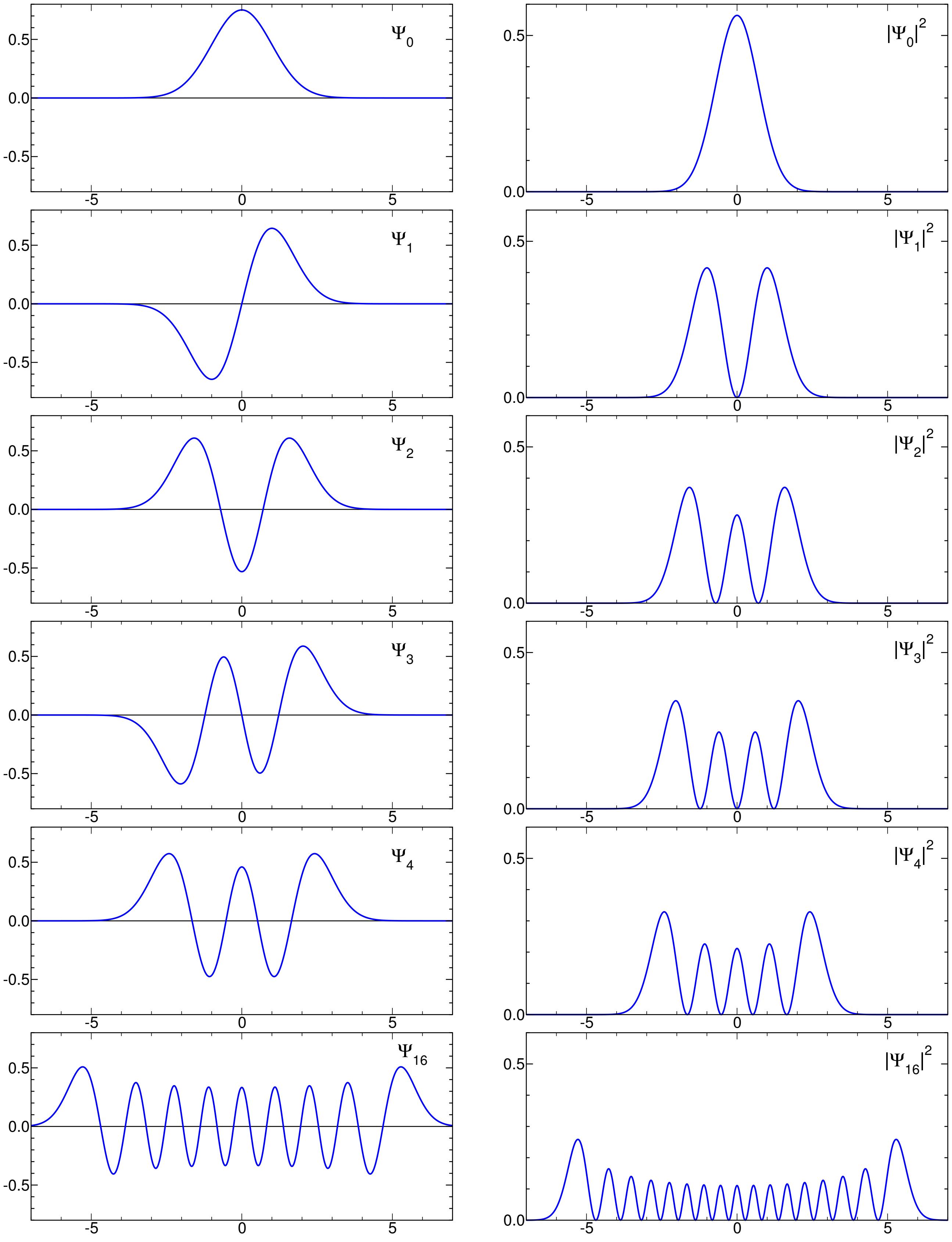

Figure 7.2.1: Harmonic oscillator stationary-state wavefunctions $\psi_n$ in units of $(m\omega/\hbar)^{1/4}$ (left column) and probability densities $|\psi_n|^2$ in units of $\sqrt{m\omega/\hbar}$ (right column) for $n = 0, 1, 2, 3, 4,$ and 16, as functions of $y = x\sqrt{m\omega/\hbar}$.

but it may not be immediately obvious how to evaluate the integral for general $n$. Fortunately, we can gain some insight into the result by noting that the Hamiltonian in eq. (7.1.18) has the very special property of being invariant under the simultaneous substitutions

$$X \leftrightarrow P, \qquad\qquad m\omega \leftrightarrow \frac{1}{m\omega}. \tag{7.2.35}$$

Since the energy eigenstates are non-degenerate, their momentum wavefunctions must be given, up to a phase $e^{i\varphi_n}$ to be determined, by making these same substitutions in eq. (7.2.28). Therefore, defining a dimensionless variable proportional to momentum,

$$v = p/\sqrt{\hbar\omega m}, \tag{7.2.36}$$

it must be that

$$\widetilde{\psi}_n(p) \;=\; e^{i\varphi_n}\frac{1}{(\pi\hbar\omega m)^{1/4}}\sqrt{\frac{1}{2^n n!}}H_n(v)e^{-v^2/2}. \tag{7.2.37}$$

The phase factor $e^{i\varphi_n}$ is needed to maintain consistency with the phase convention of the ket $|n\rangle$ that has already been fixed by eqs. (7.2.28) and (7.2.31). For the lowest few $n = 0, 1, 2, 3, \ldots$, you can do the integral in eq. (7.2.34) to check that eq. (7.2.37) is indeed true with

$$e^{i\varphi_n} = (-i)^n, \tag{7.2.38}$$

and we will prove it for general $n$ at the end of the next section.

## 7.3 Energy representation: the algebraic approach

In this section, we will use a different, and more elegant, method to solve for the stationary states of the harmonic oscillator. In this approach, due to Dirac, we use algebraic methods rather than differential equations, working directly in the energy basis.

We begin by defining dimensionless creation and annihilation operators (also known as raising and lowering or destruction operators, or together as ladder operators) by

$$a \;=\; \sqrt{\frac{m\omega}{2\hbar}}X + i\frac{1}{\sqrt{2\hbar\omega m}}P, \tag{7.3.1}$$

$$a^\dagger \;=\; \sqrt{\frac{m\omega}{2\hbar}}X - i\frac{1}{\sqrt{2\hbar\omega m}}P. \tag{7.3.2}$$

As the notation indicates, these are not Hermitian operators and so are not observables, but rather are Hermitian adjoints of each other. Equivalently, one can write the relationship as

$$X \;=\; \sqrt{\frac{\hbar}{2m\omega}}(a^\dagger + a), \tag{7.3.3}$$

$$P \;=\; i\sqrt{\frac{\hbar\omega m}{2}}(a^\dagger - a). \tag{7.3.4}$$

Using the canonical commutation relations $[X, P] = i\hbar$, we find that

$$[a, a^\dagger] \;=\; 1. \tag{7.3.5}$$

We also can compute

$$a^\dagger a \;=\; \frac{m\omega}{2\hbar}X^2 + \frac{1}{2\hbar\omega m}P^2 + \frac{i}{2\hbar}[X, P] \;=\; H/\hbar\omega - 1/2,$$

so that the Hamiltonian is simply[†]

$$H \;=\; \hbar\omega(a^\dagger a + 1/2). \tag{7.3.6}$$

Using eq. (7.3.5), one then obtains

$$[H, a] \;=\; -\hbar\omega\, a, \qquad\qquad [H, a^\dagger] \;=\; \hbar\omega\, a^\dagger. \tag{7.3.7}$$

Now suppose that we have an energy eigenstate $|E\rangle$, with energy eigenvalue $E$. Then consider the Hamiltonian acting on the state $a^\dagger|E\rangle$:

$$H\left(a^\dagger|E\rangle\right) \;=\; a^\dagger H|E\rangle + [H, a^\dagger]|E\rangle \;=\; (E + \hbar\omega)a^\dagger|E\rangle. \tag{7.3.8}$$

This shows that $a^\dagger|E\rangle$ is an eigenstate of $H$ with energy $E + \hbar\omega$. Repeating this $n$ times, the state $(a^\dagger)^n|E\rangle$ must be an energy eigenstate with energy $E + n\hbar\omega$. Similarly,

$$H\left(a|E\rangle\right) \;=\; (E - \hbar\omega)\, a|E\rangle, \tag{7.3.9}$$

so the state $a^n|E\rangle$ is apparently an energy eigenstate with energy $E - n\hbar\omega$.

Thus, naively it might appear that we can construct an infinite chain of energy eigenstates with both arbitrarily low and arbitrarily high energies,

$$\cdots,\quad |E{-}2\hbar\omega\rangle\,,\quad |E{-}\hbar\omega\rangle\,,\quad |E\rangle\,,\quad |E{+}\hbar\omega\rangle\,,\quad |E{+}2\hbar\omega\rangle\,,\quad \cdots. \tag{7.3.10}$$

But this cannot be true, since we proved at the end of section 7.1 that all of the energy eigenvalues of the harmonic oscillator are positive. The only way out is that all of the kets in the chain with negative energy are actually the null ket. In other words, one of the states in the chain, call it the ground state $|0\rangle$, must be the one with lowest energy, and must satisfy

$$a|0\rangle \;=\; 0, \tag{7.3.11}$$

---

[†] One could also write the equivalent form $H = \hbar\omega(a^\dagger a + a a^\dagger)/2$. In choosing to write eq. (7.3.6), we have followed the systematic protocol known as **normal ordering**. To "normal order" an operator means to rewrite it by moving all $a$ operators to the right and all $a^\dagger$ operators to the left, using $aa^\dagger = a^\dagger a + 1$ (which is equivalent to the commutation relation) as many times as necessary.

so that all of the kets $a^n|0\rangle = 0$ for $n \geq 1$ do not actually exist as physical states. Since $a^\dagger a|0\rangle = 0$, we have

$$(H/\hbar\omega - 1/2)\,|0\rangle \;=\; 0, \tag{7.3.12}$$

so the ground state must have $E_0 = \hbar\omega/2$, a result that we had also found in the previous section using the differential equation approach.

In section 6.3, we proved that bound states in one-dimensional quantum mechanics never have degenerate energies. This implies that there is a unique state $|0\rangle$ with $E_0 = \hbar\omega/2$, and unique states with $E_n = \hbar\omega(n + 1/2)$ for $n = 0, 1, 2, \ldots$, given up to normalization by $|n\rangle \propto (a^\dagger)^n|0\rangle$. Since they are non-degenerate eigenkets of a Hermitian operator ($H$), they can be normalized to form an orthobasis,

$$\langle k|n\rangle \;=\; \delta_{n,k}. \tag{7.3.13}$$

Let $|0\rangle$ have norm 1. Then we can define the other orthobasis kets by recurrence using

$$|n\rangle \;=\; c_n a^\dagger |n{-}1\rangle, \tag{7.3.14}$$

where $c_n$ are constants to be determined. We have

$$\langle n|n\rangle \;=\; |c_n|^2\,\langle n{-}1|aa^\dagger|n{-}1\rangle \;=\; n|c_n|^2\,\langle n{-}1|n{-}1\rangle\,, \tag{7.3.15}$$

where the second equality makes use of

$$aa^\dagger \;=\; a^\dagger a + 1 \;=\; H/\hbar\omega + 1/2 \tag{7.3.16}$$

and then $H\,|n{-}1\rangle \;=\; \hbar\omega(n-1/2)\,|n-1\rangle$. The norms in eq. (7.3.15) will both be equal to 1 provided that $c_n = 1/\sqrt{n}$, where we have made an arbitrary choice of phase. From this, we use eqs. (7.3.14) and (7.3.16) to deduce that

$$a^\dagger|n\rangle \;=\; \sqrt{n+1}\,|n{+}1\rangle, \tag{7.3.17}$$

$$a|n\rangle \;=\; \sqrt{n}\,|n{-}1\rangle, \tag{7.3.18}$$

so that $a^\dagger$ raises the energy of the state (or creates an energy quantum), and $a$ lowers the energy (or destroys an energy quantum). Taking the Hermitian conjugate gives

$$\langle n|\,a \;=\; \sqrt{n+1}\;\langle n{+}1|\,, \tag{7.3.19}$$

$$\langle n|\,a^\dagger \;=\; \sqrt{n}\;\langle n{-}1|\,. \tag{7.3.20}$$

An equivalent way to express these results is in terms of the matrix elements of the $a^\dagger$ and $a$ operators in the energy eigenstate orthobasis,

$$\langle k|a^\dagger|n\rangle \;=\; \sqrt{n+1}\,\delta_{k,n+1}, \qquad\qquad \langle k|a|n\rangle \;=\; \sqrt{n}\,\delta_{k,n-1}. \tag{7.3.21}$$

Another consequence is that the $n$th excited state can be written in terms of $n$ creation operators acting on the ground state,

$$|n\rangle \;=\; \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle. \tag{7.3.22}$$

The Hamiltonian is sometimes written as $H = \hbar\omega(N+1/2)$, where $N = a^\dagger a$ is called the number operator. It is clearly Hermitian, and satisfies

$$N|n\rangle \;=\; n|n\rangle, \tag{7.3.23}$$

and so is an observable that just measures the number of energy quanta in the state.

To illustrate the power of the algebraic method used in this section, suppose for example that we wanted to evaluate the matrix element $\langle 3|X^3|2\rangle$. We have, using eq. (7.3.3),

$$\begin{aligned}
\langle 3|X^3|2\rangle \;&=\; \left(\frac{\hbar}{2m\omega}\right)^{3/2} \langle 3|(a^\dagger + a)^3|2\rangle & (7.3.24)\\
&=\; \left(\frac{\hbar}{2m\omega}\right)^{3/2} \langle 3|\left(\cancel{a^{\dagger 3}} + a^{\dagger 2}a + a^\dagger a a^\dagger + a a^{\dagger 2} + \cancel{a^\dagger a^2} + \cancel{a a^\dagger a} + \cancel{a^2 a^\dagger} + \cancel{a^3}\right)|2\rangle\,. & (7.3.25)
\end{aligned}$$

Here, we have crossed out terms that can be immediately seen to give no contribution by simply counting quanta created and destroyed. For the first term, we start in the ket on the far right with 2 quanta, and create 3 more, so it is proportional to $|5\rangle$, which by orthonormality has vanishing inner product with the bra $\langle 3|$ on the left. Likewise, the fifth, sixth, and seventh terms cannot contribute because we start with 2 quanta and then (in various orders) create 1 more but destroy 2, resulting in a state with 1, which again has vanishing inner product with the 3-quanta bra. And, the last term immediately vanishes because $a^3|2\rangle = 0$. The remaining three terms do not vanish, but can be easily evaluated with nothing but simple arithmetic, by applying eqs. (7.3.17) and (7.3.18) repeatedly.

In contrast, evaluation of the same matrix element in the position representation approach of section 7.2 would read

$$\begin{aligned}
\langle 3|X^3|2\rangle \;=\; \int_{-\infty}^{\infty} dx \left(\frac{m\omega}{\pi\hbar 2^6 (3!)^2}\right)^{1/4} H_3\left(x\sqrt{m\omega/\hbar}\right) e^{-m\omega x^2/2\hbar}\, x^3 \\
\left(\frac{m\omega}{\pi\hbar 2^4 (2!)^2}\right)^{1/4} H_2\left(x\sqrt{m\omega/\hbar}\right) e^{-m\omega x^2/2\hbar}.
\end{aligned} \tag{7.3.26}$$

This is certainly doable, but less pleasant.

Let us see how to connect the energy and position representations, by writing $a$ and $a^\dagger$ as differential operators in the latter. In the position representation, $X \to x$ and $P \to -i\hbar d/dx$, so from eqs. (7.3.1) and (7.3.2) we get

$$a \;=\; \frac{1}{\sqrt{2}}\left(y + \frac{d}{dy}\right), \qquad a^\dagger \;=\; \frac{1}{\sqrt{2}}\left(y - \frac{d}{dy}\right), \tag{7.3.27}$$

where $y = x/b = x\sqrt{m\omega/\hbar}$ as before. Therefore, the condition $a|0\rangle = 0$ gives

$$\left(y + \frac{d}{dy}\right)\psi_0(y) \;=\; 0. \tag{7.3.28}$$

This implies $d\psi_0/\psi_0 = -y dy$, or $d(\ln\psi_0) = -d(y^2/2)$, so that upon integration, $\ln\psi_0 = -y^2/2 + \ln A_0$, where $\ln A_0$ is a constant of integration. Thus,

$$\psi_0(y) \;=\; A_0 e^{-y^2/2} \;=\; A_0 e^{-m\omega x^2/2\hbar}, \tag{7.3.29}$$

in agreement with eq. (7.2.30) after fixing the normalization constant $A_0 = (m\omega/\pi\hbar)^{1/4}$. Now, combining eqs. (7.3.22) and (7.3.27), we have

$$\psi_n(x) \;=\; \langle x|n\rangle \;=\; \frac{1}{\sqrt{n!}}\left[\frac{1}{\sqrt{2}}\left(y - \frac{d}{dy}\right)\right]^n \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-y^2/2}. \tag{7.3.30}$$

Comparing this with eq. (7.2.28) yields the general form for the Hermite polynomials given in eq. (7.2.19), as promised.

The energy and momentum representations can be connected in a similar way. The momentum representations are $P \to p$ and $X \to i\hbar d/dp$, from which one finds

$$a = \frac{i}{\sqrt{2}}\left(v + \frac{d}{dv}\right), \qquad a^\dagger \;=\; -\frac{i}{\sqrt{2}}\left(v - \frac{d}{dv}\right), \tag{7.3.31}$$

where $v = p/\sqrt{\hbar\omega m}$ is the dimensionless rescaled momentum, as in eqs. (7.2.36) and(7.2.37). Evaluating eq. (7.2.34) for the special case $n = 0$, we then find

$$\widetilde{\psi}_0(p) \;=\; \frac{1}{(\pi\hbar\omega m)^{1/4}} e^{-v^2/2} \tag{7.3.32}$$

for the ground state. Using eq. (7.3.31) in eq. (7.3.22), we obtain

$$\widetilde{\psi}_n(p) \;=\; \langle p|n\rangle = \frac{1}{\sqrt{n!}}\left[-\frac{i}{\sqrt{2}}\left(v - \frac{d}{dv}\right)\right]^n \frac{1}{(\pi\hbar\omega m)^{1/4}} e^{-v^2/2}. \tag{7.3.33}$$

Now comparing to eq. (7.2.19), we obtain eq. (7.2.37) with $e^{i\varphi_n} = (-i)^n$, as promised.

## 7.4 Coherent states of the harmonic oscillator

Consider a macroscopic harmonic oscillator of the type one might encounter in a lab experiment in an introductory physics course, with mass $m = 0.5$ kg, $\omega = 4$ radians/second, and amplitude $x_0 = 0.1$ meters. Classically, the energy can be obtained as the potential energy at the extremum of the displacement, when $p = 0$ and $x = x_0$, so that $E = m\omega^2 x_0^2/2 = 0.04$ J. The energy quantum associated with the oscillator is $\hbar\omega = 4.22 \times 10^{-34}$ J. Therefore, we expect macroscopic oscillator states to have enormous numbers of energy quanta, something like $n = E/\hbar\omega \approx 10^{32}$ in this example.

Is the energy eigenstate $|n = 10^{32}\rangle$ classical-like? The answer is clearly no, since in any energy eigenstate, $\langle X(t)\rangle = 0$ and $\langle P(t)\rangle = 0$ for all times $t$, while the classical trajectory, eq. (7.1.5), involves macroscopic oscillations. We would like to find quantum states that are approximately classical, by which we mean that $\langle X(t)\rangle \approx x_{\rm cl}(t)$ and $\langle P(t)\rangle \approx p_{\rm cl}(t)$ should be as close as possible to the solutions in eq. (7.1.5). The states that accomplish this are called coherent (or quasi-classical) states, and the argument just given ensures that they cannot be energy eigenstates.

Instead of energy eigenstates, with the wisdom of foresight, let us try kets $|\alpha\rangle$ that are eigenstates of the lowering operator $a$ with eigenvalue $\alpha$,

$$a\,|\alpha\rangle \;=\; \alpha\,|\alpha\rangle\,. \tag{7.4.1}$$

Since $a$ is not Hermitian, we have no reason to expect that the allowed eigenvalues $\alpha$ will be real, and indeed it will turn out to be very important that they are complex in general. For the same reason, we also cannot expect that the set of all states $|\alpha\rangle$ for different $\alpha$ will form an orthobasis or even be orthogonal, and again they are not.

To construct such coherent states, try an arbitrary linear combination of energy eigenstates,

$$|\alpha\rangle \;=\; \sum_{n=0}^{\infty} c_n\,|n\rangle\,, \tag{7.4.2}$$

with coefficients $c_n$ to be determined. Then,

$$a\,|\alpha\rangle \;=\; \sum_{n=0}^{\infty} c_n a\,|n\rangle \;=\; \sum_{n=1}^{\infty} c_n\sqrt{n}\,|n\!-\!1\rangle\,, \tag{7.4.3}$$

where in the second equality we have used eq. (7.3.18), and started the sum from $n = 1$ by exploiting the fact that the $n = 0$ term vanishes. If we now relabel $n \to n + 1$, and require that the result

$$a\,|\alpha\rangle \;=\; \sum_{n=0}^{\infty} c_{n+1}\sqrt{n+1}\,|n\rangle \tag{7.4.4}$$

is equal to $\alpha\sum_{n=0}^{\infty} c_n|n\rangle$, we obtain the recurrence relation

$$c_{n+1} \;=\; \frac{\alpha}{\sqrt{n+1}}c_n. \tag{7.4.5}$$

Starting with $c_0$, we get $c_1 = \alpha c_0$, $c_2 = \alpha^2 c_0/\sqrt{2}$, etc., or in general $c_n = \alpha^n c_0/\sqrt{n!}$. Therefore, up to normalization,

$$|\alpha\rangle \;\;=\;\; c_0 \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{7.4.6}$$

To fix $c_0$, we require $\langle\alpha|\alpha\rangle = 1$, or

$$1 \;=\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} c_0^* \frac{\alpha^{*m}}{\sqrt{m!}} c_0 \frac{\alpha^n}{\sqrt{n!}} \langle m|n\rangle \;=\; |c_0|^2 \sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!} \;=\; |c_0|^2 e^{|\alpha|^2}, \tag{7.4.7}$$

where the second equality uses the orthonormality $\langle m|n\rangle = \delta_{n,m}$ to collapse the sum over $m$. Therefore, $c_0 = e^{-|\alpha|^2/2}$ (up to the usual arbitrary global phase), and the normalized coherent state with complex eigenvalue $\alpha$ is

$$|\alpha\rangle \;=\; e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \tag{7.4.8}$$

Using eq. (7.3.22), this can be rewritten as

$$|\alpha\rangle \;=\; e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{(\alpha a^\dagger)^n}{n!} |0\rangle \;=\; e^{-|\alpha|^2/2} e^{\alpha a^\dagger} |0\rangle. \tag{7.4.9}$$

If a harmonic oscillator is in a coherent state $|\alpha\rangle$, and the energy is measured, all results $E_n = \hbar\omega(n + 1/2)$ can occur, with probabilities

$$\mathcal{P}_n \;=\; |\langle n|\alpha\rangle|^2 \;=\; \frac{|\alpha|^{2n}}{n!} e^{-|\alpha|^2}. \tag{7.4.10}$$

This is a Poisson distribution with mean $|\alpha|^2$, and does not depend on the phase of $\alpha$. Since

$$\mathcal{P}_n \;=\; \frac{|\alpha|^2}{n} \mathcal{P}_{n-1}, \tag{7.4.11}$$

we see that the probability increases with $n$ as long as $n < |\alpha|^2$, but decreases thereafter. Therefore, the most probable result of an energy measurement is of order $\hbar\omega|\alpha|^2$. This shows that for a macroscopic oscillator like the one posed at the beginning of this section, $|\alpha| \sim \sqrt{n} \sim 10^{16}$ in a coherent state. The expectation value of the energy is

$$\langle\alpha|H|\alpha\rangle \;=\; \hbar\omega\Big[(\langle\alpha|\, a^\dagger)(a\,|\alpha\rangle) + 1/2\Big] \;=\; \hbar\omega\Big[(\langle\alpha|\, \alpha^*)(\alpha\,|\alpha\rangle) + 1/2\Big] \;=\; \Big(|\alpha|^2 + \frac{1}{2}\Big)\hbar\omega, \tag{7.4.12}$$

giving a similar measure of the average energy.

Next let us compute the uncertainty in the energy. First, we need

$$\langle\alpha|H^2|\alpha\rangle \;=\; (\hbar\omega)^2\Big[\langle\alpha|a^\dagger a a^\dagger a|\alpha\rangle + \langle\alpha|a^\dagger a|\alpha\rangle + \frac{1}{4}\langle\alpha|\alpha\rangle\Big] \tag{7.4.13}$$

$$\;=\; (\hbar\omega)^2 \big(|\alpha|^4 + 2|\alpha|^2 + 1/4\big)\,. \tag{7.4.14}$$

Therefore

$$\Delta H \;=\; \sqrt{\langle H^2\rangle - \langle H\rangle^2} \;=\; \hbar\omega|\alpha|. \tag{7.4.15}$$

This is a very small energy uncertainty compared to the energy expectation value,

$$\Delta H/\langle H\rangle = 1/|\alpha| \ll 1, \tag{7.4.16}$$

since we found $1/|\alpha|$ of order $10^{-16}$ in our numerical example. Thus, the energy distribution is peaked extremely sharply about the classical value.

We can likewise compute the expectation value and uncertainty of the position. First,

$$\langle\alpha|X|\alpha\rangle = \sqrt{\frac{\hbar}{2m\omega}}\,\langle\alpha|(a^\dagger + a)|\alpha\rangle = \sqrt{\frac{\hbar}{2m\omega}}(\alpha^* + \alpha) = \sqrt{\frac{2\hbar}{m\omega}}\mathrm{Re}[\alpha]. \tag{7.4.17}$$

Also, one has

$$\begin{aligned}\langle\alpha|X^2|\alpha\rangle &= \frac{\hbar}{2m\omega}\,\langle\alpha|(a^\dagger + a)^2|\alpha\rangle = \frac{\hbar}{2m\omega}\,\langle\alpha|(a^{\dagger 2} + 2a^\dagger a + a^2 + 1)|\alpha\rangle && (7.4.18)\\ &= \frac{\hbar}{2m\omega}\left[(\alpha^* + \alpha)^2 + 1\right]. && (7.4.19)\end{aligned}$$

It follows that

$$\Delta X = \sqrt{\langle\alpha|X^2|\alpha\rangle - \langle\alpha|X|\alpha\rangle^2} = \sqrt{\frac{\hbar}{2m\omega}}. \tag{7.4.20}$$

Remarkably, this does not depend on $\alpha$ at all, and in fact is exactly the same as for the ground state, making it ridiculously tiny by macroscopic standards. The interpretation of $\Delta X/\langle X\rangle \ll 1$ is that the position wavefunction is extremely sharply peaked about its expectation value. In a similar way, one can compute the expectation value and the uncertainty of the momentum:

$$\begin{aligned}\langle\alpha|P|\alpha\rangle &= \sqrt{2m\hbar\omega}\,\mathrm{Im}[\alpha], && (7.4.21)\\ \Delta P &= \sqrt{\frac{m\hbar\omega}{2}}. && (7.4.22)\end{aligned}$$

The latter is again independent of $\alpha$ and is the same as in the ground state. Putting these results together, the product of uncertainties in position and momentum is

$$(\Delta X)(\Delta P) = \hbar/2, \tag{7.4.23}$$

which is exactly as small as it could possibly be, consistent with the uncertainty relation.

The preceding results suggest that there is some close relation between the ground state and the coherent states, since they have exactly the same uncertainties in position and momentum. To see this, let us work out the wavefunction of the coherent state $|\alpha\rangle$. We start with

$$\psi_\alpha(x) = \langle x|\alpha\rangle = e^{-|\alpha|^2/2}\,\langle x|e^{\alpha a^\dagger}|0\rangle \tag{7.4.24}$$

from eq. (7.4.9). Next, we write

$$\alpha a^\dagger = A + B \tag{7.4.25}$$

where, from the definition of $a^\dagger$ in eq. (7.3.2),

$$A = \alpha\sqrt{\frac{m\omega}{2\hbar}}X, \qquad B = -i\frac{\alpha}{\sqrt{2\hbar\omega m}}P. \tag{7.4.26}$$

Since $[A, B] = \alpha^2/2$ is a constant, we can use the Baker–Campbell–Hausdorff formula eq. (2.4.32), repeated here for convenience as $e^{A+B} = e^A e^B e^{-[A,B]/2}$, to obtain

$$\psi_\alpha(x) = e^{-|\alpha|^2/2}e^{-\alpha^2/4}\exp\left(\sqrt{\frac{m\omega}{2\hbar}}\alpha x\right)\langle x|\exp\left(-i\frac{\alpha}{\sqrt{2\hbar\omega m}}P\right)|0\rangle\,. \tag{7.4.27}$$

The next step is to notice that the exponential operator inside the matrix element has the same form as the translation operator $T(a)$ defined in eq. (5.2.1), with $a$ replaced by the (complex, in general) number $\alpha\sqrt{\hbar/2\omega m}$. Therefore, we can invoke eq. (5.2.7) to find

$$\langle x|\exp\left(-i\frac{\alpha}{\sqrt{2\hbar\omega m}}P\right)|0\rangle = \psi_0\left(x - \alpha\sqrt{\frac{\hbar}{2\omega m}}\right), \tag{7.4.28}$$

where $\psi_0(x)$ is the ground state wavefunction, found in eq. (7.2.30). Using this in eq. (7.4.27), and rearranging the exponentials, gives

$$\psi_\alpha(x) = e^{-(|\alpha|^2+\alpha^2)/2}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp\left[-\frac{m\omega}{2\hbar}\left(x - \alpha\sqrt{\frac{2\hbar}{\omega m}}\right)^2\right]. \tag{7.4.29}$$

Now, writing $\alpha$ in terms of its real and imaginary parts, and then expressing them in terms of the position and momentum expectation values using eqs. (7.4.17) and (7.4.21), the wavefunction takes the simple form

$$\psi_\alpha(x) = e^{i\theta}e^{i\langle P\rangle x/\hbar}\psi_0(x - \langle X\rangle), \tag{7.4.30}$$

where $\theta = -\text{Re}[\alpha]\text{Im}[\alpha]$. Up to the physically irrelevant global phase provided by $\theta$, the coherent state wavefunction is the same as the ground state, but displaced by the position expectation value and multiplied by the position-dependent phase factor that encodes the momentum expectation value.

So far, we have considered a coherent state $|\alpha\rangle$ at a fixed time $t$. Now let us investigate the time evolution of the state and its properties. Since the Hamiltonian is independent of time, the solution of the Schrödinger equation tells us that at time $t$ the state is

$$e^{-iHt/\hbar}|\alpha\rangle = e^{-iHt/\hbar}e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}e^{-i\omega(n+1/2)t}|n\rangle \tag{7.4.31}$$

$$= e^{-i\omega t/2}e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{(\alpha e^{-i\omega t})^n}{\sqrt{n!}}|n\rangle \tag{7.4.32}$$

$$= e^{-i\omega t/2}|\alpha e^{-i\omega t}\rangle. \tag{7.4.33}$$

This shows that the coherent state just evolves to another coherent state in which $\alpha$ has changed[†] to $\alpha e^{-i\omega t}$. Thus, in the time-dependent coherent state, the magnitude of $\alpha$ does not change, and the phase of $\alpha$ rotates linearly with time, at the natural angular frequency of the harmonic oscillator.

Suppose that at time $t = 0$, we start with $\alpha = |\alpha| e^{-i\phi_0}$, so that at a general time $t$,

$$\alpha \;=\; |\alpha| e^{-i(\omega t + \phi_0)} \tag{7.4.34}$$

Then, using the results from eqs. (7.4.17) and (7.4.21), we get

$$\begin{aligned} \langle X \rangle (t) &= \sqrt{\frac{2\hbar}{m\omega}}\, \text{Re}\left[|\alpha| e^{-i(\omega t+\phi_0)}\right] \;=\; x_0 \cos(\omega t + \phi_0), && (7.4.35)\\ \langle P \rangle (t) &= \sqrt{2m\hbar\omega}\, \text{Im}\left[|\alpha| e^{-i(\omega t+\phi_0)}\right] \;=\; -m\omega x_0 \sin(\omega t + \phi_0), && (7.4.36) \end{aligned}$$

where we have defined

$$x_0 = \sqrt{\frac{2\hbar}{m\omega}} |\alpha|. \tag{7.4.37}$$

This establishes the connection between the classical trajectory of the harmonic oscillator and the parameter $\alpha$ of the coherent state that most closely resembles it. Not only is the coherent state for a harmonic oscillator as close as possible to the classical limit, but we see, by comparing to eq. (7.1.5), that the expectation values of the position and momentum evolve in exactly the same way as their classical counterparts. The uncertainties do not change with time. To summarize the properties of a coherent state,

$$\begin{aligned} \langle H \rangle &= \hbar\omega(|\alpha|^2 + 1/2) & \Delta H &= \hbar\omega|\alpha|, && (7.4.38)\\ \langle X \rangle (t) &= x_{\text{cl}}(t), & \Delta X &= \sqrt{\hbar/2m\omega}, && (7.4.39)\\ \langle P \rangle (t) &= p_{\text{cl}}(t), & \Delta P &= \sqrt{m\hbar\omega/2}. && (7.4.40) \end{aligned}$$

The harmonic oscillator potential evidently has a remarkable "focusing" property, such that the coherent state wavefunctions do not spread out at all as they evolve in time, unlike the case for the free particle Gaussian states as seen in section 6.2.

There is a simple way to prepare a coherent state for the harmonic oscillator. Suppose we temporarily apply a constant (position-independent) force $f$, acting in the positive $x$ direction, so that the new Hamiltonian is

$$H_f \;=\; \frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2 - fX. \tag{7.4.41}$$

---

[†] The state ket has also acquired an irrelevant global phase $e^{-i\omega t/2}$. In contrast, the complex phase $e^{-i\omega t}$ multiplying $\alpha$ (inside the ket symbol) is certainly physically relevant, as is clear from eqs. (7.4.17) and (7.4.21).

By defining a shifted position operator

$$X' = X - \frac{f}{m\omega^2}, \tag{7.4.42}$$

the Hamiltonian can be rewritten as

$$H_f = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X'^2 - \frac{f^2}{2m\omega^2}. \tag{7.4.43}$$

The constant shift in the position operator does not affect the commutation relation,

$$[X', P] = i\hbar, \tag{7.4.44}$$

so the solution of the eigenvalue problem for $H_f$ proceeds exactly as for the Hamiltonian with $f = 0$. The energy eigenvalues will be the same, but lowered by the constant amount $-f^2/2m\omega^2$. More importantly for our present purposes, the ground state $|0\rangle_f$ of $H_f$ is defined by the property that it is annihilated by the operator

$$a' = \sqrt{\frac{m\omega}{2\hbar}}X' + i\frac{1}{\sqrt{2\hbar\omega m}}P = a - \frac{f}{\sqrt{2\hbar\omega^3 m}}, \tag{7.4.45}$$

where $a$ is the annihilation operator for $H$ with $f = 0$. This implies that the state $|0\rangle_f$ is exactly a coherent state of the original Hamiltonian $H$:

$$a\,|0\rangle_f = \alpha\,|0\rangle_f\,, \tag{7.4.46}$$

where

$$\alpha = \frac{f}{\sqrt{2\hbar\omega^3 m}} \tag{7.4.47}$$

is a real number.

To summarize, we can prepare a coherent state of the harmonic oscillator as follows. First, apply a constant force $f$, shifting the equilibrium position of the mass to the point $x_0 = f/m\omega^2$. Next, allow the oscillator to settle into the ground state of the new Hamiltonian $H_f$, with the same uncertainties in position and momentum as the ground state of $H$. Finally, we release the mass by removing the force $f$. At the instant that the mass is released, it will find itself in a coherent state of $H$, with $\alpha$ initially real and given by eq. (7.4.47). It will therefore evolve in time as we have already seen, remaining in a coherent state as $\alpha$ acquires a non-trivial phase $e^{-i\omega t}$. The subsequent behavior is as close as possible to what would happen in classical mechanics if we displaced the oscillator and then released it from rest.

## 7.5 Three-dimensional harmonic oscillator

We now return to the case of a three-dimensional harmonic oscillator, with Hamiltonian given by eq. (7.1.17). Rewriting it without the tildes (which just denoted a particular choice of rectangular coordinates), this can be rewritten as

$$H = H_x + H_y + H_z, \tag{7.5.1}$$

where

$$H_x = \frac{P_x^2}{2m} + \frac{1}{2} m\omega_x^2 X^2 = \hbar\omega_x (a_x^\dagger a_x + 1/2), \tag{7.5.2}$$

with exactly analogous expressions for $H_y$ and $H_z$. Here, we have made use of the experience of section 7.3 to write

$$a_x = \sqrt{\frac{m\omega_x}{2\hbar}} X + i\frac{1}{\sqrt{2\hbar\omega_x m}} P_x, \tag{7.5.3}$$

$$a_y = \sqrt{\frac{m\omega_y}{2\hbar}} Y + i\frac{1}{\sqrt{2\hbar\omega_y m}} P_y, \tag{7.5.4}$$

etc. Now $H_x$ and $H_y$ and $H_z$ all commute with each other, and are Hermitian. One can therefore find an orthobasis consisting of their common eigenstates, denoted $|n_x, n_y, n_z\rangle$, where operators with the $x$ subscript just ignore the $n_y$ and $n_z$ labels, and similarly for operators with the $y$ and $z$ subscripts. Thus,

$$H_x |n_x, n_y, n_z\rangle = E_x |n_x, n_y, n_z\rangle, \qquad H_y |n_x, n_y, n_z\rangle = E_y |n_x, n_y, n_z\rangle, \tag{7.5.5}$$

$$H_z |n_x, n_y, n_z\rangle = E_z |n_x, n_y, n_z\rangle. \tag{7.5.6}$$

(This can also be viewed as the tensor product of three one-dimensional harmonic oscillators.) From section 7.3, we have already solved the eigenvalue problem for each $H_j$, and so we know that the allowed energies for each of $j = x, y, z$ are

$$E_j = \hbar\omega_j (n_j + 1/2), \qquad n_j = 0, 1, 2, \ldots, \tag{7.5.7}$$

and that the corresponding eigenkets satisfy, for example:

$$a_x |n_x, n_y, n_z\rangle = \sqrt{n_x}\, |n_x - 1, n_y, n_z\rangle, \tag{7.5.8}$$

$$a_x^\dagger |n_x, n_y, n_z\rangle = \sqrt{n_x + 1}\, |n_x + 1, n_y, n_z\rangle, \tag{7.5.9}$$

$$a_y |n_x, n_y, n_z\rangle = \sqrt{n_y}\, |n_x, n_y - 1, n_z\rangle, \tag{7.5.10}$$

etc. Since the labels $n_x$, $n_y$, and $n_z$ uniquely specify the states of an orthobasis, $H_x$, $H_y$, and $H_z$ form a CSCO for this problem.

The state $|n_x, n_y, n_z\rangle$ has total energy eigenvalue

$$E \,=\, E_x + E_y + E_z \,=\, \hbar\omega_x(n_x + 1/2) + \hbar\omega_y(n_y + 1/2) + \hbar\omega_z(n_z + 1/2), \qquad (7.5.11)$$

and its wavefunction is just a product of the wavefunctions for the one-dimensional harmonic oscillator,

$$\psi_{n_x,n_y,n_z}(\vec{r}\,) \,=\, \langle \vec{r}\,|n_x, n_y, n_z\rangle \,=\, \psi_{n_x,\omega_x}(x)\,\psi_{n_y,\omega_y}(y)\,\psi_{n_z,\omega_z}(z), \qquad (7.5.12)$$

where each of the $\psi_{n_j,\omega_j}$ functions can be obtained from eq. (7.2.28) in the obvious way.

In the special case of the isotropic 3-dimensional harmonic oscillator, $\omega_x = \omega_y = \omega_z = \omega$, the system is invariant under arbitrary rotations about the origin. This symmetry is reflected in a degeneracy of the energy spectrum, which is now

$$E_n = \hbar\omega(n + 3/2), \qquad (7.5.13)$$

for states $|n, u_n\rangle$, where $n = n_x + n_y + n_z$ and $u_n = 1, \ldots, g_n$ is a degeneracy label for the energy eigenvalue $E_n$. To find the degeneracy $g_n$ of each energy level, we need to know how many ways there are to choose three non-negative integers that add up to $n$. The ground state with $n = 0$ is unique, as it can only be achieved with the lone combination $n_x = n_y = n_z = 0$, so $g_0 = 1$. The first excited energy level can be achieved in three different ways, by taking one of $n_x, n_y, n_z$ to be 1, and the others 0, so $g_1 = 3$. Similar straightforward counting reveals that $g_2 = 6$, and $g_3 = 10$. More generally, this combinatorics problem can be solved using the dots and lines trick: one arranges $n$ dots in a row, and splits them into three groups by placing two vertical lines as shown below.

$$\underbrace{\bullet\,\bullet\cdots\bullet}_{n_x}\big|\underbrace{\bullet\,\bullet\cdots\bullet}_{n_y}\big|\underbrace{\bullet\,\bullet\cdots\bullet}_{n_z} \qquad (7.5.14)$$

The degeneracy of the energy eigenvalue $E_n$ is therefore the number of ways of arranging a row of $n + 2$ objects, of which $n$ are identical and 2 are identical,

$$g_n \,=\, \frac{(n+2)!}{n!\;2!} \,=\, \frac{1}{2}(n+1)(n+2). \qquad (7.5.15)$$

There is another natural choice of CSCO that one can use for the isotropic three-dimensional harmonic oscillator, consisting of the total Hamiltonian $H$ and two observables associated with the angular momentum about the minimum of the potential. We will discuss this method in section 9.5.

# 8 Angular momentum and its representations

## 8.1 The eigenvalue problem for angular momentum

At the end of section 5.3, we introduced the commutation relation algebra for the components of a generic angular momentum operator $\vec{J} = \hat{x}J_x + \hat{y}J_y + \hat{z}J_z$, defined to satisfy

$$[J_a,\, J_b] \;=\; i\hbar\epsilon_{abc}J_c, \qquad (a, b, c = x, y, z), \tag{8.1.1}$$

with $c$ implicitly summed over. Let us now find the simultaneous eigenvectors and the corresponding eigenvalues of the compatible observables $J^2 = J_x^2 + J_y^2 + J_z^2$ and $J_z$.

We start by observing that the eigenvalues of $J^2$ must be non-negative. To see this, note that for any non-null ket $|\psi\rangle$,

$$\langle\psi|J^2|\psi\rangle \;=\; \|J_x\,|\psi\rangle\|^2 + \|J_y\,|\psi\rangle\|^2 + \|J_z\,|\psi\rangle\|^2 \;\geq\; 0, \tag{8.1.2}$$

where the possibility of equality exists only because each of the kets $J_x\,|\psi\rangle$, $J_y\,|\psi\rangle$, and $J_z\,|\psi\rangle$ could be null. Now, if

$$J^2\,|\psi\rangle = \lambda\,|\psi\rangle\,, \tag{8.1.3}$$

then $\langle\psi|J^2|\psi\rangle = \lambda\,\langle\psi|\psi\rangle \geq 0$, so $\lambda \geq 0$. For reasons to become clear shortly, it turns out to be convenient to give $\lambda$ the name $\hbar^2 j(j+1)$, by defining $j = \sqrt{\lambda/\hbar^2 + 1/4} - 1/2$. Since $\lambda \geq 0$, it follows that $j \geq 0$ also.

Because $J^2$ and $J_z$ are compatible observables, Theorem 2.7.1 says that there must be an orthobasis consisting of common eigenstates, $|j, m, u_{j,m}\rangle$, which satisfy orthonormality and completeness relations

$$\langle j', m', u'_{j',m'}|j, m, u_{j,m}\rangle \;=\; \delta_{j,j'}\,\delta_{m,m'}\,\delta_{u'_{j',m'},u_{j,m}}, \tag{8.1.4}$$

$$\sum_j \sum_m \sum_{u_{j,m}} |j, m, u_{j,m}\rangle\langle j, m, u_{j,m}| \;=\; I. \tag{8.1.5}$$

Here $u_{j,m}$ is a possible degeneracy label[†] that will be important in particular cases, but plays no role in the following discussion and so will be suppressed for simplicity. We therefore seek to solve the eigenvalue problem

$$J^2\,|j, m\rangle \;=\; \hbar^2 j(j+1)\,|j, m\rangle\,, \tag{8.1.6}$$

$$J_z\,|j, m\rangle \;=\; \hbar m\,|j, m\rangle\,. \tag{8.1.7}$$

---

[†] Soon [just before eq. (8.1.28)], we will learn that the orthobasis can actually always be chosen in such a way that the $u_{j,m}$ do not depend on $m$. But we do not know that yet.

for the allowed eigenvalues $j$ and $m$, recognizing that there may be more than one sector of such states, distinguished by the suppressed degeneracy label. In the special case of orbital angular momentum, we have already found, by requiring the position wavefunction to be single-valued (in particular, for $\phi = 0$ and $2\pi$), that the allowed eigenvalues of $L_z$ are $\hbar m$, where $m$ is an integer. However, spin (intrinsic) angular momentum is not associated with a position wavefunction, so that argument does not apply, and $m$ need not be an integer, as we will see.

It is very useful (and not just for the present purpose of solving the eigenvalue problem) to define the **angular momentum raising and lowering operators**, also sometimes known as **ladder operators** for angular momentum,

$$J_+ = J_x + iJ_y, \qquad J_- = J_x - iJ_y, \tag{8.1.8}$$

which will play a role similar to $a^\dagger$ and $a$ for the harmonic oscillator. They are not Hermitian, but can easily be shown to satisfy the useful identities,

$$\begin{aligned}
(J_+)^\dagger &= J_-, && (8.1.9)\\
[J_z,\, J_+] &= \hbar J_+, && (8.1.10)\\
[J_z,\, J_-] &= -\hbar J_-, && (8.1.11)\\
[J_+,\, J_-] &= 2\hbar J_z, && (8.1.12)\\
J_+J_- &= J^2 - J_z^2 + \hbar J_z, && (8.1.13)\\
J_-J_+ &= J^2 - J_z^2 - \hbar J_z. && (8.1.14)
\end{aligned}$$

From the last two equations, we get

$$J^2 = \frac{1}{2}J_+J_- + \frac{1}{2}J_-J_+ + J_z^2. \tag{8.1.15}$$

Each of $J_+$, $J_-$, and $J_z$ commute with the total angular momentum squared,

$$[J^2, J_+] = [J^2, J_-] = [J^2, J_z] = 0. \tag{8.1.16}$$

We now derive some useful facts by studying the ket $J_+|j,m\rangle$. Since $J^2$ commutes with $J_+$,

$$J^2\left(J_+|j,m\rangle\right) = \hbar^2 j(j+1)\left(J_+|j,m\rangle\right). \tag{8.1.17}$$

Also,

$$\begin{aligned}
J_z\left(J_+|j,m\rangle\right) &= [J_z, J_+]|j,m\rangle + J_+J_z|j,m\rangle = \hbar J_+|j,m\rangle + \hbar m J_+|j,m\rangle && (8.1.18)\\
&= \hbar(m+1)\left(J_+|j,m\rangle\right). && (8.1.19)
\end{aligned}$$

Thus, $J_+|j,m\rangle$ is an eigenstate of both $J^2$ and $J_z$, with eigenvalues $\hbar^2 j(j+1)$ and $\hbar(m+1)$, respectively. Due to the fact that the $|j,m,u_{j,m}\rangle$ were chosen as an orthobasis, we can conclude that either $J_+|j,m\rangle$ is proportional to $|j,m+1\rangle$ within each sector labeled by $u_{j,m}$, or else it is the null ket. This is why $J_+$ is called a raising operator; it raises the $J_z$ eigenvalue. To find out whether $J_+|j,m\rangle$ is the null ket, we compute its squared norm,

$$\begin{aligned} \| J_+|j,m\rangle \|^2 &= \langle j,m|J_-J_+|j,m\rangle \;=\; \langle j,m|\left(J^2 - J_z^2 - \hbar J_z\right)|j,m\rangle \\ &= \langle j,m|\left[\hbar^2 j(j+1) - (\hbar m)^2 - \hbar(\hbar m)\right]|j,m\rangle \\ &= \hbar^2[j(j+1) - m(m+1)]. \end{aligned} \tag{8.1.20}$$

From this we learn two useful things. First, the general properties of the inner product require that the squared norm is non-negative, so allowed $j, m$ must satisfy $j(j+1) - m(m+1) = (j-m)(j+m+1) \geq 0$, and since we learned at the beginning of this section that $j \geq 0$, we can conclude that for every non-null ket $|j,m\rangle$,

$$m \leq j. \tag{8.1.21}$$

The second useful result from eq. (8.1.20) is the constant of proportionality between $J_+|j,m\rangle$ and the unit-normalized ket $|j,m{+}1\rangle$,

$$J_+|j,m\rangle \;=\; \hbar\sqrt{j(j+1) - m(m+1)}\,|j,m{+}1\rangle. \tag{8.1.22}$$

Here we had to make an arbitrary and unavoidable choice of phase; eq. (8.1.22) can be taken as the definition of the relative phase between $|j,m\rangle$ and $|j,m{+}1\rangle$. It follows from eq. (8.1.22) that $J_+|j,j\rangle$ is actually the null ket.

Everything in the previous paragraph can be repeated for $J_-|j,m\rangle$. We learn that, for every non-null ket $|j,m\rangle$,

$$m \geq -j \tag{8.1.23}$$

due to the requirement of positive squared norm, and that $J_-|j,-j\rangle$ is the null ket, and that

$$J_-|j,m\rangle \;=\; \hbar\sqrt{j(j+1) - m(m-1)}\,|j,m{-}1\rangle. \tag{8.1.24}$$

This justifies calling $J_-$ the lowering operator for (the $z$ component of) angular momentum.

In eq. (8.1.24), we have made another choice of phase, and we should be careful to make sure that it is consistent with the phase choice made in eq. (8.1.22). This can be checked by using eqs. (8.1.22) and (8.1.24) to compute $J_+J_-|j,m\rangle$ and $J_-J_+|j,m\rangle$, and noting that they are equivalent to the results obtained using eqs. (8.1.13) and (8.1.14), respectively. A nice feature of eqs. (8.1.22) and (8.1.24) is that the coefficients on the right-hand sides are real and non-negative.

Combining eqs. (8.1.21) and (8.1.23) tells us that, for any given $j$, the only possible values of $m$ for physical states $|j, m\rangle$ are in the range

$$-j \leq m \leq j. \tag{8.1.25}$$

We are now ready to prove that $j = n/2$ for some integer $n$. If we compute $(J_+)^p|j, m\rangle$, it must be proportional to $|j, m + p\rangle$, according to eq. (8.1.22). For some finite $p$, this must be the ket $|j, j\rangle$, because otherwise we could keep acting with $J_+$ to find non-null kets with arbitrarily large $J_z$ eigenvalue, which would contradict of eq. (8.1.25). (Here, we are relying on the fact that eq. (8.1.22) tells us that $|j, j\rangle$ is the unique non-null ket that is annihilated by $J_+$.) Therefore, $j = m + p$ for some non-negative integer $p$. Similarly, if we compute $(J_-)^q|j, m\rangle$, it must be proportional to $|j, m - q\rangle$, according to eq. (8.1.24). For some finite $q$, this must be the ket $|j, -j\rangle$; otherwise, we would again contradict eq. (8.1.25). Therefore, $-j = m - q$ for some non-negative integer $q$. Putting the results together gives $2j = p + q$, a non-negative integer. The allowed values are

$$j \;=\; 0,\; 1/2,\; 1,\; 3/2,\; 2,\; \ldots. \tag{8.1.26}$$

The fact that $j = m+p$ for some integer $p$, together with eq. (8.1.25), also tells us which values of $m$ can give non-null kets $|j\rangle$. For each $j$, they are the $2j + 1$ values

$$m \;=\; -j,\; -j+1,\; \ldots,\; j-1,\; j. \tag{8.1.27}$$

For the special case that $\vec{J} = \vec{L}$, we already had found that $m$ must be an integer, so in the case of orbital angular momentum the allowed values of $j = l$ are also restricted to the non-negative integers $0, 1, 2, \ldots$. The case of half-integer[‡] $j$ must correspond to something other than orbital angular momenta.

For a given state $|j, m, u_{j,m}\rangle$, the operations of $J^2$, $J_z$, $J_+$, and $J_-$, given by eqs. (8.1.6), (8.1.7), (8.1.22), and (8.1.24) are independent of $u_{j,m}$, and can change $m$, but not $j$. This has two important implications.

First, if we have a single representative state with a certain $j$ and $u_{j,m}$, then all other states with the same $j$ and $u_{j,m}$ but other values of $m$ are obtained by acting repeatedly with $J_+$ or $J_-$. Therefore, the degeneracy label $u_{j,m}$ does not actually depend on $m$, as foreshadowed in the footnote following eq. (8.1.5). Thus, for any quantum system, the orthobasis of common eigenstates of $J^2$ and $J_z$ can be labeled as $|j, m, u_j\rangle$, with a slight simplification of eqs. (8.1.4)

---

[‡] The standard term "half-integer" means an odd integer divided by 2. It might be more logical to call this "half-odd-integer" or "integer-plus-half", but it is difficult to fight tradition.

and (8.1.5),

$$\begin{aligned} \langle j', m', u'_{j'} | j, m, u_j \rangle &= \delta_{j,j'}\, \delta_{m,m'}\, \delta_{u'_{j'},u_j}, & (8.1.28) \\ \sum_j \sum_m \sum_{u_j} |j, m, u_j\rangle\langle j, m, u_j| &= I. & (8.1.29) \end{aligned}$$

Typically, $u_j$ corresponds to the eigenvalue(s) of some observable(s) that, together with $J^2$ and $J_z$, form a CSCO for the quantum system. If the Hamiltonian is invariant under rotations, then $H$ can be taken to be one of those observables. The group of $2j + 1$ orthobasis states with common $j$ and $u_j$ and varying $m = -j, \ldots, j$ is called an **angular momentum multiplet**.

The second observation is that, because the operations of the angular momentum operators $J^2$, $J_z$, $J_+$, and $J_-$ (and $J_x$, $J_y$) in the orthobasis $|j, m, u_j\rangle$ do not depend on $u_j$ at all, we can work out their matrix representations and operations on kets for each relevant value of $j$ just once, and the results will be applicable to any quantum system with that $j$.

For a $j = 0$ subspace, we have $J^2 = 0$ and $J_z = J_+ = J_- = J_x = J_y = 0$, so the state space has only one non-null ket $|j = 0, m = 0\rangle$. All angular momentum operators acting on $|j = 0, m = 0\rangle$ give the null ket, so that they are all represented by the $1 \times 1$ matrix 0.

## 8.2 $j = 1/2$ representation and spin

For a $j = 1/2$ subspace, the orthobasis consists of two states with $m = \pm 1/2$, with corresponding two-component vector representations

$$|j=1/2,\ m=1/2\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |j=1/2,\ m=-1/2\rangle \leftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{8.2.1}$$

The matrix representations of the angular momentum operators can now be constructed from eqs. (8.1.6)–(8.1.8), (8.1.22), and (8.1.24), with the results

$$\begin{aligned} J^2 &\leftrightarrow \frac{3\hbar^2}{4}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & J_z &\leftrightarrow \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, & (8.2.2) \\ J_x &\leftrightarrow \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & J_y &\leftrightarrow \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. & (8.2.3) \\ J_+ &\leftrightarrow \hbar\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, & J_- &\leftrightarrow \hbar\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, & (8.2.4) \end{aligned}$$

Some special operator identities that hold only in the $j = 1/2$ case are

$$J_x^2 = J_y^2 = J_z^2 = \hbar^2/4, \qquad J_+^2 = J_-^2 = 0, \qquad (j = 1/2). \tag{8.2.5}$$

As noted after eq. (8.1.27), the fact that $m = \pm 1/2$ is not an integer implies that $\vec{J}$ cannot be a pure orbital angular momentum.

The case of purely intrinsic angular momentum for a single particle is called **spin**, and we write $\vec{J} = \vec{S}$ and $j = s$ in that case. For $s = 1/2$, we say that the particle is (or has) spin-1/2, and we often rewrite eq. (8.2.1) using a simplified notation for the $S_z$ eigenstates,

$$|\uparrow\rangle \;\leftrightarrow\; \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad\qquad |\downarrow\rangle \;\leftrightarrow\; \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{8.2.6}$$

referred to as spin-up and spin-down, with eigenvalues $m_s = 1/2$ and $-1/2$, respectively.

From eqs. (8.2.2) and (8.2.3), one can deduce that the components of the spin operator written in ket-bra form are

$$S_x = \frac{\hbar}{2}\Big(|\downarrow\rangle\langle\uparrow| + |\uparrow\rangle\langle\downarrow|\Big), \tag{8.2.7}$$

$$S_y = \frac{i\hbar}{2}\Big(|\downarrow\rangle\langle\uparrow| - |\uparrow\rangle\langle\downarrow|\Big), \tag{8.2.8}$$

$$S_z = \frac{\hbar}{2}\Big(|\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|\Big), \tag{8.2.9}$$

and a standard notation for their matrix representation is

$$\vec{S} \;\leftrightarrow\; \frac{\hbar}{2}\vec{\sigma}, \tag{8.2.10}$$

where the components of the vector $\vec{\sigma}$ are known as the **Pauli matrices**, defined by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{8.2.11}$$

The Pauli matrices obey the commutation and anticommutation relations

$$[\sigma_a, \sigma_b] = 2i\epsilon_{abc}\sigma_c, \tag{8.2.12}$$

$$\{\sigma_a, \sigma_b\} = 2\delta_{ab} \tag{8.2.13}$$

for $a, b = x, y, z$, and

$$\mathrm{Tr}[\sigma_a] = 0, \tag{8.2.14}$$

$$\mathrm{Det}[\sigma_a] = -1. \tag{8.2.15}$$

For any spatial vector $\vec{v} = \hat{x}v_x + \hat{y}v_y + \hat{z}v_z$, we have

$$\vec{v}\cdot\vec{\sigma} = \begin{pmatrix} v_z & v_x - iv_y \\ v_x + iv_y & -v_z \end{pmatrix}, \tag{8.2.16}$$

and a formula useful for simplifications,

$$(\vec{v}\cdot\vec{\sigma})(\vec{w}\cdot\vec{\sigma}) = \vec{v}\cdot\vec{w} + i(\vec{v}\times\vec{w})\cdot\vec{\sigma}, \tag{8.2.17}$$

with the $2 \times 2$ identity matrix understood in the first term on the right side. In particular,

$$(\vec{v} \cdot \vec{\sigma})(\vec{v} \cdot \vec{\sigma}) \;=\; \vec{v} \cdot \vec{v} \,=\, v_x^2 + v_y^2 + v_z^2 \,=\, v^2 \tag{8.2.18}$$

is proportional to the identity matrix.

For a spin-1/2 particle like an electron, the complete state can be written as a linear combination of kets with $S_z$ eigenvalues $m_s = \hbar/2$ and $-\hbar/2$, as

$$|\psi\rangle \;=\; |\psi_\uparrow, \uparrow\rangle \,+\, |\psi_\downarrow, \downarrow\rangle\,, \tag{8.2.19}$$

where $\psi_\uparrow$ and $\psi_\downarrow$ represent the non-spin degrees of freedom corresponding to the classical motion of the particle in three dimensions. This state can also be represented as a two-component column vector of kets, called a **spinor**,

$$|\psi\rangle \;=\; \begin{pmatrix} |\psi_\uparrow\rangle \\ |\psi_\downarrow\rangle \end{pmatrix}. \tag{8.2.20}$$

Acting on these two-component spinor kets, the spin operator is again represented by $\vec{S} \leftrightarrow \frac{\hbar}{2}\vec{\sigma}$. The bra corresponding to eq. (8.2.20) is a two-component row spinor,

$$\langle\psi| \;=\; \begin{pmatrix} \langle\psi_\uparrow| & \langle\psi_\downarrow| \end{pmatrix}. \tag{8.2.21}$$

The inner product of two states $|\psi\rangle$ and $|\chi\rangle$ is

$$\langle\chi|\psi\rangle \;=\; \begin{pmatrix} \langle\chi_\uparrow| & \langle\chi_\downarrow| \end{pmatrix} \begin{pmatrix} |\psi_\uparrow\rangle \\ |\psi_\downarrow\rangle \end{pmatrix} \;=\; \langle\chi_\uparrow|\psi_\uparrow\rangle + \langle\chi_\downarrow|\psi_\downarrow\rangle\,, \tag{8.2.22}$$

and unit normalization means

$$1 \;=\; \langle\psi|\psi\rangle \;=\; \langle\psi_\uparrow|\psi_\uparrow\rangle + \langle\psi_\downarrow|\psi_\downarrow\rangle\,. \tag{8.2.23}$$

For example, the two-component spinor position wavefunction for an electron is

$$\langle\vec{r}|\psi\rangle \;=\; \begin{pmatrix} \langle\vec{r}|\psi_\uparrow\rangle \\ \langle\vec{r}|\psi_\downarrow\rangle \end{pmatrix} \;=\; \begin{pmatrix} \psi_\uparrow(\vec{r}) \\ \psi_\downarrow(\vec{r}) \end{pmatrix}. \tag{8.2.24}$$

These correspond to orthobasis eigenstates of the CSCO consisting of the observables ($\vec{R}$, $S_z$). If we impose the usual unit normalization condition for the state, the two spin-component wavefunctions are required to satisfy

$$1 \;=\; \int d^3\vec{r}\;\, |\psi_\uparrow(\vec{r})|^2 + |\psi_\downarrow(\vec{r})|^2, \tag{8.2.25}$$

but they are otherwise independent, in general.

## 8.3 Matrix representation for $j = 1$

For a $j = 1$ subspace, the orthobasis of $J_z$ eigenstates consists of three states $|j, m\rangle$ with $j = 1$ and $m = 1, 0, -1$, with corresponding vector representations

$$|1,1\rangle \leftrightarrow \begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad |1,0\rangle \leftrightarrow \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad |1,-1\rangle \leftrightarrow \begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{8.3.1}$$

The matrix representations of the angular momentum operators can now be constructed from eqs. (8.1.6)–(8.1.8), (8.1.22), and (8.1.24), with the results

$$J^2 \leftrightarrow 2\hbar^2 \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \qquad J_z \leftrightarrow \hbar \begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix}, \tag{8.3.2}$$

$$J_+ \leftrightarrow \sqrt{2}\hbar \begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}, \qquad J_- \leftrightarrow \sqrt{2}\hbar \begin{pmatrix}0&0&0\\1&0&0\\0&1&0\end{pmatrix}, \tag{8.3.3}$$

$$J_x \leftrightarrow \frac{\hbar}{\sqrt{2}} \begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}, \qquad J_y \leftrightarrow \frac{\hbar}{\sqrt{2}} \begin{pmatrix}0&-i&0\\i&0&-i\\0&i&0\end{pmatrix}. \tag{8.3.4}$$

As an example of the use of these matrices, suppose that a system is in an eigenstate of $J_z$, and we want to know the possible results and their probabilities if $J_x$ is measured. To answer this question, we first find the eigenvalues and eigenstates of $J_x$, expressed in the orthobasis of $J_z$ eigenvectors of eq. (8.3.1). Since there is nothing special about the $x$ direction as opposed to the $z$ direction, we know even without computing the characteristic equation for $J_x$ that its eigenvalues must be the same as $J_z$, namely $\hbar$, 0, and $-\hbar$. Then, solving for the eigenvectors of the matrix representation for $J_x$, one finds the normalized kets

$$|J_x = \hbar\rangle = \frac{1}{2}|1,1\rangle + \frac{1}{\sqrt{2}}|1,0\rangle + \frac{1}{2}|1,-1\rangle, \tag{8.3.5}$$

$$|J_x = 0\rangle = \frac{1}{\sqrt{2}}|1,1\rangle - \frac{1}{\sqrt{2}}|1,-1\rangle, \tag{8.3.6}$$

$$|J_x = -\hbar\rangle = \frac{1}{2}|1,1\rangle - \frac{1}{\sqrt{2}}|1,0\rangle + \frac{1}{2}|1,-1\rangle. \tag{8.3.7}$$

As a check, these kets are mutually orthogonal, as required (Theorem 2.6.5) by the fact that they are eigenkets of a Hermitian operator with different eigenvalues. Then, for example, the probabilities that a measurement in the state $|1,1\rangle$ will yield the results $J_x = \hbar$, $J_x = 0$, and $J_x = -\hbar$ are, by applying Postulate 4,

$$\mathcal{P}_{J_x=\hbar} = |\langle J_x = \hbar|1,1\rangle|^2 = 1/4, \tag{8.3.8}$$

$$\mathcal{P}_{J_x=0} = |\langle J_x = 0|1,1\rangle|^2 = 1/2, \tag{8.3.9}$$

$$\mathcal{P}_{J_x=-\hbar} = |\langle J_x = -\hbar|1,1\rangle|^2 = 1/4. \tag{8.3.10}$$

## 8.4 Matrix representation for arbitrary $j$

For a subspace with arbitrary $j$, the orthobasis of $J_z$ eigenstates consists of $2j+1$ states $|j,m\rangle$ with $m = j, j-1, \ldots, -j+1, -j$. The corresponding column vector representations of these states, in that order, are

$$|j,\, j\rangle \;\leftrightarrow\; \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad |j,\, j-1\rangle \;\leftrightarrow\; \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \qquad \cdots, \qquad |j,\, -j\rangle \;\leftrightarrow\; \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{8.4.1}$$

Again, eqs. (8.1.6), (8.1.7), (8.1.22), and (8.1.24) provide the matrix elements of the angular momentum operators. The $(2j+1)\times(2j+1)$ matrix representations have the forms

$$J^2 \;\leftrightarrow\; \hbar^2 j(j+1)\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}, \qquad J_z \;\leftrightarrow\; \hbar\begin{pmatrix} j & 0 & \ldots & 0 \\ 0 & j-1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -j \end{pmatrix}, \tag{8.4.2}$$

$$J_+ \;\leftrightarrow\; \hbar\begin{pmatrix} 0 & \bullet & 0 & \ldots & 0 & 0 \\ 0 & 0 & \bullet & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & \bullet \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix}, \qquad J_- \;\leftrightarrow\; \hbar\begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ \bullet & 0 & 0 & \ldots & 0 & 0 \\ 0 & \bullet & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & \bullet & 0 \end{pmatrix}, \tag{8.4.3}$$

where the $\bullet$ represent the only non-zero elements, which lie just above and below the main diagonal for $J_+$ and $J_-$, respectively. The operators $J_x$ and $J_y$ have matrix elements

$$\begin{aligned} (J_x)_{m',m} &= \frac{\hbar}{2}\left[\sqrt{j(j+1)-m(m-1)}\,\delta_{m,m'+1} + \sqrt{j(j+1)-m(m+1)}\,\delta_{m,m'-1}\right], \quad (8.4.4) \\ (J_y)_{m',m} &= i\frac{\hbar}{2}\left[\sqrt{j(j+1)-m(m-1)}\,\delta_{m,m'+1} - \sqrt{j(j+1)-m(m+1)}\,\delta_{m,m'-1}\right], \quad (8.4.5) \end{aligned}$$

and so have the forms

$$J_x \;\leftrightarrow\; \hbar\begin{pmatrix} 0 & \bullet & 0 & \ldots & 0 & 0 \\ \bullet & 0 & \bullet & \ldots & 0 & 0 \\ 0 & \bullet & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & \bullet \\ 0 & 0 & 0 & \ldots & \bullet & 0 \end{pmatrix}, \qquad J_y \;\leftrightarrow\; i\hbar\begin{pmatrix} 0 & -\bullet & 0 & \ldots & 0 & 0 \\ \bullet & 0 & -\bullet & \ldots & 0 & 0 \\ 0 & \bullet & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -\bullet \\ 0 & 0 & 0 & \ldots & \bullet & 0 \end{pmatrix}, \tag{8.4.6}$$

where each pair of $\bullet$'s on opposite sides of the main diagonal are equal, consistent with the Hermiticity of these operators. Note that the $\bullet$'s are also all real and positive in the phase convention we have chosen.

## 8.5 Matrix representations for unitary rotation operators

In section 5.3, we discussed the unitary operators that act on the Hilbert space to generate rotations parameterized by $\vec{\alpha} = \hat{n}\alpha$, where the unit vector $\hat{n}$ is the axis of rotation, and $\alpha$ is the magnitude of the rotation angle,

$$U(\vec{\alpha}) \;=\; \exp\left(-i\vec{\alpha}\cdot\vec{J}/\hbar\right). \tag{8.5.1}$$

The matrix representations for these unitary operators are called **Wigner functions**, after Eugene P. Wigner, and traditionally denoted as $D^{(j)}(\vec{\alpha})$. They are defined by

$$\langle j', m'|U(\vec{\alpha})|j, m\rangle \;=\; \delta_{j,j'} D^{(j)}_{m'm}(\vec{\alpha}) \qquad (m, m' = -j, \ldots, j). \tag{8.5.2}$$

By completeness of the angular momentum eigenstates,

$$U(\vec{\alpha})\,|j,m\rangle = \sum_{m'=-j}^{j} |j,m'\rangle\, D^{(j)}_{m'm}(\vec{\alpha}). \tag{8.5.3}$$

Note that each $D^{(j)}$ is a $(2j+1)\times(2j+1)$ dimensional matrix, which can mix different $m$ values but keeps $j$ fixed.

The set of matrices $D^{(j)}_{m'm}$ for fixed $j$ form an **irreducible representation** of the rotation group, with the following consequence. If you do one rotation $\vec{\alpha}$ followed by another $\vec{\beta}$, the result is always some other rotation $\vec{\gamma}$,

$$U(\vec{\gamma}) \;=\; U(\vec{\beta})U(\vec{\alpha}). \tag{8.5.4}$$

Then, applying completeness gives

$$D^{(j)}_{m'm}(\vec{\gamma}) \;=\; \sum_{m''=-j}^{j} D^{(j)}_{m'm''}(\vec{\beta}) D^{(j)}_{m''m}(\vec{\alpha}). \tag{8.5.5}$$

The unitarity of the operator $U(\vec{\alpha})$ and the fact that $U(\vec{\alpha})^{-1} = U(-\vec{\alpha})$ implies

$$D^{(j)}_{m'm}(-\vec{\alpha}) \;=\; \left[D^{(j)}_{mm'}(\vec{\alpha})\right]^*. \tag{8.5.6}$$

For small $j$, the exponentiation in eq. (8.5.1) can be done explicitly in the matrix representation. For the trivial $j = 0$ case, we have $D^{(0)}(\vec{\alpha}) = e^0 = 1$, the unit $1\times 1$ matrix, independent of $\alpha$. This corresponds to the fact that states with 0 angular momentum are invariant under rotations.

For $j = 1/2$,

$$D^{(1/2)}(\vec{\alpha}) \;=\; \exp(-i\vec{\alpha}\cdot\vec{\sigma}/2) \;=\; \sum_{k=0}^{\infty} \frac{1}{k!}\big(-i\vec{\alpha}\cdot\vec{\sigma}/2\big)^k. \tag{8.5.7}$$

Using eq. (8.2.18), we have $(\vec{\alpha}\cdot\vec{\sigma})^2 = \alpha^2$, so that the terms in eq. (8.5.7) have a recurring matrix structure, which allows us to resum the even and odd terms of the infinite sum separately,

$$D^{(1/2)}(\vec{\alpha}) = \cos(\alpha/2)I - i\hat{n}\cdot\vec{\sigma}\sin(\alpha/2), \tag{8.5.8}$$

where $I$ is the $2\times 2$ unit matrix.

As an example, suppose we have a spin-1/2 that has been measured to be along the $\hat{z}$ direction, so that the state ket is $|\psi\rangle = |\uparrow\rangle$, or in the matrix representation,

$$|\psi\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{8.5.9}$$

Now let us rotate this state by an angle $\theta$ about the $\hat{y}$ axis. The matrix representation of the rotation is

$$D^{(1/2)}(\hat{y}\theta) = \cos(\theta/2)I - i\sigma_y\sin(\theta/2) = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \tag{8.5.10}$$

and so the rotated state has

$$|\psi'\rangle \leftrightarrow \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}, \tag{8.5.11}$$

or $|\psi'\rangle = \cos(\theta/2)\,|\uparrow\rangle + \sin(\theta/2)\,|\downarrow\rangle$. You can check that this is indeed an eigenstate of the rotated operator

$$J'_z = \cos\theta J_z + \sin\theta J_x \leftrightarrow \frac{\hbar}{2}\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \tag{8.5.12}$$

with eigenvalue $+\hbar/2$. If we prepare the system in a spin eigenstate, then the probability to find the spin oriented in a direction at an angle $\theta$ with respect to the original direction is

$$\mathcal{P} = |\langle\psi'|\psi\rangle|^2 = \cos^2(\theta/2). \tag{8.5.13}$$

As a check, this is 1 for $\theta = 0$ or $2\pi$, and 0 for $\theta = \pi$. The probability to find the spin oriented in some particular direction at a right angle to the original spin direction is $\cos^2(\pi/4) = 1/2$.

A surprising feature of eq. (8.5.8) is that a rotation about any axis by an angle $\alpha = 2\pi$ gives, instead of the identity matrix:

$$D^{(1/2)}(\hat{n}2\pi) = -I \tag{8.5.14}$$

for a spin-1/2 system. Thus, the $j = 1/2$ state always acquires a minus sign when continuously rotated by an angle $\alpha = 2\pi$, even though such a $2\pi$ rotation corresponds to no change at all for rotations of classical objects. If the state was in an eigenstate of a particular component $\hat{m}\cdot\vec{J}$, then it will still be after the $2\pi$ rotation, but with a minus sign phase change.

For general $j$, the unitary matrix for a rotation about the $z$ axis is easy to obtain, because $J_z$ has a diagonal matrix representation, making the exponentiation easy,

$$D^{(j)}(\hat{z}\alpha) = \exp\left[-i\alpha \begin{pmatrix} j & 0 & \cdots & 0 \\ 0 & j-1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -j \end{pmatrix}\right] = \begin{pmatrix} e^{-i\alpha j} & 0 & \cdots & 0 \\ 0 & e^{-i\alpha(j-1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{i\alpha j} \end{pmatrix}. \tag{8.5.15}$$

For example, in the $j = 1$ case,

$$D^{(1)}(\hat{z}\alpha) = \text{diag}(e^{-i\alpha},\, 1,\, e^{i\alpha}). \tag{8.5.16}$$

The special case of a rotation through an angle $2\pi$ gives

$$D^{(j)}(\hat{z}2\pi) = (-1)^{2j} I, \tag{8.5.17}$$

which is equal to the identity matrix for integer $j$, but is equal to $-I$ for half-integer $j$, generalizing what was found for $j = 1/2$. For this reason, it is impossible to define a single-valued continuous position wavefunction (like the spherical harmonics introduced in the next section) for non-integer $j$.

For rotations about other axes $\hat{n} \neq \hat{z}$, the algebraic form of matrices $D^{(j)}(\vec{\alpha})$ for general $j$ can be considerably more complicated. The resummation of the infinite sum in the exponential may depend on matrix recurrence relations of higher order. For an example of a harder case that can still be done straightforwardly in closed form, one can evaluate $D^{(1)}(\hat{x}\alpha)$ by using

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}^3 = 2 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{8.5.18}$$

to resum the exponential series, with the result:

$$D^{(1)}(\hat{x}\alpha) = \begin{pmatrix} \cos^2(\alpha/2) & -\frac{i}{\sqrt{2}}\sin(\alpha) & -\sin^2(\alpha/2) \\ -\frac{i}{\sqrt{2}}\sin(\alpha) & \cos(\alpha) & -\frac{i}{\sqrt{2}}\sin(\alpha) \\ -\sin^2(\alpha/2) & -\frac{i}{\sqrt{2}}\sin(\alpha) & \cos^2(\alpha/2) \end{pmatrix}. \tag{8.5.19}$$

Similar expressions for rotations about the $\hat{y}$ and $\hat{z}$ axis for $j = 1$ are left as an exercise.

For general $j$, the numerical value of any $D^{(j)}(\vec{\alpha})$ can always be obtained by exponentiation of the $(2j+1) \times (2j+1)$ matrix corresponding to eq. (8.5.1). There is also a remarkable formula, due to Wigner, for rotations about the $y$-axis,

$$\begin{aligned} D^{(j)}_{m'm}(\hat{y}\beta) = \sum_k & (-1)^{k-m+m'} [\cos(\beta/2)]^{2j-2k+m-m'} [\sin(\beta/2)]^{2k-m+m'} \\ & \frac{[(j+m)!\,(j-m)!\,(j+m')!\,(j-m')!]^{1/2}}{k!\,(k-m+m')!\,(j+m-k)!\,(j-m'-k)!}, \end{aligned} \tag{8.5.20}$$

where the sum is over integers $k$ for which all of the factorials in the denominator have non-negative arguments, in other words $k$ is summed from $\max(0, m-m')$ to $\min(j+m, j-m')$. The proof is omitted here.

The Wigner formula for the matrix $D^{(j)}(\hat{y}\beta)$ is especially useful because it can be used to construct a general rotation matrix. The reason is that a general rotation can always be built up as a product of three sequential rotations about two fixed orthogonal axes. Suppose that we first rotate by an angle $\alpha$ about the $\hat{z}$ axis, then about the $\hat{y}$ axis by an angle $\beta$, and finally by an angle $\gamma$ about the $\hat{z}$ axis again.[†] The change in coordinates induced by this rotation is

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} c_\gamma & -s_\gamma & 0 \\ s_\gamma & c_\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_\beta & 0 & s_\beta \\ 0 & 1 & 0 \\ -s_\beta & 0 & c_\beta \end{pmatrix} \begin{pmatrix} c_\alpha & -s_\alpha & 0 \\ s_\alpha & c_\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{8.5.21}$$

where $c_\alpha = \cos\alpha$, $s_\alpha = \sin\alpha$, etc. The components of vector operators including $\vec{R}$, $\vec{P}$, and $\vec{J}$ rotate in the same way as eq. (8.5.21). The corresponding unitary rotation operator is

$$U(\alpha, \beta, \gamma) = U(\hat{z}\gamma)\, U(\hat{y}\beta)\, U(\hat{z}\alpha), \tag{8.5.22}$$

so that the Wigner rotation matrix for a multiplet with angular momentum $j$ is

$$D^{(j)}(\alpha, \beta, \gamma) = D^{(j)}(\hat{z}\gamma)\, D^{(j)}(\hat{y}\beta)\, D^{(j)}(\hat{z}\alpha). \tag{8.5.23}$$

The matrices for the first and last rotations about the fixed $z$ axis are simple, being just given by eq. (8.5.15).

## 8.6 Orbital angular momentum representation in spherical coordinates

The orbital angular momentum operators $L_x$, $L_y$, and $L_z$ were introduced in section 5.3. We now consider the representation of these operators in terms of the position wavefunctions in spherical coordinates. Recall that in the position orthobasis,

$$\psi(r, \theta, \phi) = \langle r, \theta, \phi | \psi \rangle . \tag{8.6.1}$$

---

[†] There are different conventional ways of defining the three **Euler angles** needed for a general rotation. In classical mechanics, it is traditional to choose the middle rotation to be about the $\hat{x}$ axis. In quantum mechanics it is preferable to use the $\hat{y}$ axis, because $D^{(j)}(\hat{y}\beta)$ has purely real entries, as exhibited in eq. (8.5.20). There are also differing conventions for whether the axes of rotations are absolutely fixed, or whether the second and third rotation axes are the "body" axes, obtained by the previous rotations of the original (fixed) $y$ and $z$ axes. Here, our rotation axes are the absolutely fixed ones.

Using the definitions of $L_x$, $L_y$, and $L_z$ in eqs. (5.3.3), we get the rectangular coordinate representations of these as differential operators:

$$L_x \leftrightarrow -i\hbar \left( y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right), \tag{8.6.2}$$

$$L_y \leftrightarrow -i\hbar \left( z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right), \tag{8.6.3}$$

$$L_z \leftrightarrow -i\hbar \left( x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \right), \tag{8.6.4}$$

which can be converted into spherical coordinate differential operators, with the results

$$L_x \leftrightarrow i\hbar \left( \sin\phi \frac{\partial}{\partial\theta} + \frac{\cos\phi\cos\theta}{\sin\theta}\frac{\partial}{\partial\phi} \right), \tag{8.6.5}$$

$$L_y \leftrightarrow i\hbar \left( -\cos\phi \frac{\partial}{\partial\theta} + \frac{\sin\phi\cos\theta}{\sin\theta}\frac{\partial}{\partial\phi} \right), \tag{8.6.6}$$

$$L_z \leftrightarrow -i\hbar\frac{\partial}{\partial\phi}. \tag{8.6.7}$$

From these, we also obtain for the raising and lowering operators $L_\pm = L_x \pm iL_y$ as defined by eq. (8.1.8),

$$L_+ \leftrightarrow \hbar e^{i\phi} \left( \frac{\partial}{\partial\theta} + i\,\frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial\phi} \right), \tag{8.6.8}$$

$$L_- \leftrightarrow \hbar e^{-i\phi} \left( -\frac{\partial}{\partial\theta} + i\,\frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial\phi} \right), \tag{8.6.9}$$

and, using $L^2 = L_x^2 + L_y^2 + L_z^2 = (L_+L_- + L_-L_+)/2 + L_z^2$ from eq. (8.1.15), we get

$$L^2 \leftrightarrow -\hbar^2 \left( \frac{\partial^2}{\partial\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \right). \tag{8.6.10}$$

For future reference, we note that $L^2$ is closely related to the part of the Laplacian involving angular derivatives; this is useful because the kinetic energy term in the Hamiltonian involves $\nabla^2$. More precisely,

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left( r^2 \frac{\partial}{\partial r} \right) - \frac{1}{\hbar^2 r^2} L^2. \tag{8.6.11}$$

When acting on a wavefunction corresponding to an eigenstate of $L^2$ with eigenvalue $\hbar^2 l(l+1)$, this becomes

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left( r^2 \frac{\partial}{\partial r} \right) - \frac{l(l+1)}{r^2}. \tag{8.6.12}$$

We now solve the eigenvalue problem for orbital angular momentum. In doing so, we can make use of a notable feature of eqs. (8.6.5)–(8.6.10): there is no $r$ dependence in any of the angular momentum operator representations as differential operators on the wavefunctions. Thus we can consider wavefunctions for eigenstates of $L^2$ and $L_z$ in which the $r$ dependence is factored out, with an orthobasis of position eigenkets that are a tensor product of radial and angular parts,

$$|r, \theta, \phi\rangle \;=\; |r\rangle \otimes |\theta, \phi\rangle\,, \tag{8.6.13}$$

with Dirac orthonormality conditions

$$\begin{aligned} \langle r'|r\rangle &= \frac{1}{r^2}\delta(r - r'), & (8.6.14)\\ \langle \theta', \phi'|\theta, \phi\rangle &= \delta(\phi - \phi')\,\delta(\cos\theta - \cos\theta'), & (8.6.15) \end{aligned}$$

and completeness relations

$$\begin{aligned} \int_0^\infty dr\, r^2\, |r\rangle\, \langle r| &= I_r, & (8.6.16)\\ \int d\Omega\; |\theta, \phi\rangle\langle\theta, \phi| &= I_{\theta,\phi}, & (8.6.17) \end{aligned}$$

where $I_r$ and $I_{\theta,\phi}$ are the identity operators on the respective Hilbert spaces, with $I = I_r \otimes I_{\theta,\phi}$. In eq. (8.6.17), and from here on, we define

$$d\Omega \;=\; d\phi\, d(\cos\theta) \tag{8.6.18}$$

as the differential of solid angle in spherical coordinates, so that

$$\int d\Omega \cdots \;=\; \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \cdots\,. \tag{8.6.19}$$

The Hilbert space spanned by the orthobasis of angular eigenkets $|\theta, \phi\rangle$ is also spanned by an orthobasis of $(L^2, L_z)$ eigenkets $|l, m\rangle$ satisfying, from eqs. (8.1.6), (8.1.7), (8.1.22), and (8.1.24),

$$\begin{aligned} L_z\,|l, m\rangle &= \hbar m\,|l, m\rangle\,, & (8.6.20)\\ L_+\,|l, m\rangle &= \hbar\sqrt{l(l+1) - m(m+1)}\;|l, m{+}1\rangle\,, & (8.6.21)\\ L_-\,|l, m\rangle &= \hbar\sqrt{l(l+1) - m(m-1)}\;|l, m{-}1\rangle\,, & (8.6.22)\\ L^2\,|l, m\rangle &= \hbar^2 l(l+1)\,|l, m\rangle\,, & (8.6.23) \end{aligned}$$

for $l = 0, 1, 2, \ldots$ and $m = -l, \ldots, l$. The $|l, m\rangle$ also satisfy orthonormality and completeness relations

$$\begin{aligned} \langle l', m'|l, m\rangle &= \delta_{l,l'}\,\delta_{m,m'}, & (8.6.24)\\ \sum_{l=0}^{\infty}\sum_{m=-l}^{l} |l, m\rangle\langle l, m| &= I_{\theta,\phi}. & (8.6.25) \end{aligned}$$

We now define the angular wavefunctions

$$\langle \theta, \phi | l, m \rangle \;=\; Y_l^m(\theta, \phi), \tag{8.6.26}$$

called **spherical harmonics**.

The actions of the differential operators $L_z$, $L_+$, $L_-$, and $L^2$ defined in eqs. (8.6.7)–(8.6.10) on the spherical harmonic wavefunctions are expressed as

$$L_z Y_l^m(\theta, \phi) \;=\; \hbar m\, Y_l^m(\theta, \phi), \tag{8.6.27}$$

$$L_+ Y_l^m(\theta, \phi) \;=\; \hbar \sqrt{l(l+1) - m(m+1)}\, Y_l^{m+1}(\theta, \phi), \tag{8.6.28}$$

$$L_- Y_l^m(\theta, \phi) \;=\; \hbar \sqrt{l(l+1) - m(m-1)}\, Y_l^{m-1}(\theta, \phi), \tag{8.6.29}$$

$$L^2 Y_l^m(\theta, \phi) \;=\; \hbar^2 l(l+1)\, Y_l^m(\theta, \phi). \tag{8.6.30}$$

These are the position representation differential operator versions of the Hilbert space operator equations (8.6.20)–(8.6.23). For simplicity, we are using the same symbols for the differential operators (acting on wavefunctions) and the corresponding Hilbert space operators (acting on kets and bras). From eqs. (8.6.17) and (8.6.24), the spherical harmonics satisfy the orthonormality condition

$$\int d\Omega\, Y_{l'}^{m'}(\theta, \phi)^*\, Y_l^m(\theta, \phi) \;=\; \delta_{l,l'} \delta_{m,m'}, \tag{8.6.31}$$

and from eqs. (8.6.15) and (8.6.25), the completeness relation

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m(\theta', \phi')^*\, Y_l^m(\theta, \phi) \;=\; \delta(\phi - \phi')\, \delta(\cos\theta - \cos\theta'). \tag{8.6.32}$$

The full Hilbert space is now spanned by the orthobasis of tensor product kets

$$|r\rangle \otimes |l, m\rangle\,, \tag{8.6.33}$$

which describe states in which the particle is known to be at a distance $r$ from the origin, and in which $L^2$ and $L_z$ are also known to be $\hbar^2 l(l+1)$ and $\hbar m$ respectively. These orthobasis elements are an alternative to eq. (8.6.13).

Equation (8.6.27) together with eq. (8.6.7) gives the simple differential equation

$$-i\hbar \frac{\partial}{\partial \phi} Y_l^m \;=\; \hbar m Y_l^m, \tag{8.6.34}$$

which has the general solution

$$Y_l^m(\theta, \phi) \;=\; e^{im\phi} f_l^m(\theta), \tag{8.6.35}$$

where the functions $f_l^m(\theta)$ are now to be determined. We know already that the largest possible value of $m$ is $l$, so let us start with that case. Equation (8.6.28) implies $L_+ Y_l^l = 0$, or

$$\left(\frac{\partial}{\partial\theta} + i\,\frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial\phi}\right) e^{il\phi} f_l^l(\theta) \;=\; 0 \tag{8.6.36}$$

after using eq. (8.6.8). This reduces to

$$\frac{df_l^l}{d(\sin\theta)} = l\frac{f_l^l}{\sin\theta}, \tag{8.6.37}$$

which has the general solution

$$f_l^l(\theta) \;=\; e^C (\sin\theta)^l, \tag{8.6.38}$$

where $C$ is a constant of integration. Thus, we can write

$$Y_l^l(\theta,\phi) \;=\; (-1)^l \sqrt{\frac{(2l+1)!}{4\pi}}\frac{1}{2^l\, l!}\, e^{il\phi}(\sin\theta)^l, \tag{8.6.39}$$

where a multiplicative normalization factor has been included, with magnitude chosen in such a way that eq. (8.6.31) holds with $m' = m = l' = l$. The $(-1)^l$ factor is a choice of convention.

The spherical harmonics for the remaining values $m < l$ can now be obtained by acting repeatedly with $L_-$, using eq. (8.6.29), and comparing to eq. (8.6.9). First,

$$Y_l^{l-1}(\theta,\phi) \;=\; -\frac{e^{-i\phi}}{\sqrt{2l}}\left(\frac{\partial}{\partial\theta} + l\,\frac{\cos\theta}{\sin\theta}\right) Y_l^l(\theta,\phi). \tag{8.6.40}$$

Continuing in the same way, it can be shown by recursion that, for general $-l \leq m \leq l$,

$$Y_l^m(\theta,\phi) \;=\; \frac{(-1)^l}{2^l\, l!}\sqrt{\frac{(2l+1)\,(l+m)!}{4\pi\,(l-m)!}}\, e^{im\phi}\,(\sin\theta)^{-m}\frac{d^{l-m}}{d(\cos\theta)^{l-m}}(\sin\theta)^{2l}. \tag{8.6.41}$$

The normalization factor in eq. (8.6.41) ensures that eqs. (8.6.31) and (8.6.32) are satisfied. The phase convention here is determined by the choices that we have already made in eqs. (8.1.22), (8.1.24), and (8.6.39), and is called the **Condon–Shortley phase convention**. (Other phase and normalization conventions for the spherical harmonics exist, so one must be careful when comparing results from different sources.)

An equivalent way of writing the spherical harmonics is

$$Y_l^m(\theta,\phi) \;=\; \sqrt{\frac{(2l+1)\,(l-m)!}{4\pi\,(l+m)!}}\,(-1)^m e^{im\phi} P_l^m(\cos\theta), \tag{8.6.42}$$

where the $P_l^m(u)$ are called associated Legendre functions. They are solutions to the differential equations

$$\left[(1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx} + l(l+1) - \frac{m^2}{1-x^2}\right] P_l^m(x) \;=\; 0. \tag{8.6.43}$$

For non-negative $m$, the associated Legendre functions are given by

$$P_l^m(x) \;=\; (1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x), \tag{8.6.44}$$

where $P_l^0(x) = P_l(x)$ are the ordinary Legendre polynomials,

$$P_l(x) = \frac{1}{2^l\,l!}\frac{d^l}{dx^l}(x^2-1)^l. \tag{8.6.45}$$

They have a generating function

$$(1-2tx+t^2)^{-1/2} \;=\; \sum_{n=0}^{\infty} t^n P_l(x), \tag{8.6.46}$$

and satisfy the orthonormality relations

$$\int_{-1}^{1} dx\, P_{l'}(x)P_l(x) = \frac{2}{2l+1}\delta_{l,l'}. \tag{8.6.47}$$

The first few Legendre polynomials are

$$P_0(x) \;=\; 1, \qquad\qquad P_1(x) \;=\; x, \tag{8.6.48}$$

$$P_2(x) \;=\; (3x^2-1)/2, \qquad\qquad P_3(x) \;=\; (5x^3-3x)/2, \tag{8.6.49}$$

$$P_4(x) \;=\; (35x^4-30x^2+3)/8. \tag{8.6.50}$$

For negative $m$, the associated Legendre functions are

$$P_l^{-m}(x) = \frac{(l-m)!}{(l+m)!}P_l^m(x) \qquad\qquad (m<0). \tag{8.6.51}$$

With the Condon–Shortley phase convention,

$$Y_l^{-m}(\theta,\phi) \;=\; (-1)^m Y_l^m(\theta,\phi)^*. \tag{8.6.52}$$

For large $l$ and maximal $m$, $Y_l^{\pm l}(\theta,\phi) \propto (\sin\theta)^l$, which is largest in magnitude for $\theta \approx \pi$. Thus, states with large angular momentum about the $z$ axis have probability densities that are peaked near the $xy$ plane and are suppressed near the $z$ axis.

Since $Y_l^m(\theta,\phi)$ is proportional to $e^{im\phi}$, the only spherical harmonics that are independent of $\phi$ are the $m=0$ ones,

$$Y_l^0(\theta,\phi) \;=\; \sqrt{\frac{2l+1}{4\pi}}P_l(\cos\theta). \tag{8.6.53}$$

In the special case $\theta = 0$, the coordinate $\phi$ is arbitrary. This implies that $Y_l^m(0,\phi)$ cannot possibly depend on $\phi$, so it can only be non-zero if $m=0$. From eq. (8.6.53) and $P_l(1) = 1$, we therefore obtain the special value

$$Y_l^m(0,\phi) \;=\; \delta_{m,0}\sqrt{\frac{2l+1}{4\pi}}. \tag{8.6.54}$$

The spherical harmonics that are most commonly encountered in practical problems are the ones for $l \leq 2$, which are

$$Y_0^0 = \frac{1}{\sqrt{4\pi}}, \tag{8.6.55}$$

$$Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta, \qquad\qquad Y_1^{\pm 1} = \mp\sqrt{\frac{3}{8\pi}}\, e^{\pm i\phi}\, \sin\theta, \tag{8.6.56}$$

$$Y_2^0 = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1), \tag{8.6.57}$$

$$Y_2^{\pm 1} = \mp\sqrt{\frac{15}{8\pi}}\, e^{\pm i\phi}\, \sin\theta\, \cos\theta, \qquad Y_2^{\pm 2} = \sqrt{\frac{15}{32\pi}}\, e^{\pm 2i\phi}\, \sin^2\theta. \tag{8.6.58}$$

Note that $Y_l^m$ is always a polynomial of degree $l$ in $\cos\theta$, $\sin\theta$, $\cos\phi$, and $\sin\phi$.

Using completeness of the orthobasis $|l, m\rangle$ over the Hilbert space component corresponding to the angular coordinates, as expressed in eq. (8.6.25), any wavefunction can be expanded as

$$\psi(r, \theta, \phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \left(\langle r| \otimes \langle \theta, \phi|\right) |l, m\rangle \, \langle l, m|\psi\rangle \tag{8.6.59}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_l^m(\theta, \phi) \left(\langle r| \otimes \langle l, m|\right) |\psi\rangle\,. \tag{8.6.60}$$

Defining functions $F_{l,m}(r) = \left(\langle r| \otimes \langle l, m|\right) |\psi\rangle$, this can be rewritten as

$$\psi(r, \theta, \phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} F_{l,m}(r)\, Y_l^m(\theta, \phi). \tag{8.6.61}$$

To find the coefficient functions $F_{l,m}(r)$ for a given $\psi(r, \theta, \phi)$, multiply both sides of eq. (8.6.61) by $Y_{l'}^{m'}(\theta, \phi)^*$, then integrate $d\Omega$, then use the orthonormality condition eq. (8.6.31) to reduce the double sum to a single term with $l' = l$ and $m' = m$, and finally rename $(l', m') \rightarrow (l, m)$. The result is

$$F_{l,m}(r) = \int d\Omega\ Y_l^m(\theta, \phi)^*\, \psi(r, \theta, \phi). \tag{8.6.62}$$

Of course, eqs. (8.6.61) and (8.6.62) also apply if $\psi$ and $F$ have no dependence on $r$.

If the wavefunction $\psi(r, \theta, \phi)$ is normalized, so that $\int_0^\infty dr\, r^2 \int d\Omega\, |\psi(r, \theta, \phi)|^2 = 1$, then one can use the orthonormality condition eq. (8.6.31) again to obtain

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \int_0^\infty dr\, r^2 |F_{l,m}(r)|^2 = 1. \tag{8.6.63}$$

It follows that the probability of simultaneously measuring $L^2$ and $L_z$ and getting the results $\hbar^2 l(l+1)$ and $\hbar m$ is, by using Postulate 4 with $r$ playing the role of the degeneracy label,

$$\mathcal{P}_{l,m} = \int_0^\infty dr\, r^2 \Big|\left(\langle r| \otimes \langle l, m|\right) |\psi\rangle\Big|^2 = \int_0^\infty dr\, r^2 |F_{l,m}(r)|^2. \tag{8.6.64}$$

As a corollary, the probability of measuring just $L^2$ and getting $\hbar^2 l(l+1)$ is obtained by summing this over $m$, so $\mathcal{P}_l = \sum_{m=-l}^{l} \mathcal{P}_{l,m}$. And, the probability of measuring $L_z$ and getting the result $\hbar m$ is $\mathcal{P}_m = \sum_{l=|m|}^{\infty} \mathcal{P}_{l,m}$. Here, the sum over $l$ starts at $|m|$, because for smaller $l$ there are no states that have $L_z$ eigenvalue $\hbar m$.

In many cases, it is not necessary to actually do the radial integral in eq. (8.6.64) in order to evaluate the probabilities; instead one can make use of proportionalities between the various possibilities. As a simple but essential example, any wavefunction that is a function of $r$ only (with no $\theta$ or $\phi$ dependence) is proportional to $Y_0^0$, so one has $l = m = 0$, and the measurements of the compatible orbital angular momentum observables $L^2$ and $L_z$ are both certain to give 0.

For a slightly less trivial example, consider a wavefunction

$$\psi(\vec{r}) \;\; = \;\; C \sin^2\theta \, \cos^2\phi \, e^{-\alpha r}, \tag{8.6.65}$$

where $C$ and $\alpha$ are constants. (Note that we are not providing any context about whether this state is related to any particular Hamiltonian.) If $L^2$ and/or $L_z$ is measured, what are the possible outcomes and their probabilities? To answer this, we seek to write the wavefunction as a linear combination of spherical harmonics multiplied by functions of $r$ only. A useful clue is that the wavefunction is quadratic in sines and cosines of $\theta$ and $\phi$, so one should expect that it will involve $l = 2$ and $l = 0$. Indeed, one finds that

$$\sin^2\theta \, \cos^2\phi \;\; = \;\; \sqrt{\pi}\left[\frac{2}{3}Y_0^0 - \frac{2}{3\sqrt{5}}Y_2^0 + \sqrt{\frac{2}{15}}Y_2^2 + \sqrt{\frac{2}{15}}Y_2^{-2}\right]. \tag{8.6.66}$$

This implies that the measurements of $(L^2, L_z)$ can yield only the four possible pairs $(0, 0)$ and $(6\hbar^2, 0)$ and $(6\hbar^2, 2\hbar)$ and $(6\hbar^2, -2\hbar)$, with probability ratios, respectively,

$$\mathcal{P}_{l=0,m=0} : \mathcal{P}_{l=2,m=0} : \mathcal{P}_{l=2,m=2} : \mathcal{P}_{l=2,m=-2} \;\; = \;\; \left|\frac{2}{3}\right|^2 : \left|-\frac{2}{3\sqrt{5}}\right|^2 : \left|\sqrt{\frac{2}{15}}\right|^2 : \left|\sqrt{\frac{2}{15}}\right|^2. \tag{8.6.67}$$

By requiring that the sum of the probabilities is 1, it follows that

$$\mathcal{P}_{l=0,m=0} = 5/9, \qquad \mathcal{P}_{l=2,m=0} = 1/9, \qquad \mathcal{P}_{l=2,m=2} = \mathcal{P}_{l=2,m=-2} = 1/6. \tag{8.6.68}$$

Note that the radial wavefunction here was irrelevant for the angular momentum measurement probabilities, since it is common factor.

One often needs to consider wavefunctions for states that are pure eigenstates of $L^2$ and $L_z$, with fixed eigenvalues $\hbar^2 l(l + 1)$ and $\hbar m$, respectively. These will have the form

$$\psi(r, \theta, \phi) \;\; = \;\; F(r) Y_l^m(\theta, \phi). \tag{8.6.69}$$

According to eq. (8.6.12), the Laplacian acting on such a wavefunction is

$$\nabla^2 \left[F(r) Y_l^m(\theta,\phi)\right] \;=\; \frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial F}{\partial r}\right) - l(l+1)F\right] Y_l^m(\theta,\phi). \qquad (8.6.70)$$

As a consequence of the rotational invariance of the Laplacian operator, this maintains the form of an eigenfunction of $L^2$ and $L_z$ with the same eigenvalues.

We close this section with the statement and proof of an extraordinarily useful formula.

**Theorem 8.6.1 (Spherical harmonics addition formula)** *Consider any two unit vectors* $\hat{n}$ *and* $\hat{n}'$*, characterized by their spherical coordinate angles* $(\theta,\phi)$ *and* $(\theta',\phi')$*, respectively. Let us call the angle between these vectors* $\gamma$*, so that*

$$\hat{n}\cdot\hat{n}' \;=\; \cos\gamma \;=\; \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi-\phi'). \qquad (8.6.71)$$

*Then, for each* $l$*,*

$$P_l(\cos\gamma) \;=\; \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_l^m(\theta',\phi')^*\, Y_l^m(\theta,\phi). \qquad (8.6.72)$$

Proof: Since $P_l(\cos\gamma)$ is a function of $\theta$ and $\phi$, it can be written as an expansion in spherical harmonics with arguments $\theta$ and $\phi$. Since the differential operator $L^2$ is invariant under rotations, the spherical harmonics involved in the linear combination all must have the same $l$. (The reason is that we can always rotate to a coordinate system in which $\hat{n}'$ is the new $z$ direction, so that the polar angular coordinate of $\hat{n}$ is $\gamma$. In that coordinate system, $P_l(\cos\gamma)$ is proportional to a spherical harmonic with $m = 0$, an eigenfunction of the differential operator $L^2$ with eigenvalue $\hbar^2 l(l+1)$.) Likewise, $P_l(\cos\gamma)$ is also a function of $\theta'$ and $\phi'$, so it can also be written as a linear combination of spherical harmonics of those angles, with the same $l$. Therefore, it must be that $P_l(\cos\gamma)$ is a sum of terms of the form $Y_l^{m'}(\theta',\phi')^*\, Y_l^m(\theta,\phi)$. Each of these terms is proportional to $e^{i(m\phi - m'\phi')}$, but if one chooses $\phi' = \phi$, then $\cos\gamma$ is independent of $\phi$, so only terms with $m' = m$ can occur. Therefore, we must have

$$P_l(\cos\gamma) \;=\; \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta',\phi')^*\, Y_l^m(\theta,\phi), \qquad (8.6.73)$$

and the remaining task is to evaluate the coefficients $a_{lm}$.

Because $P_l(\cos\gamma)$ is real, and is unchanged if we exchange $(\theta,\phi) \leftrightarrow (\theta',\phi')$, the coefficients must satisfy $a_{lm} = a_{lm}^*$, so they are real. To learn more, consider the special case $(\theta,\phi) = (\theta',\phi')$, so that $\cos\gamma = 1$. Then, since the Legendre polynomials satisfy $P_l(1) = 1$, eq. (8.6.73) reads

$$1 \;=\; \sum_{m=-l}^{l} a_{lm}\left|Y_l^m(\theta,\phi)\right|^2. \qquad (8.6.74)$$

Integrating with respect to $d\Omega$, and using the orthonormality condition eq. (8.6.31), we get

$$4\pi = \sum_{m=-l}^{l} a_{lm}. \tag{8.6.75}$$

Next, we consider the square of eq. (8.6.73), which can be written

$$\left[P_l(\cos\gamma)\right]^2 = \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta',\phi')^* \, Y_l^m(\theta,\phi) \sum_{m'=-l}^{l} a_{lm'} Y_l^{m'}(\theta',\phi') \, Y_l^{m'}(\theta,\phi)^*. \tag{8.6.76}$$

Let us integrate over all angles $(\theta', \phi')$. To evaluate the integral of the left side, it is convenient to again use coordinates such that $\gamma$ is the polar angle, and making use of eq. (8.6.47), we have

$$\int d\Omega' \left[P_l(\cos\gamma)\right]^2 = 2\pi \int_{-1}^{1} d(\cos\gamma) \, \left[P_l(\cos\gamma)\right]^2 = \frac{4\pi}{2l+1}. \tag{8.6.77}$$

Meanwhile, the $d\Omega'$ integral of the right side of eq. (8.6.76) is evaluated by using the orthonormality condition eq. (8.6.31) again, after which only the terms with $m' = m$ contribute in the double sum. Comparing the two sides, we get

$$\frac{4\pi}{2l+1} = \sum_{m=-l}^{l} a_{lm}^2 Y_l^m(\theta,\phi)^* Y_l^m(\theta,\phi). \tag{8.6.78}$$

Integrating this with respect to $d\Omega$, and using orthonormality once again, we find

$$\frac{(4\pi)^2}{2l+1} = \sum_{m=-l}^{l} a_{lm}^2. \tag{8.6.79}$$

We now have enough information to solve for the coefficients. Using eqs. (8.6.75) and (8.6.79), we discover that

$$\sum_{m=-l}^{l} \left(a_{lm} - \frac{4\pi}{2l+1}\right)^2 = 0, \tag{8.6.80}$$

and since the left side is a sum of squares, the unique solution is obtained when each term vanishes, so that $a_{lm} = 4\pi/(2l+1)$ for all $l, m$. This concludes the proof of the spherical harmonics addition formula.

## 8.7 Parity of angular momentum eigenstates

As we noted in section 5.3, angular momentum operators have even parity, which is another way of saying that they commute with the parity operator $\Pi$. According to Theorem 2.7.1, this means that the eigenstates of angular momentum operators $J^2, J_z$ can also be chosen to be parity eigenstates.

First, let us work out the parity eigenvalues of the orbital angular momentum eigenstates $|l,m\rangle$. If $(x,y,z) \to (-x,-y,-z)$, the spherical coordinate transformations are

$$r \to r, \qquad \theta \to \pi - \theta, \qquad \phi \to \phi + \pi, \tag{8.7.1}$$

so that

$$\cos\theta \;\to\; -\cos\theta, \qquad \sin\theta \;\to\; \sin\theta, \qquad e^{im\phi} \to (-1)^m e^{im\phi}. \tag{8.7.2}$$

Using these, it follows immediately from eq. (8.6.41) that

$$Y_l^m(\pi - \theta, \phi + \pi) \;=\; (-1)^l Y_l^m(\theta,\phi). \tag{8.7.3}$$

Therefore,

$$\langle \theta,\phi|\Pi|l,m\rangle \;=\; \langle \pi-\theta,\phi+\pi|l,m\rangle \;=\; (-1)^l Y_l^m(\theta,\phi) \;=\; (-1)^l \langle \theta,\phi|l,m\rangle\,, \tag{8.7.4}$$

so

$$\Pi\,|l,m\rangle \;=\; (-1)^l\,|l,m\rangle\,. \tag{8.7.5}$$

This shows that eigenstates of orbital angular momentum are always even (odd) under parity if the quantum number $l$ is even (odd). This is true regardless of the quantum number $m$, which can be understood from the fact that the states $|l,m\rangle$ are all obtained from $|l,l\rangle$ by acting with $L_-$, which commutes with $\Pi$. [Therefore, the parity eigenvalue $(-1)^l$ could actually have been obtained by considering the spherical harmonics for only one sample value of $m$ for each $l$, for example $m = l$ using eq. (8.6.39), rather than the more general formula of eq. (8.6.41).] Since parity in spherical coordinates does not change the radial coordinate, the parity eigenvalue of an angular momentum eigenstate also does not depend on the radial wavefunction or radial quantum numbers.

Let us next consider the parity of eigenkets $|s,m_s\rangle$ of intrinsic angular momentum operators $S^2$ and $S_z$. From the fact that the lowering operator $S_-$ commutes with parity and relates kets with different $m_s$, we know that the parity eigenvalue of $|s,m_s\rangle$ cannot depend on $m_s$. We can therefore write

$$\Pi\,|s,m_s\rangle = \eta\,|s,m_s\rangle\,, \tag{8.7.6}$$

where $\eta$ is known as the **intrinsic parity** of the particle in question. For any given particle type, the intrinsic parity can be chosen to be either $+1$ or $-1$, as an arbitrary convention. This conventional choice cannot be of any practical significance at all if the number of particles of each type does not change. This is because a change in the conventional choice of intrinsic parity just amounts to flipping the parity for every bra and every ket, therefore not affecting matrix elements. So, when the number of particles does not change, one might as well just always choose $\eta = +1$, and the parity of a single particle $L^2$ eigenstate is $(-1)^l$, regardless of the spin. For a system of $N$ particles with individual orbital angular momentum quantum numbers $l_i$, the parity eigenvalue is, from eq. (8.7.5),

$$\pi \;=\; (-1)^{\sum_{i=1}^{N} l_i}. \tag{8.7.7}$$

This is the case for electrons in an atom, where the parity and the angular momenta are all defined with respect to the origin chosen to be the location of the fixed nucleus.

However, if the Hamiltonian is invariant under parity and can cause changes in the numbers of particles, it is natural and useful to adopt a convention in which the intrinsic parities of particles and antiparticles are chosen in a consistent way so that parity is conserved. Such Hamiltonians arise in quantum field theory, where the electromagnetic and strong nuclear interactions conserve parity, while the weak nuclear interactions do not but can often be treated as a perturbation. A generally accepted convention is to choose $\eta = +1$ for spin-1/2 particles (the electron, proton, neutron, muon, and quarks, for example). Then the structure of kinetic terms in relativistic quantum field theory can be used to show that their antiparticles (the positron, antiproton, antineutron, antimuon, and antiquarks) must be assigned $\eta = -1$ in the same convention.

For a composite particle $c$, made out of two particles $a$ and $b$ with intrinsic parities $\eta_a$ and $\eta_b$ in a state with relative angular momentum quantum number $L$, the consistent intrinsic parity assignment is[†]

$$\eta_c = (-1)^L \eta_a \eta_b. \tag{8.7.8}$$

Thus, mesons composed of a quark and an antiquark in an $L = 0$ bound state (for example, pions and kaons) have intrinsic parity $\eta = -1$. For a bound state of three or more particles, the situation is more complicated, but the intrinsic parities of bound states can always be defined if the Hamiltonian is invariant under space inversion. In processes governed by the electromagnetic and strong nuclear forces, one can experimentally verify parity conservation and check the consistency of the intrinsic parity assignments. In doing so, the photon and the gluon have intrinsic parity $-1$, while the Higgs particle has intrinsic parity $+1$. The $W$ and $Z$ particles are not assigned a well-defined intrinsic parity, as they are the mediators of the parity-violating weak interactions.

---

[†] Note that this differs from eq. (8.7.7), since $(-1)^L$ is not always equal to $(-1)^{l_a + l_b}$. This is because the angular momentum $L$ and the intrinsic parity $\eta_c$ in eq. (8.7.8) are defined with respect to the origin as the center-of-mass position of the two particles, rather than a fixed position as in eq. (8.7.7).

# 9 Examples with spherical symmetry

## 9.1 Stationary states with spherical symmetry

Consider a Hamiltonian of the form

$$H = \frac{P^2}{2\mu} + V(R). \tag{9.1.1}$$

In this chapter, we will write $\mu$ for the mass of the particle, to avoid confusion with the eigenvalue $\hbar m$ of $L_z$, and $R$ is the operator corresponding to the classical spherical coordinate $r$. Thus $V(R)$ is a spherically symmetric potential, with no dependence on $\theta$ or $\phi$. Likewise, the momentum squared operator

$$P^2 = P_x^2 + P_y^2 + P_z^2 = -\hbar^2\nabla^2 \tag{9.1.2}$$

is rotationally invariant, so there is no preferred direction associated with the Hamiltonian. It follows that $H$, $L^2$, and $L_z$ are compatible observables, and we can look for an orthobasis of common eigenstates

$$|E, l, m\rangle, \tag{9.1.3}$$

where $E$ is the energy of the state, the eigenvalue of the Hamiltonian. Because of eq. (8.7.5), these are also parity eigenstates, with eigenvalue $(-1)^l$.

In the following, we will work in the position wavefunction representation with

$$\psi_{E,l,m}(r,\theta,\phi) = \langle r,\theta,\phi|E,l,m\rangle = R_{E,l}(r)Y_l^m(\theta,\phi), \tag{9.1.4}$$

where $R_{E,l}(r)$ is a radial wavefunction. Note that $R_{E,l}(r)$ will depend on both $E$ and $l$ in general, but it will not depend on $m$, because the operators $L_+$ and $L_-$ raise and lower $m$ without changing $l$ or the radial dependence of the wavefunction. The time-independent Schrödinger equation in this basis is

$$\left[-\frac{\hbar^2\nabla^2}{2\mu} + V(r)\right]\psi_{E,l,m}(r,\theta,\phi) = E\psi_{E,l,m}(r,\theta,\phi). \tag{9.1.5}$$

Now, using eq. (8.6.12), this becomes

$$-\frac{\hbar^2}{2\mu}\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2}\right]R_{E,l}(r) + V(r)R_{E,l}(r) = ER_{E,l}(r), \tag{9.1.6}$$

where the spherical harmonics have been factored out.

In many cases, we will find that the radial wavefunction $R_{E,l}(r)$ has a power-law behavior $r^p$ near the origin, for an integer $p$. By requiring the total probability to be finite, it is clear that

$p$ cannot be less than $-1$; otherwise, the normalization integral $\int dr\, r^2|R_{E,l}(r)|^2$ would diverge. But in almost all cases, one can make the even stronger statement that $R_{E,l}(r)$ must be finite as $r \to 0$. This is because $\nabla^2(1/r) = -4\pi\delta(\vec{r})$, so that $p = -1$ would necessarily imply the presence of a delta function contribution to the potential $V(r)$. In the examples to follow, we will often make use of the finiteness of the radial wavefunction.

It is often useful to define a function

$$U_{E,l}(r) \;=\; rR_{E,l}(r), \tag{9.1.7}$$

in which one power of the radial coordinate has been factored out. The eigenvalue differential equation then becomes

$$\left[-\frac{\hbar^2}{2\mu}\frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2\mu r^2} + V(r)\right] U_{E,l} \;=\; EU_{E,l}, \tag{9.1.8}$$

with the nice feature that there is no term with a single $r$ derivative; that is the reason for sometimes using $U_{E,l}$ rather than $R_{E,l}$.

Indeed, eq. (9.1.8) for $U_{E,l}$ is very similar to the one-dimensional time-independent Schrödinger equation (6.3.3) with $x$ replaced by $r$. However, there are two important differences. First, the domain of the independent variable is now limited to non-negative values,

$$0 \leq r < \infty, \tag{9.1.9}$$

unlike the one-dimensional Schrödinger equation. Second, the potential has effectively been modified to

$$V_{\text{eff}}(r) \;=\; V(r) + \frac{\hbar^2 l(l+1)}{2\mu r^2}, \tag{9.1.10}$$

where the second term is a repulsive centrifugal contribution that blows up at $r = 0$, and therefore makes the wavefunction vanish there unless $l = 0$. In fact, if $V(r)$ is finite at $r = 0$, one finds from either eq. (9.1.6) or (9.1.8) that for $l \neq 0$, $R_{E,l}$ must scale like $r^l$ at very small $r$.

With these caveats, the problem of stationary states for a particle in a spherically symmetric potential has been reduced to a particular type of one-dimensional problem. The same results that we derived in section 6.3 [see the discussion surrounding eqs. (6.3.4)–(6.3.6)] for matching wavefunctions at special points holds here as for the one-dimensional problems. In particular, the radial wavefunction is always continuous, and if the potential is finite (not necessarily continuous) at a point $r = r_0$, then its first derivative with respect to $r$ is continuous there.

## 9.2 Free particle in spherical coordinates

We have already discussed the position and momentum eigenstates and wavefunctions for a free particle moving in three dimensions, in section 2.8. For position eigenstates, the CSCO used was

$(X, Y, Z)$, while for momentum eigenstates, it was $(P_x, P_y, P_z)$. Now we will discuss the same physical problem using the CSCO ($H$, $L^2$, $L_z$) with eigenvalues ($E$, $\hbar^2 l(l+1)$, $\hbar m$), starting from eq. (9.1.8) with $V(r) = 0$. We will solve the time-independent Schrödinger equation for fixed $(E, l, m)$ in the position wavefunction representation with coordinates $(r, \theta, \phi)$, where the domain may consist of only a sub-volume of the whole space. The resulting solutions can then be applied, using matching of the wavefunction, to problems in which a particle moves freely only in that sub-volume. They can also be applied to scattering problems in three dimensions.

To simplify things, define

$$k^2 = 2\mu E/\hbar^2, \tag{9.2.1}$$

$$\rho = kr. \tag{9.2.2}$$

From the discussion in section 2.8, we already know that the allowed eigenvalues $E$ are positive, so that $k$ is a real number with units of 1/[length]. Thus $\rho$ is a radial coordinate re-scaled to make it dimensionless. In terms of these, eq. (9.1.8) becomes

$$\left(\frac{d^2}{d\rho^2} + 1 - \frac{l(l+1)}{\rho^2}\right) U_l(\rho) = 0. \tag{9.2.3}$$

Note that this equation and the forms of its solutions $U_l(\rho)$ do not depend on $E$, because the dependence has been hidden in $\rho$.

Let us start with the case $l = 0$. Then eq. (9.2.3) is a familiar equation, with the familiar solutions $U_0(\rho) = \sin(\rho)$ or $\cos(\rho)$. Of these, the first solution $U_0 = \sin(\rho)$ gives a constant for $R(r)$ as $r \to 0$. However, the second solution $U_0 = \cos(\rho)$ approaches a constant as $r \to 0$. This renders it physically unacceptable if the origin is included in the spatial region under consideration, because then $R(r) \sim 1/r$, which is divergent and would require a delta function potential $V(r)$ at the origin, since $\nabla^2(1/r) = -4\pi\delta(\vec{r})$. Nevertheless, it is acceptable if we are solving for the free-particle wavefunction only in a region that does not include the origin. We therefore have two solutions for $L^2 = 0$, labeled $A$ and $B$,

$$U_0^A(kr) = \sin(kr), \qquad U_0^B(kr) = -\cos(kr), \tag{9.2.4}$$

where the minus sign is a phase choice for later convenience, and the $B$ solution is understood to be acceptable if, and only if, the origin is excluded. These solutions for $l = 0$ can be used as seeds to find solutions for $l > 0$, as we will now see.

Writing the solution for general $l$ in the form

$$U_l(\rho) = \rho^{l+1} f_l(\rho), \tag{9.2.5}$$

the differential equation (9.2.3) becomes

$$\left[\frac{d^2}{d\rho^2} + \frac{2(l+1)}{\rho}\frac{d}{d\rho} + 1\right] f_l(\rho) = 0. \tag{9.2.6}$$

Now we note that if $f_l(\rho)$ satisfies the differential equation for $l$, then

$$f_{l+1}(\rho) \equiv -\frac{1}{\rho}\frac{df_l}{d\rho} \tag{9.2.7}$$

will satisfy it for $l+1$, as can be proved by computing $d^2 f_{l+1}/d\rho^2$ and $df_{l+1}/d\rho$ and plugging into eq. (9.2.6). So, for each of our two $l = 0$ solutions from eq. (9.2.4),

$$f_0^A(\rho) = \frac{\sin\rho}{\rho}, \qquad f_0^B(\rho) = -\frac{\cos\rho}{\rho}, \tag{9.2.8}$$

by recursion we will have a solution

$$f_l = \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l f_0. \tag{9.2.9}$$

Therefore, for each value of $l$, we have two linearly independent solutions

$$R_l^A(\rho) = \rho^l f_l^A = \rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{\sin\rho}{\rho}\right) \equiv j_l(\rho), \tag{9.2.10}$$

$$R_l^B(\rho) = \rho^l f_l^B = \rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(-\frac{\cos\rho}{\rho}\right) \equiv n_l(\rho). \tag{9.2.11}$$

The functions $j_l(\rho)$ are called the **spherical Bessel functions**, and $n_l(\rho)$ are called the **spherical Neumann functions**. The lowest few are

$$j_0(\rho) = \frac{\sin\rho}{\rho}, \qquad n_0(\rho) = -\frac{\cos\rho}{\rho}, \tag{9.2.12}$$

$$j_1(\rho) = \frac{\sin\rho}{\rho^2} - \frac{\cos\rho}{\rho}, \qquad n_1(\rho) = -\frac{\cos\rho}{\rho^2} - \frac{\sin\rho}{\rho}, \tag{9.2.13}$$

$$j_2(\rho) = \left(\frac{3}{\rho^3} - \frac{1}{\rho}\right)\sin\rho - \frac{3}{\rho^2}\cos\rho, \qquad n_2(\rho) = \left(-\frac{3}{\rho^3} + \frac{1}{\rho}\right)\cos\rho - \frac{3}{\rho^2}\sin\rho. \tag{9.2.14}$$

For small $\rho$, they can be shown to behave like

$$j_l(\rho) \sim \frac{\rho^l}{(2l+1)!!}, \tag{9.2.15}$$

$$n_l(\rho) \sim -\frac{(2l-1)!!}{\rho^{l+1}}, \tag{9.2.16}$$

where the double factorial notation means

$$(2l+1)!! = (2l+1)(2l-1)\cdots(1) = (2l+1)!/(2^l l!), \tag{9.2.17}$$

with the special value $(-1)!! = 1$. Thus the $j_l$ solutions are well-behaved at the origin, but the $n_l$ solutions are not. For large $\rho$, they both fall off like $1/\rho$ multiplied by oscillating functions,

$$j_l(\rho) \sim \frac{1}{\rho}\sin(\rho - l\pi/2), \tag{9.2.18}$$

$$n_l(\rho) \sim -\frac{1}{\rho}\cos(\rho - l\pi/2). \tag{9.2.19}$$

The functions $n_l(\rho)$ are not needed for the case of a completely free particle (one that has no potential everywhere including $r = 0$). However, they are useful in situations where we piece together the free-particle solution(s) in region(s) not including the origin to other solutions that do include the origin, as for example in section 9.4.

For example, suppose we have a potential that is spherically symmetric and piece-wise constant within radial intervals, so that

$$V(r) = \begin{cases} V_0 & (0 \leq r < a_1), \\ V_1 & (a_1 < r < a_2), \\ \cdots & \cdots \\ V_N & (a_N < r < \infty). \end{cases} \tag{9.2.20}$$

In each of these regions, the constant potential $V_n$ can be absorbed into the constant energy, so the possible stationary-state solution wavefunctions are

$$\left[A_l^{(n)}\, j_l(k_n r) + B_l^{(n)}\, n_l(k_n r)\right]\, Y_l^m(\theta,\phi), \qquad \text{for } (a_n < r < a_{n+1}), \tag{9.2.21}$$

for some constants $A_l^{(n)}$ and $B_l^{(n)}$, but now with

$$k_n = \sqrt{2\mu(E - V_n)}/\hbar. \tag{9.2.22}$$

The coefficients $A_l^{(n)}$ and $B_l^{(n)}$ can be determined by matching the wavefunctions, and their first derivative with respect to $r$, at each of the points $a_n$. This procedure also simultaneously determines the energy eigenvalue $E$. In the region $0 \leq r < a_1$, the Neumann function solutions are not allowed, so one must have $B_l^{(0)} = 0$, but in all other regions $B_l^{(n)}$ is allowed to be non-zero. Note that there is a degeneracy $2l + 1$ for each of these stationary states, due to the fact that the energy does not depend on $m = -l, \ldots, l$. One can also match such wavefunctions to regions in which the potential is something more complicated.

Different linear combinations of $j_l$ and $n_l$ are useful in certain kinds of problems. The **spherical Hankel functions** defined by

$$h_l^{(1)}(\rho) = j_l(\rho) + i n_l(\rho) = -i\rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{e^{i\rho}}{\rho}\right), \tag{9.2.23}$$

$$h_l^{(2)}(\rho) = j_l(\rho) - i n_l(\rho) = i\rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{e^{-i\rho}}{\rho}\right), \tag{9.2.24}$$

correspond to outgoing and ingoing spherical waves, respectively. To understand this, note that from eqs. (9.2.18) and (9.2.19) their behavior for large $\rho = kr$ is

$$\begin{aligned} h_l^{(1)}(\rho) &\approx \frac{1}{\rho} e^{i[\rho - \pi(l+1)/2]} = i^{-l} e^{i\rho}/\rho, && (9.2.25) \\ h_l^{(2)}(\rho) &\approx \frac{1}{\rho} e^{-i[\rho + \pi(l+1)/2]} = i^{l} e^{-i\rho}/\rho, && (9.2.26) \end{aligned}$$

so that the time dependence of the corresponding wavefunctions is

$$\begin{aligned} \langle \vec{r}\,|\psi^{(1)}(t)\rangle &\propto e^{-itE/\hbar} e^{ikr}/r, && (9.2.27) \\ \langle \vec{r}\,|\psi^{(2)}(t)\rangle &\propto e^{-itE/\hbar} e^{-ikr}/r. && (9.2.28) \end{aligned}$$

These maintain constant phase at $r = (E/\hbar k)\,t + \text{constant}$ and $r = -(E/\hbar k)\,t + \text{constant}$, respectively. (The phase velocity is therefore $\omega/k = E/p = \hbar k/2m$, but recall from the discussion leading to eq. (6.2.6) that wave-packet superpositions will have a group velocity twice this, $\hbar k/m$, which is the classical value.)

For problems in which $E < V_n$ in at least one region, $k$ will be imaginary in that region. Then the solutions will involve $j_l(kr)$ and $n_l(kr)$ or $h_l^{(1)}(kr)$ and $h_l^{(2)}(kr)$, where now $k = i\sqrt{2\mu(V_n - E)}/\hbar$ is a pure imaginary number. In many such cases, a more convenient basis of solutions is provided by the **modified spherical Bessel functions**,

$$\begin{aligned} i_l(\rho) &= j_l(i\rho)/i^l = \rho^l \left(\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{\sinh\rho}{\rho}\right), && (9.2.29) \\ k_l(\rho) &= -i^l h_l^{(1)}(i\rho) = \rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{e^{-\rho}}{\rho}\right), && (9.2.30) \end{aligned}$$

where now

$$\rho = \kappa r, \qquad \kappa = k/i = \sqrt{2\mu(V_n - E)}/\hbar. \tag{9.2.31}$$

The function $i_l(\rho)$ is well-defined and useful for regions that include $r = 0$, while the function $k_l(\rho)$ is useful for regions that extend to $r = \infty$.

For the rest of this section, we consider the $(H, L^2, L_z)$ orthobasis eigenstate wavefunctions for the completely free particle whose domain includes all space, including the origin. They are

$$\langle \vec{r}\,|k, l, m\rangle = \psi_{k,l,m}(\vec{r}) = R_{k,l}(r) Y_l^m(\theta, \phi), \qquad (E = \hbar^2 k^2/2\mu), \tag{9.2.32}$$

where the radial wavefunctions $R_{k,l}(r) = A_{k,l}\, j_l(kr)$ contain a normalization constant $A_{k,l}$ to be fixed. To do so, we can use the orthonormality relations for the spherical Bessel functions,[†]

$$\int_0^\infty dr\, r^2 j_l(kr) j_l(k'r) = \frac{\pi}{2k^2}\delta(k - k'), \tag{9.2.33}$$

---

[†] Equation 9.2.33 is technically ill-defined, because the integration does not converge, even for $k \neq k'$. It should therefore be viewed as a formal relation, to be used within expressions where $k$ or $k'$ is integrated over. This is very similar to the technically ill-defined nature of the Fourier integral for the delta function in eq. (2.2.20).

for each $l$. It follows that if we choose $A_{k,l} = \sqrt{2k^2/\pi}$, so that the radial wavefunctions are

$$R_{k,l}(r) = \sqrt{\frac{2k^2}{\pi}} j_l(kr), \tag{9.2.34}$$

then we have the orthonormality relations,

$$\int dr\, r^2 \, R_{k',l}(r) R_{k,l}(r) \;=\; \delta(k - k'), \tag{9.2.35}$$

and for the full eigenstates and their wavefunctions,

$$\langle k', l', m' \,| k, l, m\rangle = \int d^3\vec{r}\;\, \psi_{k',l',m'}(\vec{r})^* \, \psi_{k,l,m}(\vec{r}) \;=\; \delta(k - k')\, \delta_{l,l'} \delta_{m,m'}. \tag{9.2.36}$$

With our choice of normalization in eqs. (9.2.34)–(9.2.36), we also have the completeness of the orthobasis in terms of an integral over $k$,

$$\int_0^\infty dk \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |k, l, m\rangle\langle k, l, m| \;=\; 1, \tag{9.2.37}$$

from which follows

$$\int_0^\infty dk \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \psi_{k,l,m}(\vec{r}\,')^* \psi_{k,l,m}(\vec{r}) \;=\; \delta^{(3)}(\vec{r} - \vec{r}\,'). \tag{9.2.38}$$

In eqs. (9.2.34)-(9.2.38), we have chosen a "wavenumber normalization" for the Dirac orthonormality of the energy eigenstates, as they contain $\delta(k-k')$ and a corresponding integration over $k$. Since $\delta(k - k') = \delta(E - E')\hbar^2 k/\mu$, a viable alternative would be to express the same results in terms of "energy normalization" kets

$$|E, l, m\rangle \;=\; \sqrt{\frac{\mu}{\hbar^2 k}}\; |k, l, m\rangle\,, \tag{9.2.39}$$

and the corresponding radial wavefunctions

$$R_{E,l}(r) = \sqrt{\frac{2\mu k}{\pi\hbar^2}}\, j_l(kr). \tag{9.2.40}$$

If one uses $|E, l, m\rangle$, $R_{E,l}(r)$, and $\psi_{E,l,m}(\vec{r})$, then the Dirac orthonormality and completeness relations have the same form as eqs. (9.2.35)-(9.2.38), but with $\delta(E - E')$ replacing $\delta(k - k')$ and $\int_0^\infty dE$ replacing $\int_0^\infty dk$.

Let us now consider the relation between the plane wavefunctions associated with momentum eigenstates $|\vec{p}\rangle$ and the spherical waves associated with the $(H, L^2, L_z)$ eigenstates $|k, l, m\rangle$. Using the completeness relation, we have

$$\langle \vec{r}|\vec{p}\rangle \;=\; \int_0^\infty dk \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \langle \vec{r}|k, l, m\rangle\langle k, l, m|\vec{p}\rangle\,. \tag{9.2.41}$$

Now, both $|\vec{p}\rangle$ and $|k,l,m\rangle$ are eigenstates of $H$ with energy eigenvalues $E = |\vec{p}|^2/2\mu$ and $\hbar^2 k^2/2\mu$, respectively. Therefore, Theorem 2.6.5 says that the last inner product in eq. (9.2.41) must vanish unless $|\vec{p}| = \hbar k$, and so it must be of the form

$$\langle k,l,m|\vec{p}\rangle \;=\; \frac{1}{(2\pi\hbar)^{3/2}}\,\delta(k-|\vec{p}|/\hbar)\,\sqrt{\frac{\pi}{2k^2}}\,C_{l,m}(\theta_{\vec{p}},\,\phi_{\vec{p}}), \tag{9.2.42}$$

for some functions $C_{l,m}$ that depend only on $l$, $m$, and the angular coordinates of the vector $\vec{p}$. (By dimensional analysis, the $C_{l,m}$ cannot depend on the magnitude $k$, because they are dimensionless and there is no other dimensionful quantity on which they could depend.) Since $k = |\vec{p}|/\hbar$ is enforced by the delta function, we can define $\vec{k} = \vec{p}/\hbar$ and use eqs. (2.8.62) and (9.2.32) to find that eq. (9.2.41) reads

$$\frac{e^{i\vec{k}\cdot\vec{r}}}{(2\pi\hbar)^{3/2}} \;=\; \frac{1}{(2\pi\hbar)^{3/2}}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{l,m}(\theta_{\vec{k}},\phi_{\vec{k}})\,j_l(kr)Y_l^m(\theta,\phi). \tag{9.2.43}$$

The claim is that the coefficient functions turn out to be

$$C_{l,m}(\theta_{\vec{k}},\phi_{\vec{k}}) \;=\; 4\pi i^l\,Y_l^m(\theta_{\vec{k}},\phi_{\vec{k}})^*, \tag{9.2.44}$$

so that we have the remarkable identity relating plane waves to spherical waves,

$$e^{i\vec{k}\cdot\vec{r}} \;=\; \sum_{l=0}^{\infty}\sum_{m=-l}^{l} 4\pi i^l\,Y_l^m(\theta_{\vec{k}},\phi_{\vec{k}})^*\,j_l(kr)Y_l^m(\theta,\phi). \tag{9.2.45}$$

Note that $(\theta_{\vec{k}},\phi_{\vec{k}})$ are the spherical coordinate angles for the vector $\vec{k}$, while $(\theta,\phi)$ are the angles for the vector $\vec{r}$ in the same coordinate system. The interpretation of eq. (9.2.45) is that a plane wave with momentum $\vec{p} = \hbar\vec{k}$ consists of a superposition of spherical waves with all allowed (quantized) values of orbital angular momentum.

To prove the claimed eq. (9.2.45), we first consider the case that $\vec{k} = k\hat{z}$. Then $\theta_{\vec{k}} = 0$, and $\vec{k}\cdot\vec{r} = kr\cos\theta$, so

$$e^{i\vec{k}\cdot\vec{r}} \;=\; e^{ikr\cos\theta}. \tag{9.2.46}$$

Because this does not depend on $\phi$ at all, its expansion in terms of spherical harmonics will only include the $m = 0$ functions $Y_l^0(\theta,\phi)$, which are proportional to the ordinary Legendre polynomials $P_l(\cos\theta)$. Therefore, in this case the expansion in eq. (9.2.45) is of the simpler form

$$e^{ikr\cos\theta} \;=\; \sum_{l=0}^{\infty} c_l\,j_l(kr)P_l(\cos\theta), \tag{9.2.47}$$

where $c_l$ are some coefficients that we need to determine. Now we can multiply both sides of eq. (9.2.47) by $P_{l'}(\cos\theta)$, and integrate with respect to $\cos\theta$, making use of the identity

$$\int_{-1}^{1} du\, P_l(u) P_{l'}(u) \;=\; \frac{2}{2l+1}\delta_{l,l'}. \tag{9.2.48}$$

The result, after relabeling $l' \to l$, is

$$c_l\, j_l(kr) \;=\; \frac{2l+1}{2}\int_{-1}^{1} d(\cos\theta)\, P_l(\cos\theta) e^{ikr\cos\theta}. \tag{9.2.49}$$

This must hold for all $r$, so we can consider the leading behavior as $r \to 0$ in particular. Using eq. (9.2.15), this gives

$$c_l \frac{(kr)^l}{(2l+1)!!} + \mathcal{O}(r^{l+1}) \;=\; \frac{2l+1}{2}\int_{-1}^{1} du\, P_l(u) e^{ikru}. \tag{9.2.50}$$

This shows that the right side evidently must be proportional to $r^l$ in the limit of small $r$. Expanding $e^{ikru}$ in a power series in $r$, using $e^{ikru} = \sum_{n=0}^{\infty}(ikru)^n/n!$, this implies the identities

$$\int_{-1}^{1} du\, u^n P_l(u) \;=\; 0 \qquad \text{(for integer } n < l\text{)}, \tag{9.2.51}$$

and the result we need (from the $n = l$ term),

$$c_l \frac{k^l}{(2l+1)!!} \;=\; \frac{2l+1}{2}\frac{(ik)^l}{l!}\int_{-1}^{1} du\, u^l P_l(u). \tag{9.2.52}$$

Using eq. (8.6.45) for the Legendre polynomials, one can obtain the integral

$$\int_{-1}^{1} du\, u^l P_l(u) \;=\; \frac{2(l!)}{(2l+1)!!}, \tag{9.2.53}$$

by integrating by parts $l$ times. Thus, eq. (9.2.52) simplifies to

$$c_l \;=\; (2l+1) i^l, \tag{9.2.54}$$

so that

$$e^{ikr\cos\theta} \;=\; \sum_{l=0}^{\infty}(2l+1) i^l\, j_l(kr) P_l(\cos\theta). \tag{9.2.55}$$

Now for the case of $\vec{k}$ in an arbitrary direction, substitute $kr\cos\theta \to \vec{k}\cdot\vec{r}$ in the preceding, or

$$\cos\theta \to \hat{k}\cdot\hat{r}. \tag{9.2.56}$$

Using the spherical harmonic addition identity, eq. (8.6.72), one finally arrives at eq. (9.2.45).

## 9.3 Particle confined to a sphere

As an application of the results of the previous section, consider a particle of mass $\mu$ that is confined within a sphere of radius $a$, but is otherwise free, so that the potential is

$$V(r) = \begin{cases} 0 & (r < a), \\ \infty & (r \geq a). \end{cases} \tag{9.3.1}$$

The stationary-state wavefunctions are

$$\psi_{E,l,m} = \begin{cases} C_{E,l,m}\, j_l(kr) Y_l^m(\theta, \phi) & (r < a), \\ 0 & (r \geq a), \end{cases} \tag{9.3.2}$$

where $C_{E,l,m}$ are normalization constants and $E = \hbar^2 k^2/2\mu$. Continuity of the wavefunction at $r = a$ requires that

$$j_l(ka) = 0, \tag{9.3.3}$$

and this boundary condition determines the allowed quantized energy levels $E_{n,l}$. Let us see how this works for $l = 0, 1, 2$.

For $l = 0$, eq. (9.3.3) becomes simply $\sin(ka)/ka = 0$, so $ka = n\pi$, where $n$ is a positive integer. The energies are therefore

$$E_{n,0} = \frac{\hbar^2 k^2}{2\mu} = \frac{\hbar^2 \pi^2 n^2}{2\mu a^2} \tag{9.3.4}$$

for $l = m = 0$, with corresponding wavefunctions

$$\psi_{n,0,0} = C_{n,0,0} \frac{\sin(n\pi r/a)}{n\pi r/a}. \tag{9.3.5}$$

Note that these wavefunctions approach a non-zero constant at $r = 0$. The number of zeros of the radial wavefunction, including the one at $r = a$, is $n$. The constant can be fixed by requiring the unit normalization condition

$$1 = 4\pi \int_0^a dr\, r^2 |\psi_{n,0,0}|^2, \tag{9.3.6}$$

which yields $C_{n,0,0} = n\sqrt{\pi/2a^3}$.

For $l = 1$, the boundary condition (9.3.3) reads

$$\tan(ka) = ka, \tag{9.3.7}$$

which is a transcendental equation that can be understood graphically and then solved numerically for $X = ka$ as shown in Figure 9.3.1. The lowest three energy solutions have

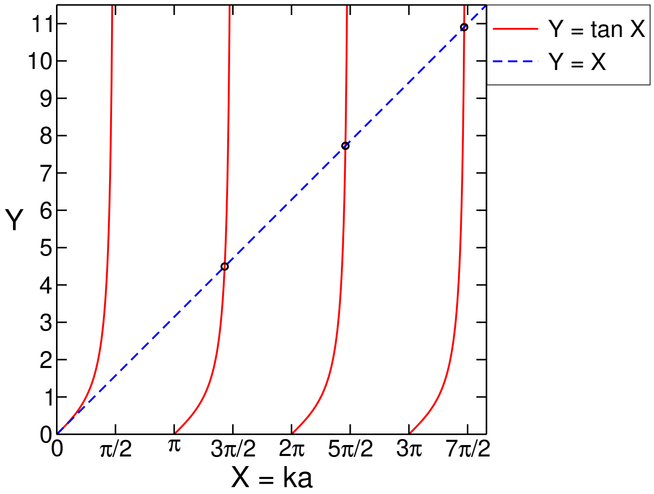

Figure 9.3.1: The solutions of the transcendental equation $X = \tan X$ for positive $X = ka$ provide the eigenvalues for $k$ for the $l = 1$ states of a particle confined to a ball of radius $a$. This graph shows the lowest three solutions $X \approx \pi(1.4303,\ 2.4590,\ 3.4709) = (4.4934,\ 7.7253,\ 10.9041)$, obtained as the intersections of $Y = \tan X$ with the line $Y = X$.

$ka/\pi \approx (1.4303,\ 2.4590,\ 3.4709,\ \ldots)$, so

$$E_{n,1} \approx \frac{\hbar^2\pi^2}{2\mu a^2}(2.0458,\ 6.0468,\ 12.0471,\ \ldots) \qquad \text{for } n = (2, 3, 4, \ldots), \tag{9.3.8}$$

where the label $n$ is again the number of zeros of the radial wavefunction. In the $l = 1$ case, $r = 0$ and $r = a$ are always both zeros, so $n \geq 2$. Each of the $l = 1$ energy levels has degeneracy 3, because the quantum number $m = -1, 0, 1$ does not affect the energy.

For $l = 2$, the boundary condition (9.3.3) becomes

$$\tan(ka) = \frac{ka}{1 - (ka)^2/3}, \tag{9.3.9}$$

which again is transcendental, but whose solutions can again be found numerically, with the results $ka/\pi \approx (1.8346,\ 2.8950,\ 3.9225,\ \ldots)$, so

$$E_{n,2} \approx \frac{\hbar^2\pi^2}{2\mu a^2}(3.3656,\ 8.3812,\ 15.3861,\ \ldots) \qquad \text{for } n = (2, 3, 4, \ldots), \tag{9.3.10}$$

where again the label $n$ is the number of zeros of the radial wavefunction $j_2(kr)$. Each of these energy levels has degeneracy 5, corresponding to $m = -2, -1, 0, 1, 2$.

The energy eigenvalues for higher angular momentum quantum number $l$ can be solved for numerically in a similar way, and have increasingly higher energies. The degeneracy of each energy level $E_{n,l}$ is $2l + 1$, corresponding to the allowed values of the $L_z$ eigenvalue $\hbar m$. The lowest few energy levels, with $E_{n,l} < 10\hbar^2\pi^2/\mu a^2$ and $l \leq 7$, are depicted in Fig. 9.3.2.

## 9.4 Particle in a spherical potential well

Consider a particle of mass $\mu$ in a spherical potential well of radius $a$, with finite depth $V_0$, so that

$$V(r) = \begin{cases} -V_0 & (r < a), \\ 0 & (r \geq a). \end{cases} \tag{9.4.1}$$

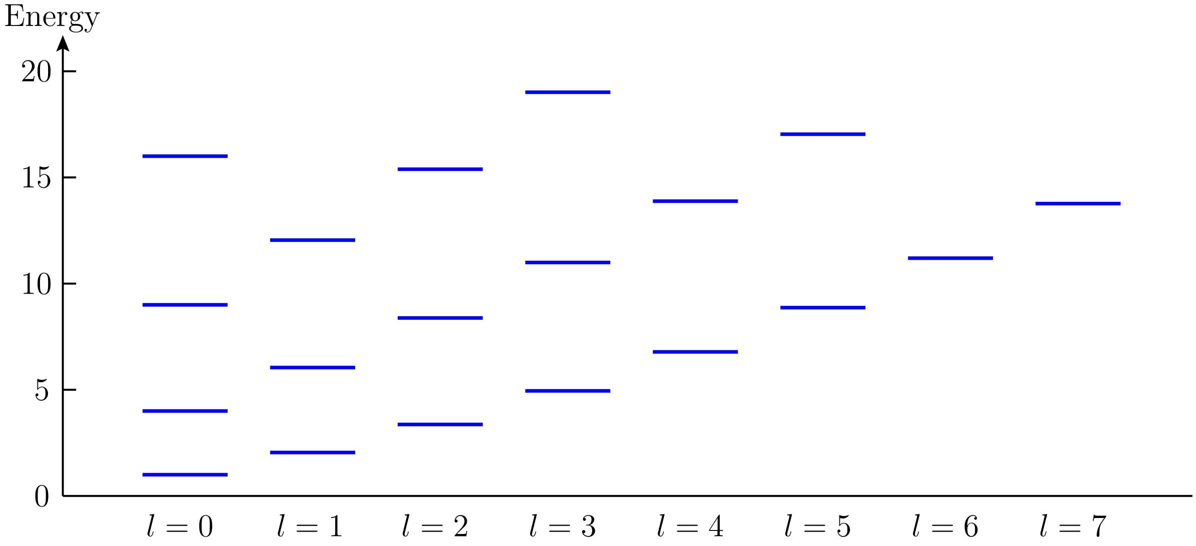

Figure 9.3.2: The lowest energy levels, in units of the ground-state energy $E_{1,0} = \hbar^2\pi^2/2\mu a^2$, for a particle of mass $\mu$ confined to a sphere of radius $a$, labeled by the angular momentum quantum number $l$. All energy eigenvalues with $E_{n,l} < 10\hbar^2\pi^2/\mu a^2$ and $l \leq 7$ are shown.

We will consider bound state solutions, which have $E < 0$. As discussed in section 9.2, the stationary-state solution for the region $r < a$ involves the ordinary spherical Bessel function $j_l(kr)$, but now with

$$k = \sqrt{2\mu(E + V_0)}/\hbar. \tag{9.4.2}$$

The Neumann functions $n_l(\rho)$ are not well-behaved at $\rho = 0$ and so do not appear for $r < a$. The solution for $r > a$ uses the modified spherical Bessel function $k_l(\kappa r)$ [see eq. (9.2.30)] with

$$\kappa = \sqrt{-2\mu E}/\hbar, \tag{9.4.3}$$

because $k_l(\rho)$ is the linear combination that is well-behaved at $\rho = \infty$. The stationary-state wavefunctions are therefore

$$\psi_{E,l,m}(\vec{r}) \;=\; \begin{cases} A\, j_l(kr)\, Y_l^m(\theta,\phi) & (r \leq a), \\ B\, k_l(\kappa r)\, Y_l^m(\theta,\phi) & (r \geq a), \end{cases} \tag{9.4.4}$$

where $A$ and $B$ are normalization constants, and

$$k^2 + \kappa^2 = 2\mu V_0/\hbar^2 \tag{9.4.5}$$

from eqs. (9.4.2) and (9.4.3).

At $r = a$, the wavefunction and its first derivative with respect to $r$ are continuous, since the potential is finite there. This gives

$$A\, j_l(ka) \;=\; B\, k_l(\kappa a), \tag{9.4.6}$$

$$kA\, j_l'(ka) \;=\; \kappa B\, k_l'(\kappa a). \tag{9.4.7}$$

Taking the ratio of these to eliminate $A$ and $B$ yields

$$\frac{k\, j_l'(ka)}{j_l(ka)} = \frac{\kappa\, k_l'(\kappa a)}{k_l(\kappa a)}, \tag{9.4.8}$$

which is a transcendental equation that can be used together with eq. (9.4.5) to solve for the allowed eigenvalues $k, \kappa$, and thus $E$. As always for a spherically symmetric potential, the allowed energies depend on $l$, but not $m$. Either equation (9.4.6) or (9.4.7) then also allows for the ratio $A/B$ to be found for each $E, l$. The remaining unknown corresponds to the overall magnitude of the constants $A$ and $B$, which can be fixed by the normalization of the wavefunction.

For example, if $l = 0$, eq. (9.4.8) yields

$$-k \cot(ka) = \kappa. \tag{9.4.9}$$

Writing dimensionless variables $X = ka$ and $Y = \kappa a$, eqs. (9.4.5) and (9.4.9) give

$$X^2 + Y^2 = 2\mu V_0 a^2/\hbar^2, \tag{9.4.10}$$

$$-X \cot X = Y. \tag{9.4.11}$$

Fortunately, these are exactly the same equations we encountered for the odd-parity solutions of the one-dimensional square well problem, with $L \to 2a$. [Compare to eqs. (6.5.12) and (6.5.18).] The same graphical and numerical analysis therefore applies. In particular, if we label the $l = 0$ stationary states by $n = 1, 2, 3, \ldots$, then the condition for the bound state $|\psi_n\rangle$ to exist is

$$\frac{\mu V_0 a^2}{\hbar^2} > \frac{\pi^2}{2}(n - 1/2)^2. \tag{9.4.12}$$

Recall that in the case of a particle in a one-dimensional square well, there is always at least one bound state, but for a sufficiently shallow potential well only the even-parity ground state exists as a bound state. For the three-dimensional spherical well, since only the analogs of the odd-parity one-dimensional square-well states exist, eq. (9.4.12) tells us that the existence of a bound state requires the potential to be sufficiently deep, $V_0 > \pi^2\hbar^2/8\mu a^2$. The more general lesson is that three-dimensional potentials may have no bound states if they are not sufficiently attractive.

## 9.5 Isotropic three-dimensional harmonic oscillator

In section 7.5, we have already encountered the isotropic three-dimensional harmonic oscillator as a special case of the general anisotropic version, and obtained its stationary-state energy levels and degeneracies. We did this using a CSCO of $(H_x, H_y, H_z)$, the individual Hamiltonians for excitations in the $x$, $y$, and $z$ directions. In this section, we will solve the problem again, this time with a CSCO consisting of $(H, L^2, L_z)$. This is possible because, in the isotropic case,

$$H \;=\; \frac{P^2}{2\mu} + \frac{1}{2}\mu\omega^2 R^2 \tag{9.5.1}$$

is invariant under all rotations and therefore commutes with $L^2$ and $L_z$. Note that we are now using $\mu$ as the symbol for the mass of the particle.

Before proceeding, let us consider what sort of answer we expect, given the results of section 7.5. At an energy level $E = \hbar\omega(n + 3/2)$ with $n = n_x + n_y + n_z$, the wavefunctions must be linear combinations of wavefunctions of the form

$$\langle x, y, z|n_x, n_y, n_z\rangle \;\propto\; H_{n_x}(x/b)H_{n_y}(y/b)H_{n_z}(z/b)\, e^{-(x^2+y^2+z^2)/2b^2}, \tag{9.5.2}$$

where the length scale in the problem is

$$b = \sqrt{\hbar/\mu\omega}, \tag{9.5.3}$$

and the $H_{n_a}$ are the Hermite polynomials. After translating to spherical coordinates, such a linear combination that is an eigenstate of $L^2$ and $L_z$ must be of the form

$$\langle r, \theta, \phi\,|n, l, m\rangle \;=\; Y_l^m(\theta, \phi)R_{n,l}(r), \tag{9.5.4}$$

where

$$R_{n,l}(r) \;=\; (\text{polynomial of degree } n \text{ in } r)\, e^{-r^2/2b^2}. \tag{9.5.5}$$

Our goal is to solve for the functions $R_{n,l}(r)$.

In section 9.1, we found the differential equation for $U_{n,l} = rR_{n,l}$, which in the present case becomes

$$\left[-\frac{\hbar^2}{2\mu}\frac{d^2}{dr^2} + \frac{1}{2}\mu\omega^2 r^2 + \frac{\hbar^2 l(l+1)}{2\mu r^2} - E\right] U_{n,l} \;=\; 0. \tag{9.5.6}$$

Inspired by eq. (9.5.5), we define dimensionless quantities $x$ and $y(x)$ by

$$x \;=\; r/b, \qquad\qquad U_{n,l} \;=\; y(x)e^{-x^2/2}. \tag{9.5.7}$$

When plugged into eq. (9.5.6), this gives

$$y'' - 2xy' + \left[\mathcal{E} - 1 - l(l+1)/x^2\right] y \;=\; 0, \tag{9.5.8}$$

where

$$\mathcal{E} \;=\; 2\mu b^2 E/\hbar^2 \;=\; 2E/\hbar\omega \tag{9.5.9}$$

is a dimensionless combination proportional to the energy eigenvalue. We already know from eq. (7.5.13) that the allowed eigenvalues are $\mathcal{E} = 2n + 3$ for non-negative integers $n$, and from eq. (9.5.5) that $y(x)$ is a polynomial, but in the following derivation we will proceed as if these facts were not known.

We now try a series solution for $y(x)$, of the form

$$y = x^q \sum_{p=0}^{\infty} c_p x^p. \tag{9.5.10}$$

Here $q$ and the $c_p$ are constants, with $c_0 \neq 0$ by definition. (Otherwise, we would adjust the value of $q$ to make it so.) Now we prepare to plug in to eq. (9.5.8) by computing

$$\begin{aligned} y'' &= \sum_{p=0}^{\infty} c_p (p+q)(p+q-1) x^{p+q-2}, && (9.5.11)\\ y/x^2 &= \sum_{p=0}^{\infty} c_p x^{p+q-2}, && (9.5.12)\\ xy' &= \sum_{p=0}^{\infty} c_p (p+q) x^{p+q} = \sum_{p=0}^{\infty} c_{p-2}(p+q-2) x^{p+q-2}, && (9.5.13)\\ y &= \sum_{p=0}^{\infty} c_{p-2} x^{p+q-2}, && (9.5.14)\end{aligned}$$

where in the last equality of each of eqs. (9.5.13) and (9.5.14) we have used the trick of relabeling $p \rightarrow p - 2$ and defining $c_{-2} = c_{-1} = 0$. The motivation behind this relabeling trick is that now all of the summands have the same powers of $x$, and so eq. (9.5.8) becomes

$$\sum_{p=0}^{\infty} x^{p+q-2} \Big\{ c_p \big[(p+q)(p+q-1) - l(l+1)\big] - c_{p-2}\big[2(p+q-2) + 1 - \mathcal{E}\big] \Big\} \;=\; 0. \tag{9.5.15}$$

For this equation to be satisfied for all $x$, each coefficient of a given power of $x$ must vanish, so the quantity in braces must vanish for each $p$.

From the first term $p = 0$, using $c_{-2} = 0$ we find

$$c_0 \big[ q(q-1) - l(l+1) \big] \;=\; 0. \tag{9.5.16}$$

Since $c_0 \neq 0$, the possible solutions are $q = l + 1$ and $q = -l$. However, the latter can be rejected on physical grounds, since it would imply that $y \sim x^{-l}$ for small $x$, which would mean $R_{n,l} \sim 1/r^{l+1}$ for small $r$, and the wavefunction would not be finite at the point $r = 0$. Therefore,

$$q \;=\; l + 1, \tag{9.5.17}$$

which implies that $R_{n,l} \sim r^l$ for small $r$. This suppression for small $r$ is consistent with the existence of the repulsive $\hbar^2 l(l+1)/2\mu r^2$ centrifugal barrier contribution to the effective potential for non-zero angular momentum, see eq. (9.1.10) or eq.(9.5.6).

The second term $p = 1$ gives us

$$c_1\big[q(q+1) - l(l+1)\big] \;=\; 0, \tag{9.5.18}$$

because $c_{-1} = 0$. Now, plugging in $q = l + 1$, this becomes $2(l + 1)c_1 = 0$. Since our knowledge of the angular momentum eigenvalue problem tells us that $l + 1$ cannot vanish, we must have

$$c_1 = 0. \tag{9.5.19}$$

For all larger $p$, the vanishing of eq. (9.5.15) implies (after using $q = l + 1$) that

$$c_p\, p(p + 2l + 1) \;=\; c_{p-2}\,(2p + 2l - 1 - \mathcal{E}). \tag{9.5.20}$$

Since $c_1 = 0$, it follows that $c_p = 0$ for all odd $p$. Defining $C_j = c_{2j}$, we have

$$y \;=\; x^{l+1}\sum_{j=0}^{\infty} C_j x^{2j}, \tag{9.5.21}$$

with, from eq. (9.5.20) by taking $p = 2(j + 1)$, the recurrence relation

$$C_{j+1} \;\;=\;\; \frac{4j + 2l + 3 - \mathcal{E}}{2(j+1)(2j+2l+3)}\, C_j. \tag{9.5.22}$$

For large $j$, we have $C_j \approx C_{j-1}/j$, so if the series does not terminate, it would behave asymptotically like $y \sim x^{l+1}\sum_j (x^2)^j/j! \sim x^{l+1}e^{x^2}$, which would imply that $R_{n,l} \sim r^l e^{-r^2/2b^2}e^{r^2/b^2} \sim r^l e^{r^2/2b^2}$. As in the case of the one-dimensional harmonic oscillator, such a solution that blows up exponentially as $r \to \infty$ is unphysical, since it cannot be normalized.

We can therefore conclude that the series must terminate, and $y(x)$ is actually a polynomial, in agreement with eq. (9.5.5). Then the recurrence relation eq. (9.5.22) implies that physically valid solutions for $y(x)$ must have $C_{k+1} = 0$ for some integer $k \geq 0$, and so satisfy

$$\mathcal{E} = 4k + 2l + 3, \tag{9.5.23}$$

or, using eq. (9.5.9),

$$E = \hbar\omega(2k + l + 3/2). \tag{9.5.24}$$

The integer $n = n_x + n_y + n_z$ must then be equal to $2k + l$, and the energy eigenkets $|E, l, m\rangle$ are linear combinations of the kets $|n_x, n_y, n_z\rangle$. An interesting feature of these results is that, unlike the particle-in-a-sphere and particle-in-a-spherical-well examples of sections 9.3 and 9.4, here the degeneracies in the energy levels are not entirely due to the rotational invariance of the problem. For a given $l$, there are $2l+1$ degenerate states with $m = -l, \ldots, l$ that have the same energy. But, there are also "accidental" energy degeneracies between states with different angular momenta, for example $k = 0, l = 2$ and $k = 1, l = 0$.

Summarizing what we now know about the three-dimensional isotropic harmonic oscillator stationary-state wavefunctions with definite $(L^2, L_z)$,

$$\psi_{k,l,m}(r,\theta,\phi) \;=\; A R_{k,l}(r)\, Y_l^m(\theta,\phi), \tag{9.5.25}$$

where $A$ is a normalization constant. The radial wavefunctions are of the form

$$R_{k,l}(r) \;=\; (r/b)^l P_{k,l}(r^2/b^2) e^{-r^2/2b^2}, \tag{9.5.26}$$

where, with $z = x^2$,

$$P_{k,l}(z) \;=\; \sum_{j=0}^{k} C_j z^j. \tag{9.5.27}$$

Using eq. (9.5.23), the recurrence relation (9.5.22) becomes

$$C_{j+1} = \frac{2(j-k)}{(j+1)(2j+2l-3)} C_j. \tag{9.5.28}$$

For the lowest few values $k = 0, 1, 2, 3$, we have (leaving the coefficient $C_0$ as an arbitrary normalization for the moment):

$$\begin{aligned}
P_{0,l}(z) &= C_0, && (9.5.29)\\
P_{1,l}(z) &= C_0\Big(1 - \frac{2}{2l+3} z\Big), && (9.5.30)\\
P_{2,l}(z) &= C_0\Big(1 - \frac{4}{2l+3} z + \frac{4}{(2l+3)(2l+5)} z^2\Big), && (9.5.31)\\
P_{3,l}(z) &= C_0\Big(1 - \frac{6}{2l+3} z + \frac{12}{(2l+3)(2l+5)} z^2 - \frac{8}{(2l+3)(2l+5)(2l+7)} z^3\Big). && (9.5.32)
\end{aligned}$$

These polynomials are proportional to **associated Laguerre polynomials**, for which there are unfortunately at least three different notational conventions in common use. Adopting here the definition of the associated Laguerre polynomial $L_k^\alpha(z)$ used by Mathematica,[†]

$$P_{k,l}(z) \;=\; L_k^{l+1/2}(z), \tag{9.5.33}$$

---

[†] The relations between the definition of the associated Laguerre polynomials used here and in some other sources is $\Gamma(N+\alpha+1) L_N^\alpha(z)_{\text{here, Mathematica}} = L_N^\alpha(z)_{\text{some other books}} = (-1)^\alpha L_{N+\alpha}^\alpha(z)_{\text{still other books}}$.

corresponding to the choice

$$C_0 \;=\; \frac{(2l+2k+1)!!}{2^k\, k!\, (2l+1)!!} \;=\; \frac{(2l+2k+1)!\, l!}{2^{2k}\, (2l+1)!\, (l+k)!\, k!}. \tag{9.5.34}$$

More generally, the conventional normalization is such that, at $z = 0$ and for any $\alpha$,

$$L_k^\alpha(0) = \frac{\Gamma(k+\alpha+1)}{k!\, \Gamma(\alpha+1)}. \tag{9.5.35}$$

Here, the **Gamma function** is defined by

$$\Gamma(z) \;=\; \int_0^\infty dt\, t^{z-1} e^{-t}, \tag{9.5.36}$$

and satisfies $\Gamma(z) = (z-1)\Gamma(z-1)$. When $z$ is an integer, then $z! = \Gamma(z+1)$. Another useful value is $\Gamma(1/2) \;=\; \sqrt{\pi}$.

An explicit, general form for the associated Laguerre polynomials is

$$L_N^\alpha(z) \;=\; \frac{1}{N!} z^{-\alpha} e^z \frac{d^N}{dz^N} \left(e^{-z} z^{N+\alpha}\right). \tag{9.5.37}$$

Note that the lower index $N$ is always an integer equal to the degree of the polynomial, but the raised index $\alpha$ need not be, as in the present application where $\alpha = l + 1/2$. They satisfy the differential equation

$$\left[z \frac{d^2}{dz^2} + (1 - z + \alpha) \frac{d}{dz} + N\right] L_N^\alpha(z) \;=\; 0, \tag{9.5.38}$$

and the orthogonality relation

$$\int_0^\infty dz\, e^{-z} z^\alpha L_N^\alpha(z) L_{N'}^\alpha(z) \;=\; \frac{\Gamma(N+\alpha+1)}{k!} \delta_{N,N'}. \tag{9.5.39}$$

The degree $N$ is also equal to the number of zeros (in the present case $N = k$), which all occur for positive real $z$.

Putting everything together, the wavefunctions for the stationary states of the three-dimensional isotropic harmonic oscillator with energies $E = \hbar\omega(2k + l + 3/2)$ are

$$\psi_{k,l,m}(r, \theta, \phi) \;=\; \sqrt{\frac{2(k!)}{b^3\, \Gamma(k + l + 3/2)}} \left(\frac{r}{b}\right)^l L_k^{l+1/2}(r^2/b^2)\, e^{-r^2/2b^2}\, Y_l^m(\theta, \phi). \tag{9.5.40}$$

The normalization factor $A$ in eq. (9.5.25) has been chosen, with the aid of eq. (9.5.39), so that orthonormality holds,

$$\langle k', l', m' | k, l, m\rangle \;=\; \int_0^\infty dr\, r^2 \int d\Omega\, \left[\psi_{k',l',m'}(r, \theta, \phi)\right]^* \psi_{k,l,m}(r, \theta, \phi) \;=\; \delta_{k,k'} \delta_{l,l'} \delta_{m,m'}. \tag{9.5.41}$$

# 10 Coulomb potential and hydrogen-like atoms

## 10.1 Bound states for hydrogen atom

One of the most important exactly solvable problems in quantum mechanics is that of a particle moving in a Coulomb potential. This includes the hydrogen atom, and more generally hydrogen-like ions, which consist of a single electron with mass $m_e$ and charge $-e$ [with the normalization as defined in Gaussian cgs metric system units, see eqs. (1.1.1) and (1.1.2)] and a heavy nucleus with mass $m_N$ and charge $Ze$, where $Z$ is an integer. As discussed in a general context in section 4.2, the problem can be separated into center-of-mass and relative degrees of freedom, where the latter use a reduced mass $\mu = m_e m_N/(m_e + m_N)$, which is very close to $m_e$. In this chapter, we neglect the small effects of special relativity and electron and nuclear spins, which will be treated in Chapter 15. So, our Hamiltonian is $H = P^2/2\mu + V(R)$, where the potential energy in the position representation is

$$V(r) \;=\; -Ze^2/r. \tag{10.1.1}$$

The special case $Z = 1$ is the hydrogen atom. Because the potential is spherically symmetric, we choose the CSCO to be $(H, L^2, L_z)$, and look for an orthobasis of eigenstates $|E, l, m\rangle$ with wavefunctions

$$\psi_{E,l,m}(r, \theta, \phi) \;=\; \langle r, \theta, \phi\,|E, l, m\rangle \;=\; R_{E,l}(r)\, Y_l^m(\theta, \phi). \tag{10.1.2}$$

The goal is to find the energy eigenvalues and the radial wavefunctions $R_{E,l}(r)$.

In this section, we will work out the properties of the bound state solutions, for which the energy eigenvalues are discrete and satisfy $E < V(\infty) = 0$, and $|R_{E,l}|$ decreases exponentially as $r \to \infty$. Unlike the case of the three-dimensional harmonic oscillator, there are also unbound energy eigenstates, for which $E$ is continuous and non-negative, and $rR_{E,l}$ oscillates with an amplitude approaching a constant for large $r$. The unbound state solutions will be found in section 10.2.

The time-independent Schrödinger equation for the radial wavefunction is

$$\left[-\frac{\hbar^2}{2\mu}\left(\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2}\right) - \frac{Ze^2}{r} - E\right] R_{E,l}(r) \;=\; 0. \tag{10.1.3}$$

A good first step is to replace $r$ by a dimensionless variable. Since $-\hbar^2/2\mu E$ has units of [length]$^2$ and is a positive number for the bound states, we define a rescaled radial coordinate

$$s \;=\; 2r/b, \qquad b \;=\; \hbar/\sqrt{-2\mu E}. \tag{10.1.4}$$

With this change of variables, eq. (10.1.3) becomes

$$\left(\frac{d^2}{ds^2} + \frac{2}{s}\frac{d}{ds} - \frac{l(l+1)}{s^2} + \frac{n}{s} - \frac{1}{4}\right) R_{E,l} \;=\; 0, \tag{10.1.5}$$

where we have introduced

$$n \;=\; \frac{Ze^2}{\hbar}\sqrt{\frac{-\mu}{2E}}, \tag{10.1.6}$$

a dimensionless quantity that parameterizes the energy eigenvalue. Note that $n$ is a real number for $E < 0$, and is pure imaginary for $E > 0$. We will soon find out [see eq. (10.1.15)] that $n$ must be a positive integer for a bound-state solution, but we do not know that yet.

The next part of our strategy is to identify, and factor out, the large-distance and short-distance behaviors of $R_{E,l}$. For $s \to \infty$, eq. (10.1.5) becomes $d^2R/ds^2 \approx R/4$, which has two linearly independent solutions, $e^{s/2}$ and $e^{-s/2}$. The first of these is unphysical, as it blows up for $s \to \infty$ and so is not normalizable. In the opposite limit $s \to 0$, the last two terms of eq. (10.1.5) can be neglected, and there is a power-law solution for $R$ proportional to $s^l$. We therefore write

$$R_{E,l} \;=\; s^l e^{-s/2} f(s), \tag{10.1.7}$$

which factors out the leading behavior in the two limits. Plugging this into eq. (10.1.5) gives

$$s\frac{d^2 f}{ds^2} + (2l + 2 - s)\frac{df}{ds} + (n - l - 1)f \;=\; 0, \tag{10.1.8}$$

which we must now solve simultaneously for the eigenfunctions $f(s)$ and the corresponding eigenvalues $n$.

Equation (10.1.8) is a special case of a famous differential equation, called the **confluent hypergeometric equation**, which we write in a general form as

$$x\frac{d^2F}{dx^2} + (c - x)\frac{dF}{dx} - aF \;=\; 0, \tag{10.1.9}$$

where $a$ and $c$ are constants. This has a unique (up to a multiplicative constant) solution that is finite as $x \to 0$, the confluent hypergeometric function, which has a series expansion

$$F(a, c, x) \;=\; 1 + \frac{a}{c}x + \frac{a(a{+}1)}{c(c{+}1)}\frac{x^2}{2!} + \frac{a(a{+}1)(a{+}2)}{c(c{+}1)(c{+}2)}\frac{x^3}{3!} + \cdots\,, \tag{10.1.10}$$

as can be verified by direct substitution into the differential equation. This series converges for all finite $|x|$, even if $a$, $c$, and $x$ are complex, provided that $c \neq 0, -1, -2, \ldots$. One can check (either from the series solution, or by direct substitution into the differential equation) that it has an integral representation

$$F(a, c, x) \;=\; \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 dt\, e^{tx}\, t^{a-1}(1-t)^{c-a-1}, \tag{10.1.11}$$

provided that $\text{Re}[a] > \text{Re}[c] > 0$. For real $x$ large and positive, it has the asymptotic form

$$F(a,c,x) \;\approx\; \frac{\Gamma(c)}{\Gamma(a)}\, x^{a-c}\, e^{x} \qquad \text{(large real } x\text{, and } a \neq 0, -1, -2, \ldots). \tag{10.1.12}$$

However, as we are about to see, the special case relevant for our present purposes will arise when $a$ is a non-positive integer; in that case, $F(a, c, x)$ is instead a polynomial in $x$ of degree $-a$, because the series in eq. (10.1.10) terminates.

For the present application to the bound states of the hydrogen-like atom, we have $x = s$ and $a = l + 1 - n$ and $c = 2l + 2$, by comparing eq. (10.1.8) to (10.1.9). Therefore, $f(s)$ in eq. (10.1.7) is equal to $F(l{+}1{-}n,\, 2l{+}2,\, s)$, up to normalization, so

$$R_{n,l} \;=\; C_{n,l}\, s^{l} e^{-s/2} F(l{+}1{-}n,\, 2l{+}2,\, s), \tag{10.1.13}$$

Here we have replaced the subscript label $E$ by the label $n$, which contains the same information, and $C_{n,l}$ is a normalization constant to be chosen later. In the large-distance limit $s \to \infty$, the asymptotic form of eq. (10.1.12) would seem to tell us

$$R_{n,l} \;\propto\; s^{-n-1} e^{s/2} \qquad (n{-}l{-}1 \;\neq\; 0, 1, 2, \ldots). \tag{10.1.14}$$

However, the exponential factor $e^{s/2}$ is the form that we already rejected as non-normalizable and therefore physically unacceptable; it would imply that no matter how far from the nucleus you look, the electron must have infinitely larger probability to be farther away. The only way to get a physically sensible bound state is to arrange for eq. (10.1.14) not to apply. In other words, it is necessary that the series solution eq. (10.1.10) for $F(l{+}1{-}n,\, 2l{+}2,\, s)$ terminates, so that instead of being proportional to $e^{s}$, it is actually a polynomial in $s$ with degree that we will call $k$. This implies that

$$n \;=\; k + l + 1. \tag{10.1.15}$$

Since $k$, being the degree of the polynomial, is a non-negative integer, and $l + 1$ is always a positive integer, $n$ must be a positive integer, called the **principal quantum number** of the hydrogen atom bound state.

Inverting eq. (10.1.6), the allowed energy eigenvalues are

$$E_n \;=\; -\left(\frac{e^4 \mu}{2\hbar^2}\right)\frac{Z^2}{n^2} \;=\; -\left(\frac{e^2}{2a_0}\right)\frac{Z^2}{n^2}, \tag{10.1.16}$$

which depend only on $n$, not $k$ and $l$ individually. To write the last expression we have defined[†]

---

[†] The definitions of the Bohr radius and Rydberg energy units used here are appropriate for the infinite nuclear mass limit, with $\mu = m_e$. An alternative definition uses, instead of the electron mass $m_e$, the reduced mass $\mu$ for the lightest isotope of hydrogen, which is smaller by a factor $m_p/(m_e + m_p) \approx 0.999453$. From here on, we ignore the small difference between $m_e$ and $\mu$, which can be restored by replacing $m_e \to \mu$ in the obvious way.

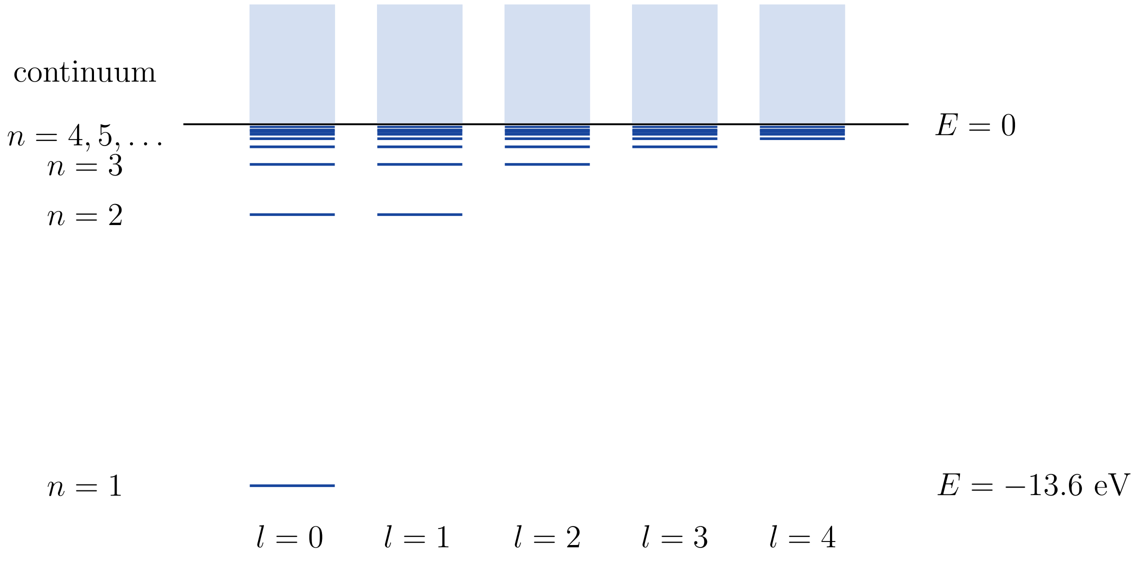

Figure 10.1.1: Stationary-state energy levels of the hydrogen atom for $l \leq 4$. The bound state levels with $E_n = -13.6\text{ eV}/n^2$ for $n = 1, 2, 3, \ldots$ are discrete and have degeneracy $n^2$, with $0 \leq l \leq n - 1$. For each non-negative integer $l$, there are also unbound continuum energy eigenstates with $E \geq 0$.

the **Bohr radius** (named after Niels Bohr),

$$a_0 \;=\; \frac{\hbar^2}{e^2 m_e} \;=\; 5.292 \times 10^{-11}\,\text{meters} \;=\; \text{Bohr radius}. \tag{10.1.17}$$

The scale of energy is therefore

$$\frac{e^4 m_e}{2\hbar^2} \;=\; \frac{e^2}{2a_0} \;=\; 2.180 \times 10^{-18}\,\text{Joules} \;=\; 13.606\,\text{eV} \;=\; \text{Rydberg}, \tag{10.1.18}$$

named after Johannes Rydberg. The ground state has $n = 1$ and $k = l = 0$. More generally, for each $l$, the lowest possible energy is obtained for $k = 0$, so that $n = l + 1$.

The energy levels of the hydrogen atom are depicted in Fig. 10.1.1. At each bound-state energy level $n$, the values of $l$ that can occur are the integers from 0 to $n - 1$, each with degeneracy $2l + 1$, which comes from $m = -l, \ldots, l$. The total degeneracy for each energy level $E_n$ is therefore

$$g_n \;=\; \sum_{l=0}^{n-1} (2l + 1) \;=\; n^2. \tag{10.1.19}$$

Because this includes states with different angular momenta, the energy degeneracy is accidental, meaning that it cannot be explained by rotational invariance alone. This was also the case for the three-dimensional isotropic harmonic oscillator. For hydrogen-like atoms, this degeneracy is slightly broken by spin and relativistic effects, as we will discuss in detail in Chapter 15.

The distinct values of $l$ are given letter codes that go back to the early history of spectroscopic analysis. This **spectroscopic notation** is as follows:

$$\begin{aligned} l = 0 \quad &\leftrightarrow \quad s \text{ "sharp"}, \\ l = 1 \quad &\leftrightarrow \quad p \text{ "principal"}, \\ l = 2 \quad &\leftrightarrow \quad d \text{ "diffuse"}, \\ l = 3 \quad &\leftrightarrow \quad f \text{ "fundamental"}, \\ l = 4 \quad &\leftrightarrow \quad g \text{ "grotesque"}, \\ l = 5 \quad &\leftrightarrow \quad h \text{ "horrendous"}, \\ \ldots \quad & \quad \ldots \end{aligned} \tag{10.1.20}$$

The letter codes are in alphabetical order $f, g, h, i, k, \ldots$ for $l = 3, 4, 5, 6, 7, \ldots$, but $j$ is omitted. (And I'm just kidding about the names "grotesque" and "horrendous"; unlike the first four, I just made those up, and they do not actually have standard names to fit the standard letters.) A supposedly useful mnemonic is "**s**ober **p**hysicists **d**on't **f**ind **g**iraffes **h**iding **i**n **k**itchens". The hydrogen atom states are often referred to in the notation $nl$, but with $l = 0, 1, 2, 3, 4, 5, \ldots$ replaced by the spectroscopic code[‡] letter $s, p, d, f, g, h \ldots$, so that the ground state is called $1s$ and the first excited states are $2s$ and $2p$, and the second excited states are $3s$, $3p$, and $3d$.

Returning to the radial wavefunction, eq. (10.1.13) tells us that

$$R_{n,l} \;=\; C_{n,l}\, s^l e^{-s/2} F_{n,l}, \tag{10.1.21}$$

where we are now adopting the shorter notation

$$F_{n,l} \;=\; F(l{+}1{-}n,\, 2l{+}2, s), \tag{10.1.22}$$

which is a polynomial of degree $k = n - l - 1$. Recalling that $s = 2r/b$ from eq. (10.1.4), the exponential factor tells us that the spatial support of wavefunctions is set, in terms of the Bohr radius, by the length scale

$$b \;=\; na_0/Z. \tag{10.1.23}$$

Equations (10.1.16) and (10.1.23) show that the magnitude of the binding energy scales like $Z^2$, while the characteristic size of a given wavefunction's support scales like $1/Z$. In the classical limit $\hbar \to 0$, the Bohr radius goes to 0 and the binding energy of the ground state goes to $-\infty$, in accord with the discussion of the classical instability of atoms in section 1.1.

---

[‡] In some other contexts, the letter codes for orbital angular momentum in eq. (10.1.20) are capitalized.

For the lowest few energy states, the polynomials are, from eq. (10.1.10),

$$F_{1,0} = 1, \tag{10.1.24}$$

$$F_{2,1} = 1, \qquad F_{2,0} = 1 - \frac{s}{2}, \tag{10.1.25}$$

$$F_{3,2} = 1, \qquad F_{3,1} = 1 - \frac{s}{4}, \qquad F_{3,0} = 1 - s + \frac{s^2}{6}, \tag{10.1.26}$$

$$F_{4,3} = 1, \qquad F_{4,2} = 1 - \frac{s}{6}, \qquad F_{4,1} = 1 - \frac{s}{2} + \frac{s^2}{20}, \qquad F_{4,0} = 1 - \frac{3s}{2} + \frac{s^2}{2} - \frac{s^3}{24}. \tag{10.1.27}$$

More generally, in terms of the associated Laguerre polynomials $L_N^\alpha(x)$ already defined explicitly by eq. (9.5.37), it can be shown that

$$F_{n,l} \;=\; \frac{(n-l-1)!\,(2l+1)!}{(n+l)!}\, L^{2l+1}_{n-l-1}(s). \tag{10.1.28}$$

The lower index on $L_N^\alpha$ is always equal to the degree of the associated Laguerre polynomial, which is also the number of its zeros, in this case $N = k - 1$. Note, however, that in the case of the isotropic three-dimensional harmonic oscillator the upper index $\alpha$ was always half-integer [see eq. (9.5.33)], while here $\alpha = 2l + 1$ is always a positive odd integer.

From Theorem 2.6.5, we know that the kets $|n, l, m\rangle$ and $|n', l', m'\rangle$ are orthogonal whenever any of $n$, $l$, or $m$ differ from $n'$, $l'$, or $m'$, respectively. The orthogonality conditions $\delta_{l,l'}$ and $\delta_{m,m'}$ are already enforced by the proportionality of the wavefunctions to spherical harmonics. It follows that the associated Laguerre polynomials must also satisfy an orthogonality relation of the form (taking $x = ns = 2Zr/a_0$):

$$\int_0^\infty dx\, x^{2l+2}\, L^{2l+1}_{n-l-1}(x/n)\, L^{2l+1}_{n'-l-1}(x/n')\, e^{-x(1/n+1/n')/2} \;=\; \delta_{n,n'}\, B_{n,l}, \tag{10.1.29}$$

for $n, n' = 1, 2, 3, \ldots$ and $l = 0, 1, \ldots, \min(n, n') - 1$. Although not obvious, it can be checked that this is true, with

$$B_{n,l} \;=\; \frac{2\,(n+l)!\, n^{2l+4}}{(n-l-1)!}\,. \tag{10.1.30}$$

Note that eq. (10.1.29) is a quite different orthogonality relation for associated Laguerre polynomials than the one that was useful for the three-dimensional harmonic oscillator, eq. (9.5.39).

Putting together eqs. (10.1.4), (10.1.21), (10.1.23), and (10.1.28), we obtain the radial wavefunctions of the hydrogen-like atom bound states,

$$R_{n,l}(r) \;=\; A_{n,l}\, e^{-Zr/na_0} \left(\frac{2Zr}{na_0}\right)^l L^{2l+1}_{n-l-1}(2Zr/na_0), \tag{10.1.31}$$

where we have introduced a new normalization constant factor

$$A_{n,l} \;=\; \left(\frac{2Z}{a_0}\right)^{3/2} \frac{n^l}{\sqrt{B_{n,l}}} \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{2}{n^2}\sqrt{\frac{(n-l-1)!}{(n+l)!}}, \tag{10.1.32}$$

chosen so that

$$\int_0^\infty dr\, r^2\, R_{n,l}(r) R_{n',l}(r) \;=\; \delta_{n,n'}. \tag{10.1.33}$$

The full wavefunction [obtained by including the $Y_l^m(\theta,\phi)$ factor according to eq. (10.1.2)] is

$$\psi_{n,l,m}(r,\theta,\phi) \;=\; R_{n,l}(r) Y_l^m(\theta,\phi). \tag{10.1.34}$$

An arbitrary choice of phase has been made in $A_{n,l}$, as usual. With this sensible choice, the radial wavefunctions $R_{n,l}$ are all real.

We now have everything necessary to evaluate the wavefunctions for the lowest few energy levels. For the ground state,

$$R_{1,0}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} 2e^{-Zr/a_0}, \tag{10.1.35}$$

and for the first excited states,

$$R_{2,0}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{2\sqrt{2}} \left(2 - \frac{Zr}{a_0}\right) e^{-Zr/2a_0}, \tag{10.1.36}$$

$$R_{2,1}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{2\sqrt{6}} \frac{Zr}{a_0} e^{-Zr/2a_0}, \tag{10.1.37}$$

and for the second excited states,

$$R_{3,0}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{2}{81\sqrt{3}} \left[27 - 18\frac{Zr}{a_0} + 2\left(\frac{Zr}{a_0}\right)^2\right] e^{-Zr/3a_0}, \tag{10.1.38}$$

$$R_{3,1}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{2\sqrt{2}}{81\sqrt{3}} \left[6\frac{Zr}{a_0} - \left(\frac{Zr}{a_0}\right)^2\right] e^{-Zr/3a_0}, \tag{10.1.39}$$

$$R_{3,2}(r) \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{2\sqrt{2}}{81\sqrt{15}} \left(\frac{Zr}{a_0}\right)^2 e^{-Zr/3a_0}. \tag{10.1.40}$$

These radial wavefunctions, and the corresponding radial probability densities $r^2|R_{n,l}|^2$, are graphed in Figure 10.1.2 for $Z = 1$. For future reference, we also note that the hydrogen atom radial wavefunction at $r = 0$ can be evaluated, using eqs. (10.1.28), 10.1.31), and (10.1.32), as

$$R_{n,l}(0) \;=\; 2\left(\frac{Z}{na_0}\right)^{3/2} \delta_{l,0}. \tag{10.1.41}$$

In particular, it is only non-zero for states with orbital angular momentum $l = 0$.

Putting in the spherical harmonics with the Condon–Shortley phase convention, the wavefunctions for the lowest few energy levels are

$$\psi_{1,0,0} \;=\; \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{\sqrt{\pi}} e^{-Zr/a_0}, \tag{10.1.42}$$

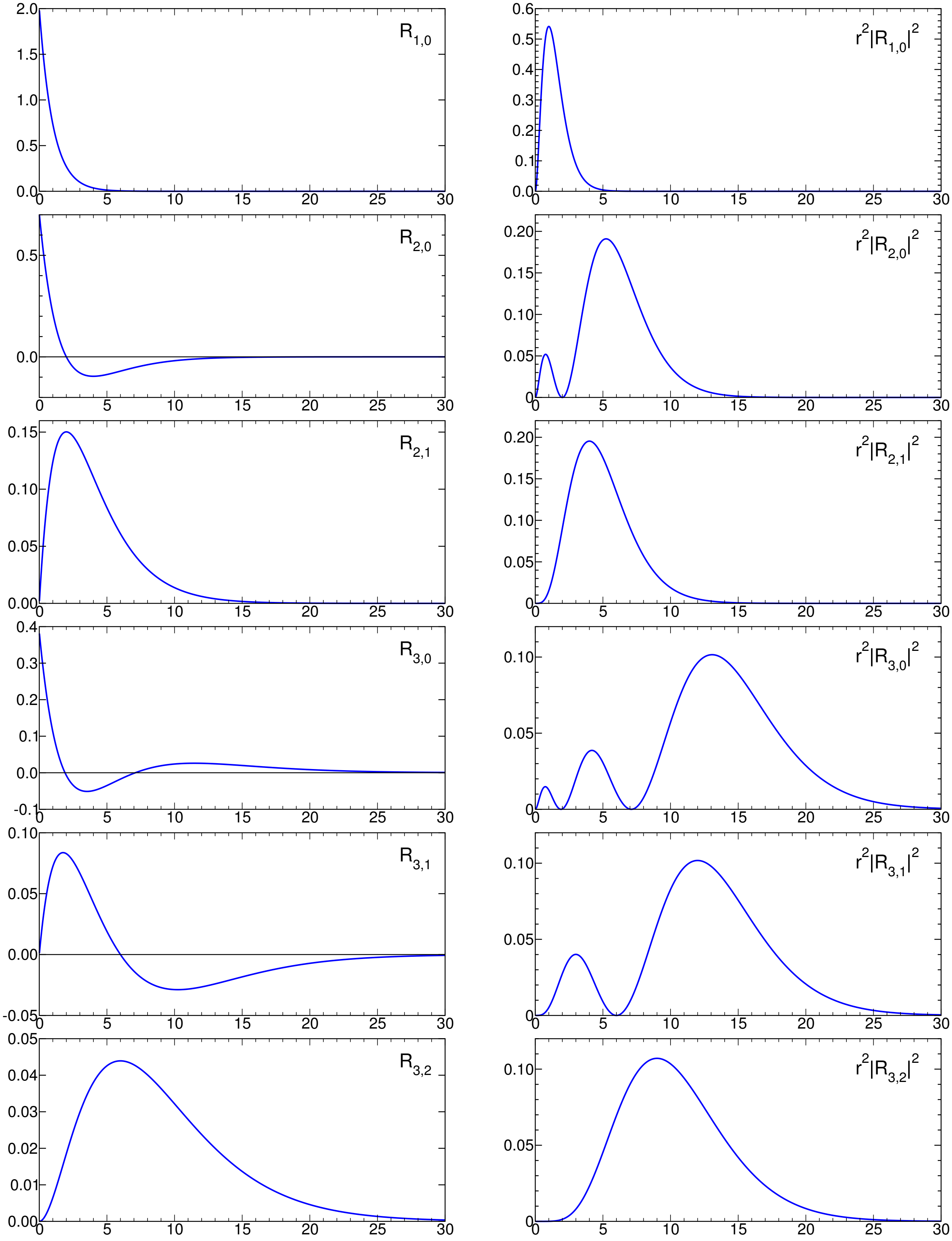

Figure 10.1.2: Radial wavefunctions $R_{n,l}$ in units of $1/a_0^{3/2}$ (left column) and radial probability densities $r^2|R_{n,l}|^2$ in units of $1/a_0$ (right column) for the $Z = 1$ hydrogen atom stationary states with $n = 1, 2, 3$, as functions of $r/a_0$. Note the differing vertical scales.

for the ground state, and

$$\psi_{2,0,0} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{4\sqrt{2\pi}} \left(2 - \frac{Zr}{a_0}\right) e^{-Zr/2a_0}, \tag{10.1.43}$$

$$\psi_{2,1,0} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{4\sqrt{2\pi}} \frac{Zr}{a_0} e^{-Zr/2a_0} \cos\theta, \tag{10.1.44}$$

$$\psi_{2,1,\pm 1} = \mp \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{8\sqrt{\pi}} \frac{Zr}{a_0} e^{-Zr/2a_0} \sin\theta \, e^{\pm i\phi}, \tag{10.1.45}$$

for the first excited states, and

$$\psi_{3,0,0} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{81\sqrt{3\pi}} \left[27 - 18\frac{Zr}{a_0} + 2\left(\frac{Zr}{a_0}\right)^2\right] e^{-Zr/3a_0}, \tag{10.1.46}$$

$$\psi_{3,1,0} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{\sqrt{2}}{81\sqrt{\pi}} \left[6\frac{Zr}{a_0} - \left(\frac{Zr}{a_0}\right)^2\right] e^{-Zr/3a_0} \cos\theta, \tag{10.1.47}$$

$$\psi_{3,1,\pm 1} = \mp \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{81\sqrt{\pi}} \left[6\frac{Zr}{a_0} - \left(\frac{Zr}{a_0}\right)^2\right] e^{-Zr/3a_0} \sin\theta \, e^{\pm i\phi}, \tag{10.1.48}$$

$$\psi_{3,2,0} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{81\sqrt{6\pi}} \left(\frac{Zr}{a_0}\right)^2 e^{-Zr/3a_0} (3\cos^2\theta - 1), \tag{10.1.49}$$

$$\psi_{3,2,\pm 1} = \mp \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{81\sqrt{\pi}} \left(\frac{Zr}{a_0}\right)^2 e^{-Zr/3a_0} \sin\theta\cos\theta \, e^{\pm i\phi}, \tag{10.1.50}$$

$$\psi_{3,2,\pm 2} = \left(\frac{Z}{a_0}\right)^{3/2} \frac{1}{162\sqrt{\pi}} \left(\frac{Zr}{a_0}\right)^2 e^{-Zr/3a_0} \sin^2\theta \, e^{\pm 2i\phi}, \tag{10.1.51}$$

for the second excited states.

Let us now work out the expectation value of $R^p$ in the ground state, for integer $p$,

$$\langle 1,0,0|R^p|1,0,0\rangle = \int_0^\infty dr\, r^2 \int d\Omega\, r^p \left(\frac{Z}{a_0}\right)^3 \frac{1}{\pi} e^{-2Zr/a_0}. \tag{10.1.52}$$

Using $\int d\Omega = 4\pi$, this evaluates to

$$\langle 1,0,0|R^p|1,0,0\rangle = \begin{cases} \frac{1}{2}(p+2)!\, \left(\frac{a_0}{2Z}\right)^p & (p \geq -2), \\ \infty & (p \leq -3). \end{cases} \tag{10.1.53}$$

In particular, plugging in $p = 0$ gives $\langle 1,0,0|1|1,0,0\rangle = 1$ (confirming the correct normalization of the wavefunction), while $p = 1$ gives

$$\langle 1,0,0|R|1,0,0\rangle = \langle R\rangle = \frac{3a_0}{2Z}, \tag{10.1.54}$$

and $p = -1$ gives

$$\langle 1,0,0|\, 1/R\, |0,0,1\rangle = Z/a_0. \tag{10.1.55}$$

(Note that this is not equal to $1/\langle R\rangle$.) We can also directly compute, for the ground state,

$$\langle 1,0,0|P^2|1,0,0\rangle \;=\; \int_0^\infty dr\, r^2 \int d\Omega\, \psi^*_{1,0,0}(-\hbar^2\nabla^2)\psi_{1,0,0} \;=\; \hbar^2 Z^2/a_0^2. \tag{10.1.56}$$

As a check, the expectation value of the Hamiltonian in the ground state is

$$\langle H\rangle \;=\; \langle P^2\rangle/2m_e - Ze^2\langle 1/R\rangle \;=\; -Z^2e^2/2a_0 \;=\; -Z^2 \text{ Rydberg}, \tag{10.1.57}$$

in agreement with the energy eigenvalue.

Later, we will want more general expressions for the expectation values $\langle R^p\rangle$ in arbitrary hydrogen-like atom stationary bound states $|n,l,m\rangle$, for various integer powers $p$. A helpful tool to find these for all integer $p \geq -1$, and for all integer $p \leq -3$ in terms of the single seed case $p = -2$, is the **Kramers–Pasternack recurrence formula**, which relates the expectation values for any three consecutive integer powers of the radial coordinate $R$. It is

$$\frac{q+1}{n^2}\langle R^q\rangle \;-\; (2q+1)\frac{a_0}{Z}\langle R^{q-1}\rangle \;+\; \frac{q}{4}\left[(2l+1)^2 - q^2\right]\frac{a_0^2}{Z^2}\langle R^{q-2}\rangle \;=\; 0, \tag{10.1.58}$$

valid for all $q > -2l-1$.

The proof of eq. (10.1.58) is far from obvious, but goes as follows. Start from the differential equation for the radial wavefunction $R_{n,l}(s)$ with $s = 2Zr/na_0$, as given in eq. (10.1.5). Multiply by $4s^{q+3}R'_{n,l} + 2(1-q)s^{q+2}R_{n,l}$ (this is perhaps the most non-obvious part), and integrate with respect to $s$. Then, expand the integrand and eliminate all derivatives of $R_{n,l}$ using the integration-by-parts identities[§]

$$\int_0^\infty ds\, \frac{d}{ds}\left(s^{q+3}R'^2_{n,l}\right) \;=\; 0, \tag{10.1.59}$$

$$\int_0^\infty ds\, \frac{d}{ds}\left(s^{q+2}R_{n,l}R'_{n,l}\right) \;=\; 0, \tag{10.1.60}$$

$$\int_0^\infty ds\, \frac{d}{ds}\left(s^{p}R^2_{n,l}\right) \;=\; 0, \tag{10.1.61}$$

with $p = q+1$, $q+2$, and $q+3$ in the last equation. The result is

$$\int_0^\infty ds\, \left\{(q+1)s^{q+2} - 2(2q+1)ns^{q+1} + q\left[(2l+1)^2 - q^2\right]s^q\right\} R^2_{n,l} \;=\; 0. \tag{10.1.62}$$

Finally, using the change-of-integration-variable relation,

$$\langle R^p\rangle \;=\; \int_0^\infty dr\, r^{p+2}\left[R_{n,l}(r)\right]^2 \;=\; (na_0/2Z)^{p+3}\int_0^\infty ds\, s^{p+2}\left[R_{n,l}(s)\right]^2, \tag{10.1.63}$$

from eqs. (10.1.4), and (10.1.23), we arrive at eq. (10.1.58).

---

[§] The $'$ denotes a derivative with respect to $s$. The boundary terms in eqs. (10.1.59)–(10.1.61) vanish for $q > -2l-1$, because $R_{n,l}$ scales like $s^l$ for $s \to 0$, and like $e^{-s/2}$ for $s \to \infty$.

Using $q = 0$ in the Kramers–Pasternack formula of eq. (10.1.58), and $\langle R^0 \rangle = \langle 1 \rangle = 1$, we immediately find

$$\langle 1/R \rangle \;=\; \frac{Z}{n^2 a_0}, \tag{10.1.64}$$

in agreement with the $n = 1$ special case in eq. (10.1.55). Then, using $q = 1$, one obtains

$$\langle R \rangle \;=\; \frac{a_0}{2Z}[3n^2 - l(l + 1)]. \tag{10.1.65}$$

Using this as a measure of the "size" of the state $|n, l, m\rangle$, we note that it does not depend on the magnetic quantum number $m$, and

- shrinks with larger nuclear charge $Z$,
- grows with larger $n$, for fixed $l$,
- shrinks with larger $l$, for fixed $n$.

The last two of these features can be compared visually to the probability density distributions in the right column of Figure 10.1.2.

Applying eq. (10.1.58) with $q = 2$, we get

$$\langle R^2 \rangle \;=\; \frac{a_0^2}{Z^2}\frac{n^2}{2}\left[5n^2 - 3l(l + 1) + 1\right]. \tag{10.1.66}$$

One can apply eq. (10.1.58) recursively to find $\langle R^q \rangle$ for any desired positive integer $q$. However, for negative $q$, one finds an obstacle, that $\langle 1/R^2 \rangle$ cannot be determined by the recurrence relation alone. In section 13.6, we will use another method to find

$$\langle 1/R^2 \rangle \;=\; \frac{Z^2}{a_0^2 n^3 (l + 1/2)}. \tag{10.1.67}$$

With this as a seed, all results for $\langle R^p \rangle$ with $p \leq -3$ can then be determined by the recurrence relation. In particular, using $q = -1$ in eq. (10.1.58) yields[¶]

$$\langle 1/R^3 \rangle \;=\; \frac{Z^3}{a_0^3 n^3 l(l + 1)(l + 1/2)}. \tag{10.1.68}$$

The expectation value of $P^2$ can also be found by relating it to the Hamiltonian,

$$\langle P^2 \rangle \;=\; 2m_e\big(\langle H \rangle + Ze^2 \langle 1/R \rangle\big) \;=\; m_e\left(\frac{e^2}{2a_0}\right)\frac{2Z^2}{n^2} \;=\; \frac{\hbar^2 Z^2}{a_0^2 n^2}, \tag{10.1.69}$$

where we have used the known energy eigenvalues and the result of eq. (10.1.64).

---

[¶] Note that $\langle 1/R^3 \rangle$ diverges for $l = 0$. More generally, $\langle R^p \rangle = \int_0^\infty dr\, r^{2+p} |R_{n,l}(r)|^2$ diverges if $p \leq -2l - 3$, because $|R_{nl}| \propto r^l$ for small $r$.

The characteristic size of the hydrogen atom with $Z = 1$ is set by the Bohr radius $a_0 = \hbar^2/m_e e^2$. This could have been estimated by a dimensional analysis construction from the available quantities in the problem, by requiring it to have units of length, and demanding that it vanish in each of the following three extreme limits: (1) the classical limit $\hbar \to 0$, and (2) the limit that the electron mass $m_e$ is taken very large, and (3) the limit of large $e^2$ so that the attractive electrical force between the nucleus and the electron is large.

However, if we add the speed of light $c$ to our toolbox of quantities, then there is a dimensionless quantity that we can form out of the available parameters in the problem, namely the **fine structure constant**,

$$\alpha = \frac{e^2}{\hbar c} \approx 1/137.036. \tag{10.1.70}$$

So, one can construct another length scale, which also turns out to be significant, the **reduced**[‖] **Compton wavelength** of the electron,

$$\lambdabar_e \;=\; \alpha a_0 \;=\; \frac{\hbar}{m_e c} \approx 3.862 \times 10^{-13}\,\text{meters}. \tag{10.1.71}$$

Named for Arthur Compton's studies of the scattering of electrons by high-energy photons, this is also the (very small) length scale at which virtual electron-positron pair production becomes important for understanding the energy levels of electron bound states. To see this, note that according to Einstein's famous formula relating energy and mass, the amount of energy needed to make such a pair is $\Delta E = 2m_e c^2$. For ultra-relativistic particles, energy is related to momentum by $\Delta E \sim c\Delta p$, so $e^- e^+$ pair production becomes important for $\Delta p > 2m_e c$. From the uncertainty principle $(\Delta x)(\Delta p) \geq \hbar/2$, confinement of the electron to a size $\Delta x \sim \hbar/4m_e c \sim \lambdabar_e/4$ will result in large enough fluctuations $\Delta p$ to cause virtual $e^- e^+$ pair production to be an issue. Furthermore, if we call $v = \sqrt{\langle P^2\rangle}/m_e$ the order of magnitude of the "velocity" of electrons in a hydrogen-like atom state with principal quantum number $n$, then

$$v \;\sim\; \frac{Z\hbar}{n a_0 m_e} \;=\; Z\alpha c/n. \tag{10.1.72}$$

These considerations show that the approximation we have made in this section works only because the fine structure constant is small. That is what ensures $\lambdabar_e \ll a_0$ and $v \ll c$, so that electron-positron pair production and other relativistic corrections to the Hamiltonian can be neglected, to first approximation. However, for atoms with very large $Z$, the speeds of atomic electrons increase, and relativistic effects become important.

A related point is that the binding energy of the hydrogen atom is very small compared to $m_e c^2 = 0.511$ MeV, the rest energy of the electron. In terms of the fine structure constant,

$$1\ \text{Rydberg} \;=\; \frac{e^2}{2a_0} \;=\; \frac{e^4 m_e}{2\hbar^2} \;=\; \frac{1}{2}\alpha^2 m_e c^2. \tag{10.1.73}$$

---

[‖] The ordinary Compton wavelength of the electron is defined as $\lambda_e = h/m_e c = 2\pi\lambdabar_e$.

To estimate (very roughly) the size of the relativistic corrections to the energies, we can take the total classical relativistic energy of an electron with momentum $p$, and subtract off the rest energy, then expand in small $p$,

$$E_{\text{relativistic}} \;=\; \sqrt{m_e^2c^4 + p^2c^2} - m_ec^2 \;=\; \frac{p^2}{2m_e} - \frac{p^4}{8m_e^3c^2} + \cdots\,, \tag{10.1.74}$$

where we have used the binomial expansion $\sqrt{1+x} = 1 + x/2 - x^2/8 + \cdots$ for small $x = p^2/m_e^2c^2$. The first term is just the non-relativistic kinetic energy that we have been using in the quantum Hamiltonian. So, we can estimate that relativistic effects should make a difference of order

$$\Delta E_{\text{relativistic}} \;\sim\; -\frac{\langle P^2\rangle^2}{8m_e^3c^2} \;=\; -\frac{\alpha^2Z^4}{4n^4}\,\text{Rydberg}. \tag{10.1.75}$$

It turns out that a correct relativistic analysis must take into account spin. We will carry this out in section 15.1, with a final result in eq. (15.1.27). Our crude estimate in eq. (10.1.75) is parametrically correct in the sense that the fine structure effects indeed modify the binding energies by amounts that are suppressed by $\alpha^2$ and by powers of $n$, but we will see that the other numerical details are different than the naive estimate of eq. (10.1.75).

When an electron transitions between states of the hydrogen atom, it releases or absorbs a photon with energy equal to the difference in energy levels, $\Delta E = E_n - E_{n'}$. This is equal to $\hbar\omega$ where $\omega$ is the angular frequency of the photon. Therefore, taking $Z = 1$ for the remainder of this section,

$$\omega_{n,n'} \;=\; \frac{13.6\,\text{eV}}{\hbar}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right), \tag{10.1.76}$$

or equivalently in terms of wavelength,

$$\lambda_{n,n'} \;=\; (9.11\times10^{-8}\,\text{meters})\frac{n'^2}{1 - n'^2/n^2}. \tag{10.1.77}$$

This formula was found empirically by Rydberg in 1888, generalizing work by Johann Balmer. For transitions between the ground state $n' = 1$ and the states with $n \geq 2$, these spectral lines are called the Lyman series, after Theodore Lyman. They are all in the ultraviolet range,

$$\text{Lyman series: } \omega_{n,1} = (2.067\times10^{16}\,\text{s}^{-1})\left(1 - \frac{1}{n^2}\right), \qquad (n = 2, 3, 4, \ldots, \infty), \tag{10.1.78}$$

with wavelengths from $\lambda_{2,1} = 1.216\times10^{-7}$ meters to $\lambda_{\infty,1} = 9.11\times10^{-8}$ meters. The latter wavelength, corresponding to a photon emitted when a free electron is captured to the ground state of a hydrogen atom, or absorbed in the process of ionizing a hydrogen atom that was initially in its ground state, is called the Lyman limit. The spectral line with $n' = 1$ and $n = 2$

is called the Lyman-alpha line, and is important in astronomy as an absorption feature when observing very distant sources, as a tool to learn about intervening intergalactic gas.

The Balmer series comes from transitions between $n' = 2$ and all higher-energy states $n \geq 3$,

$$\text{Balmer series:}\;\; \omega_{n,2} = (5.168 \times 10^{15}\,\text{s}^{-1})\left(1 - \frac{4}{n^2}\right), \qquad (n = 3, 4, 5, \ldots, \infty), \tag{10.1.79}$$

with wavelengths ranging from $\lambda_{3,2} = 6.56 \times 10^{-7}$ meters to $\lambda_{\infty,2} = 3.97 \times 10^{-7}$ meters. The Balmer lines are particularly interesting because four of them can be seen by the human eye. For $n = 3$ the Balmer line is red, which is responsible for the reddish color of supernova remnants and star-forming regions, including the Orion nebula. The $n = 4$, 5, and 6 Balmer lines are greenish-blue, blue, and violet, respectively, while $n = 7, \ldots, \infty$ are in the ultraviolet. The remaining named series are all entirely in the infrared, and are called Paschen (transitions between $n' = 3$ and $n \geq 4$), Brackett (transitions between $n' = 4$ and $n \geq 5$), Pfund (transitions between $n' = 5$ and $n \geq 6$), and Humphreys (transitions between $n' = 6$ and $n \geq 7$), with higher $n'$ series not named. There is significant overlap between the infrared series of spectral lines. Within each series, the spectral lines are called $\alpha$, $\beta$, etc. in order of decreasing wavelength.

## 10.2 Unbound states of Coulomb potentials

We now consider the unbound stationary states of the Coulomb problem, with $E \geq 0$. These include the ionized states of the hydrogen-like atom, describing an electron that is influenced by the nucleus but not localized near it. The electron's wavefunction can again be split into the product of an angular part consisting of a spherical harmonic and a radial wavefunction,

$$\langle \vec{r}|\Phi_{E,l,m}\rangle \;=\; \Phi_{E,l,m}(r,\theta,\phi) \;=\; \widetilde{R}_{E,l}(r) Y_l^m(\theta,\phi) \qquad (E \geq 0). \tag{10.2.1}$$

We write $\Phi_{E,l,m}$ and $\widetilde{R}_{E,l}$ here to distinguish them from the bound-state wavefunctions $\psi_{n,l,m}$ and $R_{n,l}$ of the preceding section. For the unbound states, each $E \geq 0$ can have any non-negative integer value of $l$, as depicted in Fig. 10.1.1.

The differential equation satisfied by the radial wavefunction for unbound states is the same as eq. (10.1.5), with the important differences that the rescaled radial coordinate $s$ and the quantity $n$, defined in eqs. (10.1.4) and (10.1.6) respectively, are imaginary for $E > 0$. Accordingly, we make the replacement[†]

$$n \to \frac{i}{ka_0}, \tag{10.2.2}$$

---

[†] In this section, we set $\mu = m_e$ and $Z = 1$ for simplicity, with the understanding that general $Z$ can always be restored by the replacement $a_0 \to a_0/Z$.

which defines a continuous real wavenumber parameter $k$ with units of 1/[length]. Comparing to the definition of $n$ in eq. (10.1.6) gives

$$ka_0 = \sqrt{\frac{2\hbar^2 E}{m_e e^4}}, \tag{10.2.3}$$

or, equivalently,

$$E \;=\; \frac{1}{2}(ka_0)^2 \text{ Rydberg}. \tag{10.2.4}$$

Then, from eq. (10.1.4) the rescaled radial coordinate used in the previous section becomes

$$s \;=\; 2ikr. \tag{10.2.5}$$

It follows that, up to normalization, we can obtain the solution in terms of the confluent hypergeometric function defined in eq. (10.1.10), by simply re-using eq. (10.1.13) with the preceding substitutions for $n$ and $s$. Since $s$ is now imaginary, there is no problem with having an asymptotic behavior like $e^{s/2}$ for large $s$, and therefore no restriction that the confluent hypergeometric function must be a polynomial; that is why there is no discrete quantization of $k$. The result is

$$\widetilde{R}_{E,l}(r) \;=\; C_{k,l}\,(2kr)^l\, e^{-ikr}\, F(l{+}1{+}\frac{i}{ka_0},\, 2l{+}2,\, 2ikr), \tag{10.2.6}$$

where $C_{k,l}$ is a normalization constant to be determined, which we will choose to be real.

Despite the appearance of the phase factor $e^{-ikr}$ and the complex arguments of the confluent hypergeometric function, $\widetilde{R}_{E,l}(r)$ as given in eq. (10.2.6) is real. This follows immediately from a property of the confluent hypergeometric function,

$$F(a,c,z) \;=\; e^z F(c-a, c, -z), \tag{10.2.7}$$

which can in turn be proved quickly from the integral representation of eq. (10.1.11) by using the change of integration variable $t \to 1-t$.

Let us now consider the behavior of the radial wavefunction $\widetilde{R}_{E,l}$ for large $r$. This can be done by using the asymptotic form for the confluent hypergeometric function for complex arguments,

$$F(a,c,z) \;\approx\; \frac{\Gamma(c)}{\Gamma(c-a)}(-z)^{-a} + \frac{\Gamma(c)}{\Gamma(a)} e^z z^{a-c} \qquad \text{(large } |z|\text{)}, \tag{10.2.8}$$

which generalizes eq. (10.1.12) for real arguments. Applying this to eq. (10.2.6), the two terms turn out to give contributions that are complex conjugates of each other, as required by the previous paragraph. After some simplification, one finds the asymptotic form

$$\widetilde{R}_{E,l}(r) \;\approx\; C_{k,l}\,\frac{(2l+1)!\, e^{-\pi/2ka_0}}{|\Gamma(l+1+i/ka_0)|}\,\frac{1}{kr}\,\sin\Bigl(kr + \frac{1}{ka_0}\ln(2kr) - \frac{\pi}{2}l - \delta\Bigr) \tag{10.2.9}$$

for large $kr$, where

$$\delta = \mathrm{Arg}\big[\Gamma\big(l+1+i/ka_0\big)\big]. \tag{10.2.10}$$

The radial probability density $r\widetilde{R}_{E,l}(r)$ approaches a sinusoidal oscillation with a constant amplitude at large distances, but with a phase shift that depends logarithmically on $kr$.

In order to determine the normalization constant, we first define the orthonormality properties of the states. Since the energy $E \geq 0$ is continuous, we will use a Dirac orthonormality condition for unbound states $|E, l, m\rangle$, along with the ordinary discrete orthonormality of the bound states $|n, l, m\rangle$ of the previous section. So,

$$\langle E', l', m'|E, l, m\rangle = \delta(E-E')\,\delta_{l,l'}\,\delta_{m,m'}, \tag{10.2.11}$$

$$\langle n', l', m'|E, l, m\rangle = 0, \tag{10.2.12}$$

$$\langle n', l', m'|n, l, m\rangle = \delta_{n,n'}\,\delta_{l,l'}\,\delta_{m,m'}. \tag{10.2.13}$$

In terms of the radial wavefunctions for each $l$, these become

$$\int_0^\infty dr\, r^2\, \widetilde{R}_{E',l}(r)\widetilde{R}_{E,l}(r) = \delta(E-E'), \tag{10.2.14}$$

$$\int_0^\infty dr\, r^2\, \widetilde{R}_{E',l}(r)R_{n,l}(r) = 0, \tag{10.2.15}$$

$$\int_0^\infty dr\, r^2\, R_{n',l}(r)R_{n,l}(r) = \delta_{n,n'}. \tag{10.2.16}$$

Now, for the purposes of normalizing the radial wavefunction in eq. (10.2.9), note that when $E = E'$, one need only consider the asymptotic form at very large $r$, because contributions to $\int_0^\infty dr\, r^2\, |\widetilde{R}_{E,l}(r)|^2$ from any finite range in $r$ contribute only an infinitesimal fraction of the total. In the very large $r$ limit, the logarithmic variation in the phase shift can be neglected, and the normalization problem is the same as for the simpler wavefunctions

$$f_k(r) = A\,\frac{\sin(kr+\beta)}{kr}, \tag{10.2.17}$$

for some phase shift $\beta$, where

$$A = C_{k,l}\,\frac{(2l+1)!\,e^{-\pi/2ka_0}}{|\Gamma(l+1+i/ka_0)|}. \tag{10.2.18}$$

The norm of this wavefunction is infinite, but integrating over a finite range gives

$$\int_0^D dr\, r^2\, f_{k'}(r)f_k(r) = \frac{A^2}{2kk'}\left\{\frac{\sin[(k-k')D]}{k-k'} + \frac{\sin[2\beta]-\sin[2\beta+(k+k')D]}{k+k'}\right\}. \tag{10.2.19}$$

In the formal limit $D \to \infty$, one can interpret the first term in the braces as a delta function distribution [see eq. (2.2.22)], while the remaining part remains bounded for all $k$ and $k'$, and vanishes for an infinite number of choices $D = 2\pi q/(k+k')$, for arbitrarily large integers $q$. This allows us to interpret, when inserted in any expression in which $k$ or $k'$ is integrated over,

$$\int_0^\infty dr\, r^2\, f_{k'}(r) f_k(r) \;\;=\;\; A^2 \frac{\pi}{2k^2}\delta(k-k') \;=\; A^2 \frac{\pi\hbar^2}{4m_e k}\delta(E-E'). \qquad (10.2.20)$$

Thus we can adopt the energy normalization for Dirac orthonormality, by taking $A^2 = 4m_e k/\pi\hbar^2$, allowing us to solve eq. (10.2.18) for $C_{k,l}$. Using this in eq. (10.2.6), the final result for the unbound energy eigenstate state radial wavefunction is

$$\widetilde{R}_{E,l}(r) \;\;=\;\; \frac{2}{\hbar}\sqrt{\frac{m_e k}{\pi}}\frac{|\Gamma(l{+}1{+}i/ka_0)|e^{\pi/2ka_0}}{(2l+1)!}\,(2kr)^l\, e^{-ikr}\, F(l{+}1{+}\frac{i}{ka_0},\, 2l+2,\, 2ikr), \quad (10.2.21)$$

where $k$ is related to $E$ by eq. (10.2.3) or eq. (10.2.4). Although this is not the simplest result one might have hoped for, we again remark that at least it is real, despite naive appearances.

The completeness relation corresponding to the energy eigenstates of eqs. (10.2.11)-(10.2.13) contains both a sum over bound states and an integral over unbound states:

$$\sum_{n=1}^{\infty}\sum_{l=0}^{n-1}\sum_{m=-l}^{l} |n,l,m\rangle\langle n,l,m| \,+\, \int_0^\infty dE \sum_{l=0}^{\infty}\sum_{m=-l}^{l} |E,l,m\rangle\langle E,l,m| \;\;=\;\; I. \qquad (10.2.22)$$

This mixture of Dirac and ordinary orthonormality and completeness will be crucial in the evaluation of the ground state energy of the hydrogen atom in an electric field (the quadratic Stark effect), in section 13.7.

For a general potential $V(r) = -Ze^2/r$, the results above can be obtained by making the replacement $a_0 \to a_0/Z$ everywhere in the preceding discussion. This includes the case of a repulsive potential with $Z < 0$, but with one qualitative difference: the bound states $|n,l,m\rangle$ do not exist, and the orthonormality and completeness relations therefore include only the unbound states with continuous positive $E$.

# 11 Addition of angular momenta

## 11.1 Statement of the problem

In quantum mechanical problems, we often deal with two or more distinct angular momenta. The sum of two independent angular momentum operators,

$$\vec{J} = \vec{J}_1 + \vec{J}_2, \tag{11.1.1}$$

is also an angular momentum operator. This expresses the fact that if all components of $\vec{J}_1$ commute with all components of $\vec{J}_2$, and if the components of $\vec{J}_1$ and $\vec{J}_2$ each satisfy the commutator algebra (5.3.32), then so will the components of $\vec{J}$.

Since $J_1^2$, $J_2^2$, $J_{1z}$, and $J_{2z}$ are compatible operators, we can find an orthobasis of common eigenkets for them, labeled

$$|j_1\, j_2\, m_1\, m_2\rangle \equiv |j_1\, m_1\rangle \otimes |j_2\, m_2\rangle\,, \tag{11.1.2}$$

with eigenvalues $\hbar^2 j_1(j_1{+}1)$, $\hbar^2 j_2(j_2{+}1)$, $\hbar m_1$, and $\hbar m_2$, respectively. We call this the **product orthobasis** for two angular momenta, because its elements consist of the tensor products of elements of the orthobases for the individual angular momenta. Here we have suppressed any degeneracy labels, which might correspond to different radial wavefunctions, for example. For fixed $j_1$ and $j_2$, there are $2j_1 + 1$ allowed values of $m_1$, and $2j_2 + 1$ allowed values of $m_2$,

$$m_1 = j_1,\ j_1{-}1, \ldots,\ -j_1{+}1,\ -j_1, \tag{11.1.3}$$

$$m_2 = j_2,\ j_2{-}1, \ldots,\ -j_2{+}1,\ -j_2. \tag{11.1.4}$$

So, there are $(2j_1 + 1)(2j_2 + 1)$ orthobasis kets of the form eq. (11.1.2) for fixed $j_1$ and $j_2$.

Another set of compatible operators is $J_1^2$, $J_2^2$, $J^2$, and $J_z$. They also have a set of common eigenkets that form a different orthobasis, whose kets we can write as

$$|j_1\, j_2\, j\, m\rangle\,. \tag{11.1.5}$$

This is called the **total angular momentum orthobasis**. As we will see in the following sections, for fixed $j_1$ and $j_2$, the possible values of $j$ range from a maximum of $j_1 + j_2$ to a minimum of $|j_1 - j_2|$, with integer increments,

$$j = j_1{+}j_2,\ j_1{+}j_2{-}1,\ \ldots,\ |j_1{-}j_2|{+}1,\ |j_1{-}j_2|. \tag{11.1.6}$$

Intuitively, the extremes for $j$ correspond to the two angular momenta aligned in the same and in opposite directions, respectively. Then, for each $j$, there are $2j{+}1$ allowed values

$$m = j,\ j{-}1,\ \ldots, -j{+}1,\ -j. \tag{11.1.7}$$

As a check, the number of total angular momentum basis states for fixed $j_1$ and $j_2$ is

$$\sum_{j=|j_1-j_2|}^{j_1+j_2} (2j+1) \quad = \quad (2j_1+1)(2j_2+1), \tag{11.1.8}$$

matching the result for the number of product basis kets.

There is a potentially annoying problem of notation to be addressed here, because both orthobases are labeled by four numbers. In many cases, this will not cause confusion, but what if it does? Our solution to this problem will be to use a colon to separate the last two (magnetic quantum number) labels for the product angular momentum orthobasis whenever there is a chance of confusion. All other labels in both the product and total angular momentum bases are separated by no punctuation, or by a comma when it is typographically convenient (or just suits our mood). Also, there will be many occasions in which the labels $j_1$ and $j_2$ that are common to both orthobases are fixed, and understood from context. In that case, we can suppress those labels and call the total angular momentum orthobasis kets $|jm\rangle$ or $|j,m\rangle$, and the product orthobasis kets $|m_1 m_2\rangle$ or $|m_1, m_2\rangle$, but use the notation $|m_1 \!:\! m_2\rangle$ if there is a significant chance of confusion with the total angular momentum orthobasis.

Because they are both bases, each element of the total angular momentum orthobasis must be a linear combination of the product orthobasis kets, and vice versa. In general, the problem of **addition of angular momenta** is to evaluate the coefficients appearing in these linear combinations.

We will start with two useful and common special cases before taking on the most general case. First, in section 11.2, we consider the case of two spins $s_1 = s_2 = 1/2$, for two particles with orbital angular momenta absent or disregarded. Next, in section 11.3, we will take up the example of a single particle with arbitrary orbital angular momentum $l$ and spin $s = 1/2$, which can be combined to form the total angular momentum of the particle. In section 11.4 we will discuss the general case of addition of two arbitrary angular momenta. These results can be applied recursively to combine any number of angular momenta.

## 11.2 Addition of $s_1 = 1/2$ and $s_2 = 1/2$

Consider two spins with $s_1 = s_2 = 1/2$. Spin magnitudes are always fixed, so we suppress those labels, and write the four product orthobasis kets as

$$|{\uparrow\uparrow}\rangle\,, \quad |{\uparrow\downarrow}\rangle\,, \quad |{\downarrow\uparrow}\rangle\,, \quad |{\downarrow\downarrow}\rangle\,. \tag{11.2.1}$$

Here the first ↑ or ↓ label on each ket stands for the eigenvalue $\hbar/2$ or $-\hbar/2$ of $S_{1z}$, and the second label similarly stands for the eigenvalue of $S_{2z}$. We say that the individual spins in this basis are either "up" or "down" relative to our choice of the $\hat{z}$ direction. Now, define the total spin operator by

$$\vec{S} = \vec{S}_1 + \vec{S}_2. \tag{11.2.2}$$

The product orthobasis kets are also eigenkets of $S_z$, since

$$S_z |m_1 m_2\rangle = S_{1z} |m_1 m_2\rangle + S_{2z} |m_1 m_2\rangle = \hbar(m_1 + m_2) |m_1 m_2\rangle , \tag{11.2.3}$$

so that $S_z |{\uparrow\uparrow}\rangle = \hbar |{\uparrow\uparrow}\rangle$, and $S_z |{\uparrow\downarrow}\rangle = S_z |{\downarrow\uparrow}\rangle = 0$, and $S_z |{\downarrow\downarrow}\rangle = -\hbar |{\downarrow\downarrow}\rangle$. Choosing a representation in which

$$|{\uparrow\uparrow}\rangle \leftrightarrow \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \qquad |{\uparrow\downarrow}\rangle \leftrightarrow \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \qquad |{\downarrow\uparrow}\rangle \leftrightarrow \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \qquad |{\downarrow\downarrow}\rangle \leftrightarrow \begin{pmatrix}0\\0\\0\\1\end{pmatrix}, \tag{11.2.4}$$

we have the matrix representation

$$S_z \leftrightarrow \hbar \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}. \tag{11.2.5}$$

However, the product orthobasis kets are not all eigenstates of $S^2$. The operation of $S^2$ on them can be obtained from

$$\begin{aligned} S^2 &= (\vec{S}_1 + \vec{S}_2) \cdot (\vec{S}_1 + \vec{S}_2) = S_1^2 + S_2^2 + 2\vec{S}_1 \cdot \vec{S}_2 \\ &= \frac{3}{2}\hbar^2 + S_{1+}S_{2-} + S_{1-}S_{2+} + 2S_{1z}S_{2z}, \end{aligned} \tag{11.2.6}$$

where in the last equality we have taken advantage of the fact that all of the kets in the state space under consideration are eigenkets of both $S_1^2$ and $S_2^2$ with eigenvalues $3\hbar^2/4$, and then applied eq. (8.1.8) to write the result in terms of the angular momentum raising and lowering operators. Now, recall from eqs. (8.1.22) and (8.1.24) that, for any $j = 1/2$ system,

$$J_+ |{\uparrow}\rangle = 0, \qquad J_+ |{\downarrow}\rangle = \hbar |{\uparrow}\rangle , \tag{11.2.7}$$

$$J_- |{\downarrow}\rangle = 0, \qquad J_- |{\uparrow}\rangle = \hbar |{\downarrow}\rangle . \tag{11.2.8}$$

Applying these for each of $J = S_1$ and $S_2$, we obtain from eq. (11.2.6),

$$S^2 |{\uparrow\uparrow}\rangle = 2\hbar^2 |{\uparrow\uparrow}\rangle , \tag{11.2.9}$$

$$S^2 |{\uparrow\downarrow}\rangle = \hbar^2 |{\uparrow\downarrow}\rangle + \hbar^2 |{\downarrow\uparrow}\rangle , \tag{11.2.10}$$

$$S^2 |{\downarrow\uparrow}\rangle = \hbar^2 |{\uparrow\downarrow}\rangle + \hbar^2 |{\downarrow\uparrow}\rangle , \tag{11.2.11}$$

$$S^2 |{\downarrow\downarrow}\rangle = 2\hbar^2 |{\downarrow\downarrow}\rangle . \tag{11.2.12}$$

In matrix representation form, this reads

$$S^2 \;\leftrightarrow\; \hbar^2 \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \tag{11.2.13}$$

This matrix has eigenvalues $2\hbar^2$, $2\hbar^2$, $2\hbar^2$, and 0, which correspond to $S^2 = \hbar^2 s(s+1)$ with $s = 1$ and $s = 0$. The addition of two spins with $s_1 = s_2 = 1/2$ thus gives a triplet of states $|1, m\rangle$ with total spin $s = 1$ and $m = -1, 0, 1$, plus a singlet state with total spin $s = 0$, denoted $|0, 0\rangle$. By finding the eigenkets corresponding to the eigenvalue pairs $(s, m)$, we can summarize the information about the addition of two spin-1/2 systems as

$$\begin{aligned} &\text{total spin basis } |s, m\rangle && \text{product basis } |m_1 m_2\rangle \\ &|1, 1\rangle \;=\; && |{\uparrow\uparrow}\rangle, && (11.2.14) \\ &|1, 0\rangle \;=\; && \frac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle + |{\downarrow\uparrow}\rangle\right), && (11.2.15) \\ &|1, -1\rangle \;=\; && |{\downarrow\downarrow}\rangle, && (11.2.16) \\ &|0, 0\rangle \;=\; && \frac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle - |{\downarrow\uparrow}\rangle\right). && (11.2.17) \end{aligned}$$

Both orthobases have 4 members, and we can of course invert the relationship for the $m = 0$ states, to find

$$|{\uparrow\downarrow}\rangle \;=\; \frac{1}{\sqrt{2}}\left(|1, 0\rangle + |0, 0\rangle\right), \tag{11.2.18}$$

$$|{\downarrow\uparrow}\rangle \;=\; \frac{1}{\sqrt{2}}\left(|1, 0\rangle - |0, 0\rangle\right). \tag{11.2.19}$$

As a way of expressing the fact that the tensor product of two spin-1/2 systems gives a sum of angular momenta 0 and 1, we write

$$\frac{1}{2} \otimes \frac{1}{2} \;=\; 0_A \oplus 1_S, \tag{11.2.20}$$

Here, the $\otimes$ represents the addition of angular momenta, while the $\oplus$ indicates the combinations of total angular momenta. The subscripts $A$ and $S$ in this notation are a reminder that the $s = 0$ singlet state is antisymmetric under exchange of the two spins, while each of the $s = 1$, $m = 1, 0, -1$ triplet states is symmetric, as can be seen in eqs. (11.2.14)-(11.2.17).

An important practical question is: which orthobasis should we use? The answer depends on the problem under consideration. As a general rule, it is convenient to use a basis in which the operators most important to us are diagonal. For example, suppose that we have a Hamiltonian of the form

$$H \;=\; b_1 S_{1z} + b_2 S_{2z}, \tag{11.2.21}$$

which could occur if both spins are interacting with a magnetic field, but not with each other. This Hamiltonian is diagonal in the product basis, where the kets are already eigenvalues of $S_{1z}$ and $S_{2z}$, but it is diagonal in the total angular momentum basis only if $b_1 = b_2$. On the other hand, suppose that our Hamiltonian is of the form

$$H = a\vec{S}_1 \cdot \vec{S}_2, \tag{11.2.22}$$

which comes from the magnetic moments of the particles interacting with each other. In that case, we can use a common trick, by writing

$$\vec{S}_1 \cdot \vec{S}_2 = \frac{1}{2}\big(S^2 - S_1^2 - S_2^2\big). \tag{11.2.23}$$

The right side shows that this operator is diagonal in the total spin basis, where it evaluates to

$$\vec{S}_1 \cdot \vec{S}_2 = \frac{\hbar^2}{2}\left[s(s+1) - 3/2\right], \tag{11.2.24}$$

with $s = 1$ for the triplet states and $s = 0$ for the singlet state. In other cases, a judgment can be made about which is the most convenient basis, often based on which part of the Hamiltonian, or some other observable of interest, is the most important.

As a classic example, the **hyperfine** splitting of the hydrogen atom ground state arises from the interaction of the electron's spin with the spin of the proton. The hyperfine Hamiltonian $H_{\text{hf}}$ has exactly the form of eq. (11.2.22), with a constant $a$ that is positive, and very small compared to the Rydberg energy scale divided by $\hbar^2$. Before taking into account fine or hyperfine effects, the hydrogen atom energy eigenstates can be given in the product spin basis labeled as

$$|\underbrace{n, l, m_l}_{\text{orbital}}, \overbrace{s_e, m_{s_e}}^{\text{electron spin}}, \underbrace{s_p, m_{s_p}}_{\text{proton spin}}\rangle = |n, l, m_l, m_{s_e}, m_{s_p}\rangle, \tag{11.2.25}$$

or we can use the the total spin basis,

$$|n, l, m_l, s_e, s_p, s, m_s\rangle = |n, l, m_l, s, m_s\rangle. \tag{11.2.26}$$

In both cases, the individual spin labels $s_e, s_p = 1/2$ are completely fixed and therefore can be suppressed, as indicated. Our present interest is the effect on the ground state with $n = 1$ and $l = 0$. Since $H_{\text{orbital}} = P^2/2m_e - e^2/R$ commutes with

$$H_{\text{hf}} = a\vec{S}_e \cdot \vec{S}_p, \tag{11.2.27}$$

they have an orthobasis of common eigenstates. The eigenvalues of $H_{\text{hf}}$ are found immediately in the total spin basis, using eq. (11.2.24), as

$$E_{\text{hf}} = \begin{cases} a\hbar^2/4 & (s = 1), \\ -3a\hbar^2/4 & (s = 0), \end{cases} \tag{11.2.28}$$

for the triplet and singlet total spin states, respectively. The energy splitting between these states is therefore $a\hbar^2$, which for the ground state of the hydrogen atom is, numerically,

$$\Delta E_{\text{hf}} = a\hbar^2 = 5.87 \times 10^{-6}\,\text{eV}, \tag{11.2.29}$$

corresponding to a wavelength $\lambda = 2\pi c/a\hbar = 0.211$ meters. This is the famous 21 centimeter line of radio astronomy. The rate for transitions between the $s = 1$ and $s = 0$ states turns out to be highly suppressed (for reasons to be discussed in section 19.5), but space is big and mostly cold, so it is a very useful observational tool in astrophysics and experimental cosmology.

There is a special consideration when the two spin-1/2 particles are identical. The **Pauli exclusion principle** says that two identical particles with half-integer spin cannot be in the same quantum state, and more generally that the quantum state describing two identical particles must be antisymmetric under their exchange. Such particles are called **fermions**, after Enrico Fermi, while particle with integer spin are called **bosons** after Satyendra Nath Bose. For example, suppose that the orbital wavefunctions of two spin-1/2 fermions are $\psi_a(\vec{r})$ and $\psi_b(\vec{r})$, corresponding to single-particle kets (neglecting spins for the moment) $|\psi_a\rangle$ and $|\psi_b\rangle$. From these, one can form symmetric and antisymmetric combinations

$$|\psi_a, \psi_b\rangle_S = \frac{1}{\sqrt{2}}\Big(|\psi_a, \psi_b\rangle + |\psi_b, \psi_a\rangle\Big), \tag{11.2.30}$$

$$|\psi_a, \psi_b\rangle_A = \frac{1}{\sqrt{2}}\Big(|\psi_a, \psi_b\rangle - |\psi_b, \psi_a\rangle\Big), \tag{11.2.31}$$

where, on the right-hand sides, the first and second entries in each ket correspond to particle labels 1 and 2, respectively. Then the allowed states constructed from the tensor product of the orbital and total spin states must be either symmetric in orbital kets and antisymmetric in spin kets, or vice versa,

$$|\psi_a, \psi_b\rangle_S \otimes \frac{1}{\sqrt{2}}\Big(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\Big) \qquad (s = 0), \tag{11.2.32}$$

$$|\psi_a, \psi_b\rangle_A \otimes |\uparrow\uparrow\rangle \qquad (s = 1,\ m_s = 1), \tag{11.2.33}$$

$$|\psi_a, \psi_b\rangle_A \otimes \frac{1}{\sqrt{2}}\Big(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle\Big) \qquad (s = 1,\ m_s = 0), \tag{11.2.34}$$

$$|\psi_a, \psi_b\rangle_A \otimes |\downarrow\downarrow\rangle \qquad (s = 1,\ m_s = -1). \tag{11.2.35}$$

Other kets, such as $|\psi_a, \psi_b\rangle_S \otimes |\uparrow\uparrow\rangle$, do not exist as physical states. For example, the ground state of helium, with two electrons, has an orbital part that is symmetric under interchange of the positions of the two electrons. The spin state is therefore the antisymmetric total spin $s = 0$ combination.

## 11.3 Addition of orbital angular momentum and spin 1/2

In this section, we consider the combination of two angular momenta $j_1$ and $j_2$, with $j_1$ arbitrary and $j_2 = 1/2$. This applies, for example, to the case where we are combining the orbital $\vec{L}$ and intrinsic $\vec{S}$ angular momenta for a single spin-1/2 particle such as the electron. We will frame our discussion in that context.

As we saw in section 8.2, the spin states of an electron can be associated with a two-component spinor, with the spin operator $\vec{S}$ represented by $2 \times 2$ matrices $\frac{\hbar}{2}\vec{\sigma}$. In the position wavefunction spinor representation corresponding to the CSCO $(\vec{R}, S^2, S_z)$, the orbital angular momentum operators $\vec{L}$ and $L^2$ are represented by differential operators proportional to the unit matrix in the spin sector, for example [compare eqs. (8.6.7) and (8.6.8)]

$$L_z \leftrightarrow -i\hbar \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{\partial}{\partial \phi}, \qquad L_+ \leftrightarrow \hbar e^{i\phi} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \left( \frac{\partial}{\partial \theta} + i \frac{\cos\theta}{\sin\theta} \frac{\partial}{\partial \phi} \right), \tag{11.3.1}$$

etc. However, it is often more useful in problems with spherical symmetry to replace $\vec{R}$ in the CSCO with the radial coordinate operator $R$ and the operators $L^2$ and $L_z$, so that all of the angular momentum operators are represented by matrices rather than differential operators on wavefunctions.

The angular momentum product orthobasis consists of common eigenkets of the compatible operators ($L^2$, $S^2$, $L_z$, $S_z$). The corresponding basis kets can be labeled $|l\, s\, m_l\, m_s\rangle$, where we have suppressed one or more degeneracy labels (which might correspond to the radial wavefunction). Then, also suppressing the labels $l, s$, the product basis states are

$$|m_l, \uparrow\rangle \qquad \text{and} \qquad |m_l, \downarrow\rangle\,, \tag{11.3.2}$$

where $m_s = 1/2$ and $-1/2$ are represented by ↑ and ↓, respectively. Using $\vec{J} = \vec{L} + \vec{S}$, we have another choice of compatible operators, $(L^2, S^2, J^2, J_z)$, associated to the total angular momentum orthobasis eigenkets

$$|l\, s\, j\, m\rangle = |j,\, m\rangle\,. \tag{11.3.3}$$

Our goal is to express the kets of eq. (11.3.3) as linear combinations of the product basis kets (11.3.2), for each $l$.

If $l = 0$, then things are very easy; the product basis kets $|0, \uparrow\rangle$ and $|0, \downarrow\rangle$ are already eigenstates of $J^2$ with eigenvalue $3\hbar^2/4$, and of $J_z$ with eigenvalues $\pm\hbar/2$, respectively. Therefore, the total angular momentum basis kets have $j = 1/2$ and $m = \pm 1/2$, and are $|\frac{1}{2}, \frac{1}{2}\rangle = |0, \uparrow\rangle$ and $|\frac{1}{2}, -\frac{1}{2}\rangle = |0, \downarrow\rangle$.

For $l \neq 0$, we begin with some preliminary counting, in order to know what to expect. There are $(2l+1)2$ product basis kets $|m_l, \uparrow\rangle$ and $|m_l, \downarrow\rangle$. They are all eigenkets of $J_z = L_z + S_z$, and the largest eigenvalue of $J_z$ is $\hbar m = \hbar(l + 1/2)$. This implies that there must be, in the total angular momentum basis, a multiplet with $j = l + 1/2$, which will have $2(l + 1/2) + 1 = 2l + 2$ basis elements. Exactly one of those will have $m = l - 1/2$, but we also know that there are two linearly independent states with that eigenvalue in the product basis, namely $|l, \downarrow\rangle$ and $|l-1, \uparrow\rangle$. Therefore, there must also be a multiplet with $j = l-1/2$, which has $2(l-1/2)+1 = 2l$ orthobasis members. Since we have accounted for all $4l + 2$ linearly independent kets, we have established that the tensor product of angular momentum $l$ with angular momentum $1/2$ must consist of states with total angular momenta $j = l+1/2$ and $l-1/2$, and no others. In notation similar to eq. (11.2.20), this is expressed as

$$l \otimes \frac{1}{2} \,=\, (l - 1/2) \oplus (l + 1/2). \tag{11.3.4}$$

Unlike eq. (11.2.20), there are no $A$ or $S$ subscripts here, because one cannot define antisymmetrization or symmetrization with respect to angular momenta that are are not the same.

To construct a complete map between the total and product orthobases, let us start with the state $|m_l, m_s\rangle = |l, \uparrow\rangle$. Acting on this with $J_z$ gives

$$J_z\, |l, \uparrow\rangle \,=\, (L_z + S_z)\, |l, \uparrow\rangle \,=\, \hbar(l + 1/2)\, |l, \uparrow\rangle\,, \tag{11.3.5}$$

and this is clearly the unique state with the largest possible $J_z$ eigenvalue. Since this is an eigenstate of $J^2$ and $J_z$ with quantum numbers $j = l+1/2$ and $m = l+1/2$, up to a phase that we can set arbitrarily,

$$|l{+}1/2,\, l{+}1/2\rangle \,=\, |l, \uparrow\rangle\,. \tag{11.3.6}$$

Now we can use this to construct all of the other states with $j = l + 1/2$, by acting repeatedly with the lowering operator $J_- = L_- + S_-$. From eq. (8.1.24),

$$J_-\, |l{+}1/2,\, l{+}1/2\rangle \,=\, \hbar\sqrt{2l+1}\, |l{+}1/2,\, l - 1/2\rangle\,, \tag{11.3.7}$$

so we get

$$|l{+}1/2,\, l{-}1/2\rangle \;=\; \frac{1}{\hbar\sqrt{2l+1}}(S_- + L_-)\, |l, \uparrow\rangle\,, \tag{11.3.8}$$

or, using eq. (8.1.24) again to evaluate the action of each of $S_-$ and $L_-$,

$$|l{+}1/2,\, l{-}1/2\rangle \;=\; \frac{1}{\sqrt{2l+1}}\Big(|l, \downarrow\rangle + \sqrt{2l}\, |l{-}1, \uparrow\rangle\Big). \tag{11.3.9}$$

This is our second total angular momentum orthobasis eigenket. Acting with $J_-$ on it in a similar way gives

$$|l{+}1/2,\, l{-}3/2\rangle \;=\; \frac{1}{\sqrt{2l+1}}\Big(\sqrt{2}\, |l{-}1,\, \downarrow\rangle + \sqrt{2l-1}\, |l{-}2,\, \uparrow\rangle\Big). \tag{11.3.10}$$

Using the same strategy, by induction we obtain all of the $j = l + 1/2$ kets,

$$|l{+}1/2,\, m\rangle \;=\; \frac{1}{\sqrt{2l+1}}\Big(\sqrt{l{-}m{+}1/2}\,|m{+}1/2,\,\downarrow\rangle + \sqrt{l{+}m{+}1/2}\,|m{-}1/2,\,\uparrow\rangle\Big) \qquad (11.3.11)$$

for all $m = -l{-}1/2, \ldots, l{+}1/2$.

We are done finding the states with total angular momentum $j = l{+}1/2$, but now we must find the states with $j = l{-}1/2$. Each state with $j = l{-}1/2$ must be a linear combination of the form

$$|l{-}1/2,\, m\rangle \;=\; a\,|m{+}1/2, \downarrow\rangle + b\,|m{-}1/2, \uparrow\rangle\,. \qquad (11.3.12)$$

We know this because the product orthobasis kets on the right side are the only two that have the correct eigenvalue $m$ of $J_z/\hbar$. Now, we can appeal to Theorem 2.6.5, which assures us that, since $J^2$ is Hermitian, the state $|l{-}1/2,\, m\rangle$ must be orthogonal to the state $|l{+}1/2,\, m\rangle$. Carrying out the inner product of eq. (11.3.11) with eq. (11.3.12), we learn that

$$a\sqrt{l - m + 1/2} \;\;=\;\; -b\sqrt{l + m + 1/2}. \qquad (11.3.13)$$

As an arbitrary phase convention, we choose $a$ real and positive. Requiring unit normalization of the ket then uniquely determines that for $j = l - 1/2$,

$$|l{-}1/2,\, m\rangle \;=\; \frac{1}{\sqrt{2l+1}}\Big(\sqrt{l{+}m{+}1/2}\,|m{+}1/2,\,\downarrow\rangle - \sqrt{l{-}m{+}1/2}\,|m{-}1/2,\,\uparrow\rangle\Big), \qquad (11.3.14)$$

for $m = -l{+}1/2, \ldots, l{-}1/2$.

That concludes our task. To recap, in eqs. (11.3.11) and (11.3.14), we have obtained the $J^2, J_z$ orthobasis kets as linear combinations of the product orthobasis kets.

As the simplest non-trivial concrete example, which we will have occasion to use later (at the end of section 15.3), consider the addition of an angular momentum $l = 1$ to a spin-1/2:

$$1 \otimes \frac{1}{2} \;\;=\;\; \frac{3}{2} \oplus \frac{1}{2}. \qquad (11.3.15)$$

Applying eq. (11.3.11), we have the $j = 3/2$ total angular momentum basis states

$$\big|\tfrac{3}{2}, \tfrac{3}{2}\big\rangle \;\;=\;\; |1, \uparrow\rangle\,, \qquad (11.3.16)$$

$$\big|\tfrac{3}{2}, \tfrac{1}{2}\big\rangle \;\;=\;\; \sqrt{\tfrac{1}{3}}\;|1, \downarrow\rangle + \sqrt{\tfrac{2}{3}}\;|0, \uparrow\rangle\,, \qquad (11.3.17)$$

$$\big|\tfrac{3}{2}, -\tfrac{1}{2}\big\rangle \;\;=\;\; \sqrt{\tfrac{2}{3}}\;|0, \downarrow\rangle + \sqrt{\tfrac{1}{3}}\;|{-1}, \uparrow\rangle\,, \qquad (11.3.18)$$

$$\big|\tfrac{3}{2}, -\tfrac{3}{2}\big\rangle \;\;=\;\; |{-1}, \downarrow\rangle\,, \qquad (11.3.19)$$

and applying eq. (11.3.14) we get the $j = 1/2$ total angular momentum basis states

$$\big|\tfrac{1}{2}, \tfrac{1}{2}\big\rangle \;\;=\;\; \sqrt{\tfrac{2}{3}}\;|1, \downarrow\rangle - \sqrt{\tfrac{1}{3}}\;|0, \uparrow\rangle\,, \qquad (11.3.20)$$

$$\big|\tfrac{1}{2}, -\tfrac{1}{2}\big\rangle \;\;=\;\; \sqrt{\tfrac{1}{3}}\;|0, \downarrow\rangle - \sqrt{\tfrac{2}{3}}\;|{-1}, \uparrow\rangle\,. \qquad (11.3.21)$$

## 11.4 The general case and Clebsch–Gordan coefficients

Now consider the general case of addition of angular momenta $\vec{J}_1$ and $\vec{J}_2$. Let us apply reasoning similar to the counting that led to eq. (11.3.4). The largest eigenvalue of $J_z = J_{1z} + J_{2z}$ is $\hbar(j_1 + j_2)$, and there is only one such state, $|j_1 : j_2\rangle$, so there must be exactly one total angular momentum multiplet with $j = j_1 + j_2$. (For the remainder of this section, product angular momentum basis kets will always be distinguished by use of a colon.) The subspace with $J_z$ eigenvalue $\hbar(j_1 + j_2 - 1)$ has dimension 2, spanned by the two states in the product basis $|j_1 - 1 : j_2\rangle$ and $|j_1 : j_2 - 1\rangle$. One linear combination of these is found in the $j = j_1+j_2$ multiplet, so there must also be exactly one total angular momentum multiplet with $j = j_1 + j_2 - 1$. Similarly, the subspace with $J_z$ eigenvalue $\hbar(j_1 + j_2 - 2)$ has dimension 3, spanned by the product basis kets $|j_1 - 2 : j_2\rangle$ and $|j_1 - 1 : j_2 - 1\rangle$ and $|j_1 : j_2 - 2\rangle$. Two linear combinations of these will occur in the $j = j_1 + j_2$ and $j = j_1 + j_2 - 1$ multiplets that we already know about, so there must also be exactly one total angular momentum multiplet with $j = j_1 + j_2 - 2$.

Continuing in this way, one finds that the tensor product of a multiplet with $J_1^2$ eigenvalue $\hbar^2 j_1(j_1 + 1)$ and a multiplet with $J_2^2$ eigenvalue $\hbar^2 j_2(j_2 + 1)$ must consist of a sum of multiplets with $J^2 = \hbar^2 j(j+1)$, with $j$ taking on the values from $|j_1-j_2|$ to $j_1+j_2$, with integer increments,

$$j_1 \otimes j_2 \;=\; |j_1 - j_2| \oplus \cdots \oplus (j_1 + j_2). \tag{11.4.1}$$

Each of the total angular momentum basis kets $|j_1\, j_2\, j, m\rangle$ is a linear combination of the product basis kets $|j_1\, j_2\, m_1 \!:\! m_2\rangle$. By the completeness of the latter, we can write

$$|j_1\, j_2\, j, m\rangle \;=\; \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1\, j_2\, m_1 \!:\! m_2\rangle\; \langle j_1\, j_2\, m_1 \!:\! m_2 | j_1\, j_2\, j, m\rangle \tag{11.4.2}$$

$$\;=\; \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1\, j_2\, m_1 \!:\! m_2\rangle\; C^{j_1 j_2 j}_{m_1 m_2 m} \tag{11.4.3}$$

where the inner products

$$C^{j_1 j_2 j}_{m_1 m_2 m} \;\equiv\; \langle j_1\, j_2\, m_1 \!:\! m_2 | j_1\, j_2\, j, m\rangle \tag{11.4.4}$$

are known as **Clebsch–Gordan coefficients**, after mathematicians Alfred Clebsch and Paul Gordan. Various different notations and conventions for them appear in the literature. We will usually use the $C$ notation for these inner products, as a way of saving space, with commas inserted between the superscripts or subscripts when it helps to make the meaning clear.

The Clebsch–Gordan coefficients obey selection rules; they can be non-zero only if

$$|j_1 - j_2| \leq j \leq j_1 + j_2, \quad \text{(triangle condition)}, \tag{11.4.5}$$

$$j_1 + j_2 - j \text{ is an integer}, \tag{11.4.6}$$

$$m = m_1 + m_2. \tag{11.4.7}$$

The first two of these simply restate eq. (11.4.1), and the last follows from evaluating matrix elements of $J_z = J_{1z} + J_{2z}$. In general, the Clebsch–Gordan coefficients are subject to phase ambiguities, which are resolved here by adopting a phase convention that $C^{j_1 j_2 j}_{m_1 m_2 m}$ is real and positive for $j_1 \geq j_2$ when $m = j$ and $m_1 = j_1$. (In the case that $j_1 = j_2$, one should just arbitrarily chose one of the two equal angular momenta for the label 1.) Then, because the actions of $J_-$ and $J_{1-}$ and $J_{2-}$ only involve real coefficients, as given by eq. (8.1.24), we find the useful result that all of the Clebsch–Gordan coefficients will be real. If one simultaneously flips the signs of $m_1$, $m_2$, and $m$, then they satisfy

$$C^{j_1 j_2 j}_{m_1 m_2 m} = (-1)^{j_1+j_2-j}\, C^{j_1 j_2 j}_{-m_1,-m_2,-m} \tag{11.4.8}$$

in our phase convention.

One can also use completeness with respect to the orthobasis elements $|j_1\, j_2\, j, m\rangle$, to obtain the inverse relation to eq. (11.4.3), which gives each product orthobasis ket as a linear combination of the total angular momentum orthobasis kets,

$$\begin{aligned} |j_1\, j_2\, m_1 \!:\! m_2\rangle &= \sum_{j=|j_1-j_2|}^{j_1+j_2} \sum_{m=-j}^{j} |j_1\, j_2\, j, m\rangle\; \langle j_1\, j_2\, j, m | j_1\, j_2\, m_1 \!:\! m_2\rangle && (11.4.9) \\ &= \sum_{j=|j_1-j_2|}^{j_1+j_2} \sum_{m=-j}^{j} |j_1\, j_2\, j, m\rangle\; C^{j_1 j_2 j}_{m_1 m_2 m}. && (11.4.10) \end{aligned}$$

The matrix elements on the right are actually the complex conjugates of the Clebsch–Gordan coefficients, but in our phase convention, all of them are real anyway.

As eqs. (11.4.3) and (11.4.10) demonstrate, results for total angular momentum orthobasis states in terms of product states, or vice versa, can always be expressed in terms of Clebsch–Gordan coefficients. For example, the content of the results for $1 \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2}$ in eqs. (11.3.16)-(11.3.21) can be expressed as

$$C^{1,\frac{1}{2},\frac{3}{2}}_{1,\frac{1}{2},\frac{3}{2}} = 1, \qquad C^{1,\frac{1}{2},\frac{3}{2}}_{1,-\frac{1}{2},\frac{1}{2}} = \sqrt{\frac{1}{3}}, \qquad C^{1,\frac{1}{2},\frac{3}{2}}_{0,\frac{1}{2},\frac{1}{2}} = \sqrt{\frac{2}{3}}, \tag{11.4.11}$$

$$C^{1,\frac{1}{2},\frac{1}{2}}_{1,-\frac{1}{2},\frac{1}{2}} = \sqrt{\frac{2}{3}}, \qquad C^{1,\frac{1}{2},\frac{1}{2}}_{0,\frac{1}{2},\frac{1}{2}} = -\sqrt{\frac{1}{3}}, \tag{11.4.12}$$

and others related to them by eq. (11.4.8). One can easily find published or online tables of Clebsch–Gordan coefficients, and software implementations of them, but it is a good idea to read the fine print to make sure of the phase convention being used.

In the remainder of this section, we will often suppress the labels $j_1, j_2$, which are the same for all of the kets. We will also assume that $j_1 \geq j_2$, in order to implement our phase convention. (Otherwise, one can simply reverse the roles of $j_1$ and $j_2$.)

To find the total angular momentum orthobasis states, and thus the Clebsch–Gordan coefficients, in a general case, we can use the following recipe. Start with the the state with the highest possible $m$, which is $m = j_1 + j_2 = j$. Since there is only one product orthobasis state with $J_{1z} + J_{2z}$ eigenvalue equal to $\hbar(j_1 + j_2)$, we have, using our phase convention choice,

$$|j_1{+}j_2,\, j_1{+}j_2\rangle \;=\; |j_1\,{:}\,j_2\rangle. \tag{11.4.13}$$

Now we follow the strategy of repeatedly acting with the lowering operator $J_- = J_{1-} + J_{2-}$ to find new states $|j_1 + j_2, m\rangle$. Using eq. (8.1.24) gives

$$J_-|j_1{+}j_2,\, j_1{+}j_2\rangle \;=\; \hbar\sqrt{2(j_1 + j_2)}|j_1{+}j_2,\, j_1{+}j_2{-}1\rangle, \tag{11.4.14}$$

which also can be evaluated as

$$J_{1-}|j_1\,{:}\,j_2\rangle + J_{2-}|j_1\,{:}\,j_2\rangle \;=\; \hbar\sqrt{2j_1}|j_1{-}1\,{:}\,j_2\rangle + \hbar\sqrt{2j_2}|j_1\,{:}\,j_2{-}1\rangle. \tag{11.4.15}$$

Therefore,

$$|j_1{+}j_2,\, j_1{+}j_2{-}1\rangle \;=\; \frac{1}{\sqrt{j_1 + j_2}}\Big(\sqrt{j_2}|j_1\,{:}\,j_2{-}1\rangle + \sqrt{j_1}|j_1{-}1\,{:}\,j_2\rangle\Big). \tag{11.4.16}$$

[The special cases $j_1 = j_2 = 1/2$ and $j_1 = l, j_2 = 1/2$ were previously found in eqs. (11.2.15) and (11.3.9), respectively.] Continuing to act with $J_-$, we similarly find all of the orthobasis kets $|j_1{+}j_2, m\rangle$. Eventually we will obtain $m = -j_1 - j_2$, which is then annihilated by $J_-$.

Next, we proceed to obtain the states that have total $j = j_1 + j_2 - 1$. We start by writing the most general possible expression for the state with the largest possible $m$,

$$|j_1{+}j_2{-}1,\, j_1{+}j_2{-}1\rangle \;=\; a|j_1\,{:}\,j_2{-}1\rangle + b|j_1{-}1\,{:}\,j_2\rangle. \tag{11.4.17}$$

The coefficients $a$ and $b$ can be uniquely identified by requiring that this ket is orthogonal to $|j_1{+}j_2,\, j_1{+}j_2{-}1\rangle$, that the ket is normalized so that $|a|^2 + |b|^2 = 1$, and that $a$ is real and positive according to our phase convention. This gives

$$|j_1{+}j_2{-}1,\, j_1{+}j_2{-}1\rangle \;=\; \frac{1}{\sqrt{j_1 + j_2}}\Big(\sqrt{j_1}|j_1\,{:}\,j_2{-}1\rangle - \sqrt{j_2}|j_1{-}1\,{:}\,j_2\rangle\Big). \tag{11.4.18}$$

[Again, we had previously derived this in the special cases $j_1 = j_2 = 1/2$ and $j_1 = l, j_2 = 1/2$, in eqs. (11.2.17) and (11.3.14), respectively.] Now, we again apply the operator $J_-$ repeatedly, to obtain all of the other states $|j_1{+}j_2{-}1,\, m\rangle$.

Continuing on our vaunted quest, we proceed to the states with $j = j_1 + j_2 - 2$, starting again with the maximum value of $m$. This must be some linear combination of the three product orthobasis kets with $m = m_1 + m_2 = j_1 + j_2 - 2$, which we can write as

$$|j_1{+}j_2{-}2,\, j_1{+}j_2{-}2\rangle \;=\; a|j_1\,{:}\,j_2{-}2\rangle + b|j_1{-}1\,{:}\,j_2{-}1\rangle + c|j_1{-}2\,{:}\,j_2\rangle. \tag{11.4.19}$$

The three unknown coefficients $a$, $b$, $c$ can be obtained by demanding that this be orthogonal to both of the kets $|j_1{+}j_2,\, j_1{+}j_2{-}2\rangle$ and $|j_1{+}j_2{-}1,\, j_1{+}j_2{-}2\rangle$, as required by Theorem 2.6.5 (since all three are eigenvectors of $J^2$ with different eigenvalues), and that the ket is normalized, so $|a|^2 + |b|^2 + |c|^2 = 1$, and that $a$ is real and positive according to our phase convention. The remaining states $|j_1{+}j_2{-}2,\, m\rangle$ are then obtained by acting repeatedly with $J_-$.

The preceding illustrates the general procedure, which is shown schematically in Figure 11.4.1. Each column represents the repeated action of $J_-$ on the state with the highest $m$ for a given $j$. After completing each column, we move to the next column by first writing a candidate ket $|j, j\rangle$ as a linear combination of the product orthobasis kets that have $m_1 + m_2 = m = j$. The coefficients are determined by requiring orthogonality to all of the previously obtained total angular momentum orthobasis kets with that value of $m$, and total angular momentum $j + 1$ or larger. Demanding that the ket be normalized and obey our phase convention, the state is then uniquely determined, and the rest of the states $|j,\, m\rangle$ follow by applying $J_-$ repeatedly. The process iterates by moving to the next column with $j$ lower by 1. Eventually, we will finish the last column of orthobasis kets with $j = |j_1 - j_2|$, and there will be none more.

The identities necessary and sufficient to carry out the calculations just described can be expressed directly in terms of the Clebsch–Gordan coefficients. These formulas will have other uses, too. First,

$$\begin{aligned}\sqrt{j(j+1) - m(m-1)}\, C^{j_1\, j_2\, j}_{m_1, m_2, m-1} \;&=\; \sqrt{j_1(j_1+1) - m_1(m_1+1)}\, C^{j_1\, j_2\, j}_{m_1+1, m_2, m}\\ &\quad + \sqrt{j_2(j_2+1) - m_2(m_2+1)}\, C^{j_1\, j_2\, j}_{m_1, m_2+1, m}, \qquad (11.4.20)\end{aligned}$$

which follows from evaluating

$$\langle m_1 \!:\! m_2 | J_- | j, m\rangle = \langle m_1 \!:\! m_2 |\, (J_{1-} + J_{2-})\, | j, m\rangle = (\langle j, m | (J_{1+} + J_{2+}) | m_1 \!:\! m_2\rangle)^* \quad (11.4.21)$$

using eqs. (8.1.22) and (8.1.24). From the reality of the Clebsch–Gordan coefficients in our chosen convention, and the total angular momentum orthonormality relations with product basis completeness, we also have

$$\sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} C^{j_1\, j_2\, j}_{m_1\, m_2\, m} C^{j_1\, j_2\, j'}_{m_1\, m_2\, m'} \;=\; \delta_{j,j'}\, \delta_{m,m'}. \qquad (11.4.22)$$

The Clebsch–Gordan coefficient results can alternatively be derived, or checked, using the recurrence relations that follow similarly from matrix elements of $J_+$,

$$\begin{aligned}\sqrt{j(j+1) - m(m+1)}\, C^{j_1\, j_2\, j}_{m_1, m_2, m+1} \;&=\; \sqrt{j_1(j_1+1) - m_1(m_1-1)}\, C^{j_1\, j_2\, j}_{m_1-1, m_2, m}\\ &\quad + \sqrt{j_2(j_2+1) - m_2(m_2-1)}\, C^{j_1\, j_2\, j}_{m_1, m_2-1, m}, \qquad (11.4.23)\end{aligned}$$

$$\begin{tikzcd}
\text{Start} \arrow[r] & {|j_1+j_2, j_1+j_2\rangle} \arrow[d, "J_-"] & & & \\
& {|j_1+j_2, j_1+j_2-1\rangle} \arrow[d, "J_-"] \arrow[r, "\text{orthog.}"] & {|j_1+j_2-1, j_1+j_2-1\rangle} \arrow[d, "J_-"] & & \\
& {|j_1+j_2, j_1+j_2-2\rangle} \arrow[d, "J_-"] & {|j_1+j_2-1, j_1+j_2-2\rangle} \arrow[d, "J_-"] \arrow[r, "\text{orthog.}"] & {|j_1+j_2-2, j_1+j_2-2\rangle} \arrow[d, "J_-"] & \\
& \vdots \arrow[d, "J_-"] & \vdots \arrow[d, "J_-"] & \vdots \arrow[d, "J_-"] & \cdots \\
& {|j_1+j_2, -j_1-j_2+2\rangle} \arrow[d, "J_-"] & {|j_1+j_2-1, -j_1-j_2+2\rangle} \arrow[d, "J_-"] & {|j_1+j_2-2, -j_1-j_2+2\rangle} & \\
& {|j_1+j_2, -j_1-j_2+1\rangle} \arrow[d, "J_-"] & {|j_1+j_2-1, -j_1-j_2+1\rangle} & & \\
& {|j_1+j_2, -j_1-j_2\rangle} & & &
\end{tikzcd}$$

Figure 11.4.1: Schematic of a plan to compute the total angular momentum orthobasis kets $|j_1\, j_2\, j, m\rangle$, abbreviated here as $|j, m\rangle$, in terms of product orthobasis kets, when two angular momenta $j_1$ and $j_2$ are combined. The process starts at the upper left with $j = m = j_1 + j_2$. Each column has fixed $j$, and is constructed from top to bottom using $J_-$. Then the next column is started by constructing the state with maximum $m = j$, by requiring it to be orthogonal to all of the previously found kets that have that same $m$. The process ends after the rightmost column with $j = |j_1 - j_2|$ is finished, with $m = -|j_1 - j_2|$. The results of this process provide the Clebsch–Gordan coefficients $\langle m_1 \!:\! m_2 | j, m\rangle = \langle j_1\, j_2\, m_1 \!:\! m_2 | j_1\, j_2\, j, m\rangle = C^{j_1\, j_2\, j}_{m_1\, m_2\, m}$.

and the product basis orthonormality relations

$$\sum_{j=|j_1-j_2|}^{j_1+j_2} \sum_{m=-j}^{j} C^{j_1\, j_2\, j}_{m_1\, m_2\, m} C^{j_1\, j_2\, j}_{m_1'\, m_2'\, m} \;=\; \delta_{m_1,m_1'}\, \delta_{m_2,m_2'}, \tag{11.4.24}$$

which follow from completeness of the total angular momentum orthobasis.

Besides the examples already done in the previous sections, consider the special case of adding two angular momenta $j_1 = 1$ and $j_2 = 1$. The resulting allowed values of $j$ are 0, 1, and 2. By following the procedure summarized in Figure 11.4.1, one finds the $j = 2$ states

$$|2,2\rangle \;=\; |1\!:\!1\rangle\,, \tag{11.4.25}$$

$$|2,1\rangle \;=\; \frac{1}{\sqrt{2}}\big(|1\!:\!0\rangle + |0\!:\!1\rangle\big), \tag{11.4.26}$$

$$|2,0\rangle \;=\; \frac{1}{\sqrt{6}}\big(|1\!:\!-1\rangle + 2\,|0\!:\!0\rangle + |-1\!:\!1\rangle\big), \tag{11.4.27}$$

$$|2,-1\rangle \;=\; \frac{1}{\sqrt{2}}\big(|0\!:\!-1\rangle + |-1\!:\!0\rangle\big), \tag{11.4.28}$$

$$|2,-2\rangle \;=\; |-1\!:\!-1\rangle\,, \tag{11.4.29}$$

which are all symmetric under $m_1 \leftrightarrow m_2$, followed by the $j = 1$ states

$$|1,1\rangle \;=\; \frac{1}{\sqrt{2}}\big(|1\!:\!0\rangle - |0\!:\!1\rangle\big), \tag{11.4.30}$$

$$|1,0\rangle \;=\; \frac{1}{\sqrt{2}}\big(|1\!:\!-1\rangle - |-1\!:\!1\rangle\big), \tag{11.4.31}$$

$$|1,-1\rangle \;=\; \frac{1}{\sqrt{2}}\big(|0\!:\!-1\rangle - |-1\!:\!0\rangle\big), \tag{11.4.32}$$

which are each antisymmetric under the same exchange, and finally the $j = 0$ state

$$|0,0\rangle \;=\; \frac{1}{\sqrt{3}}\big(|1\!:\!-1\rangle - |0\!:\!0\rangle + |-1\!:\!1\rangle\big), \tag{11.4.33}$$

which is symmetric. To summarize,

$$1 \otimes 1 \;=\; 2_S \oplus 1_A \oplus 0_S, \tag{11.4.34}$$

where the $S$ and $A$ subscripts indicate the symmetry or antisymmetry of the total angular momentum orthobasis kets under exchange of the two product basis eigenvalues $m_1$ and $m_2$.

More generally, for the combination of two equal angular momenta $j_1 = j_2 = j$, the symmetry and antisymmetry properties for exchange of $m_1$ and $m_2$ are summarized by

$$j \otimes j \;=\; (2j)_S \oplus (2j-1)_A \oplus (2j-2)_S \oplus \cdots \oplus 0, \tag{11.4.35}$$

with alternating $S$ and $A$, so that the singlet on the right is symmetric if $j$ is an integer and antisymmetric if $j$ is a half-integer.

For example, if $j_1 = j_2 = 3/2$, and we call the resulting total angular momentum quantum number $J$, then you can show that the $J = 3, m_J = 3$ and $J = 3, m_J = 0$ and $J = 0, m_J = 0$ states are, in terms of product basis states $|m_1 : m_2\rangle$,

$$|3,3\rangle = |\tfrac{3}{2} : \tfrac{3}{2}\rangle, \tag{11.4.36}$$

$$|3,0\rangle = \frac{1}{\sqrt{20}}\Big(|\tfrac{3}{2} : -\tfrac{3}{2}\rangle + |-\tfrac{3}{2} : \tfrac{3}{2}\rangle + 3|\tfrac{1}{2} : -\tfrac{1}{2}\rangle + 3|-\tfrac{1}{2} : \tfrac{1}{2}\rangle\Big), \tag{11.4.37}$$

$$|0,0\rangle = \frac{1}{2}\Big(|\tfrac{3}{2} : -\tfrac{3}{2}\rangle - |-\tfrac{3}{2} : \tfrac{3}{2}\rangle - |\tfrac{1}{2} : -\tfrac{1}{2}\rangle + |-\tfrac{1}{2} : \tfrac{1}{2}\rangle\Big), \tag{11.4.38}$$

respectively. It is left as an exercise to work out the other total angular momentum orthobasis states $|J, m_J\rangle$ in this example.

One can also combine three (or more) angular momenta, to obtain an orthobasis of kets that are eigenstates of the observables $J^2$ and $J_z$ for the total angular momentum $\vec{J} = \vec{J}_1 + \vec{J}_2 + \vec{J}_3$. This can be done recursively, by first combining $\vec{J}_1$ and $\vec{J}_2$ into an angular momentum operator $\vec{J}_{12}$, and then combining the results with $\vec{J}_3$. For example, combining three spin-1/2 angular momenta results in

$$\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = (1 \oplus 0) \otimes \frac{1}{2} = (1 \otimes \frac{1}{2}) \oplus (0 \otimes \frac{1}{2}) = (3/2)_S \oplus (1/2)_a \oplus (1/2)_b. \tag{11.4.39}$$

The subscript $S$ on the $j = 3/2$ multiplet indicates that the states in it are symmetric under exchange of any two of the three individual spins, as we will soon check. There are also two $j = 1/2$ multiplets in the result, distinguished by degeneracy labels $a$ and $b$. To check the multiplicities of states, we note that on the left side of eq. (11.4.39) there are $2^3 = 8$ product orthobasis states, since each individual spin has 2 values of $m_s = \pm 1/2$. On the right, the multiplicities of the total angular momentum states are $4+2+2 = 8$. To explicitly construct the total angular momentum orthobasis kets, one can first use eqs. (11.2.14)–(11.2.17) to construct the $J_{12}^2, J_{12z}$ eigenstate kets. Combining the resulting $j_{12} = 1$ states with the third spin, one finds, by using eq. (11.3.11), that the $j = 3/2$ states are

$$|\tfrac{3}{2}, \tfrac{3}{2}\rangle = |\uparrow\uparrow\uparrow\rangle\,, \tag{11.4.40}$$

$$|\tfrac{3}{2}, \tfrac{1}{2}\rangle = \frac{1}{\sqrt{3}}\big(|\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\downarrow\uparrow\uparrow\rangle\big), \tag{11.4.41}$$

$$|\tfrac{3}{2}, -\tfrac{1}{2}\rangle = \frac{1}{\sqrt{3}}\big(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle\big), \tag{11.4.42}$$

$$|\tfrac{3}{2}, -\tfrac{3}{2}\rangle = |\downarrow\downarrow\downarrow\rangle\,, \tag{11.4.43}$$

and, by using eq. (11.3.14), the $j = 1/2$ states

$$|\tfrac{1}{2}, \tfrac{1}{2}, a\rangle = \frac{1}{\sqrt{6}}\big(2\,|\uparrow\uparrow\downarrow\rangle - |\uparrow\downarrow\uparrow\rangle - |\downarrow\uparrow\uparrow\rangle\big), \tag{11.4.44}$$

$$|\tfrac{1}{2}, -\tfrac{1}{2}, a\rangle = \frac{1}{\sqrt{6}}\big(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle - 2\,|\downarrow\downarrow\uparrow\rangle\big). \tag{11.4.45}$$

From combining the $j_{12} = 0$ states from eq. (11.2.17) with the third spin, we have the additional $j = 1/2$ total angular momentum basis states

$$|\tfrac{1}{2}, \tfrac{1}{2}, b\rangle \;=\; \frac{1}{\sqrt{2}}\big(|{\uparrow\downarrow\uparrow}\rangle - |{\downarrow\uparrow\uparrow}\rangle\big), \tag{11.4.46}$$

$$|\tfrac{1}{2}, -\tfrac{1}{2}, b\rangle \;=\; \frac{1}{\sqrt{2}}\big(|{\uparrow\downarrow\downarrow}\rangle - |{\downarrow\uparrow\downarrow}\rangle\big). \tag{11.4.47}$$

The explicit forms for the $j = 3/2$ states show that they are indeed each symmetric under exchange of any two of the three individual spins. For the $j = 1/2$ states, the ones labeled $a$ are symmetric, and the ones labeled $b$ are antisymmetric, under exchange of the first two spins. However, these $j = 1/2$ states do not have a complete symmetry or antisymmetry under exchange of every pair of spins. Also note that there is no totally antisymmetric combination of three or more spin 1/2 states, simply because in the product basis there are only two values $\uparrow$ and $\downarrow$ available, so that every product orthobasis ket is symmetric under interchange of at least one pair of spins.

Similarly, combining four spin-1/2 angular momenta gives

$$\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} \;=\; (1 \oplus 0) \otimes (1 \oplus 0) \;=\; 2_S \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 0. \tag{11.4.48}$$

Here the counting is that the number of product orthobasis states on the left is $2^4 = 16$, while the counting of multiplicities of total angular momentum orthobasis states on the right is $5 + 3 + 3 + 3 + 1 + 1 = 16$. Only the total spin 2 multiplet is totally symmetric, and there is no totally antisymmetric multiplet. The other multiplets on the right side of eq. (11.4.48) have mixed symmetry properties under exchange of the spins.

In the case of $N$ electrons in an atom, the angular momentum eigenstates are often given in spectroscopic notation, defined as follows. First, combine all of the individual spin operators $\vec{S}_i$ to obtain the total spin angular momentum operator $\vec{S} = \sum_i \vec{S}_i$. The eigenvalues of the operator $S^2 = \vec{S} \cdot \vec{S}$ are then denoted by $\hbar^2 S(S+1)$ where $S$ is a number.[†] Clearly, if $N$ is even, then the number $S$ must be an integer with $0 \leq S \leq N/2$. If $N$ is odd, then the number $S$ must be half-integer, with $1/2 \leq S \leq N/2$. Next, combine all of the individual orbital angular momenta operators $\vec{L}_i$ to obtain the total orbital angular momentum operator $\vec{L} = \sum_i \vec{L}_i$. The eigenvalues of the operator $L^2 = \vec{L} \cdot \vec{L}$ are likewise denoted $\hbar^2 L(L + 1)$, where the quantum number $L$ is always an integer. Finally, the operators $\vec{S}$ and $\vec{L}$ are combined to form the total angular momentum operator $\vec{J}$, and the operators $J^2$ and $J_z$ have eigenvalues denoted $\hbar^2 J(J+1)$ and $\hbar m_J$, respectively. The traditional notation for a multiplet of common eigenstates of the

---

[†] Following a common but potentially confusing notation, in the case of combined angular momenta, capital letters are often used both for the names of the operators as well as the corresponding quantum numbers.

observables $S^2$, $L^2$, and $J^2$ is then

$$^{2S+1}L_J, \tag{11.4.49}$$

where $S$, $L$, and $J$ are the quantum numbers, but with $L$ replaced by the capital letter code $S$, $P$, $D$, $F$, $G, \ldots$, according to whether the number $L$ is $0, 1, 2, 3, 4, \ldots$, as indicated in eq. (10.1.20). The degeneracy, or multiplicity, of each group of states denoted by $^{2S+1}L_J$ is $2J + 1$, since the eigenvalue $m_J$ can take on the values $-J$, $-J+1, \ldots, J-1, J$.

For example, a single electron always has $S = 1/2$. For $L = 0$, it has $J = 1/2$, and for $L \geq 1$ it can have $J = L \pm 1/2$, as we saw in section 11.3. So, the list of total angular momentum multiplets for a single electron in an atom is

$$^2S_{1/2},\ ^2P_{1/2},\ ^2P_{3/2},\ ^2D_{3/2},\ ^2D_{5/2},\ ^2F_{5/2},\ ^2F_{7/2},\ ^2G_{7/2},\ ^2G_{9/2}, \ldots. \tag{11.4.50}$$

For two electrons, the possible total spin quantum numbers are $S = 0$ and $1$, as we saw in section 11.2. The orbital angular momenta $\vec{L}_1$ and $\vec{L}_2$ with quantum numbers $l_1$ and $l_2$ can be combined into $L = |l_1 - l_2|, \ldots, l_1 + l_2$. Therefore, the possible total angular momentum eigenstates for two electrons are

$$^1S_0,\ ^1P_1,\ ^1D_2,\ ^1F_3,\ ^1G_4, \ldots \tag{11.4.51}$$

for $S = 0$, and

$$^3S_1,\ ^3P_0,\ ^3P_1,\ ^3P_2,\ ^3D_1,\ ^3D_2,\ ^3D_3,\ ^3F_2,\ ^3F_3,\ ^3F_4,\ \ldots \tag{11.4.52}$$

for $S = 1$. However, as we will discuss in more detail in section 16.1, Fermi–Dirac statistics requires that the total state must be antisymmetric under exchange of the two electrons. In the special case that the electrons have the same radial wavefunction and the same orbital angular momentum quantum number $l_1 = l_2$, then one can use eq. (11.4.35) to see that if $L$ is odd (an antisymmetric position wavefunction) then only the symmetric spin combination $S = 1$ is allowed, and if $L$ is even (a symmetric position wavefunction), then only the antisymmetric spin combination $S = 0$ is allowed.

## 11.5 How spherical harmonics combine

The spherical harmonics are the position representations of the orbital angular momentum eigenstates of $L^2$ and $L_z$. In this section, we will learn how the addition of angular momentum technology is realized in terms of the spherical harmonics.

Let us start by considering fixed $l_1$, $l_2$, $l$, and $m$. We then define the following function of $\theta, \phi$, as a linear combination of products of spherical harmonics weighted by the Clebsch–Gordan coefficients defined by eq. (11.4.4),

$$\Psi_l^m(\theta,\phi) \;\equiv\; \sum_{m_1,m_2} Y_{l_1}^{m_1}(\theta,\phi)\, Y_{l_2}^{m_2}(\theta,\phi)\, C^{l_1 l_2 l}_{m_1 m_2 m}, \tag{11.5.1}$$

where the sums are over all $m_1$ and $m_2$ such that the Clebsch–Gordan coefficient does not vanish. Our goal is to find another simple expression for $\Psi_l^m(\theta,\phi)$. We claim that, as its labeling suggests, this wavefunction is an eigenstate of the differential operators $L^2$ and $L_z$, with eigenvalues $\hbar^2 l(l+1)$ and $\hbar m$, respectively, and it is therefore proportional to $Y_l^m(\theta,\phi)$.

To prove this claim, we will check the action of the differential operators $L_z$, $L_+$, and $L_-$ on it. First, by applying eq. (8.6.27), we find

$$L_z \Psi_l^m \;=\; \sum_{m_1,m_2} (\hbar m_1 + \hbar m_2)\, Y_{l_1}^{m_1}\, Y_{l_2}^{m_2}\, C^{l_1 l_2 l}_{m_1 m_2 m}. \tag{11.5.2}$$

The Clebsch–Gordan coefficient enforces that all non-zero contributions have $m_1 + m_2 = m$, so

$$L_z \Psi_l^m \;=\; \hbar m \Psi_l^m. \tag{11.5.3}$$

Next, applying eq. (8.6.29) we find

$$\begin{aligned} L_- \Psi_l^m \;=\; \hbar \sum_{m_1,m_2} \Big[ &\sqrt{l_1(l_1+1) - m_1(m_1-1)}\, Y_{l_1}^{m_1-1}\, Y_{l_2}^{m_2} \\ &+ \sqrt{l_2(l_2+1) - m_2(m_2-1)}\, Y_{l_1}^{m_1}\, Y_{l_2}^{m_2-1} \Big]\, C^{l_1 l_2 l}_{m_1 m_2 m}. \end{aligned} \tag{11.5.4}$$

Since we are summing over all $m_1$ and $m_2$, we can use the trick of relabeling $m_1 \to m_1 + 1$ in the first term and $m_2 \to m_2 + 1$ in the second term, to obtain

$$\begin{aligned} L_- \Psi_l^m \;=\; \hbar \sum_{m_1,m_2} \Big[ &\sqrt{l_1(l_1+1) - m_1(m_1+1)}\, C^{l_1 l_2 l}_{m_1+1, m_2, m} \\ &+ \sqrt{l_2(l_2+1) - m_2(m_2+1)}\, C^{l_1 l_2 l}_{m_1, m_2+1, m} \Big]\, Y_{l_1}^{m_1}\, Y_{l_2}^{m_2}. \end{aligned} \tag{11.5.5}$$

Now the Clebsch–Gordan recurrence relation eq. (11.4.20) turns this into

$$L_- \Psi_l^m \;=\; \hbar \sum_{m_1,m_2} \sqrt{l(l+1) - m(m-1)}\, C^{l_1 l_2 l}_{m_1, m_2, m-1}\, Y_{l_1}^{m_1}\, Y_{l_2}^{m_2}, \tag{11.5.6}$$

or, using the definition of eq. (11.5.1), simply,

$$L_- \Psi_l^m \;=\; \hbar \sqrt{l(l+1) - m(m-1)}\, \Psi_l^{m-1}. \tag{11.5.7}$$

In an exactly analogous way, one finds

$$L_+ \Psi_l^m \;=\; \hbar \sqrt{l(l+1) - m(m+1)}\, \Psi_l^{m+1}. \tag{11.5.8}$$

Taken together, eqs. (11.5.3), (11.5.7) and (11.5.8), combined with $L^2 = (L_+L_- + L_-L_+)/2 + L_z^2$ from eq. (8.1.15), show that this wavefunction also obeys

$$L^2 \Psi_l^m \;=\; \hbar^2 l(l+1)\Psi_l^m. \tag{11.5.9}$$

Since the spherical harmonic $Y_l^m$ is the unique $(\theta, \phi)$ function with $L^2$ eigenvalue $\hbar^2 l(l+1)$ and $L_z$ eigenvalue $\hbar m$, eqs. (11.5.3) and (11.5.9) show that $\Psi_l^m$ must be proportional to it, and

$$\Psi_l^m(\theta,\phi) \;=\; c_{l_1,l_2,l,m}\, Y_l^m(\theta,\phi) \tag{11.5.10}$$

as claimed, for some constant $c_{l_1,l_2,l,m}$ (independent of $\theta, \phi$).

Furthermore, by acting with $L_-$ on both sides of the previous equation, we obtain

$$L_-\Psi_l^m \;=\; c_{l_1,l_2,l,m}\, L_- Y_l^m \;=\; c_{l_1,l_2,l,m}\hbar\,\sqrt{l(l+1) - m(m-1)}\, Y_l^{m-1}, \tag{11.5.11}$$

but also from eq. (11.5.7) this is equal to

$$L_-\Psi_l^m \;=\; \hbar\sqrt{l(l+1) - m(m-1)}\, c_{l_1,l_2,l,m-1}\, Y_l^{m-1}. \tag{11.5.12}$$

Comparing these informs us that $c_{l_1,l_2,l,m} = c_{l_1,l_2,l,m-1}$ for all $-l+1 \leq m \leq l$, so the constant of proportionality does not actually depend on $m$, and we therefore drop that label.

All of the preceding just served to prove that we can write

$$\sum_{m_1,m_2} Y_{l_1}^{m_1}(\theta,\phi) Y_{l_2}^{m_2}(\theta,\phi) C_{m_1 m_2 m}^{l_1 l_2 l} \;=\; c_{l_1 l_2 l}\, Y_l^m(\theta,\phi) \tag{11.5.13}$$

for some proportionality constant $c_{l_1 l_2 l}$. Our remaining task is to identify this constant. Fortunately, there is a trick to do this very easily; just consider the special case $\theta = 0$. Using the fact found in eq. (8.6.54),

$$Y_l^m(0,\phi) \;=\; \delta_{m,0}\sqrt{\frac{2l+1}{4\pi}}, \tag{11.5.14}$$

the double sum in eq. (11.5.13) collapses to a single term with $m_1 = m_2 = 0$, and it reads

$$\sqrt{\frac{2l_1+1}{4\pi}}\sqrt{\frac{2l_2+1}{4\pi}}\, C_{000}^{l_1 l_2 l} \;=\; c_{l_1 l_2 l}\sqrt{\frac{2l+1}{4\pi}}. \tag{11.5.15}$$

Solving this for $c_{l_1 l_2 l}$, eq. (11.5.13) becomes

$$\sum_{m_1,m_2} Y_{l_1}^{m_1}(\theta,\phi) Y_{l_2}^{m_2}(\theta,\phi) C_{m_1 m_2 m}^{l_1 l_2 l} \;=\; \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l+1)}} C_{000}^{l_1 l_2 l} Y_l^m(\theta,\phi). \tag{11.5.16}$$

This accomplishes our goal.

To derive a related interesting and useful identity, start with eq. (11.5.16), multiply by $C^{l_1 l_2 l}_{m_1' m_2' m}$, then sum over all $l, m$, then apply the orthogonality of the Clebsch–Gordan coefficients of eq. (11.4.24) on the left side, and finally relabel $m_1' \rightarrow m_1$ and $m_2' \rightarrow m_2$. The result is

$$Y^{m_1}_{l_1}(\theta,\phi) Y^{m_2}_{l_2}(\theta,\phi) = \sum_{l=|l_1-l_2|}^{l_1+l_2} \sum_{m=-l}^{l} \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l+1)}} C^{l_1 l_2 l}_{000} C^{l_1 l_2 l}_{m_1 m_2 m} Y^m_l(\theta,\phi). \qquad (11.5.17)$$

The product of any two spherical harmonics is thus an appropriately weighted sum over the spherical harmonics selected by the corresponding addition of angular momenta.

Another closely related identity follows. Start from eq. (11.5.17), multiply by $Y^{m'}_{l'}(\theta,\phi)^*$, integrate over the angular coordinates, then apply the orthonormality of the spherical harmonics using eq. (8.6.31) on the right side, and finally relabel $l' \rightarrow l$ and $m' \rightarrow m$. The result is

$$\int d\Omega \big[Y^m_l(\theta,\phi)\big]^* Y^{m_1}_{l_1}(\theta,\phi) Y^{m_2}_{l_2}(\theta,\phi) = \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l+1)}} C^{l_1 l_2 l}_{000} C^{l_1 l_2 l}_{m_1 m_2 m}. \qquad (11.5.18)$$

This is the **Wigner–Eckart formula for spherical harmonics**. It is a special case of the general Wigner–Eckart theorem (a statement about matrix elements of operators, and applicable not just to orbital angular momentum) to be derived in section 12.3. Remarkably, the integral only depends on the magnetic quantum numbers $m_1$, $m_2$, and $m$ through the coefficient $C^{l_1 l_2 l}_{m_1 m_2 m}$, so the Clebsch–Gordan selection rules (11.4.5)–(11.4.7) also govern which integrals of this type can be non-zero. This formula can be used to evaluate the integral of any product of three spherical harmonics (with or without complex conjugates), simply by making use of eq. (8.6.52).

# 12 Tensor operators and rules they obey

## 12.1 Irreducible tensor operators

Consider a finite spatial rotation generated by a unitary operator, so that a state ket $|\psi\rangle$ is transformed to $|\psi'\rangle = U(\vec{\alpha})\,|\psi\rangle$, as discussed in section 5.3. Recall, from the discussion surrounding eqs. (2.5.25)–(2.5.30), that for any operator $A$ we can consider the corresponding rotated version defined by $A' = U(\vec{\alpha})AU(\vec{\alpha})^\dagger$, so that any matrix element $\langle\chi|A|\psi\rangle$ is equal to the matrix element of the rotated operator for the rotated states, $\langle\chi'|A'|\psi'\rangle$. We will now study operators that transform under rotations in a simple and useful way, specifically, as irreducible representations of the rotation group.

An **irreducible tensor operator** (also known as a **spherical tensor operator**) of order $j$ is a collection of $2j+1$ component operators labeled by an index $m$,

$$T^{(j)}_m, \qquad (m = -j, \ldots, j), \tag{12.1.1}$$

which, by definition, are required to obey the following transformation rule for every unitary rotation operator $U(\vec{\alpha})$,

$$U(\vec{\alpha})\, T^{(j)}_m\, U(\vec{\alpha})^\dagger \;=\; \sum_{m'=-j}^{j} T^{(j)}_{m'} D^{(j)}_{m'm}(\vec{\alpha}), \tag{12.1.2}$$

with the Wigner rotation matrices $D^{(j)}(\vec{\alpha})$ as defined in eq. (8.5.2). Using the form of $U(\vec{\alpha})$ for an infinitesimal rotation in terms of the total angular momentum operator $\vec{J}$ in eq. (8.5.1), and the actions of $J_z$, $J_+$, and $J_-$ on angular momentum eigenstate as given in eqs. (8.1.7), (8.1.22), and (8.1.24), which are reproduced here for convenience,

$$\begin{aligned}
J_z\,|j,m\rangle &= \hbar m\,|j,m\rangle\,, &(12.1.3)\\
J_+\,|j,m\rangle &= \hbar\sqrt{j(j+1)-m(m+1)}\;|j,m{+}1\rangle\,, &(12.1.4)\\
J_-\,|j,m\rangle &= \hbar\sqrt{j(j+1)-m(m-1)}\;|j,m{-}1\rangle\,, &(12.1.5)
\end{aligned}$$

it is left as an exercise to show that the components of an irreducible tensor operator must obey the commutation relations

$$\begin{aligned}
\big[J_z,\, T^{(j)}_m\big] &= \hbar m\, T^{(j)}_m, &(12.1.6)\\
\big[J_+,\, T^{(j)}_m\big] &= \hbar\sqrt{j(j+1)-m(m+1)}\, T^{(j)}_{m+1}, &(12.1.7)\\
\big[J_-,\, T^{(j)}_m\big] &= \hbar\sqrt{j(j+1)-m(m-1)}\, T^{(j)}_{m-1}. &(12.1.8)
\end{aligned}$$

The similarity of eqs. (12.1.6)–(12.1.8) to eqs. (12.1.3)–(12.1.5) suggests that the operators $T^{(j)}_m$ "carry" angular momentum as indicated by the quantum number labels $j$ and $m$. Indeed this intuition is valid, as we will explain in the following, with the most precise statement to be given in eq. (12.3.9).

The simplest type of irreducible tensor operator is obtained for $j = 0$ with a single component, and is called a **scalar operator**. A scalar operator $S$ obeys $U(\vec{\alpha})\, S\, U(\vec{\alpha})^\dagger = S$, or equivalently it commutes with the total angular momentum operators $J_z$, $J_+$, $J_-$, and therefore also with $J_x$ and $J_y$ and $J^2$. Consider a state $|\alpha, j, m\rangle$ that is an eigenstate of $J^2$ and $J_z$ with eigenvalues $\hbar^2 j(j+1)$ and $\hbar m$, with $\alpha$ a degeneracy label. It follows that the state $S\,|\alpha, j, m\rangle$ is also an eigenstate of $J^2$ and $J_z$ with the same eigenvalues. In that sense, the scalar operator $S$ carries no total angular momentum.

Irreducible tensor operators with $j = 1/2$, or more generally with half-integer $j$, are called **spinor operators**. However, these cannot be observables, because they are necessarily double-valued; under a rotation by $2\pi$ they acquire a minus sign, just as we saw for matrix representations of rotations of states with half-integer angular momentum in eqs. (8.5.14) and (8.5.17). Observables can be constructed by taking the product of an even number of spinor operators, using Theorem 12.1.1 below. We will not discuss them further here, although they do play a very important role in quantum field theories with fermions.

An irreducible tensor operator with order $j = 1$ and three components $m = -1, 0, 1$ is called a **vector operator** $V_m^{(1)}$. These can be used to define the familiar rectangular $x, y, z$ components of $\vec{V} = \hat{x}V_x + \hat{y}V_y + \hat{z}V_z$, as follows:

$$V_x = \big(V_{-1}^{(1)} - V_1^{(1)}\big)/\sqrt{2}, \tag{12.1.9}$$

$$V_y = i\big(V_{-1}^{(1)} + V_1^{(1)}\big)/\sqrt{2}, \tag{12.1.10}$$

$$V_z = V_0^{(1)}. \tag{12.1.11}$$

These can be shown, using eqs. (12.1.6)–(12.1.8) with $j = 1$, to obey the commutation relations

$$[J_a, V_b] = i\hbar\epsilon_{abc}V_c, \qquad (a, b, c = x, y, z). \tag{12.1.12}$$

For example, the rectangular coordinate position operators found in $\vec{R} = \hat{x}X + \hat{y}Y + \hat{z}Z$ can be expressed in terms of components of an irreducible tensor operator $R^{(1)}$, as

$$R_0^{(1)} = Z, \tag{12.1.13}$$

$$R_1^{(1)} = -(X + iY)/\sqrt{2}, \tag{12.1.14}$$

$$R_{-1}^{(1)} = (X - iY)/\sqrt{2}, \tag{12.1.15}$$

and it is left as an exercise to check that the requisite commutation relations (12.1.6)–(12.1.8) are indeed satisfied. Besides the position operator, other examples of vector operators are the momentum operator $\vec{P}$, the orbital angular momentum $\vec{L}$, the spin $\vec{S}$, and the total angular momentum operator itself, $\vec{J} = \vec{L} + \vec{S}$. Note that the angular momentum raising and lowering operators defined by eq. (8.1.8) are actually given in the irreducible tensor component formalism by $J_+ = -\sqrt{2}J_1^{(1)}$ and $J_- = \sqrt{2}J_{-1}^{(1)}$, respectively, while $J_z = J_0^{(1)}$.

Given any two irreducible tensor operators, one can multiply them to construct others. The way to do this is specified by the following theorem.

**Theorem 12.1.1 (Product rules for tensor operators)** *If $T^{(j_1)}$ and $S^{(j_2)}$ are irreducible tensor operators of orders $j_1$ and $j_2$ respectively, and $C^{j_1\,j_2\,j}_{m_1\,m_2\,m}$ are the Clebsch–Gordan coefficients defined in eq. (11.4.4), then*

$$W_m^{(j)} = \sum_{m_1}\sum_{m_2} T_{m_1}^{(j_1)} S_{m_2}^{(j_2)} C^{j_1\,j_2\,j}_{m_1\,m_2\,m} \qquad (m = -j, \ldots, j) \tag{12.1.16}$$

*is an irreducible tensor operator of order $j$. Furthermore,*

$$T_{m_1}^{(j_1)} S_{m_2}^{(j_2)} = \sum_j \sum_m W_m^{(j)} C^{j_1\,j_2\,j}_{m_1\,m_2\,m}. \tag{12.1.17}$$

The proof is left as an exercise.

As an example, consider two vector operators $V^{(1)}$ and $U^{(1)}$. Then, using the results of eqs. (11.4.25)–(11.4.33) to extract the necessary Clebsch–Gordan coefficients for $1\otimes 1 = 0\oplus 1\oplus 2$, we apply eq. (12.1.16) to construct the following three tensor operators, of orders 0, 1, and 2:

$$W_0^{(0)} = \left(V_1^{(1)}U_{-1}^{(1)} + V_{-1}^{(1)}U_1^{(1)} - V_0^{(1)}U_0^{(1)}\right)/\sqrt{3}, \tag{12.1.18}$$

and

$$W_{\pm 1}^{(1)} = \pm\left(V_{\pm 1}^{(1)}U_0^{(1)} - V_0^{(1)}U_{\pm 1}^{(1)}\right)/\sqrt{2}, \tag{12.1.19}$$

$$W_0^{(1)} = \left(V_1^{(1)}U_{-1}^{(1)} - V_{-1}^{(1)}U_1^{(1)}\right)/\sqrt{2}, \tag{12.1.20}$$

and

$$W_{\pm 2}^{(2)} = V_{\pm 1}^{(1)}U_{\pm 1}^{(1)}, \tag{12.1.21}$$

$$W_{\pm 1}^{(2)} = \left(V_{\pm 1}^{(1)}U_0^{(1)} + V_0^{(1)}U_{\pm 1}^{(1)}\right)/\sqrt{2}, \tag{12.1.22}$$

$$W_0^{(2)} = \left(V_1^{(1)}U_{-1}^{(1)} + V_{-1}^{(1)}U_1^{(1)} + 2V_0^{(1)}U_0^{(1)}\right)/\sqrt{6}. \tag{12.1.23}$$

Now, $W_0^{(0)}$ is a scalar operator, equal to $-\vec{V}\cdot\vec{U}/\sqrt{3}$. It is also not hard to check that $W_m^{(1)}$ forms a vector operator, and using eqs. (12.1.9)–(12.1.11), the corresponding rectangular components are found to be those of $i(\vec{V}\times\vec{U})/\sqrt{2}$. The $j = 2$ tensor operator can be rewritten, in terms of the rectangular components of $\vec{V}$ and $\vec{U}$, as

$$W_{\pm 2}^{(2)} = (V_x \pm iV_y)(U_x \pm iU_y)/2, \tag{12.1.24}$$

$$W_{\pm 1}^{(2)} = \left[(\mp V_x - iV_y)U_z + V_z(\mp U_x - iU_y)\right]/2, \tag{12.1.25}$$

$$W_0^{(2)} = (2V_zU_z - V_xU_x - V_yU_y)/\sqrt{6}. \tag{12.1.26}$$

A different basis for these $j = 2$ operators consists of the five operators

$$\begin{aligned} &V_xU_y + V_yU_x, \qquad V_xU_z + V_zU_x, \qquad V_yU_z + V_zU_y, \\ &V_xU_x - V_yU_y, \qquad 2V_zU_z - V_xU_x - V_yU_y, \end{aligned} \tag{12.1.27}$$

which can also be recognized as a basis for the entries of the traceless symmetric 3 × 3 matrix

$$\mathcal{W}_{ab} = (V_aU_b + V_bU_a)/2 - \delta_{ab}\vec{V}\cdot\vec{U}/3, \qquad (a, b = x, y, z). \tag{12.1.28}$$

Therefore, $W_m^{(2)}$ and $\mathcal{W}_{ab}$ are the same operators in different guises. In terms of rectangular components, the product of $V_a$ and $U_b$ decomposes to a linear combination of the irreducible tensor operators with $j = 0, 1, 2$, as

$$V_aU_b = \delta_{ab}\vec{V}\cdot\vec{U}/3 + \epsilon_{abc}(\vec{V}\times\vec{U})_c/2 + \mathcal{W}_{ab}. \tag{12.1.29}$$

This can be checked using $\epsilon_{abc}(\vec{V}\times\vec{U})_c = V_aU_b - V_bU_a$. Thus $V_aU_b$ is an example of a **reducible tensor operator**.[†]

As a further specialization, we can take $\vec{V} = \vec{U} = \vec{R}$, the position operator. The resulting $j = 2$ spherical tensor is called the **quadrupole moment operator**:

$$Q_{\pm 2}^{(2)} = (X \pm iY)^2/2, \tag{12.1.30}$$

$$Q_{\pm 1}^{(2)} = (\mp X - iY)Z, \tag{12.1.31}$$

$$Q_0^{(2)} = (2Z^2 - X^2 - Y^2)/\sqrt{6}, \tag{12.1.32}$$

with the same content as the rectangular-component traceless symmetric tensor

$$\mathcal{Q}_{ab} = R_aR_b - \delta_{ab}R^2/3. \tag{12.1.33}$$

We will encounter this operator again in our study of absorption and emission of electromagnetic radiation, in section 19.5.

## 12.2 Selection rules for scalar and vector operators and the Landé projection formula

As remarked in the previous section, scalar operators do not change the total angular momentum quantum numbers of the states they act on. Furthermore, their matrix elements between total angular momentum $(J^2, J_z)$ eigenstates do not depend on the magnetic quantum number. Both properties are incorporated in the following:

---

[†] A **Cartesian tensor operator** $T_{a_1,\ldots,a_n}$ with $n \geq 2$ indices $a_i = x, y, z$ is generically (if no special symmetry, antisymmetry, or trace conditions are imposed) reducible, in the sense that its entries can be written as linear combinations of irreducible tensor operators with more than one $j$.

**Theorem 12.2.1 (Scalar operator selection rule)** *If $S$ is a scalar operator, and $|\alpha jm\rangle$ are eigenstates of the total angular momentum operators $J^2$ and $J_z$, with additional labels $\alpha$, then*

$$\langle\beta j'm'|S|\alpha jm\rangle \;=\; \delta_{j',j}\delta_{m',m}\,\langle\beta j\|S\|\alpha j\rangle\,, \tag{12.2.1}$$

*where $\langle\beta j\|S\|\alpha j\rangle$ does not depend on $m$.*

As a matter of terminology, the object $\langle\beta j\|S\|\alpha j\rangle$ is called a **reduced matrix element**. This is not a matrix element in the usual sense, but rather just a symbol denoting a number that depends on $\alpha$, $\beta$, and $j$ and on the choice of the operator $S$, but not on $m$ or $m'$.

To prove Theorem 12.2.1, consider $\langle\beta j'm'|[J_z, S]|\alpha jm\rangle$, which is 0 from the definition of a scalar operator. Evaluating each $J_z$ acting on the bra or ket to which it is adjacent, this immediately tells us that $(\hbar m' - \hbar m)\,\langle\beta j'm'|S|\alpha jm\rangle = 0$, so the matrix element in question can only be non-zero if $m' = m$. Similarly, the matrix element of $[J^2, S] = 0$ tells us that $[\hbar^2 j'(j'+1) - \hbar^2 j(j+1)]\,\langle\beta j'm'|S|\alpha jm\rangle = 0$, so the matrix element can only be non-zero if $j' = j$, since the other root $j' = -j-1$ is impossible because $j$ and $j'$ are both non-negative. It remains to show that the matrix element is independent of $m$. This follows from

$$\langle\beta, j, m{+}1|S|\alpha, j, m{+}1\rangle \;=\; \langle\beta, j, m|S|\alpha, j, m\rangle\,, \tag{12.2.2}$$

which can be obtained by using the following facts: $|\alpha, j, m{+}1\rangle$ is proportional to $J_+\,|\alpha, j, m\rangle$; and $\langle\beta, j, m|$ is proportional to $\langle\beta, j, m{+}1|\,J_+$ with the same constant of proportionality, namely $\hbar\sqrt{j(j+1) - m(m+1)}$; and $[J_+, S] = 0$ for a scalar operator.

Theorem 12.2.1 is a powerful selection rule which can be applied, for example, if $S$ is a part of the Hamiltonian, which is often invariant under rotations and therefore a scalar operator. (Examples of exceptions include a charged particle in the presence of an external electric or magnetic field, which will always pick out a special direction that destroys the rotational symmetry. For example, if there is a constant external electric field $\vec{E}$, the Hamiltonian will contain a term proportional to $\vec{E}\cdot\vec{R}$, which is not a scalar operator. This is because $\vec{E}$, being constant, does not transform as a vector operator in the sense we have defined.) This theorem will even be useful to us in the seemingly trivial case that $S$ is the identity operator.

We now turn our attention to vector operators, which in this section we will treat in terms of their rectangular coordinate components $V_a$ for $a = x, y, z$ with $\vec{V} = \hat{x}V_x + \hat{y}V_y + \hat{z}V_z$. By definition, these satisfy

$$[J_a, V_b] \;=\; i\hbar\epsilon_{abc}V_c. \tag{12.2.3}$$

One useful observation is that the operator $\vec{J}\cdot\vec{V} = \vec{V}\cdot\vec{J}$ does not depend on the ordering, since each $J_a$ commutes with the corresponding component $V_a$. Furthermore, it is a scalar operator:

$$\left[J_a,\, \vec{J}\cdot\vec{V}\right] \;=\; 0. \tag{12.2.4}$$

One can also check straightforwardly that

$$\big[J^2,\, V_a\big] \;=\; i\hbar\epsilon_{abc}(V_b J_c + J_c V_b). \tag{12.2.5}$$

From this, and the identity $\epsilon_{abc}\epsilon_{dec} = \delta_{ad}\delta_{be} - \delta_{ae}\delta_{bd}$, one can establish the double commutator relation

$$\big[J^2, \big[J^2, V_a\big]\big] \;=\; 2\hbar^2 \Big(J^2 V_a + V_a J^2 - 2J_a(\vec{J}\cdot\vec{V})\Big)\,. \tag{12.2.6}$$

We now exploit eq. (12.2.6) by evaluating both sides as a matrix element between orthobasis states $\langle\beta, j', m'|$ and $|\alpha, j, m\rangle$, where $\alpha$ and $\beta$ are degeneracy labels. Each instance of $J^2$ can be immediately evaluated as either $\hbar^2 j(j+1)$ or $\hbar^2 j'(j'+1)$, leading to

$$\begin{aligned}\hbar^4\Big([j(j+1) - j'(j'+1)]^2 - 2[j(j+1) + j'(j'+1)]\Big)\,\langle\beta, j', m'|V_a|\alpha, j, m\rangle \;=&\\ -4\hbar^2\,\langle\beta, j', m'|J_a(\vec{J}\cdot\vec{V})|\alpha, j, m\rangle\,.&\end{aligned} \tag{12.2.7}$$

The matrix element on the right can be separated, by using completeness, to get

$$\langle\beta, j', m'|J_a(\vec{J}\cdot\vec{V})|\alpha, j, m\rangle \;=\; \sum_{\gamma, j'', m''} \langle\beta, j', m'|J_a|\gamma j'' m''\rangle\,\langle\gamma j'' m''|\vec{J}\cdot\vec{V}|\alpha, j, m\rangle\,. \tag{12.2.8}$$

Here, the first matrix element vanishes unless $\gamma = \beta$ and $j'' = j'$, while according to Theorem 12.2.1 the second matrix element vanishes unless $j'' = j$ and $m'' = m$. Therefore, only one term in the completeness sum contributes, and eq. (12.2.7) simplifies to a remarkably useful formula,

$$\begin{aligned}\hbar^2(j + j' + 2)(j + j')[(j - j')^2 - 1]\,\langle\beta, j', m'|V_a|\alpha, j, m\rangle \;=&\\ -4\delta_{jj'}\,\langle j, m'|J_a|j, m\rangle\,\langle\beta, j, m|\vec{J}\cdot\vec{V}|\alpha, j, m\rangle\,,&\end{aligned} \tag{12.2.9}$$

where the irrelevant label $\beta$ has been removed from the first matrix element on the right.

One important application of eq. (12.2.9) is the special case $j' = j$. Then it immediately reduces to the following result due to Alfred Landé:

**Theorem 12.2.2 (Vector operator projection rule)** *If an operator $\vec{V}$ transforms as a vector with respect to rotations generated by the angular momentum operator $\vec{J}$, then the* **Landé projection formula** *holds:*

$$\langle\beta, j, m'|\vec{V}|\alpha, j, m\rangle \;=\; \langle j, m'|\vec{J}|j, m\rangle\,\frac{\langle\beta, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle}{\hbar^2 j(j+1)} \tag{12.2.10}$$

*for $j \neq 0$, where the reduced matrix element*

$$\langle\beta, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle \;\equiv\; \langle\beta, j, m|\vec{J}\cdot\vec{V}|\alpha, j, m\rangle \tag{12.2.11}$$

*does not depend on $m$ or $m'$. For the special case $j = 0$,*

$$\langle\beta, 0, 0|\vec{V}|\alpha, 0, 0\rangle \;=\; 0. \tag{12.2.12}$$

Because $\vec{J}\cdot\vec{V}$ is a scalar operator, Theorem 12.2.1 reminds us that the right side of eq. (12.2.11) does not actually depend on $m$, as claimed. So, in applications, we can pick any convenient value for $m$ to do the evaluation of $\langle\beta, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle$. (The most convenient choice is usually either $m = j$ or $m = 0$.) Of course, if $j = 0$ then eq. (12.2.10) fails, but in that case $j = m = m' = 0$, and the vanishing of $\langle\beta, 0, 0|\vec{V}|\alpha, 0, 0\rangle$ follows from spherical symmetry. The $j = 0$ states do not have any special direction along which the matrix element of a vector operator could point.

Intuitively, the Landé projection formula says that, within each subspace of fixed $j$, all vector operators are proportional to each other, and in particular to the angular momentum operator. As a check, if we take $\vec{V} = \vec{J}$, then $\vec{J}\cdot\vec{V} = J^2$, and eq. (12.2.10) is satisfied. The reason for the word "projection" in the name is that it implies that fixed-$j$ matrix elements of $\vec{V}$ are equal to those of $\vec{J}(\vec{J}\cdot\vec{V})/J^2$, the geometric projection of $\vec{V}$ onto the angular momentum operator direction. The projection formula is a particularly useful special case of the Wigner–Eckart theorem discussed in the next section.

Another important application of eq. (12.2.9) occurs if we take $\vec{V} = \vec{R}$, the position operator. As a notable example, this will be useful for evaluating absorption and emission of electromagnetic radiation in the electric dipole approximation (in section 19.4). For this reason, the conditions on the matrix elements of the components of $\vec{R}$ that we are about to derive are often called the **dipole selection rules**.

When considering matrix elements of $\vec{R}$, we can take $\vec{J}$ to be just the orbital angular momentum operator $\vec{L}$, since that is what generates rotations for $\vec{R}$; the spin is not relevant. Further, $\vec{L}\cdot\vec{R} = (\vec{R}\times\vec{P})\cdot\vec{R} = 0$, so the right side of eq. (12.2.9) vanishes, and it gives simply

$$\hbar^2(l + l' + 2)(l + l')[(l - l')^2 - 1]\ \langle\beta, l', m'|\vec{R}\,|\alpha, l, m\rangle \;=\; 0. \tag{12.2.13}$$

Therefore, for the matrix element to be non-zero, we need

$$(l + l' + 2)(l + l')[(l - l')^2 - 1] \;=\; 0. \tag{12.2.14}$$

Now, $l+l'+2$ cannot vanish because $l$ and $l'$ are always non-negative. Also, the condition $l+l' = 0$ can only be satisfied if $l = l' = 0$, but then $\langle\beta, l', m'|\vec{R}|\alpha, l, m\rangle$ can only be $\langle\beta, 0, 0|\vec{R}|\alpha, 0, 0\rangle$, which must vanish anyway due to the spherical symmetry of the state wavefunctions. From eq. (12.2.14), the only other possibility for a non-vanishing matrix element is $(l - l')^2 = 1$. Thus, we have the first dipole selection rule,

$$\langle\beta, l', m'|\vec{R}|\alpha, l, m\rangle \;\neq\; 0 \quad \text{requires} \quad l' - l = \pm 1. \tag{12.2.15}$$

Since this holds for the whole vector $\vec{R}$, it is also true for the matrix elements of each of its components, $X$, $Y$, and $Z$. The remaining dipole selection rules are

$$\langle \beta, l', m' | Z | \alpha, l, m\rangle \neq 0 \quad \text{requires } m' = m, \tag{12.2.16}$$

$$\langle \beta, l', m' | (X + iY) | \alpha, l, m\rangle \neq 0 \quad \text{requires } m' = m + 1, \tag{12.2.17}$$

$$\langle \beta, l', m' | (X - iY) | \alpha, l, m\rangle \neq 0 \quad \text{requires } m' = m - 1. \tag{12.2.18}$$

To prove these, we can suppress the labels $\alpha, \beta$, which play no role. Taking the matrix element of the commutation relation $[L_z, Z] = 0$ gives $\langle l', m' | [L_z, Z] | l, m\rangle = (\hbar m' - \hbar m) \langle l', m' | Z | l, m\rangle = 0$. This shows that if $\langle l', m' | Z | l, m\rangle$ is to be non-zero, we need $m' = m$. A simple alternative proof uses the $\phi$-dependence of the position representations of the spherical harmonics and $X$, $Y$, and $Z$. We have $\langle l', m' | Z | l, m\rangle \propto \int_0^{2\pi} d\phi \, \big(e^{-im'\phi}\big)\big(1\big)\big(e^{im\phi}\big) \propto \delta_{m,m'}$, and similarly, $\langle l', m' | (X \pm iY) | l, m\rangle \propto \int_0^{2\pi} d\phi \, \big(e^{-im'\phi}\big)\big(e^{\pm i\phi}\big)\big(e^{im\phi}\big) \propto \delta_{m',m\pm 1}$. Equations (12.2.17) and (12.2.18) also imply that the corresponding matrix elements of $X$ and $Y$ vanish unless $m' = m \pm 1$.

As a last application of eq. (12.2.9), consider the selection rules for the matrix elements $\langle \beta, j, m' | \vec{V} | \alpha, j, m\rangle$ of a general vector operator. The only difference compared to the special case of $\vec{R}$ that led to eq. (12.2.15) is that now we must admit the possibility that the right side of eq. (12.2.9) does not vanish for $j = j'$. Therefore, the **general vector operator selection rules** are

$$\langle \beta, j', m' | \vec{V} | \alpha, j, m\rangle \neq 0 \quad \text{requires } \; j' - j = 0, \pm 1. \tag{12.2.19}$$

$$\langle \beta, j', m' | V_z | \alpha, j, m\rangle \neq 0 \quad \text{requires } m' = m, \tag{12.2.20}$$

$$\langle \beta, j', m' | (V_x + iV_y) | \alpha, j, m\rangle \neq 0 \quad \text{requires } m' = m + 1, \tag{12.2.21}$$

$$\langle \beta, j', m' | (V_x - iV_y) | \alpha, j, m\rangle \neq 0 \quad \text{requires } m' = m - 1. \tag{12.2.22}$$

The last three can be obtained by using $[J_z, V_z] = 0$ and $[J_+, V_+] = 0$ and $[J_-, V_-] = 0$.

## 12.3 The Wigner–Eckart Theorem and selection rules for tensor operators

In this section we will state and prove the **Wigner–Eckart Theorem** for matrix elements of irreducible tensor operators. Developed independently by Wigner and Carl Eckart, this is a generalization of three simpler results that we have already given: eq. (11.5.18) for spherical harmonics, Theorem 12.2.1 for scalar operators, and Theorem 12.2.2 for vector operators. The reader may wish to contemplate the parallels between the following derivations and those used in section 11.5 for spherical harmonics.

Consider an irreducible tensor operator $T^{(j_1)}_{m_1}$ with $m_1 = -j_1, \ldots, j_1$, and a multiplet of angular momentum eigenstates $|\alpha j_2 m_2\rangle$ with $m_2 = -j_2, \ldots, j_2$, where $\alpha$ is a degeneracy label.

Now we construct the states

$$|\gamma j m\rangle \;=\; \sum_{m_1,m_2} T^{(j_1)}_{m_1}\,|\alpha j_2 m_2\rangle\; C^{j_1\,j_2\,j}_{m_1\,m_2\,m}, \tag{12.3.1}$$

for $j = |j_1 - j_2|, \ldots, j_1 + j_2$ and $m = -j, \ldots, j$, where $C^{j_1\,j_2\,j}_{m_1\,m_2\,m}$ are Clebsch–Gordan coefficients as defined by eq. (11.4.4). We claim that, as the labeling indicates, eq. (12.3.1) is an eigenstate of $J^2$ and $J_z$ with eigenvalues $\hbar^2 j(j+1)$ and $\hbar m$, respectively. (It is not necessarily normalized to unity.) To prove this, it suffices to show that

$$J_z\,|\gamma j m\rangle \;=\; \hbar m\,|\gamma j m\rangle\,, \tag{12.3.2}$$

$$J_-\,|\gamma j m\rangle \;=\; \hbar\sqrt{j(j+1) - m(m-1)}\,|\gamma, j, m-1\rangle\,, \tag{12.3.3}$$

$$J_+\,|\gamma j m\rangle \;=\; \hbar\sqrt{j(j+1) - m(m+1)}\,|\gamma, j, m+1\rangle\,, \tag{12.3.4}$$

as then the result for $J^2 = (J_+J_- + J_-J_+)/2 + J_z^2$ follows.

To prove eq. (12.3.2), we use eq. (12.1.6) to obtain

$$J_z\,|\gamma j m\rangle = \sum_{m_1,m_2} T^{(j_1)}_{m_1}(\hbar m_1 + J_z)\,|\alpha j_2 m_2\rangle\; C^{j_1\,j_2\,j}_{m_1\,m_2\,m}. \tag{12.3.5}$$

Then, use $J_z\,|\alpha j_2 m_2\rangle = \hbar m_2\,|\alpha j_2 m_2\rangle$, and eq. (12.3.2) follows by noting that the Clebsch–Gordan coefficient enforces that only $m_1 + m_2 = m$ contributes in the sum.

To prove eq. (12.3.3), we similarly use eq. (12.1.8) and then eq. (12.1.5) to find

$$\begin{aligned} J_-\,|\gamma j m\rangle \;=\; \hbar \sum_{m_1,m_2} C^{j_1\,j_2\,j}_{m_1\,m_2\,m}\Big(&\sqrt{j_1(j_1+1) - m_1(m_1-1)}\,T^{(j_1)}_{m_1-1}\,|\alpha j_2 m_2\rangle \\ &+\sqrt{j_2(j_2+1) - m_2(m_2-1)}\,T^{(j_1)}_{m_1}\,|\alpha, j_2, m_2-1\rangle\Big). \end{aligned} \tag{12.3.6}$$

Now, since we are summing over all $m_1$ and $m_2$, we can use the trick of relabeling the indices according to $m_1 \to m_1 + 1$ in the first term and $m_2 \to m_2 + 1$ in the second term. The result is

$$J_-\,|\gamma j m\rangle \;=\; \hbar \sum_{m_1,m_2} T^{(j_1)}_{m_1}\,|\alpha j_2 m_2\rangle\,\Big(\sqrt{j_1(j_1+1) - m_1(m_1+1)}\,C^{j_1\,j_2\,j}_{m_1+1,\,m_2,\,m} \tag{12.3.7}$$

$$+\sqrt{j_2(j_2+1) - m_2(m_2+1)}\,C^{j_1\,j_2\,j}_{m_1,\,m_2+1,\,m}\Big). \tag{12.3.8}$$

Equation (12.3.3) then follows immediately from using the Clebsch–Gordan recurrence relation eq. (11.4.20). The proof of eq. (12.3.4) is completely analogous.

Having established that the states $|\gamma j m\rangle$ defined by eq. (12.3.1) are really eigenstates of $J^2$ and $J_z$, we can now solve for each term in the sum on the right side of that equation. Starting from eq. (12.3.1), multiply by $C^{j_1\,j_2\,j}_{m_1'\,m_2'\,m}$, sum over all $j$ and $m$, and then use the Clebsch–Gordan orthonormality relation eq. (11.4.24) to collapse the $m_1, m_2$ sums to a single term, and finally relabel $m_1' \to m_1$ and $m_2' \to m_2$. The result is

$$T^{(j_1)}_{m_1} |\alpha j_2 m_2\rangle \;=\; \sum_{j,m} |\gamma j m\rangle \, C^{j_1\, j_2\, j}_{m_1\, m_2\, m}. \tag{12.3.9}$$

This is the precise version of the intuitive statement that $T^{(j_1)}_{m_1}$ carries angular momentum quantum numbers $j_1$ and $m_1$, and that it imparts them to the states it acts on, yielding a sum of states that occur in the addition of angular momenta $j_1$ and $j_2$.

Now acting on eq. (12.3.9) with $\langle \beta j' m'|$, and then relabeling $j' \to j$ and $m' \to m$, we obtain

$$\langle \beta j m | T^{(j_1)}_{m_1} | \alpha j_2 m_2\rangle \;=\; \langle \beta j m | \gamma j m\rangle \, C^{j_1\, j_2\, j}_{m_1\, m_2\, m}. \tag{12.3.10}$$

By its construction, the state $|\gamma j m\rangle$ depends on $\alpha$ and the choice of tensor operator $T^{(j_1)}$ and $j$. However, the matrix element $\langle \beta j m | \gamma j m\rangle$ clearly does not depend on $m_1$ or $m_2$, and applying Theorem 12.2.1 with $S$ taken to be the identity operator, we learn that it does not actually depend on $m$ either. Therefore, we have proved a powerfully general result:

**Theorem 12.3.1 (Wigner–Eckart)** *If $T^{(j_1)}$ is an irreducible tensor operator, and $|\alpha j_2 m_2\rangle$ and $|\beta j m\rangle$ are eigenstates of $J^2$ and $J_z$ with eigenvalues as labeled, then*[†]

$$\langle \beta j m | T^{(j_1)}_{m_1} | \alpha j_2 m_2\rangle \;=\; \langle \beta j \| T^{(j_1)} \| \alpha j_2\rangle \, C^{j_1\, j_2\, j}_{m_1\, m_2\, m}, \tag{12.3.11}$$

*where the reduced matrix element $\langle \beta j \| T^{(j_1)} \| \alpha j_2\rangle$ does not depend on $m$, $m_1$, or $m_2$.*

For $j_1 = m_1 = 0$, so that $T^{(0)}_0 = S$ is a scalar operator, we can use the fact that $C^{0 j j'}_{0 m m'} = \delta_{j,j'}\delta_{m,m'}$ to recover Theorem 12.2.1 as a special case of the Wigner–Eckart Theorem. This is hardly surprising, since we just used the former in the proof of the latter. The Landé projection formula, eq. (12.2.10), can be shown to be equivalent to the special case obtained for vector operators, when $j_1 = 1$. The triple spherical harmonic integral relation eq. (11.5.18) is equivalent to the special case in which the generator of rotations $\vec{J}$ is taken to be the orbital angular momentum $\vec{L}$.

One way to use the Wigner–Eckart Theorem is as a labor-saving device. If we can calculate the matrix element on the left side of eq. (12.3.11) for just one instance $(m_1, m_2, m)$ for which it is non-zero, then we can infer the reduced matrix element. Once that is done, all other cases of $(m_1, m_2, m)$ with the same $\alpha, \beta, j_1, j_2$, and $j$ will be known.

The Wigner–Eckart Theorem also provides selection rules, since it implies that the matrix element can only be non-zero if the Clebsch–Gordan coefficient $C^{j_1\, j_2\, j}_{m_1\, m_2\, m}$ is non-zero. This means

---

[†] Some sources define the reduced matrix element with other normalization conventions, most commonly by including a factor of $1/\sqrt{2j+1}$ (and sometimes a different sign) on the right side of eq. (12.3.11).

that the same selection rules apply, so that $\langle\beta j m|T^{(j_1)}_{m_1}|\alpha j_2 m_2\rangle$ can be non-zero only if

$$j = |j_1 - j_2|, \ldots, j_1 + j_2 - 1, j_1 + j_2, \tag{12.3.12}$$

$$m = m_1 + m_2. \tag{12.3.13}$$

These **tensor operator selection rules** generalize the results in eqs. (12.2.19)–(12.2.22) that we obtained for the special case of vector operators.

# 13 Stationary-state perturbation theory

## 13.1 Perturbative expansion for energy eigenstates

Given a Hamiltonian $H$ that does not depend on time, we would like to solve the time-independent Schrödinger equation,

$$H|\psi_n\rangle \;=\; E_n|\psi_n\rangle, \tag{13.1.1}$$

as an eigenvalue problem for the stationary states $|\psi_n\rangle$ and their energies $E_n$. Unfortunately, in many cases this is too hard to do exactly. However, suppose that it is possible to split the Hamiltonian into two parts,

$$H \;=\; H_0 + \lambda W, \tag{13.1.2}$$

where $H_0$ is a simpler Hamiltonian for which we already know how to solve the eigenvalue problem, and $\lambda$ is a small parameter, so that $W$ can be treated as a perturbation. Then we can hope to find an approximate solution for the full Hamiltonian $H$, by expanding in $\lambda$. We will now work out how to do this. This is called **time-independent perturbation theory** or **stationary-state perturbation theory**.

To begin, assume that we have found the complete set of eigenstates $|n\rangle$ and eigenvalues $\mathcal{E}_n$ for $H_0$, so that

$$H_0|n\rangle \;=\; \mathcal{E}_n|n\rangle. \tag{13.1.3}$$

In general, the states $|n\rangle$ for a given $\mathcal{E}_n$ might be degenerate, but for now we assume they are not. For each unperturbed stationary state $|n\rangle$ and energy $\mathcal{E}_n$, we want to solve for the corresponding $|\psi_n\rangle$ and energy $E_n$. The connection between the unperturbed and perturbed stationary-state solutions is assumed to be smoothly continuous in $\lambda$ in order for the following strategy to work. In particular, there should be a one-to-one correspondence between the known $|n\rangle$ and the desired $|\psi_n\rangle$ for all $\lambda$, as long as it is not too large.

The perturbation theory strategy is to write the unknown full eigenstates and energy eigenvalues each as a power series in $\lambda$,

$$|\psi_n\rangle \;=\; |\psi_n^{(0)}\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \lambda^3|\psi_n^{(3)}\rangle + \cdots, \tag{13.1.4}$$

$$E_n \;=\; E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \lambda^3 E_n^{(3)} + \cdots. \tag{13.1.5}$$

Now, $\lambda = 0$ must recover the unperturbed solutions, so

$$|\psi_n^{(0)}\rangle \;=\; |n\rangle, \tag{13.1.6}$$

$$E_n^{(0)} \;=\; \mathcal{E}_n. \tag{13.1.7}$$

The unperturbed stationary states are a complete orthobasis, with

$$\sum_n |n\rangle\langle n| \;=\; I, \tag{13.1.8}$$

$$\langle n'|n\rangle \;=\; \delta_{n,n'}. \tag{13.1.9}$$

This is always true, because $H_0$ is assumed to be an observable.

We could also choose unit normalization for the eigenkets $|\psi_n\rangle$, but we won't. It is much more convenient to choose the following normalization condition for them instead:

$$\langle n|\psi_n\rangle = 1 \qquad \text{(for each } n\text{)}, \tag{13.1.10}$$

which will result in $\langle\psi_n|\psi_n\rangle \neq 1$. After obtaining the solutions for $|\psi_n\rangle$, we can re-normalize them later, as a last step. Plugging eq. (13.1.4) into eq. (13.1.10), we have

$$\langle n|\left(|n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots\right) \;=\; 1. \tag{13.1.11}$$

Since $\langle n|n\rangle = 1$, in order to have a solution for general $\lambda$, it is necessary and sufficient that the coefficient of $\lambda^k$ in eq. (13.1.11) vanishes for each $k \geq 1$, which implies

$$\langle n|\psi_n^{(k)}\rangle \;=\; 0 \qquad \text{(for } k = 1, 2, 3, \ldots\text{)}. \tag{13.1.12}$$

Thus, the normalization choice eq. (13.1.10) ensures that all of the corrections to each state are orthogonal to the corresponding unperturbed state. This is extremely important, and will be used repeatedly in the following.

The time-independent Schrödinger equation (13.1.1) now reads

$$\begin{aligned}(H_0 + \lambda W)\left(|n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots\right) \;=&\\ \left(\mathcal{E}_n + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \cdots\right)&\left(|n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots\right).\end{aligned} \tag{13.1.13}$$

Expanding both sides, and matching the coefficient of each power $\lambda^k$, we get

$$\lambda^0: \qquad H_0|n\rangle \;=\; \mathcal{E}_n|n\rangle, \tag{13.1.14}$$

$$\lambda^1: \qquad H_0|\psi_n^{(1)}\rangle + W|n\rangle \;=\; \mathcal{E}_n|\psi_n^{(1)}\rangle + E_n^{(1)}|n\rangle, \tag{13.1.15}$$

$$\lambda^2: \qquad H_0|\psi_n^{(2)}\rangle + W|\psi_n^{(1)}\rangle \;=\; \mathcal{E}_n|\psi_n^{(2)}\rangle + E_n^{(1)}|\psi_n^{(1)}\rangle + E_n^{(2)}|n\rangle, \tag{13.1.16}$$

and, in general,

$$\lambda^k: \qquad H_0|\psi_n^{(k)}\rangle + W|\psi_n^{(k-1)}\rangle \;=\; \mathcal{E}_n|\psi_n^{(k)}\rangle + \sum_{j=1}^{k-1} E_n^{(j)}|\psi_n^{(k-j)}\rangle + E_n^{(k)}|n\rangle. \tag{13.1.17}$$

Equation (13.1.14) is, of course, just a repetition of eq. (13.1.3).

Consider first the effect of the perturbation at order $\lambda^1$. Taking the inner product of eq. (13.1.15) with $\langle n|$ gives

$$\langle n|H_0|\psi_n^{(1)}\rangle + \langle n|W|n\rangle \;=\; \mathcal{E}_n\langle n|\psi_n^{(1)}\rangle + E_n^{(1)}\langle n|n\rangle. \tag{13.1.18}$$

Now using $\langle n|H_0 = \mathcal{E}_n\langle n|$ on the first term, we see that it and the first term on the right side are equal, and both vanish due to eq. (13.1.12). Therefore, using the orthonormality condition eq. (13.1.9) on the last term, eq. (13.1.18) simplifies to

$$E_n^{(1)} \;=\; \langle n|W|n\rangle, \tag{13.1.19}$$

so that

$$E_n \;=\; \mathcal{E}_n + \lambda\langle n|W|n\rangle + \cdots . \tag{13.1.20}$$

To first order in the expansion in $\lambda$, the energy shift in the state $|n\rangle$ is simply the expectation value of the Hamiltonian perturbation. This is the most important and commonly used result from stationary-state perturbation theory.

To find the first-order correction to the energy eigenkets, we note that completeness of the $H_0$ eigenstates allows us to write

$$|\psi_n^{(1)}\rangle \;=\; \sum_{m\neq n} |m\rangle\langle m|\psi_n^{(1)}\rangle, \tag{13.1.21}$$

where we have made good use of eq. (13.1.12) to dispense with the $m = n$ term. We now need to find the coefficients $\langle m|\psi_n^{(1)}\rangle$ for $m \neq n$. To do so, take the inner product of $\langle m|$ acting on eq. (13.1.15),

$$\langle m|H_0|\psi_n^{(1)}\rangle + \langle m|W|n\rangle \;=\; \mathcal{E}_n\langle m|\psi_n^{(1)}\rangle + E_n^{(1)}\langle m|n\rangle. \tag{13.1.22}$$

The last term vanishes due to the orthonormality of the unperturbed stationary states eq. (13.1.9), and the first term on the left can be simplified using $\langle m|H_0 = \mathcal{E}_m\langle m|$. Therefore,

$$(\mathcal{E}_n - \mathcal{E}_m)\,\langle m|\psi_n^{(1)}\rangle \;=\; \langle m|W|n\rangle \qquad (m \neq n). \tag{13.1.23}$$

Now, assuming that there are no degeneracies, so that $\mathcal{E}_m \neq \mathcal{E}_n$ for $m \neq n$, we can solve for $\langle m|\psi_n^{(1)}\rangle$, and eq. (13.1.21) becomes

$$|\psi_n^{(1)}\rangle \;=\; \sum_{m\neq n} |m\rangle\frac{\langle m|W|n\rangle}{\mathcal{E}_n - \mathcal{E}_m}. \tag{13.1.24}$$

We have thus obtained the perturbed energy eigenstates at first order in the expansion in $\lambda$,

$$|\psi_n\rangle \;=\; |n\rangle + \lambda\sum_{m\neq n} |m\rangle\frac{\langle m|W|n\rangle}{\mathcal{E}_n - \mathcal{E}_m} + \cdots . \tag{13.1.25}$$

We can now see why the normalization condition eq. (13.1.10) was so useful and important; it gave us eq. (13.1.12), which in turn allowed us to exclude $m = n$ from the sum, which avoided a term with a disastrous vanishing denominator[†] in $|\psi_n^{(1)}\rangle$, and in similar expressions to follow.

Having solved the perturbation theory problem at order $\lambda^1$, we can continue iteratively to obtain the perturbed energies and their eigenstates at any desired order in $\lambda$. We will now find the necessary recurrence relations in general. At order $\lambda^k$, the inner product of $\langle n|$ with eq. (13.1.17) is

$$\langle n|H_0|\psi_n^{(k)}\rangle + \langle n|W|\psi_n^{(k-1)}\rangle \;=\; \mathcal{E}_n\,\langle n|\psi_n^{(k)}\rangle + \sum_{j=1}^{k-1} E_n^{(j)}\,\langle n|\psi_n^{(k-j)}\rangle + E_n^{(k)}. \qquad (13.1.26)$$

The first term on the left vanishes because of eq. (13.1.12), after pulling out $H_0 = \mathcal{E}_n$. On the right side of the equality, using eq. (13.1.12) again ensures that all of the terms vanish except the last one, where we have used $\langle n|n\rangle = 1$. So, we find the remarkable result

$$E_n^{(k)} \;=\; \langle n|W|\psi_n^{(k-1)}\rangle, \qquad (13.1.27)$$

that the $\lambda^k$ correction to each energy eigenvalue is always obtained in terms of the $\lambda^{k-1}$ correction to the corresponding ket.

Also, using completeness, we can write

$$|\psi_n^{(k)}\rangle \;=\; \sum_{m\neq n} |m\rangle\langle m|\psi_n^{(k)}\rangle, \qquad (13.1.28)$$

where once again eq. (13.1.12) has been used to eliminate the $m = n$ term. To find the coefficients $\langle m|\psi_n^{(k)}\rangle$ for $m \neq n$, we take $\langle m|$ acting on eq. (13.1.17), which gives

$$(\mathcal{E}_n - \mathcal{E}_m)\,\langle m|\psi_n^{(k)}\rangle \;=\; \langle m|W|\psi_n^{(k-1)}\rangle - \sum_{j=1}^{k-1} E_n^{(j)}\langle m|\psi_n^{(k-j)}\rangle, \qquad (13.1.29)$$

where $\langle m|n\rangle = 0$ has been used to eliminate the term proportional to $E_n^{(k)}$. The right-hand side only involves expressions from orders less than $k$. Therefore, eqs. (13.1.27), (13.1.28), and (13.1.29) can be used to solve iteratively for $E_n^{(k)}$ and then $|\psi_n^{(k)}\rangle$ at each successive level $k$.

Let us apply the preceding for $k = 2$. Evaluating eq. (13.1.27) using eq. (13.1.24) gives

$$E_n^{(2)} \;=\; \sum_{m\neq n} \frac{|\langle m|W|n\rangle|^2}{\mathcal{E}_n - \mathcal{E}_m}. \qquad (13.1.30)$$

We have thus obtained the second-order corrections to the energies. Also, eq. (13.1.29) gives

$$\langle m|\psi_n^{(2)}\rangle \;=\; \frac{1}{\mathcal{E}_n - \mathcal{E}_m}\Big(\langle m|W|\psi_n^{(1)}\rangle - E_n^{(1)}\langle m|\psi_n^{(1)}\rangle\Big), \qquad (13.1.31)$$

---

[†] Of course, the denominator is still a disaster if the unperturbed states have energy degeneracies, so that $\mathcal{E}_m = \mathcal{E}_n$ for some $m \neq n$. This issue is addressed in section 13.6.

or, using eqs. (13.1.19) and (13.1.24),

$$\langle m|\psi_n^{(2)}\rangle = \sum_{p\neq n} \frac{\langle m|W|p\rangle\langle p|W|n\rangle}{(\mathcal{E}_n - \mathcal{E}_p)(\mathcal{E}_n - \mathcal{E}_m)} - \frac{\langle m|W|n\rangle\langle n|W|n\rangle}{(\mathcal{E}_n - \mathcal{E}_m)^2} \tag{13.1.32}$$

for $n \neq m$. Summarizing the results to order $\lambda^2$:

$$E_n = \mathcal{E}_n + \lambda\langle n|W|n\rangle + \lambda^2 \sum_{m\neq n} \frac{|\langle m|W|n\rangle|^2}{\mathcal{E}_n - \mathcal{E}_m} + \cdots, \tag{13.1.33}$$

and

$$\begin{aligned} |\psi_n\rangle &= |n\rangle + \lambda \sum_{m\neq n} |m\rangle \frac{\langle m|W|n\rangle}{\mathcal{E}_n - \mathcal{E}_m} \\ &\quad + \lambda^2 \sum_{m\neq n} |m\rangle \left( \sum_{p\neq n} \frac{\langle m|W|p\rangle\langle p|W|n\rangle}{(\mathcal{E}_n - \mathcal{E}_p)(\mathcal{E}_n - \mathcal{E}_m)} - \frac{\langle m|W|n\rangle\langle n|W|n\rangle}{(\mathcal{E}_n - \mathcal{E}_m)^2} \right) + \cdots. \end{aligned} \tag{13.1.34}$$

Don't forget that, by design, this ket $|\psi_n\rangle$ does not have unit norm, but now it can be renormalized as $|\psi_n\rangle/\sqrt{\langle\psi_n|\psi_n\rangle}$, if desired. In doing so, it often makes sense to expand in $\lambda$ and drop the terms of order $\lambda^3$ or higher. From eq. (13.1.34), we find using the orthonormality of the unperturbed kets that

$$\langle\psi_n|\psi_n\rangle = 1 + \lambda^2 \sum_{m\neq n} \frac{|\langle m|W|n\rangle|^2}{(\mathcal{E}_n - \mathcal{E}_m)^2} + \cdots, \tag{13.1.35}$$

so that, working consistently to second order in $\lambda$, the normalized eigenstate can be obtained by just multiplying the first term $|n\rangle$ in eq. (13.1.34) by

$$1/\sqrt{\langle\psi_n|\psi_n\rangle} = 1 - \frac{\lambda^2}{2} \sum_{m\neq n} \frac{|\langle m|W|n\rangle|^2}{(\mathcal{E}_n - \mathcal{E}_m)^2} + \cdots. \tag{13.1.36}$$

It is also worth noting that, using eq. (13.1.27), the energy eigenvalue to all orders in $\lambda$ is

$$E_n = \mathcal{E}_n + \sum_{k=1}^{\infty} \lambda^k \langle n|W|\psi_n^{(k-1)}\rangle = \mathcal{E}_n + \lambda\langle n|W \sum_{k=1}^{\infty} \lambda^{k-1}|\psi_n^{(k-1)}\rangle, \tag{13.1.37}$$

which can be rewritten

$$E_n = \mathcal{E}_n + \lambda\langle n|W|\psi_n\rangle. \tag{13.1.38}$$

The all-orders energy $E_n$ is therefore known exactly in terms of the all-orders state ket $|\psi_n\rangle$. Of course, the latter may well be only known partially as a perturbative expansion.

So far, we have used the parameter $\lambda$ as a way of keeping track of the order of perturbation theory. In any given term, the power of $\lambda$ is the same as the number of matrix elements of

$W$, so we can simply absorb $\lambda$ into the definition of $W$ without losing any information. (The convergence of perturbation theory still relies on the assumption that the properties of $W$ allow it to be treated as a small correction. For example, $W$ may contain one or more small dimensionless parameters that effectively play the role of $\lambda$.) Equivalently, taking $\lambda = 1$, we have

$$\begin{aligned} E_n &= \mathcal{E}_n + E_n^{(1)} + E_n^{(2)} + E_n^{(3)} + \cdots , && (13.1.39) \\ |\psi_n\rangle &= |n\rangle + |\psi_n^{(1)}\rangle + |\psi_n^{(2)}\rangle + |\psi_n^{(3)}\rangle + \cdots . && (13.1.40) \end{aligned}$$

The preceding results can be expressed more compactly with the following simplifying notation,

$$\begin{aligned} W_{mn} &= \langle m|W|n\rangle, && (13.1.41) \\ \mathcal{E}_{nm} &= \mathcal{E}_n - \mathcal{E}_m. && (13.1.42) \end{aligned}$$

Then, the energy eigenvalue corrections are summarized as

$$\begin{aligned} E_n^{(1)} &= W_{nn}, && (13.1.43) \\ E_n^{(2)} &= \sum_{m\neq n} \frac{|W_{mn}|^2}{\mathcal{E}_{nm}}, && (13.1.44) \\ E_n^{(3)} &= \sum_{m\neq n}\sum_{p\neq n} \frac{W_{nm}W_{mp}W_{pn}}{\mathcal{E}_{np}\mathcal{E}_{nm}} - W_{nn}\sum_{m\neq n}\frac{|W_{mn}|^2}{\mathcal{E}_{nm}^2}, && (13.1.45) \end{aligned}$$

etc., and for the (un-normalized) energy eigenstates,

$$\begin{aligned} |\psi_n^{(1)}\rangle &= \sum_{m\neq n} |m\rangle \frac{W_{mn}}{\mathcal{E}_{nm}}, && (13.1.46) \\ |\psi_n^{(2)}\rangle &= \sum_{m\neq n} |m\rangle \left(\sum_{p\neq n}\frac{W_{mp}W_{pn}}{\mathcal{E}_{np}\mathcal{E}_{nm}} - \frac{W_{nn}W_{mn}}{\mathcal{E}_{nm}^2}\right), && (13.1.47) \end{aligned}$$

etc. All higher order corrections likewise involve powers of matrix elements of $W$ in the numerators, and powers of unperturbed energy differences $\mathcal{E}_{nm}$'s in the denominators. This means that the perturbative expansion, as we have constructed it here, may fail to converge for a state $|n\rangle$ if there is some other state $|m\rangle$ such that $\mathcal{E}_{nm}$ is zero, or small compared to the matrix elements of $W$. To treat such cases, we will need to use degenerate perturbation theory (section 13.6), or almost-degenerate perturbation theory (section 13.8), respectively.

## 13.2 Simple examples of perturbation theory

Consider the one-dimensional harmonic oscillator with a linear perturbation:

$$H_0 = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X^2, \qquad W = -fX. \qquad (13.2.1)$$

Here $f$ is a constant force, which plays the role of the expansion parameter $\lambda$ in the discussion of section 13.1. Before solving this as a problem in perturbation theory for small $f$, we recall that it is special in that the stationary states can actually be found exactly. As already noted at the end of section 7.4, this is because the total Hamiltonian can be rewritten, by completing the square, as

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2\Big(X - \frac{f}{m\omega^2}\Big)^2 - \frac{f^2}{2m\omega^2}, \tag{13.2.2}$$

so that after a shift of variables $X \to X + f/m\omega^2$, the new $H$ will have the same $m$, $\omega$, and energy eigenvalues, except that the latter will be offset by a constant $-f^2/2m\omega^2$. Thus we have

$$E_n = \hbar\omega(n + 1/2) - f^2/2m\omega^2, \tag{13.2.3}$$

exactly. Furthermore, the corresponding energy eigenstates must be

$$|\psi_n\rangle = T(f/m\omega^2)|n\rangle, \tag{13.2.4}$$

where $T(a)$ is the translation operator of eq. (5.2.1), and $|n\rangle$ are the stationary states of $H_0$. We can expand this exact result to linear order in $f$, using eq. (7.3.4) for the momentum operator in terms of harmonic oscillator creation and annihilation operators, to get

$$T(f/m\omega^2) = I + \frac{f}{\sqrt{2m\hbar\omega^3}}(a^\dagger - a) + \mathcal{O}(f^2), \tag{13.2.5}$$

so that the eigenstates of $H$ are

$$|\psi_n\rangle = |n\rangle + \frac{f}{\sqrt{2m\hbar\omega^3}}\Big(\sqrt{n+1}|n{+}1\rangle - \sqrt{n}|n{-}1\rangle\Big) + \mathcal{O}(f^2) \tag{13.2.6}$$

in terms of the eigenstates of $H_0$.

Now let us use solve the same problem using the technology of perturbation theory that we have developed. The unperturbed energy eigenvalues and eigenstates are $\mathcal{E}_n = \hbar\omega(n + 1/2)$ and $|n\rangle$. The general matrix elements needed for the perturbative expansion are

$$\begin{aligned} W_{n'n} = \langle n'|W|n\rangle &= -f\langle n'|X|n\rangle = -f\sqrt{\frac{\hbar}{2m\omega}}\langle n'|(a^\dagger + a)|n\rangle \\ &= -f\sqrt{\frac{\hbar}{2m\omega}}\Big(\sqrt{n+1}\,\delta_{n',n+1} + \sqrt{n}\,\delta_{n',n-1}\Big). \end{aligned} \tag{13.2.7}$$

Now, since this vanishes for $n' = n$, the first-order correction to the energies are all 0. From eqs. (13.1.43) and (13.1.44), we get, through second order,

$$E_n^{(1)} = 0, \tag{13.2.8}$$

$$E_n^{(2)} = \frac{\big|-f\sqrt{\hbar(n+1)/2m\omega}\big|^2}{\mathcal{E}_n - \mathcal{E}_{n+1}} + \frac{\big|-f\sqrt{\hbar n/2m\omega}\big|^2}{\mathcal{E}_n - \mathcal{E}_{n-1}}, \tag{13.2.9}$$

where only two terms in the sum contributed. Now $\mathcal{E}_n - \mathcal{E}_{n+1} = -\hbar\omega$ and $\mathcal{E}_n - \mathcal{E}_{n-1} = \hbar\omega$, so eq. (13.2.9) simplifies to

$$E_n^{(2)} = -f^2/2m\omega^2, \tag{13.2.10}$$

and we have

$$E_n = \hbar\omega(n + 1/2) - f^2/2m\omega^2 + \cdots . \tag{13.2.11}$$

This agrees with the exact result of eq. (13.2.3). In fact, this comparison shows that the possible higher-order terms (indicated by the ellipsis here) must actually conspire to vanish. The first-order perturbative correction to the stationary state $|n\rangle$, applying eq. (13.1.46), also only has two terms in the sum,

$$\begin{aligned} |\psi_n^{(1)}\rangle &= -f\sqrt{\frac{\hbar}{2m\omega}}\left(\frac{\sqrt{n+1}|n{+}1\rangle}{-\hbar\omega} + \frac{\sqrt{n}|n{-}1\rangle}{\hbar\omega}\right) && (13.2.12)\\ &= \frac{f}{\sqrt{2m\hbar\omega^3}}\left(\sqrt{n+1}|n{+}1\rangle - \sqrt{n}|n{-}1\rangle\right). && (13.2.13)\end{aligned}$$

Again this agrees with the exact result in eq. (13.2.4), after the expansion for linear order in $f$ as in eq. (13.2.6).

As a second example, less amenable to an easy exact calculation, consider the one-dimensional potential

$$V(x) = \begin{cases} V_0 & (|x| < a/2), \\ 0 & (a/2 < |x| < L/2), \\ \infty & (|x| > L/2), \end{cases} \tag{13.2.14}$$

illustrated in Figure 13.2.1.

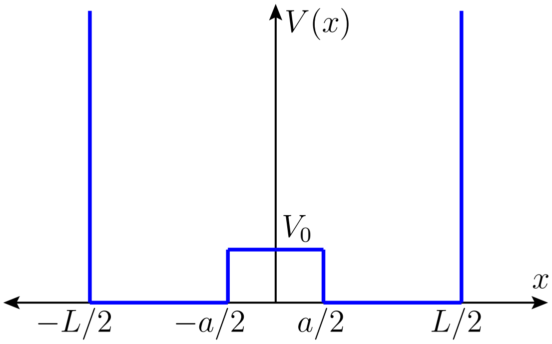

Figure 13.2.1: An infinite one-dimensional square well potential with width $L$, with a symmetric bump of height $V_0$ and width $a$ to be treated as a perturbation.

This is an infinite square well with a perturbation bump in the middle, so we define $H = H_0 + W$ where, using a hybrid operator/position-representation notation,

$$H_0 = \frac{P^2}{2m} + \begin{cases} 0 & (|x| < L/2), \\ \infty & (|x| > L/2), \end{cases} \tag{13.2.15}$$

and the perturbation Hamiltonian is the bump potential

$$W = \begin{cases} V_0 & (|x| < a/2), \\ 0 & (|x| > a/2). \end{cases} \tag{13.2.16}$$

The infinite square well problem defined by $H_0$ was solved in section 6.4. The wavefunctions $\langle x|n\rangle = \phi_n(x)$ for the unperturbed $H_0$ eigenstates $|n\rangle$ were found to be

$$\phi_n(x) = \sqrt{\frac{2}{L}}\cos(n\pi x/L) \qquad (n = 1, 3, 5, \ldots), \tag{13.2.17}$$

$$\phi_n(x) = \sqrt{\frac{2}{L}}\sin(n\pi x/L) \qquad (n = 2, 4, 6, \ldots), \tag{13.2.18}$$

with, in both cases, energies

$$\mathcal{E}_n = \frac{\hbar^2\pi^2 n^2}{2mL^2}. \tag{13.2.19}$$

Applying eq. (13.1.19) or its equivalent eq. (13.1.43), the leading correction to the energy eigenvalue for the state $n$ is

$$E_n^{(1)} = V_0 \int_{-a/2}^{a/2} dx\, |\phi_n(x)|^2. \tag{13.2.20}$$

For the ground state with $n = 1$, the energy eigenvalue to first order in $V_0$ evaluates to

$$E_1 = \frac{\hbar^2\pi^2}{2mL^2} + V_0\left(\frac{a}{L} + \frac{1}{\pi}\sin(\pi a/L)\right). \tag{13.2.21}$$

As a check, note that if $a = L$, then the perturbation simply adds $V_0$ to the energy eigenvalue, independent of $L$. This is just a constant shift in the Hamiltonian.

For the change in the ground-state wavefunction, we apply eq. (13.1.46). This requires us to find $W_{k1}/\mathcal{E}_{1k}$, which for odd $k \neq 1$ is

$$\frac{\langle k|W|1\rangle}{\mathcal{E}_1 - \mathcal{E}_k} = \frac{2mL^2}{\hbar^2\pi^2(1-k^2)}\frac{2V_0}{L}\int_{-a/2}^{a/2} dx\, \cos\left(\frac{\pi k x}{L}\right)\cos\left(\frac{\pi x}{L}\right) \tag{13.2.22}$$

$$= \frac{4V_0 mL^2}{\hbar^2\pi^3(1-k^2)}\left[\frac{\sin(\pi a(k-1)/2L)}{k-1} + \frac{\sin(\pi a(k+1)/2L)}{k+1}\right]. \tag{13.2.23}$$

For even $k$ we have instead

$$\frac{\langle k|W|1\rangle}{\mathcal{E}_1 - \mathcal{E}_k} = 0, \tag{13.2.24}$$

because of

$$\int_{-a/2}^{a/2} dx\, \sin\left(\frac{\pi k x}{L}\right)\cos\left(\frac{\pi x}{L}\right) = 0. \tag{13.2.25}$$

This reflects (sorry not sorry about the pun) the parity selection rule of eq. (5.4.17). Since the perturbation Hamiltonian has even parity $\pi_W = 1$, and $\pi_k = (-1)^{k-1}$ for a state $|k\rangle$, the parity selection rule tells us that $\langle k|W|n\rangle$ vanishes if $k+n$ is odd, and in particular that $\langle k|W|1\rangle = 0$ for even $k$. It follows that the corrected ground-state wavefunction is

$$\psi_1(x) = \sqrt{\frac{2}{L}}\cos\left(\frac{\pi x}{L}\right) + \psi_1^{(1)}(x), \tag{13.2.26}$$

with

$$\begin{aligned}\psi_1^{(1)}(x) &= \sum_{k=3,5,7,\ldots} \frac{\langle k|W|1\rangle}{\mathcal{E}_1 - \mathcal{E}_k}\phi_k(x) &(13.2.27)\\ &= \sum_{k=3,5,7,\ldots} \frac{4V_0 m L^2}{\hbar^2\pi^3(1-k^2)}\left[\frac{\sin(\pi a(k-1)/2L)}{k-1} + \frac{\sin(\pi a(k+1)/2L)}{k+1}\right]\sqrt{\frac{2}{L}}\cos\left(\frac{\pi k x}{L}\right). &(13.2.28)\end{aligned}$$

The corrections to the energies and wavefunctions of the other stationary states are left as an exercise.

For a third example, let us estimate the effect of the proton size on the energy levels of the hydrogen atom. In doing so, we will treat the proton (not completely realistically) as a ball of uniform charge density with radius $a_p$. For all $r \geq a_p$, the electric field is the same as if the proton were a point charge. For $r \leq a_p$, an elementary application of Gauss' Law in electrostatics shows that the electric field rises linearly with $r$, so that the potential energy of the electron is

$$V(r) = \begin{cases} -e^2\left(\dfrac{3}{2a_p} - \dfrac{r^2}{2a_p^3}\right), & (r \leq a_p),\\ -\dfrac{e^2}{r}, & (r \geq a_p),\end{cases} \tag{13.2.29}$$

as shown in Figure 13.2.2.

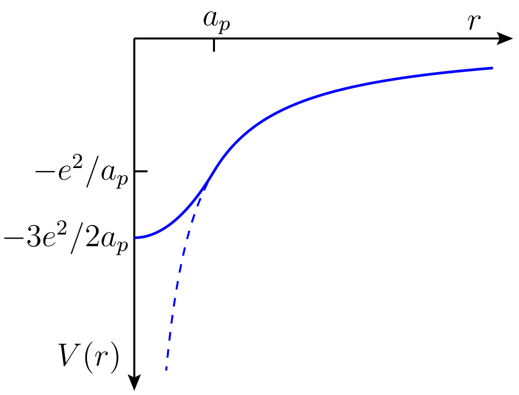

Figure 13.2.2: The solid line shows the shape of the potential energy for the electron in the hydrogen atom, with the proton modeled as a sphere of radius $a_p$ with uniform charge density. The dashed line shows the potential with the proton modeled as a point, as in section 10.1.

Therefore, compared to our treatment in section 10.1, the perturbation to the Hamiltonian is

$$W = \begin{cases} e^2\left(\frac{1}{r} - \frac{3}{2a_p} + \frac{r^2}{2a_p^3}\right), & (r \leq a_p), \\ 0, & (r \geq a_p), \end{cases} \tag{13.2.30}$$

in the position representation. So, the first-order correction to the ground state energy is

$$\begin{aligned} E^{(1)}_{1,0,0} &= \langle 1,0,0|W|1,0,0\rangle & (13.2.31) \\ &= 4\pi \int_0^{a_p} dr\, r^2 \left(\frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0}\right)^2 e^2 \left(\frac{1}{r} - \frac{3}{2a_p} + \frac{r^2}{2a_p^3}\right), & (13.2.32) \end{aligned}$$

where the $4\pi$ factor comes from the angular integration. Within the radial integral, $e^{-2r/a_0} \approx 1$ is an excellent approximation, since $a_p \ll a_0$. The integration then yields

$$E^{(1)}_{1,0,0} = \frac{2e^2 a_p^2}{5a_0^3} = \frac{4a_p^2}{5a_0^2} \text{ Rydberg} \approx 2.1 \times 10^{-10} \text{ Rydberg}, \tag{13.2.33}$$

where we have used $a_p \approx 8.5 \times 10^{-16}$ meters in the last approximation.

In the preceding, we used the small size of the proton compared to the Bohr radius to effectively replace the radial wavefunction by its (very nearly constant) value at $r = 0$, with $|R_{1,0}(r)|^2 \to |R_{1,0}(0)|^2$. For a general $n, l, m$ state, the same procedure will give

$$E^{(1)}_{n,l,m} = |R_{n,l}(0)|^2 \int d\Omega\, |Y_l^m(\theta,\phi)|^2 \int_0^{a_p} dr\, r^2\, e^2 \left(\frac{1}{r} - \frac{3}{2a_p} + \frac{r^2}{2a_p^3}\right). \tag{13.2.34}$$

Now, $\int d\Omega |Y_l^m|^2 = 1$, and $R_{n,l}(0) = 0$ unless $l = 0$, which also implies $m = 0$. From eq. (10.1.41), we know that the hydrogen atom radial wavefunction evaluated at $r = 0$ is $R_{n,l}(0) = \delta_{l,0} 2/(na_0)^{3/2}$. So, we find

$$E^{(1)}_{n,l,m} = \delta_{l,0}\, \delta_{m,0}\, \frac{4a_p^2}{5a_0^2 n^3} \text{ Rydberg}. \tag{13.2.35}$$

Thus the finite proton radius only affects the energies of the states with no angular momentum, and with decreasing importance for larger principal quantum number $n$.

The preceding calculation is just an estimate, as the proton is certainly not a ball of uniform charge density. A more sophisticated calculation can be used to define the effective charge radius of the proton, which can then be compared to various experimental determinations of the same quantity. This has recently been the subject of some interest and controversy, because experimental determinations of the proton charge radius inferred from electron scattering seem to be slightly larger than those obtained from measuring the energy levels of muonic hydrogen, in which a muon replaces the electron. The muon is used because its much larger mass gives it a smaller Bohr radius than the electron, leading to a larger effect in eq. (13.2.35) by a factor $m_\mu^2/m_e^2 \approx 42753$; the heavier muon is much more likely to be found inside the proton than an electron, and therefore provides a better probe.

## 13.3 Helium atom ground state from first-order perturbation theory

In this section, we consider the helium atom consisting of two electrons each with mass $m_e$ and charge $-e$, and a heavy nucleus with charge $Ze = +2e$ at the origin, as pictured in Figure 13.3.1.

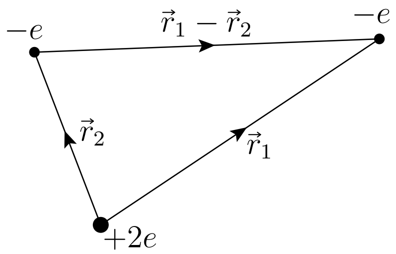

Figure 13.3.1: The helium atom consists of two electrons with mass $m_e$ and charge $-e$ at positions $\vec{r}_1$ and $\vec{r}_2$, and a very heavy nucleus of charge $+2e$ at the origin.

The Hamiltonian is[†]

$$H = H_1 + H_2 + H_{12}, \tag{13.3.1}$$

in which, in the position representation,

$$H_1 = -\frac{\hbar^2}{2m_e}\nabla_1^2 - \frac{Ze^2}{r_1}, \qquad H_2 = -\frac{\hbar^2}{2m_e}\nabla_2^2 - \frac{Ze^2}{r_2}, \tag{13.3.2}$$

$$H_{12} = \frac{e^2}{|\vec{r}_1 - \vec{r}_2|}, \tag{13.3.3}$$

where $Z = 2$, and $\vec{r}_1$ and $\vec{r}_2$ describe the position coordinates of the two electrons, and $\nabla_1^2$ and $\nabla_2^2$ are the corresponding Laplacians. If the electrons did not interact with each other, then $H_1$ and $H_2$ would just be the Hamiltonians for hydrogen-like atoms with $Z = 2$. The electron-electron interaction term $H_{12}$ is the source of complications, and in this section we treat it as the perturbation $W$, although it is not particularly small.

The lowest energy eigenstate of the unperturbed Hamiltonian $H_0 = H_1 + H_2$ is simply the tensor product of two hydrogen-like atomic ground states with $(n, l, m) = (1, 0, 0)$,

$$|(1,0,0),(1,0,0)\rangle = |1,0,0\rangle_1 \otimes |1,0,0\rangle_2. \tag{13.3.4}$$

The position wavefunction for the unperturbed state is

$$\psi(\vec{r}_1, \vec{r}_2) = \langle \vec{r}_1|1,0,0\rangle_1 \, \langle \vec{r}_2|1,0,0\rangle_2 = \psi_{1,0,0}(\vec{r}_1)\, \psi_{1,0,0}(\vec{r}_2), \tag{13.3.5}$$

where, as we found in eq. (10.1.42),

$$\psi_{1,0,0}(\vec{r}) = \sqrt{\frac{Z^3}{\pi a_0^3}}\, e^{-Zr/a_0}. \tag{13.3.6}$$

---

[†] For simplicity, we ignore small effects due to spin interactions and the finite mass of the nucleus. Also, we ignore the fact that the electrons are identical; this issue turns out to not affect the following treatment of the ground state because the electron spins will be oriented oppositely, but it does affect the excited states of helium, as we will discuss later, in section 16.3.

The corresponding unperturbed total energy eigenvalue for $H_1 + H_2$ is, including the contributions of both electrons,

$$\mathcal{E}_0 \;=\; 2\left(-\frac{Z^2e^2}{2a_0}\right) \;=\; -8\,\text{Rydberg} \;=\; -108.85\,\text{eV}. \tag{13.3.7}$$

This turns out to be about 30% lower than the experimental value (the minimum energy needed to completely ionize the helium atom). The reason is that the electron-electron Coulomb repulsive potential $H_{12}$ raises the energy, by an amount that we now estimate.

At first order in perturbation theory, the ground-state energy correction is

$$\begin{aligned} E^{(1)} &= \langle (1,0,0),(1,0,0)|H_{12}|(1,0,0),(1,0,0)\rangle && (13.3.8)\\ &= \int d^3\vec{r}_1 \int d^3\vec{r}_2\, |\psi_{1,0,0}(\vec{r}_1)|^2 |\psi_{1,0,0}(\vec{r}_2)|^2 \frac{e^2}{|\vec{r}_1-\vec{r}_2|}. && (13.3.9)\end{aligned}$$

We therefore have

$$E^{(1)} \;=\; e^2\left(\frac{Z^3}{\pi a_0^3}\right)^2 I, \tag{13.3.10}$$

with

$$I \;=\; \int_0^\infty dr_1\, r_1^2 \int d\Omega_1 \int_0^\infty dr_2\, r_2^2 \int d\Omega_2\, e^{-2Zr_1/a_0} e^{-2Zr_2/a_0} \frac{1}{|\vec{r}_1-\vec{r}_2|}. \tag{13.3.11}$$

Our challenge now is to evaluate the integral $I$. To do so, we use the valuable identity

$$\frac{1}{|\vec{r}_1-\vec{r}_2|} \;=\; \sum_{l=0}^{\infty} \frac{r_{\text{min}}^l}{r_{\text{max}}^{l+1}} P_l(\cos\gamma), \tag{13.3.12}$$

in which $P_l(x)$ are the Legendre polynomials, $\gamma$ is the angle between $\vec{r}_1$ and $\vec{r}_2$, and

$$r_{\text{max}} = \max(r_1, r_2), \qquad\qquad r_{\text{min}} = \min(r_1, r_2). \tag{13.3.13}$$

The identity (13.3.12), which may be familiar from the multipole expansion of problems with azimuthal symmetry in electrostatics, follows from the generating function eq. (8.6.46). The spherical harmonics addition formula eq. (8.6.72) tells us that

$$P_l(\cos\gamma) \;=\; \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^m(\theta_2,\phi_2)^*\, Y_l^m(\theta_1,\phi_1). \tag{13.3.14}$$

Using eqs. (13.3.12) and (13.3.14) in the integral $I$, we therefore have $\int d\Omega_1\, Y_l^m(\theta_1,\phi_1) \;=\; 0$ unless $l = m = 0$, and similarly for the $d\Omega_2$ integral. So, we can replace

$$P_l(\cos\gamma) \;\rightarrow\; 4\pi\, Y_0^0(\theta_2,\phi_2)^*\, Y_0^0(\theta_1,\phi_1)\delta_{l,0} \;=\; 4\pi\left(1/\sqrt{4\pi}\right)^2 \delta_{l,0} \;=\; \delta_{l,0}, \tag{13.3.15}$$

eliminating all of the angular dependences of the integrand. So, using $\int d\Omega_1 = \int d\Omega_2 = 4\pi$,

$$I = (4\pi)^2 \int_0^\infty dr_1\, r_1^2 \int_0^\infty dr_2\, r_2^2\, e^{-2Zr_1/a_0} e^{-2Zr_2/a_0} \frac{1}{r_{\text{max}}}. \tag{13.3.16}$$

Separating the $dr_2$ integral into two parts, according to whether $r_2 < r_1$ or $r_2 > r_1$, now gives

$$I = (4\pi)^2 \int_0^\infty dr_1 \left( \int_0^{r_1} dr_2\, r_1 r_2^2 + \int_{r_1}^\infty dr_2\, r_1^2 r_2 \right) e^{-2Zr_1/a_0} e^{-2Zr_2/a_0}, \tag{13.3.17}$$

which can be done straightforwardly. The two parts turn out to contribute equally to the total,

$$I = \frac{5\pi^2 a_0^5}{8Z^5}. \tag{13.3.18}$$

Putting this into eq. (13.3.10), the first-order correction to the ground-state energy is

$$E^{(1)} = \frac{5Ze^2}{8a_0}. \tag{13.3.19}$$

Combining this with eq. (13.3.7), and plugging in $Z = 2$, the total ground-state energy for helium estimated from first-order perturbation theory is

$$E = \mathcal{E}_0 + E^{(1)} = (-8 + 5/2)\,\text{Rydberg} = -74.83\,\text{eV}. \tag{13.3.20}$$

Note that the first-order correction is not suppressed by an adjustable small parameter compared to the unperturbed energy; it is 5/16 as large in magnitude. This might cause worry that the perturbative expansion is not converging fast enough to be trusted. Nevertheless, our final result differs by only about 5.3% from the experimental result of −79.01 eV. In section 14.4, we will use a non-perturbative technique, the variational method, to do even better.

## 13.4 Brillouin–Wigner perturbation theory

The standard organization of stationary-state perturbation theory summarized in eqs. (13.1.39)–(13.1.47) has an alternative, called **Brillouin–Wigner perturbation theory** after Léon Brillouin and Wigner, which sometimes has better numerical convergence behavior. To derive it, we start with the eigenvalue problem in the form

$$(H_0 + W - E_n)\,|\psi_n\rangle = 0, \tag{13.4.1}$$

using the same notations as in section 13.1, but with $\lambda = 1$ from the start. Now, act with the unperturbed bra $\langle m|$, then use $\langle m|H_0 = \langle m|\mathcal{E}_m$, and rearrange the result to get

$$\langle m|\psi_n\rangle = \frac{\langle m|W|\psi_n\rangle}{E_n - \mathcal{E}_m}. \tag{13.4.2}$$

Note the simultaneous presence in the denominator of the full energy eigenvalue $E_n$ and the unperturbed energy $\mathcal{E}_m$. From completeness of the unperturbed states $|m\rangle$, we also have

$$|\psi_n\rangle \;=\; \sum_m |m\rangle\langle m|\psi_n\rangle \;=\; |n\rangle\langle n|\psi_n\rangle + \sum_{m\neq n} |m\rangle\langle m|\psi_n\rangle\,. \tag{13.4.3}$$

As in section 13.1.47, we choose the normalization condition $\langle n|\psi_n\rangle = 1$, and use eq. (13.4.2) in (13.4.3) to obtain

$$|\psi_n\rangle \;=\; |n\rangle + \sum_{m\neq n} |m\rangle \frac{\langle m|W|\psi_n\rangle}{E_n - \mathcal{E}_m}. \tag{13.4.4}$$

This has a straightforward iterative solution,

$$\begin{aligned} |\psi_n\rangle \;=\; & |n\rangle + \sum_{m\neq n} |m\rangle \frac{1}{E_n - \mathcal{E}_m} \langle m|W|n\rangle \\ & + \sum_{m\neq n}\sum_{p\neq n} |m\rangle \frac{1}{E_n - \mathcal{E}_m} \langle m|W|p\rangle \frac{1}{E_n - \mathcal{E}_p} \langle p|W|n\rangle + \cdots, \end{aligned} \tag{13.4.5}$$

as can be checked by plugging it in. To find an equation for the corresponding energy eigenvalues, we now use eq. (13.1.38) to obtain

$$\begin{aligned} E_n \;=\; & \mathcal{E}_n + \langle n|W|n\rangle + \sum_{m\neq n} \langle n|W|m\rangle \frac{1}{E_n - \mathcal{E}_m} \langle m|W|n\rangle \\ & + \sum_{m\neq n}\sum_{p\neq n} \langle n|W|m\rangle \frac{1}{E_n - \mathcal{E}_m} \langle m|W|p\rangle \frac{1}{E_n - \mathcal{E}_p} \langle p|W|n\rangle + \cdots. \end{aligned} \tag{13.4.6}$$

Equations (13.4.5) and (13.4.6) summarize Brillouin–Wigner perturbation theory. They have a simpler structure than the standard perturbation theory of section 13.1, but have the disadvantage that they are not yet a complete solution, since the full energy $E_n$ appears on both sides of eq. (13.4.6).

At first order in Brillouin–Wigner perturbation theory, only the first two terms on the right-hand side of eq. (13.4.6) are included, and the solution for $E_n$ is the same as in standard perturbation theory. At second order or beyond, the equation for $E_n$ is a non-linear algebraic equation, and often can only be obtained numerically as an approximation. However, this can be a good thing, because the results so obtained are often more accurate than those found from the standard perturbation theory of section 13.1 at the same order. One way of thinking about this is that, at a given order in perturbation theory, the $E_n$ appearing in the denominators of Brillouin–Wigner perturbation theory contain more complete information than the $\mathcal{E}_n$ appearing in the denominators of standard perturbation theory. Once $E_n$ has been obtained by solving eq.(13.4.6) truncated at a specific order in the expansion, it can be plugged into eq. (13.4.5) to obtain the corresponding energy eigenstate.

## 13.5 Dalgarno–Lewis method for simplifying perturbation theory

The formulas of stationary-state perturbation theory found in section 13.1 for the state, and for the energy beyond first order, contain sums over all of the unperturbed states. For example, the second-order correction to the energy contains the sum

$$E_n^{(2)} = \sum_{m \neq n} \frac{|\langle m|W|n\rangle|^2}{\mathcal{E}_n - \mathcal{E}_m} \tag{13.5.1}$$

While this is a well-defined formal solution, in practice it is sometimes hard to compute such sums directly. It is especially difficult if some of the unperturbed states form a continuum, in which case the corresponding part of the summation will actually be an integration. Fortunately, there is an elegant method, due to A. Dalgarno and J.T. Lewis, that allows these sums to be recast into a form that can often be solved exactly. Even when an exact solution cannot be obtained, the **Dalgarno–Lewis method** still allows for a systematic approximation, often by numerically solving a differential equation.

Suppose that, given $|n\rangle$, one can find an operator $A$ that satisfies

$$[A, H_0]\,|n\rangle \;=\; (W - c)\,|n\rangle\,, \tag{13.5.2}$$

where $H_0$ and $W$ are the unperturbed and perturbation parts of the Hamiltonian, respectively, and $c$ is a constant. In fact, $c$ is not arbitrary, because after acting on both sides of eq. (13.5.2) with $\langle n|$, the left side vanishes, which implies $c = \langle n|W|n\rangle = E_n^{(1)}$. Also, since only the commutator of $A$ appears in eq. (13.5.2), one can always add a constant multiple of the identity operator to it, to ensure that

$$\langle n|A|n\rangle \;=\; 0. \tag{13.5.3}$$

Note that the required operator $A$ is different for each unperturbed state $|n\rangle$. Now, for every other unperturbed orthobasis state $|m\rangle$, eq. (13.5.2) implies

$$\langle m|W|n\rangle \;=\; \langle m|[A, H_0]|n\rangle + c\,\langle m|n\rangle \;=\; (\mathcal{E}_n - \mathcal{E}_m)\,\langle m|A|n\rangle\,, \tag{13.5.4}$$

where we have used $\langle m|n\rangle = 0$ by orthonormality of the unperturbed $H_0$ eigenstates. The Dalgarno–Lewis idea is to use eq. (13.5.4) to cancel the energy denominators in the results of perturbation theory, by expressing matrix elements of $W$ in terms of matrix elements of $A$. The resulting expressions can then be simplified using the completeness relation for the unperturbed states.

For example, the first-order correction to the state is, from eq. (13.1.24),

$$|\psi_n^{(1)}\rangle \;=\; \sum_{m \neq n} |m\rangle \frac{\langle m|W|n\rangle}{\mathcal{E}_n - \mathcal{E}_m} \;=\; \sum_{m \neq n} |m\rangle\langle m|A|n\rangle \;=\; \sum_{m} |m\rangle\,\langle m|A|n\rangle\,. \tag{13.5.5}$$

The last equality relies on the fact that the additional term with $m = n$ vanishes, because of eq. (13.5.3). Now, from the completeness relation $\sum_m |m\rangle\langle m| = I$, we conclude that the first-order correction to the state $|n\rangle$ is simply

$$|\psi_n^{(1)}\rangle = A\,|n\rangle\,. \tag{13.5.6}$$

In words, the desired operator $A$ turns the unperturbed state into the first-order correction to the state. Similarly, the second-order correction to the energy is, from eq. (13.5.1),

$$E_n^{(2)} = \sum_{m\neq n} \langle n|W|m\rangle\,\langle m|A|n\rangle = \sum_m \langle n|W|m\rangle\,\langle m|A|n\rangle = \langle n|WA|n\rangle\,, \tag{13.5.7}$$

where we have again used eq. (13.5.3) to account for the vanishing of the additional term with $m = n$, and then used completeness. As a check, this can be rewritten as

$$E_n^{(2)} = \langle n|W|\psi_n^{(1)}\rangle\,, \tag{13.5.8}$$

in accord with eq. (13.1.27) with $k = 2$. In a similar way, it is not too hard to show that even the third-order correction to the energy can be written as

$$E_n^{(3)} = \langle\psi_n^{(1)}|W|\psi_n^{(1)}\rangle - \langle\psi_n^{(1)}|\psi_n^{(1)}\rangle\langle n|W|n\rangle\,, \tag{13.5.9}$$

without energy-difference denominators.

To make use of the Dalgarno–Lewis method, one must either find the operator $A$, or equivalently (and more directly) solve for the ket $A|n\rangle = |\psi_n^{(1)}\rangle$. Rewriting eq. (13.5.2), the key equation to be solved is

$$(\mathcal{E}_n - H_0)\,|\psi_n^{(1)}\rangle = W|n\rangle - |n\rangle\,\langle n|W|n\rangle\,. \tag{13.5.10}$$

In favorable cases, this can be solved exactly for $|\psi_n^{(1)}\rangle$ or the corresponding wavefunction; we will see an example of this in section 13.7 when we treat the ground state of the hydrogen atom in an electric field (the quadratic Stark effect). Even when this is not feasible, it is often possible to cast eq. (13.5.10) in the position representation as a differential equation that can be solved approximately, perhaps numerically. This is often faster and simpler than trying to do the sums over all other states $|m\rangle$ with energy denominators $\mathcal{E}_n - \mathcal{E}_m$, especially if there are an infinite number of such states. The great advantage of the Dalgarno–Lewis method is that eq. (13.5.10) only involves the eigenvalues and matrix elements of the unperturbed state $|n\rangle$ that one is considering. Once $|\psi_n^{(1)}\rangle$ has been found, the second and third-order corrections to the energy follow immediately from eqs. (13.5.8) and (13.5.9), respectively.

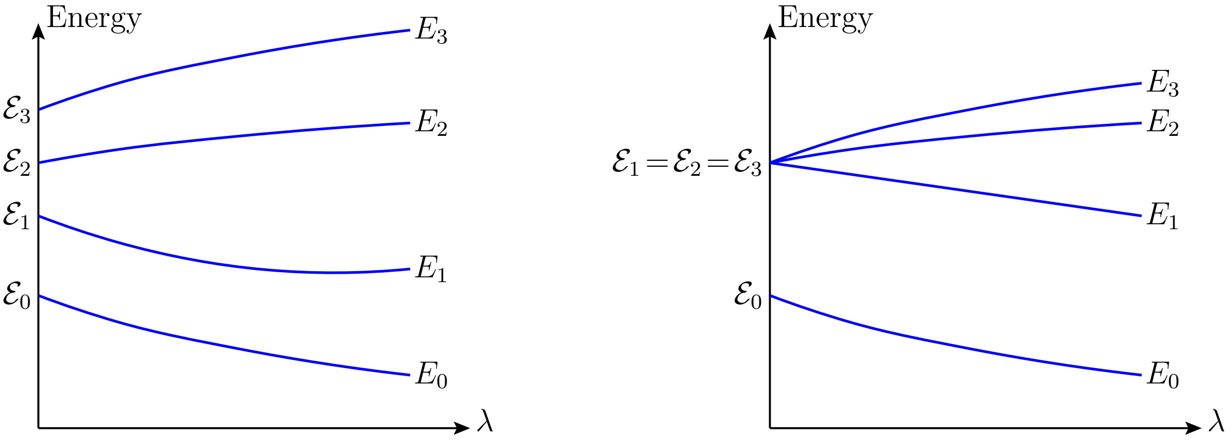

Figure 13.6.1: Perturbed energies as a function of the expansion parameter $\lambda$, for non-degenerate (left) and degenerate (right) perturbation theory. In degenerate perturbation theory, for each of the perturbed states $|\psi_1\rangle$, $|\psi_2\rangle$, $|\psi_3\rangle$ there is an ambiguity in choosing the corresponding unperturbed states $|1\rangle$, $|2\rangle$, $|3\rangle$, since arbitrary linear combinations of them have the same $H_0$ eigenvalue $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{E}_3$. This ambiguity is resolved (at least in part) by choosing the unperturbed states to diagonalize the perturbation Hamiltonian $W$ within each subspace of degenerate $\mathcal{E}_n$.

## 13.6 Degenerate perturbation theory

As noted in section 13.1, a special problem arises when unperturbed states $|n\rangle$ have degeneracies, because then the energy differences $\mathcal{E}_{nm} = \mathcal{E}_n - \mathcal{E}_m$ appearing in the denominators of eqs. (13.1.44)–(13.1.47) will vanish. In this section, we show how to deal with this difficulty.

The bad denominator problem is related to the fact that perturbation theory requires a smooth one-to-one correspondence between the unperturbed states and the perturbed states as the expansion parameter $\lambda$ varies. Specifically, this is what allowed us to demand that the unperturbed state $|n\rangle$ and the corrections to it $|\psi_n^{(k)}\rangle$ are orthogonal, as we did in eq. (13.1.12). This in turn was used to exclude the terms in the sums that would otherwise have had vanishing energy difference denominators. In the degenerate case, the essential problem is that for a given perturbed state $|\psi_n\rangle$ there is an ambiguity in deciding which of the degenerate unperturbed states $|n\rangle$ corresponds to it and therefore can be excluded from the sums, since arbitrary linear combinations of them have the same $H_0$ eigenvalue $\mathcal{E}_n$. This is illustrated in Figure 13.6.1, which compares the situations for non-degenerate and degenerate unperturbed states.

To resolve the problem, we need to answer the following question: if a group of unperturbed states $\{|n\rangle\}$ have the same unperturbed energy $\mathcal{E}_n$, then which linear combinations of them are the $|\psi_n^{(0)}\rangle$? The problem of avoiding bad (vanishing) denominators in perturbation theory is the same as the problem of choosing the correct linear combinations of unperturbed energy states.

Consider what happens when we return to eq. (13.1.23), rewritten here for convenience,

$$(\mathcal{E}_n - \mathcal{E}_m) \langle m|\psi_n^{(1)}\rangle \;=\; \langle m|W|n\rangle \qquad (m \neq n), \tag{13.6.1}$$

but now allowing for the possibility that some unperturbed states have degenerate unperturbed energies. Whenever $\mathcal{E}_m \neq \mathcal{E}_n$, we can still solve

$$\langle m|\psi_n^{(1)}\rangle \;=\; \frac{W_{mn}}{\mathcal{E}_{nm}}, \qquad (\mathcal{E}_m \neq \mathcal{E}_n). \tag{13.6.2}$$

However, for $\mathcal{E}_m = \mathcal{E}_n$, we instead learn only that $\langle m|W|n\rangle$ must vanish for $m \neq n$, in order to have a consistent perturbative solution. It is therefore required that the matrix elements of $W$ form a diagonal matrix when restricted to each subspace with degenerate unperturbed energies,

$$\langle m|W|n\rangle \;=\; \delta_{nm} W_{nn} \qquad (\text{for } \mathcal{E}_m = \mathcal{E}_n). \tag{13.6.3}$$

Fortunately, this is always possible to arrange by a suitable choice of the orthobasis of stationary states $|n\rangle$. This is assured by Theorem 2.6.6, since $W$ is Hermitian.

Equation 13.6.3 is a necessary (but not always sufficient, as we will see) requirement on the choice of unperturbed states. The point is that by choosing a "good" basis of unperturbed states in this way, the perturbation Hamiltonian matrix $W_{mn}$ does not connect pairs of distinct unperturbed states for which the energy-difference denominators would vanish.

In practice, one must solve the eigenvalue problem for $g_{\mathcal{E}_n} \times g_{\mathcal{E}_n}$ matrices, formed by the matrix elements of the operator $W$ restricted to each $g_{\mathcal{E}_n}$-dimensional degenerate subspace. Once we choose the states $|n\rangle$ as the resulting orthonormal eigenvectors, then it follows from eq. (13.1.19) that the eigenvalues are the corresponding first-order corrections to the energies,

$$E_n^{(1)} = W_{nn}, \tag{13.6.4}$$

the same result as for the non-degenerate case. Note that it is only necessary to diagonalize $W$ within the subspaces for degenerate $\mathcal{E}_n$. In particular, for each unperturbed state $|n\rangle$ that is non-degenerate, there is nothing to do; $W$ restricted to that subspace is simply a 1 × 1 matrix, so $|n\rangle$ is automatically an eigenvector, with eigenvalue $W_{nn} = \langle n|W|n\rangle$.

Often, only the first-order energy corrections are required; then eq. (13.6.4) is sufficient together with eq. (13.6.3), and in that case you can skip the complications of the next few paragraphs, from here to eq. (13.6.15).

If one needs the first-order corrections to the state ket $|\psi_n^{(1)}\rangle$ and the second-order corrections to the energies $E_n^{(2)}$, then one must confront the fact that eq. (13.6.1) provides no information about $\langle m|\psi_n^{(1)}\rangle$ for $\mathcal{E}_m = \mathcal{E}_n$. By completeness,

$$|\psi_n^{(1)}\rangle \;=\; \sum_m{}' \, |m\rangle \frac{W_{mn}}{\mathcal{E}_{nm}} + \sum_m{}'' \, |m\rangle \langle m|\psi_n^{(1)}\rangle \,, \tag{13.6.5}$$

where we have introduced the notations, always for fixed $n$,

$$\sum_m{}' \quad = \quad \text{sum over all } m \text{ such that } \mathcal{E}_m \neq \mathcal{E}_n, \tag{13.6.6}$$

$$\sum_m{}'' \quad = \quad \text{sum over all } m \text{ such that } m \neq n \text{ and } \mathcal{E}_m = \mathcal{E}_n, \tag{13.6.7}$$

and used eq. (13.6.2) in the first summation. To solve for the remaining matrix elements $\langle m|\psi_n^{(1)}\rangle$ in the last summation, one must use the equations governing the perturbative expansion at order $\lambda^2$, and possibly beyond.

To see how this goes, consider eq. (13.1.29) with $k = 2$ and $\mathcal{E}_m = \mathcal{E}_n$, which yields

$$E_n^{(1)} \langle m|\psi_n^{(1)}\rangle \;=\; \langle m|W|\psi_n^{(1)}\rangle\,, \qquad \text{(for } m \neq n \text{ and } \mathcal{E}_m = \mathcal{E}_n\text{)}. \tag{13.6.8}$$

After using $E_n^{(1)} = W_{nn}$, and eq. (13.6.5) with $m$ replaced by $p$, and defining an operator $W^{(2)}$ with matrix elements

$$W^{(2)}_{mn} \;\equiv\; \sum_p{}' \frac{W_{mp}W_{pn}}{\mathcal{E}_{np}}, \tag{13.6.9}$$

eq. (13.6.8) becomes

$$W_{nn} \langle m|\psi_n^{(1)}\rangle \;=\; W^{(2)}_{mn} + \sum_p{}'' W_{mp} \langle p|\psi_n^{(1)}\rangle\,, \qquad \text{(for } m \neq n \text{ and } \mathcal{E}_m = \mathcal{E}_n\text{)}. \tag{13.6.10}$$

In the last summation, we know that $W_{mp} = \delta_{mp} W_{mm}$, because of eq. (13.6.3). Therefore, only the single term with $p = m$ in that sum survives, and eq. (13.6.10) reduces to

$$(W_{nn} - W_{mm}) \langle m|\psi_n^{(1)}\rangle \;=\; W^{(2)}_{mn}, \qquad \text{(for } m \neq n \text{ and } \mathcal{E}_m = \mathcal{E}_n\text{)}. \tag{13.6.11}$$

Now, for states that are non-degenerate at first order, so that $W_{mm} \neq W_{nn}$, we obtain

$$\langle m|\psi_n^{(1)}\rangle \;=\; \frac{W^{(2)}_{mn}}{W_{nn} - W_{mm}} \qquad \text{(for } \mathcal{E}_m = \mathcal{E}_n \text{ and } W_{mm} \neq W_{nn}\text{)}, \tag{13.6.12}$$

but in the opposite case we learn only that

$$W^{(2)}_{mn} \;=\; 0, \qquad \text{(for } m \neq n \text{ and } \mathcal{E}_m = \mathcal{E}_n \text{ and } W_{mm} = W_{nn}\text{)}. \tag{13.6.13}$$

This is a second condition that must be imposed on the choice of the unperturbed states, in addition to eq. (13.6.3). In words, we have found that we must choose the kets $|n\rangle$ so that on subspaces where both $H_0$ and $W$ are proportional to the identity, the matrix $W^{(2)}$ is also diagonal. Fortunately, this is always possible, because $W^{(2)}$ is a Hermitian matrix when restricted to this subspace. Now, in general, the second-order correction to the energy is obtained from eq. (13.1.27), and becomes simply

$$E_n^{(2)} = W_{nn}^{(2)} = \sum_m{}' \frac{|W_{mn}|^2}{\mathcal{E}_{nm}}, \tag{13.6.14}$$

where it is crucial that the unperturbed states satisfy the conditions given in both eqs. (13.6.3) and (13.6.13), and there is no vanishing denominator problem because of the fact that the sum excludes all terms with $\mathcal{E}_m = \mathcal{E}_n$.

There may still be some ambiguity in the matrix elements $\langle m|\psi_n^{(1)}\rangle$, if for some $m \neq n$ one has $\mathcal{E}_m = \mathcal{E}_n$ and $W_{mm} = W_{nn}$ and $W_{nn}^{(2)} = W_{mm}^{(2)}$, so that the degeneracy is not completely lifted even at second order. In that case, one must use the information from third order, that is, $k = 3$ in eq. (13.1.17). One then finds a third condition on the choice of unperturbed states $|n\rangle$, that the operator $W^{(3)}$ with matrix elements

$$W_{mn}^{(3)} \equiv \sum_q{}' \sum_p{}' \frac{W_{mq}W_{qp}W_{pn}}{\mathcal{E}_{nq}\mathcal{E}_{np}} - W_{nn} \sum_p{}' \frac{W_{mp}W_{pn}}{\mathcal{E}_{np}^2} \tag{13.6.15}$$

is diagonal on each subspace of states where the operators $H_0$, $W$, and $W^{(2)}$ are proportional to the identity operator. If the degeneracy of unperturbed states is still not removed at third order (that is, $W^{(3)}$ has degenerate eigenvalues when restricted to that subspace of states), one must continue to $k = 4$ in eq. (13.1.17), etc. These considerations from $k \geq 3$ affect the determination of $|\psi_n^{(1)}\rangle$ (through $\langle m|\psi_n^{(1)}\rangle$) even though they seemingly involve higher orders in perturbation theory, but they do not affect the determination of $E_n^{(2)}$ given by eq. (13.6.14).

Practical applications of degenerate perturbation theory often only need $E_n^{(1)}$ as given in eq. (13.6.4), and therefore one only needs to worry about choosing the unperturbed states to satisfy eq. (13.6.3). Examples will appear in sections 13.7 and 13.8 and in Chapter 15. To conclude this section, we exploit first-order perturbation theory to infer the exact matrix element $\langle 1/R^2\rangle$ for the states $|n, l, m\rangle$ of the hydrogen atom, fulfilling a promise made in section 10.1. Consider the Hamiltonian $H_\lambda = H_0 + \lambda/R^2$, where $H_0 = P^2/2m_e - e^2/R$ is the unperturbed hydrogen atom Hamiltonian, and $\lambda$ is a small parameter. In this case, the operator $1/R^2$ commutes with the angular momentum operators $L^2$ and $L_z$, so it is diagonal in the orthobasis of states $|n, l, m\rangle$ for fixed $n$ and different $l, m$. Therefore $|n, l, m\rangle$ is a good basis for degenerate first-order perturbation theory in $\lambda$, and the first-order correction to the energy eigenvalue for each state is simply proportional to the expectation value that we seek,

$$E_{n,l,m}^{(1)} = \lambda\langle 1/R^2\rangle_{n,l,m}. \tag{13.6.16}$$

Meanwhile, the exact energy eigenvalues $E_{\lambda,k,l,m}$ of $H_\lambda$ are the solutions to the radial wavefunction eigenvalue problem

$$\left[-\frac{\hbar^2}{2m_e}\left(\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}-\frac{l(l+1)}{r^2}\right)+\frac{\lambda}{r^2}\right]R_{k,l} \;=\; E_{\lambda,k,l,m}R_{k,l}, \tag{13.6.17}$$

where the integer $k$ is the radial quantum number (the number of zeros of $R_{k,l}(r)$ for $0 < r < \infty$). Now we can define a quantity $l'$ by

$$\frac{\hbar^2}{2m_e}l'(l'+1) \;=\; \frac{\hbar^2}{2m_e}l(l+1)+\lambda, \tag{13.6.18}$$

or, equivalently,

$$l' \;=\; \sqrt{(l+1/2)^2+2m_e\lambda/\hbar^2}-1/2. \tag{13.6.19}$$

Note that $l'$ need not be an integer, unlike $l$. Nevertheless, the analysis of the radial wavefunction eigenvalue problem that led to eq. (10.1.16) goes through, but with $l$ replaced everywhere by $l'$, and $n$ replaced by $k+l'+1$. In particular, it is crucial that $k$ is a non-negative integer, in order that the wavefunction is normalizable. The result is

$$E_{\lambda,k,l,m} \;=\; -\frac{e^2}{2a_0(k+l'+1)^2}. \tag{13.6.20}$$

Expanding in small $\lambda$, for fixed integers $k$ and $l$,

$$E_\lambda \;=\; E_{\lambda=0}+\lambda\left.\frac{dE_\lambda}{d\lambda}\right|_{\lambda=0}+\cdots, \tag{13.6.21}$$

and by comparing to eq. (13.6.16), we find

$$\langle 1/R^2\rangle_{n,l,m} \;=\; \frac{d}{d\lambda}\left.\left(\frac{-e^2}{2a_0(k+l'+1)^2}\right)\right|_{\lambda=0} \;=\; \frac{e^2}{a_0(k+l+1)^3}\left.\frac{dl'}{d\lambda}\right|_{\lambda=0}. \tag{13.6.22}$$

From eq. (13.6.19), we have

$$\left.\frac{dl'}{d\lambda}\right|_{\lambda=0} \;=\; \frac{m_e}{\hbar^2(l+1/2)}. \tag{13.6.23}$$

Plugging eq. (13.6.23) into eq. (13.6.22), and using $n=k+l+1$, we obtain

$$\langle 1/R^2\rangle_{n,l,m} \;=\; \frac{m_e e^2}{\hbar^2 a_0 n^3(l+1/2)} \;=\; \frac{1}{a_0^2 n^3(l+1/2)}. \tag{13.6.24}$$

Equation (10.1.67) follows by restoring the general nuclear charge $Z$, using the rule $a_0 \to a_0/Z$.

The same sort of perturbation theory trick can be used to check the result we found for $\langle 1/R\rangle_{n,l,m}$ in eq. (10.1.64). Consider $H_Z = P^2/2m_e - Ze^2/R$, the Hamiltonian for general $Z$. We rewrite this as $H_Z = H_0 + W$, where $H_0 = P^2/2m_e - e^2/R$ is the $Z=1$ Hamiltonian, and $W = (1-Z)e^2/R$ is treated as a perturbation by taking $Z$ to be close to 1. Then, for an eigenstate of $H_0$ with principal quantum number $n$, the first-order energy correction is

$$E_n^{(1)} \;=\; (1-Z)e^2 \langle 1/R \rangle_{n,l,m}, \tag{13.6.25}$$

which is proportional to the expectation value that we seek. (Again we are making use of the fact that the basis $|n, l, m\rangle$ is a good one for first-order degenerate perturbation theory, in the sense that the perturbation $1/R$ is diagonal for fixed $n$ and different $l, m$.) On the other hand, the exact energies were found in eq. (10.1.16) to be

$$E_n \;=\; -\frac{Z^2 e^2}{2a_0 n^2} \;=\;=\; -\frac{e^2}{2a_0 n^2}\left[1 + 2(Z-1) + (Z-1)^2\right]. \tag{13.6.26}$$

Comparing the terms linear in small $Z-1$ in eqs. (13.6.25) and (13.6.26), we must have

$$\langle 1/R \rangle_{n,l,m} \;=\; \frac{1}{a_0 n^2}. \tag{13.6.27}$$

This confirms eq. (10.1.64) with $Z = 1$.

## 13.7 Hydrogen atom in electric field (Stark effect)

Consider a hydrogen atom in a static external electric field $\vec{E} = E\hat{z}$. The corresponding electrostatic potential seen by the electron is $\Phi(\vec{r}) = -Ez$, giving a potential energy contribution to the Hamiltonian,

$$W \;=\; -e\Phi \;=\; eEZ, \tag{13.7.1}$$

which we will treat as a perturbation. The calculation of shifts in the atomic energy levels, known as the the Stark effect, provides a nice example of both non-degenerate second-order and degenerate first-order perturbation theory.

For the ground state $|n, l, m\rangle = |1, 0, 0\rangle$, the first-order correction to the energy vanishes,

$$E_{1,0,0}^{(1)} \;=\; \langle 1,0,0|W|1,0,0\rangle \;=\; eE\langle 1,0,0|Z|1,0,0\rangle \;=\; 0. \tag{13.7.2}$$

This can be understood as a consequence of the parity selection rule, eq. (5.4.17). More generally, the expectation value of $Z$ in any parity eigenstate is always 0, because $\Pi Z \Pi = -Z$, so that $\pi_Z = -1$, giving $\pi_Z \pi_{n,l,m}^2 = -1$. Another way to see that the matrix element vanishes is the dipole selection rule of eq. (12.2.15).

Since the first-order correction to the ground-state energy is 0, we turn to the second-order correction. Applying the general formula of eq. (13.1.44), and remembering that the unperturbed energies do not depend on $l$ and $m$, we have

$$E^{(2)}_{1,0,0} \;=\; e^2E^2\sum_{n=2}^{\infty}\sum_{l=0}^{n-1}\sum_{m=-l}^{l}\frac{|\langle n,l,m|Z|1,0,0\rangle|^2}{\mathcal{E}_1-\mathcal{E}_n}+e^2E^2\int_0^{\infty}d\mathcal{E}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{|\langle \mathcal{E},l,m|Z|1,0,0\rangle|^2}{\mathcal{E}_1-\mathcal{E}}. \tag{13.7.3}$$

Note that this includes integration over the unbound states $|\mathcal{E},l,m\rangle$ with continuous unperturbed energies $\mathcal{E}>0$ as well as summation over the bound states $|n,l,m\rangle$ with discrete unperturbed energies $\mathcal{E}_n = -e^2/2a_0n^2$, excluding $n=1$. Equation (13.7.3) simplifies considerably if we use the dipole selection rules of eqs. (12.2.15) and (12.2.16), which inform us that only the $l=1$ and $m=0$ matrix elements are non-zero. Therefore, it can be rewritten as

$$E^{(2)}_{1,0,0} \;=\; -e^2E^2\left(\sum_{n=2}^{\infty}\frac{|\langle n,1,0|Z|1,0,0\rangle|^2}{\mathcal{E}_n-\mathcal{E}_1}\;+\;\int_0^{\infty}d\mathcal{E}\,\frac{|\langle \mathcal{E},1,0|Z|1,0,0\rangle|^2}{\mathcal{E}-\mathcal{E}_1}\right), \tag{13.7.4}$$

where the overall minus sign appears because the denominators have flipped sign to make them positive. Since both the sum and the integral are clearly positive, eq. (13.7.4) is negative; the presence of the electric field lowers the hydrogen atom energy.

The evaluation of eq. (13.7.4) is not trivial, so we will study it using a series of tricks with instructive value. First, we will find a simple lower bound on its magnitude, then a simple upper bound, and then attempt a numerical approximation by including only the bound state sum contribution. Finally, we will obtain the exact value of $E^{(2)}_{1,0,0}$ by exploiting the power of the Dalgarno–Lewis method of section 13.5.

To obtain the lower bound, note that the magnitude of eq. (13.7.4) is certainly greater than that of the $n=2$ term alone, since all of the contributions have the same sign. So, using $\mathcal{E}_2-\mathcal{E}_1 = 3e^2/8a_0$,

$$|E^{(2)}_{1,0,0}| \;>\; (n\!=\!2 \text{ term only}) \;=\; \frac{8}{3}a_0E^2|\langle 2,1,0|Z|1,0,0\rangle|^2. \tag{13.7.5}$$

The relevant matrix element is

$$\langle 2,1,0|Z|1,0,0\rangle \;=\; \int_0^{2\pi}d\phi\int_{-1}^{1}d(\cos\theta)\int_0^{\infty}dr\,r^2\,(R_{21}Y^0_1)^*\,r\cos\theta\,(R_{10}Y^0_0) \;=\; \frac{128\sqrt{2}}{243}a_0. \tag{13.7.6}$$

Therefore, we have

$$|E^{(2)}_{1,0,0}| \;>\; \frac{2^{18}}{3^{11}}E^2a_0^3 \;\approx\; 1.4798E^2a_0^3 \tag{13.7.7}$$

as a lower bound on the energy shift of the ground state due to the external electric field $E$.

We also obtain a bound from the other direction, by noting that the denominators in eq. (13.7.4) with $n \neq 2$ are all larger than they would be if each instance of $\mathcal{E}_n$ or $\mathcal{E}$ were replaced by $\mathcal{E}_2$. So,

$$|E^{(2)}_{1,0,0}| \;<\; \frac{8}{3}a_0 E^2 \left( \sum_{n=2}^{\infty} |\langle n,1,0|Z|1,0,0\rangle|^2 + \int_0^\infty d\mathcal{E} |\langle \mathcal{E},1,0|Z|1,0,0\rangle|^2 \right). \tag{13.7.8}$$

Now that the annoying energy-difference denominators are gone, we can evaluate the whole quantity in parentheses using the completeness identity of eq. (10.2.22), remembering that the missing $l \neq 1$ and $m \neq 0$ terms vanish by the dipole selection rules. The result is

$$|E^{(2)}_{1,0,0}| \;<\; \frac{8}{3}a_0 E^2 \, \langle 1,0,0|Z^2|1,0,0\rangle. \tag{13.7.9}$$

Another trick: by rotational symmetry of the ground state, we can replace $Z^2$ in the expectation value by the average of $X^2$, $Y^2$, and $Z^2$, which is easier to compute. So

$$\langle 1,0,0|Z^2|1,0,0\rangle \;=\; \frac{1}{3}\langle 1,0,0|(X^2+Y^2+Z^2)|1,0,0\rangle \;=\; \frac{1}{3}\langle 1,0,0|R^2|1,0,0\rangle \;=\; a_0^2, \tag{13.7.10}$$

where, at the end, we have used eq. (10.1.53) with $p = 2$. Thus we have

$$|E^{(2)}_{1,0,0}| \;<\; \frac{8}{3}E^2 a_0^3 \;\approx\; 2.6667 a_0^3 E^2 \tag{13.7.11}$$

as an upper bound on the energy shift to complement the lower bound in eq. (13.7.7).

As an attempt to do better than the bounds in eqs. (13.7.7) and (13.7.11), one can add up the exact contributions from all bound states $n = 2, 3, 4, \ldots$ with $\mathcal{E}_n = -e^2/2a_0 n^2$ in eq. (13.7.4). It is possible, but quite non-trivial, to show that

$$e^2 \frac{|\,\langle n,1,0|Z|1,0,0\rangle\,|^2}{\mathcal{E}_n - \mathcal{E}_1} \;=\; \frac{(2n)^9 (n-1)^{2n-6}}{3(n+1)^{2n+6}}\, a_0^3. \tag{13.7.12}$$

The sum over all integers $n \geq 2$ in eq. (13.7.4) is then found to converge to a numerical value

$$E^{(2),\,\text{partial}}_{1,0,0} \;\approx\; -1.8316 E^2 a_0^3, \tag{13.7.13}$$

which is indeed between the two bounds that we have already obtained. However, this turns out to be numerically far from the true answer; as we are about to discover, the integral over unbound states contributes significantly.

Fortunately, one can obtain the correct result while avoiding having to perform directly the difficult sum plus integration, by instead using the Dalgarno–Lewis method described in section 13.5. To do this, we must first use eq. (13.5.10) to solve for the first-order correction to the ground-state wavefunction, $\psi^{(1)}_{1,0,0}(\vec{r})$. In the position representation, this equation is

$$(\mathcal{E}_1 - H_0)\psi^{(1)}_{1,0,0} \;=\; eE\, r\cos\theta\, \psi^{(0)}_{1,0,0}, \tag{13.7.14}$$

where $\mathcal{E}_1 = -e^2/2a_0$ and $\psi^{(0)}_{1,0,0} = e^{-r/a_0}/\sqrt{\pi a_0^3}$ from the results of section 10.1, and we have used the fact that $W = eEZ$ has vanishing expectation value in the unperturbed ground state. The unperturbed Hamiltonian operator is

$$H_0 \;=\; -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{r} \;=\; -\frac{a_0 e^2}{2}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right] - \frac{e^2}{r}, \quad (13.7.15)$$

where, in the second equality, we have used the fact that there is no $\phi$ dependence because of the symmetry of the problem about the $z$-axis, and $\hbar^2/2m_e = a_0e^2/2$. Given the form of eq. (13.7.14), it is reasonable to try for a solution of the form

$$\psi^{(1)}_{1,0,0} \;=\; f(r)\cos\theta\,\psi^{(0)}_{1,0,0}. \quad (13.7.16)$$

With this guess, the differential equation (13.7.14) becomes

$$\frac{1}{2}f'' + \left(\frac{1}{r} - \frac{1}{a_0}\right)f' - \frac{1}{r^2}f \;=\; \frac{E}{ea_0}r, \quad (13.7.17)$$

in which we are delighted to see that the $\theta$ dependence has disappeared. Since the inhomogeneous part is linear in $r$, we are inspired to make the further guess that $f$ is a quadratic polynomial in $r$, and are rewarded with the simple solution

$$f \;=\; -\frac{E}{e}(a_0 + r/2)r. \quad (13.7.18)$$

[This corresponds to the Dalgarno–Lewis operator $A = -\frac{E}{e}(a_0 + R/2)Z$ as the solution to eqs. (13.5.2) and (13.5.3), although we do not directly need it.]

Having successfully found $\psi^{(1)}_{1,0,0}$, the rest is relatively easy. We get, using eq. (13.5.8) in the position representation,

$$E^{(2)}_{1,0,0} = \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \int_0^\infty dr\, r^2 \left(eEr\cos\theta\right)\left(-\frac{E}{e}(a_0 + r/2)r\cos\theta\right)|\psi^{(0)}_{1,0,0}|^2, \quad (13.7.19)$$

which evaluates finally to the correct, and surprisingly simple, result

$$E^{(2)}_{1,0,0} \;=\; -\frac{9}{4}a_0^3E^2 \;=\; -2.25a_0^3E^2. \quad (13.7.20)$$

Because the leading correction to the ground state energy comes from second order in perturbation theory, and is therefore quadratic in the applied electric field, this is called the **quadratic Stark effect**.

Let us now see how to connect eq. (13.7.20) to an experimental measurement. In general, for a dipole moment $\vec{p}$ induced by an external electric field $\vec{E}$, the polarizability $\alpha$ is defined by

$$\vec{p} \;=\; \alpha\vec{E}. \quad (13.7.21)$$

As the electric field is increased, the change in the energy of the dipole is

$$dU \;=\; -\vec{p}\cdot d\vec{E} \;=\; -\alpha\, E\, dE, \tag{13.7.22}$$

so that the total energy of the induced dipole is

$$U \;=\; -\frac{1}{2}\alpha E^2. \tag{13.7.23}$$

Comparing to eq. (13.7.20), we have $-\frac{9}{4}a_0^3E^2 = -\frac{1}{2}\alpha E^2$, or

$$\alpha \;=\; \frac{9}{2}a_0^3 \;=\; 6.67\times 10^{-31}\, m^3, \tag{13.7.24}$$

which agrees well with the experimental value for the polarizability of atomic hydrogen in the ground state.

Let us now turn to the Stark effect for the first excited ($n = 2$) stationary states of the hydrogen atom. The unperturbed states are

$$|n,l,m\rangle = |2,0,0\rangle,\;\; |2,1,0\rangle,\;\; |2,1,1\rangle,\;\; \text{and}\;\; |2,1,-1\rangle, \tag{13.7.25}$$

which all have the same unperturbed energy $\mathcal{E}_2 = -e^2/8a_0$. We therefore need to apply the method of degenerate perturbation theory as outlined in section 13.6. To start, we must choose the unperturbed orthobasis of $n = 2$ states as the eigenkets of the $4\times 4$ matrix

$$W_{(l',m'),\,(l,m)} \;=\; \langle 2,l',m'|eEZ|2,l,m\rangle. \tag{13.7.26}$$

Using the dipole selection rule of eqs. (12.2.15) and (12.2.16), the only non-zero entries of this matrix come from

$$\langle 2,1,0|Z|2,0,0\rangle \;=\; \langle 2,0,0|Z|2,1,0\rangle, \tag{13.7.27}$$

which can be evaluated as

$$\begin{aligned}
\langle 2,1,0|Z|2,0,0\rangle \;&=\; \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \int_0^\infty dr\, r^2 \left(R_{21}Y_1^0\right)^* \big(r\cos\theta\big)\left(R_{20}Y_0^0\right) && (13.7.28)\\
&=\; \frac{1}{16}\int_{-1}^{1} d(\cos\theta)\, \cos^2\theta \int_0^\infty dr \left(\frac{r}{a_0}\right)^4 (2 - r/a_0)\, e^{-r/a_0} && (13.7.29)\\
&=\; -3a_0. && (13.7.30)
\end{aligned}$$

So, in the orthobasis of eq. (13.7.25), the perturbation Hamiltonian has the matrix representation

$$W \;\leftrightarrow\; -3a_0 eE \begin{pmatrix} 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \tag{13.7.31}$$

which has eigenvalues $-3a_0eE$, $+3a_0eE$, $0$, and $0$. Thus, the energy eigenvalues to first order in the perturbing electric field and the corresponding unperturbed energy states are

$$\begin{array}{ll} \underline{\text{Energies}} & \underline{\text{States}} \end{array} \tag{13.7.32}$$

$$-\frac{e^2}{8a_0} - 3a_0eE \qquad \frac{1}{\sqrt{2}}\left(|2,0,0\rangle + |2,1,0\rangle\right), \tag{13.7.33}$$

$$-\frac{e^2}{8a_0} + 3a_0eE \qquad \frac{1}{\sqrt{2}}\left(|2,0,0\rangle - |2,1,0\rangle\right), \tag{13.7.34}$$

$$-\frac{e^2}{8a_0} \qquad |2,1,1\rangle, \qquad |2,1,-1\rangle. \tag{13.7.35}$$

Because the perturbed $n = 2$ energy levels vary linearly with the applied electric field, this is called the **linear Stark effect**.

Now that we have found the leading results in perturbation theory for the $n = 1$ and $n = 2$ energy levels, it is time to confess the awful truth that in this example, the perturbative expansion does not converge! The perturbation Hamiltonian $W = eEZ$ is unbounded from below; it approaches $-\infty$ for $z \to -\infty$. This implies that in the idealized problem of a hydrogen atom in a uniform electric field filling infinite space, there are no bound states at all. Instead, the exact stationary-state solutions will have only continuous energies, with resonances near the bound state energies that we have obtained. For this reason, the perturbative expansion in powers of the electric field $E$ actually cannot converge to the supposed atomic bound states.

Fortunately, however, in any real-world version of the Stark effect, the electric field is cut off at some large distance, and the perturbation Hamiltonian remains finite. So, our effort was not at all wasted. In fact, the practically relevant results really are the perturbative ones we found, not the "exact" calculation for unbound states based on an unattainable ideal.

## 13.8 Almost-degenerate perturbation theory

The perturbative expansion can also face problems if the perturbed state is almost, but not quite, degenerate with some other states. This means that for some subset of stationary states $|n_1\rangle, |n_2\rangle, \ldots |n_N\rangle$, the differences in unperturbed energies $\mathcal{E}_{n_j} - \mathcal{E}_{n_k}$ are all small compared to the magnitudes of the matrix elements of $W$. In this case, the terms in the perturbation expansions for $E_{n_j}$ and $|\psi_{n_j}\rangle$ will be finite, but will not become smaller at higher orders, and convergence will fail. To get sensible results we must formulate **almost-degenerate perturbation theory** to evaluate them, as follows.

For inspiration, we write the perturbation Hamiltonian as

$$W = \sum_n \sum_m |n\rangle\langle n|W|m\rangle\langle m|, \tag{13.8.1}$$

by using completeness twice. Consider a projection operator onto the subspace of almost-degenerate states that we wish to consider,

$$P_{\rm deg} = \sum_k |n_k\rangle\langle n_k|. \tag{13.8.2}$$

Now we define the perturbation Hamiltonian projected onto the almost-degenerate subspace,

$$W_{\rm deg} = P_{\rm deg} W P_{\rm deg}, \tag{13.8.3}$$

which is equivalent to only including the orthobasis elements for the degenerate subspace in the summations in eq. (13.8.1). Note that $P_{\rm deg} W_{\rm deg} P_{\rm deg} = W_{\rm deg}$, since $P_{\rm deg}^2 = P_{\rm deg}$, as always for projection operators.

The strategy is to now re-divide the full Hamiltonian as

$$H = H_0' + W', \tag{13.8.4}$$

where

$$H_0' = H_0 + W_{\rm deg}, \tag{13.8.5}$$

$$W' = W - W_{\rm deg}. \tag{13.8.6}$$

One now endeavors to solve the eigenvalue problem for $H_0'$ exactly, and then treat $W'$ as the perturbation. In favorable circumstances, for example if the dimension of the almost-degenerate subspace is small, this can be done exactly. In less favorable cases, one might have to solve the eigenvalue problem for $H_0'$ in an approximation, before then applying $W'$ as the perturbation. In any case, let us refer to the unperturbed eigenstates of $H_0'$ within the almost-degenerate subspace as $|n_j'\rangle$, with unperturbed energies $\mathcal{E}_{n_j'}$. The point is that all of the $W'$ matrix elements within the almost-degenerate subspace vanish,

$$\langle n_j'|W'|n_k'\rangle = 0. \tag{13.8.7}$$

This follows because the left side can be rewritten, using $|n_k'\rangle = P_{\rm deg}|n_k'\rangle$, as

$$\langle n_j'|P_{\rm deg}(W - W_{\rm deg})P_{\rm deg}|n_k'\rangle = \langle n_j'|(W_{\rm deg} - W_{\rm deg})|n_k'\rangle = 0. \tag{13.8.8}$$

One can now fearlessly apply the usual non-degenerate perturbation theory results of eqs. (13.1.43)–(13.1.47) with $H_0'$ as the unperturbed Hamiltonian with eigenstates $|n_j'\rangle$, and $W'$ as the perturbation, because small denominators $\mathcal{E}_{n_j'} - \mathcal{E}_{n_k'}$ can never appear in the sums.

As a minimal example of all three types of stationary-state perturbation theory (non-degenerate, degenerate, almost-degenerate), consider a system with only 2 states, with

$$H_0 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \qquad W = \begin{pmatrix} 0 & c \\ c^* & 0 \end{pmatrix}, \tag{13.8.9}$$

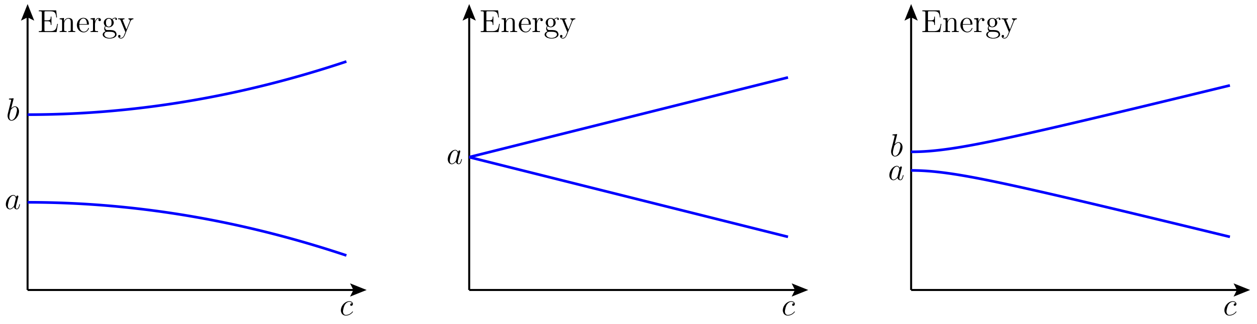

Figure 13.8.1: Perturbation theory results at leading order for energy eigenvalues of a two-state system with Hamiltonian given by eq. (13.8.9), for non-degenerate perturbation theory with $b - a \gg |c|$ (left), degenerate perturbation theory with $a = b$ (center), and almost-degenerate perturbation theory with $a \approx b$ (right). For a two-state system, almost-degenerate perturbation theory is the exact result, by definition. In all three cases, the energy eigenvalues are moved apart by the perturbation, illustrating level repulsion.

where $a, b$ are real numbers and $c$ is complex. Of course, $2\times 2$ matrices are easy, so the eigenvalues for $H = H_0 + W$ can be done exactly in this example, with the results

$$E_{1,2} = \frac{1}{2}\left(a + b \mp \sqrt{(a-b)^2 + 4|c|^2}\right). \tag{13.8.10}$$

If one applies non-degenerate perturbation theory, one readily obtains (assuming that $b > a$)

$$E_1 = a - |c|^2/(b-a) + \cdots, \qquad E_2 = b + |c|^2/(b-a) + \cdots, \tag{13.8.11}$$

with the perturbation $c$ contributing quadratically. On the other hand, if $b = a$, then the previous expansion fails, but one can apply degenerate perturbation theory, with the result

$$E_1 = a - |c| + \cdots, \qquad E_2 = a + |c| + \cdots, \tag{13.8.12}$$

in which the perturbation $c$ contributes linearly. In this simple example, the application of almost-degenerate perturbation theory is just the exact result, which behaves quadratically with $|c|$ for small $|c| \ll |a - b|$, and transitions to behaving linearly with $|c|$ for large $|c| \gg |a + b|$. These statements are illustrated in Figure 13.8.1. The fact that increasing the off-diagonal perturbation moves the energy eigenvalues apart is called **level repulsion**.

As a less minimal example of almost-degenerate perturbation theory, consider a state space of dimension 3, with matrix representations for the unperturbed and perturbation Hamiltonians

$$H_0 = \begin{pmatrix} a-\Delta & 0 & 0 \\ 0 & a+\Delta & 0 \\ 0 & 0 & b \end{pmatrix}, \qquad W = \begin{pmatrix} 0 & \delta & \epsilon \\ \delta & 0 & 0 \\ \epsilon & 0 & 0 \end{pmatrix}, \tag{13.8.13}$$

where $a, b, \Delta, \delta$, and $\epsilon$ are all real constants. The exact eigenvalues of $H = H_0 + W$ can be obtained by solving a cubic equation, but the results are complicated. However, it is worth noting that, because of the theorem in linear algebra that the sum of the eigenvalues of a matrix is equal to the trace, the exact energies must satisfy a simple sum rule,

$$E_1 + E_2 + E_3 \;=\; 2a + b, \tag{13.8.14}$$

regardless of the values of the other parameters. Now let us see what perturbation theory says.

First, let us find the energy eigenvalues by applying non-degenerate perturbation theory. Because $H_0$ is diagonal, the unperturbed stationary states are the kets represented by the unit column vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{13.8.15}$$

with

$$\mathcal{E}_1 = a - \Delta, \qquad \mathcal{E}_2 = a + \Delta, \qquad \mathcal{E}_3 = b. \tag{13.8.16}$$

From eq. (13.1.43), the first-order corrections to the energies vanish, because $W$ has no non-zero diagonal entries. Applying (13.1.44) gives the energy eigenvalues at second order in non-degenerate perturbation theory in $\delta, \epsilon$,

$$E_1 \;=\; a - \Delta + \epsilon^2/(a - b - \Delta) - \delta^2/2\Delta, \tag{13.8.17}$$

$$E_2 \;=\; a + \Delta + \delta^2/2\Delta, \tag{13.8.18}$$

$$E_3 \;=\; b + \epsilon^2/(b - a + \Delta). \tag{13.8.19}$$

As a check, these do satisfy the sum rule eq.(13.8.14).

However, if $\Delta$ is small compared to $\delta$, then the last term in each of eqs. (13.8.17) and (13.8.18) will be large, signaling a breakdown in the perturbative expansion because the unperturbed states $|1\rangle$ and $|2\rangle$ are too close in energy. In that case, one must use almost-degenerate perturbation theory. In this example,

$$P_{\text{deg}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad W_{\text{deg}} = \begin{pmatrix} 0 & \delta & 0 \\ \delta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{13.8.20}$$

and so

$$H_0' = \begin{pmatrix} a - \Delta & \delta & 0 \\ \delta & a + \Delta & 0 \\ 0 & 0 & b \end{pmatrix}, \qquad W' = \begin{pmatrix} 0 & 0 & \epsilon \\ 0 & 0 & 0 \\ \epsilon & 0 & 0 \end{pmatrix}. \tag{13.8.21}$$

Note that $W'$ is indeed 0 within the $2 \times 2$ subspace of almost-degenerate states. It is not too hard to find the unperturbed energies and eigenstates of $H_0'$,

$$\mathcal{E}_{1'} = a - \sqrt{\delta^2 + \Delta^2}, \qquad |1'\rangle = \sqrt{\frac{r_+}{2}}\,|1\rangle - \sqrt{\frac{r_-}{2}}\,|2\rangle, \tag{13.8.22}$$

$$\mathcal{E}_{2'} = a + \sqrt{\delta^2 + \Delta^2}, \qquad |2'\rangle = \sqrt{\frac{r_-}{2}}\,|1\rangle + \sqrt{\frac{r_+}{2}}\,|2\rangle, \tag{13.8.23}$$

$$\mathcal{E}_{3'} = b, \qquad |3'\rangle = |3\rangle, \tag{13.8.24}$$

where

$$r_\pm = 1 \pm \frac{\Delta}{\sqrt{\delta^2 + \Delta^2}}. \tag{13.8.25}$$

We then apply eqs. (13.1.43) and (13.1.44) with $H_0'$ and $W'$, while being careful to note that the matrix representation of $W'$ in eq. (13.8.21) is in the original basis $|1\rangle$, $|2\rangle$, $|3\rangle$, so that

$$W' = \epsilon|1\rangle\langle 3| + \epsilon|3\rangle\langle 1|. \tag{13.8.26}$$

The first-order contributions to the energies vanish because $W'$ is also off-diagonal in the basis $|1'\rangle$, $|2'\rangle$, $|3'\rangle$. The second-order energy corrections are

$$E^{(2)}_{1'} = |\langle 3'|W'|1'\rangle|^2/(\mathcal{E}_{1'} - \mathcal{E}_{3'}), \tag{13.8.27}$$

$$E^{(2)}_{2'} = |\langle 3'|W'|2'\rangle|^2/(\mathcal{E}_{2'} - \mathcal{E}_{3'}), \tag{13.8.28}$$

$$E^{(2)}_{3'} = |\langle 1'|W'|3'\rangle|^2/(\mathcal{E}_{3'} - \mathcal{E}_{1'}) \;+\; |\langle 2'|W'|3'\rangle|^2/(\mathcal{E}_{3'} - \mathcal{E}_{2'}), \tag{13.8.29}$$

resulting in the energy eigenvalues at second-order in almost-degenerate perturbation theory,

$$E_{1'} = a - \sqrt{\delta^2 + \Delta^2} + \frac{\epsilon^2 r_+}{2(a - b - \sqrt{\delta^2 + \Delta^2})}, \tag{13.8.30}$$

$$E_{2'} = a + \sqrt{\delta^2 + \Delta^2} + \frac{\epsilon^2 r_-}{2(a - b + \sqrt{\delta^2 + \Delta^2})}, \tag{13.8.31}$$

$$E_{3'} = b + \frac{\epsilon^2(b - a - \Delta)}{(b - a)^2 - \delta^2 - \Delta^2}, \tag{13.8.32}$$

now with $\epsilon$ as the sole expansion parameter. Again one can check that the sum rule eq.(13.8.14) is indeed satisfied.

The almost-degenerate perturbation theory expansion in $\epsilon$ will be a good one as long as $|\epsilon/(\mathcal{E}_{1'} - \mathcal{E}_{3'})| = |\epsilon/(a - b - \sqrt{\delta^2 + \Delta^2})|$ and $|\epsilon/(\mathcal{E}_{2'} - \mathcal{E}_{3'})| = |\epsilon/(a - b + \sqrt{\delta^2 + \Delta^2})|$ are both small. Conversely, if they are not, then even this expansion will break down, and almost-degenerate perturbation theory would, unfortunately, consist of solving the original $3 \times 3$ problem exactly.

# 14 The variational method

## 14.1 Estimate and upper bound on the ground state energy

In the previous chapter, we studied how to use perturbation theory to find approximate solutions to the Hamiltonian eigenvalue problem. In this chapter, we will consider the **variational method** (also known as the **Rayleigh–Ritz method**), which applies even in situations for which it is not possible to make a sensible perturbation series expansion. Because the variational method is not based on a perturbative expansion, there is no need to know the exact solution to a simpler unperturbed problem.

The key idea comes from an inequality satisfied by the expectation value of the Hamiltonian in any state $|\psi\rangle$ of our choice. Although the eigenstates $|E_n, u_{E_n}\rangle$ of $H$ and the corresponding energy eigenvalues $E_n$ and degeneracy labels $u_{E_n}$ are unknown, we can use completeness to write

$$\begin{aligned}\langle\psi|H|\psi\rangle \;&=\; \sum_n \sum_{u_{E_n}} \langle\psi|H|E_n, u_{E_n}\rangle\langle E_n, u_{E_n}\psi\rangle \;=\; \sum_n \sum_{u_{E_n}} E_n \,\langle\psi|E_n, u_{E_n}\rangle\,\langle E_n, u_{E_n}|\psi\rangle && (14.1.1)\\ &\geq\; E_0 \sum_n \sum_{u_{E_n}} \langle\psi|E_n, u_{E_n}\rangle\langle E_n, u_{E_n}|\psi\rangle\,, && (14.1.2)\end{aligned}$$

where we used $E_0 \leq E_n$ to get the last line. Now we can again use completeness to rewrite $\displaystyle\sum_n \sum_{u_{E_n}} \langle\psi|E_n, u_{E_n}\rangle\langle E_n, u_{E_n}|\psi\rangle = \langle\psi|\psi\rangle$, and obtain the bound

$$\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \geq E_0. \tag{14.1.3}$$

(We have not assumed the normalization condition $\langle\psi|\psi\rangle = 1$, as this is sometimes not convenient.) Equation (14.1.3) means that we can make a guess, or a class of guesses, for $|\psi\rangle$, and then evaluate the left side of eq. (14.1.3) to obtain an estimate for $E_0$ that is guaranteed not to be lower than the true value. The guess is called a **trial state** or **trial wavefunction**.

In a typical application, we improve the quality of our guess by choosing a whole family of trial states $|\psi(a)\rangle$, where $a$ denotes a continuously variable parameter (or, more generally, a set of parameters). Now we compute

$$E(a) \;=\; \frac{\langle\psi(a)|H|\psi(a)\rangle}{\langle\psi(a)|\psi(a)\rangle}, \tag{14.1.4}$$

and minimize it with respect to $a$ to obtain $a_{\text{min}}$, typically by solving

$$\frac{\partial E(a)}{\partial a} = 0. \tag{14.1.5}$$

Then $E(a_{\text{min}})$ is our corresponding best estimate and upper bound for $E_0$, and $|\psi(a_{\text{min}})\rangle$ is our best estimate for the ground state $|E_0\rangle$.

The variational method tends to forgive us if our guess is good, but not perfect. To see why, suppose the result of our best guess can be written as

$$|\psi(a_{\rm min})\rangle \;=\; |E_0\rangle + \sum_{n\neq 0} \delta_n |E_n\rangle, \tag{14.1.6}$$

where the $\delta_n$ are some unknown, but hopefully small, numbers representing the fact that our guess was not perfect. (For simplicity, take the energy eigenvalues to be not degenerate for the illustration; this is not a crucial assumption.) Then we compute

$$\langle\psi|H|\psi\rangle \;=\; E_0 + \sum_{n\neq 0} E_n |\delta_n|^2, \tag{14.1.7}$$

$$\langle\psi|\psi\rangle \;=\; 1 + \sum_{n\neq 0} |\delta_n|^2, \tag{14.1.8}$$

which yields

$$\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \;=\; E_0 + \frac{\displaystyle\sum_{n\neq 0}(E_n - E_0)|\delta_n|^2}{\displaystyle 1 + \sum_{n\neq 0}|\delta_n|^2}. \tag{14.1.9}$$

The error in our estimate for $E_0$ is seen to be quadratic, involving $|\delta_n|^2$ rather than $|\delta_n|$. The lesson here is that the estimate for the energy is better than the estimate for the state, provided only that the best-estimate errors are not too horrible, obeying $|\delta_n| < 1$. A 10% error in the state ket typically leads to only a $\sim$ 1% error in the energy.

Still, we do want to make a good guess for our parameterized family of states $|\psi(a)\rangle$, so that the best of them has a chance to be as close as possible to the true ground state. For example, we should choose trial wavefunctions $\psi(\vec{r})$ that have as few wiggles and nodes as possible, consistent with the given potential $V(\vec{r})$, in order to minimize the kinetic contribution to the Hamiltonian. Also, $|\psi(\vec{r})|$ should be larger where the potential $V(\vec{r})$ is smaller. In three-dimensional problems, in order to minimize the kinetic contribution to the energy, we should also make guesses that have $L^2$ eigenvalue $l = 0$ if the Hamiltonian has rotational symmetry.

## 14.2 Variational method for excited states

The variational method also can provide information about higher-energy states, by projecting out lower-energy states from the trial state guesses. For example, suppose that we want to obtain the energy eigenvalue for the first excited state(s) $|E_1\rangle$. If we somehow already knew the exact $|E_0\rangle$, then we could define the projector

$$P \;=\; I - |E_0\rangle\langle E_0|, \tag{14.2.1}$$

which satisfies $P^2 = P$ and $P|E_0\rangle = 0$. Then, for all trial states $P|\chi\rangle$, one can prove the bound

$$\frac{\langle\chi|PHP|\chi\rangle}{\langle\chi|P|\chi\rangle} \;\; \geq \;\; E_1, \tag{14.2.2}$$

where $E_1$ is the first excited energy eigenvalue.

In practice, we may only have a variational estimate $|\psi_{\rm est}\rangle$ for $|E_0\rangle$, for example $|\psi_{\rm est}\rangle = |\psi(a_{\rm min})\rangle$ from a previous variational estimate of the ground state. In that case, it might be that the best we can do is to instead define

$$\widetilde{P} \;\; = \;\; I - |\psi_{\rm est}\rangle\, \langle\psi_{\rm est}|\,, \tag{14.2.3}$$

(assuming that $|\psi_{\rm est}\rangle$ has been normalized to unity) and then minimize the expectation value of the Hamiltonian in a parameterized family of states $\widetilde{P}|\chi(b)\rangle$, where $|\chi(b)\rangle$ is chosen based on a prior understanding of the properties of the first excited state(s), and depends on a variational parameter $b$. Unfortunately, the accuracy of the resulting estimate for $E_1$ will be limited by the goodness of the estimate $|\psi_{\rm est}\rangle \approx |E_0\rangle$. To the extent that this is not a good approximation, there is some leakage from the ground state into the guess, and we do not have a strict bound on the result of the minimization of the Hamiltonian expectation value. For this reason, it is often computationally challenging to obtain good estimates for all but the lowest few energy eigenvalues.

Fortunately, strict bounds and variational estimates for some of the higher energy states can be obtained if the Hamiltonian is compatible with some other observable(s). Suppose that there is an observable $\Omega$ that satifies

$$[H, \Omega] \;\; = \;\; 0. \tag{14.2.4}$$

This implies that one can choose an orthobasis of common eigenstates of $H$ and $\Omega$, so that the energy eigenstates can be sub-classified by their $\Omega$ eigenvalues, denoted $\omega$. Common examples for $\Omega$ are the total angular momentum operators $J^2$, $J_z$ and/or the parity operator $\Pi$. In such cases, we can define a family of trial kets $|\psi_\omega(a)\rangle$ that are restricted to be eigenstates of $\Omega$ with a fixed eigenvalue $\omega$, so

$$\Omega\,|\psi_\omega(a)\rangle \;\; = \;\; \omega\,|\psi_\omega(a)\rangle\,. \tag{14.2.5}$$

In most examples, it will be obvious how to do this, but if necessary we can define the projection operator for each $\omega$,

$$P_\omega \;\; = \;\; \sum_{u_\omega} |\omega, u_\omega\rangle\langle\omega, u_\omega|, \tag{14.2.6}$$

where the $u_\omega$ are degeneracy labels for $\omega$, and then take $|\psi_\omega(a)\rangle = P_\omega\, |\psi(a)\rangle$. In any case, given a family of trial states $|\psi_\omega(a)\rangle$, one finds

$$E(a,\omega) \;=\; \frac{\langle\psi_\omega(a)|H|\psi_\omega(a)\rangle}{\langle\psi_\omega(a)|\psi_\omega(a)\rangle} \;\geq\; E_{\rm min}(\omega), \tag{14.2.7}$$

where $E_{\rm min}(\omega)$ is the true minimum energy eigenvalue within the subspace of states that have $\Omega = \omega$. Thus, we obtain a separate variational estimate and bound on the lowest energy eigenstate within each $\omega$ subspace.

For example, if the Hamiltonian for a spin-less particle has rotational symmetry, then we can take $\Omega = L^2$. Then, applying the variational principle in turn to the trial wavefunctions $\psi(r,\theta,\phi) = R_A(r)$ and $R_B(r)Y_1^0(\theta,\phi)$ and $R_C(r)Y_2^0(\theta,\phi)$, we will obtain estimates for the minimum energy eigenvalues and wavefunctions for the ground state with $l = 0$, and for the lowest energy state that has $l = 1$, and for the lowest energy state that has $l = 2$, respectively. (Recall that rotational invariance guarantees that the energy eigenvalues and the radial wavefunctions are actually independent of $m$, so we took $m = 0$ in each case just to be specific. Any other value of $m$ would work just as well.) It is not necessary in this case to go to the trouble of defining the projection operator in eq. (14.2.6), because the eigenstates of $L^2$ are known in terms of the spherical harmonics.

## 14.3 Examples of the variational method

As an easy example, consider a system for which we already know the exact answer: the harmonic oscillator in one dimension, with

$$H \;=\; \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X^2. \tag{14.3.1}$$

We now use a family of trial wavefunctions

$$\psi(a) \;=\; e^{-x^2/a^2}. \tag{14.3.2}$$

This is a very fortunate guess, as it just so happens to include the correct ground-state wavefunction as a special case. Now, calculate

$$\langle\psi(a)|\psi(a)\rangle \;=\; \int_{-\infty}^{\infty} dx\, e^{-2x^2/a^2} = \sqrt{\frac{\pi}{2}}a, \tag{14.3.3}$$

and

$$\begin{aligned}
\langle\psi(a)|H|\psi(a)\rangle \;&=\; \int_{-\infty}^{\infty} dx\, e^{-x^2/a^2}\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right] e^{-x^2/a^2} && (14.3.4)\\
&=\; \int_{-\infty}^{\infty} dx\,\left[\frac{\hbar^2}{2m}\left(\frac{2}{a^2} - \frac{4x^2}{a^4}\right) + \frac{1}{2}m\omega^2 x^2\right] e^{-2x^2/a^2} && (14.3.5)\\
&=\; \sqrt{\frac{\pi}{2}}\left(\frac{\hbar^2}{2ma} + \frac{1}{8}m\omega^2 a^3\right). && (14.3.6)
\end{aligned}$$

Using these in eq. (14.1.4) gives

$$E(a) = \frac{\hbar^2}{2ma^2} + \frac{m\omega^2 a^2}{8}. \tag{14.3.7}$$

The first term is the kinetic contribution, which wants $a$ to be large, and the second term is the potential contribution, which prefers $a$ to be small. Minimizing $E(a)$ with respect to $a^2$ gives

$$a^2_{\rm min} \;=\; \frac{2\hbar}{m\omega}, \tag{14.3.8}$$

and so the result of the variational method is

$$E_0 \;\leq\; E(a_{\rm min}) \;=\; \hbar\omega/2, \tag{14.3.9}$$

in agreement with the exact result. Of course, $\psi(a_{\rm min})$ is also the exact ground state wavefunction (up to the normalization factor), as a result of our fortunate guess.

For a less trivial example, consider the Hamiltonian

$$H \;=\; \frac{P^2}{2m} + \lambda X^4, \tag{14.3.10}$$

which is a cousin of the harmonic oscillator but with a steeper, quartic, potential. Not feeling very imaginative at the moment, and remembering the forgiving nature of the variational principle, we again choose as the trial wavefunction

$$\psi(a) \;=\; e^{-x^2/a^2}, \tag{14.3.11}$$

which gives

$$E(a) \;=\; \frac{\hbar^2}{2ma^2} + \frac{3}{16}\lambda a^4. \tag{14.3.12}$$

(The computation of the first term is identical to that for the previous example, as the trial wavefunction and the $P^2/2m$ part of the Hamiltonian have not changed.) Now minimizing $E(a)$ with respect to $a^2$ gives

$$a^2_{\rm min} \;=\; \left(\frac{4\hbar^2}{3\lambda m}\right)^{1/3} \;\approx\; 1.101\left(\frac{\hbar^2}{\lambda m}\right)^{1/3}, \tag{14.3.13}$$

and

$$E_{\rm min} \;=\; \frac{3}{8}\left(\frac{6\hbar^4\lambda}{m^2}\right)^{1/3} \;\approx\; 0.681\left(\frac{\hbar^4\lambda}{m^2}\right)^{1/3} \tag{14.3.14}$$

as the estimate of the ground state energy.

To do better, we invent a three-parameter trial wavefunction

$$\psi(a,b,c) \quad = \quad (1 + bx^2 + cx^4)e^{-x^2/a^2}. \tag{14.3.15}$$

We do not bother including odd powers of $x$, because we expect that the parity of the ground state is $+1$. Using a computer to minimize $E(a,b,c)$ with respect to all three of $a,b,c$ gives

$$E(a_{\rm min}, b_{\rm min}, c_{\rm min}) \quad \approx \quad 0.668 \left(\frac{\hbar^4 \lambda}{m^2}\right)^{1/3} \tag{14.3.16}$$

as a better estimate. As always, the more general wavefunction gives an estimate that is lower, and closer to the true ground-state energy eigenvalue.

The first excited state for $H$ in eq. (14.3.10) will have one node, at $x = 0$, and odd parity. To find it, we can therefore exploit eq. (14.2.7), by choosing a simple trial wavefunction

$$\psi(a) \quad = \quad xe^{-x^2/a^2}. \tag{14.3.17}$$

A straightforward calculation of the Hamiltonian's expectation value and the norm gives

$$E(a) \quad = \quad \frac{3\hbar^2}{2ma^2} + \frac{15}{16}\lambda a^4, \tag{14.3.18}$$

with minimum at

$$a^2_{\rm min} \quad = \quad \left(\frac{4\hbar^2}{5\lambda m}\right)^{1/3}, \tag{14.3.19}$$

leading to the estimate and upper bound for $E_1$ of

$$E(a_{\rm min}) \quad = \quad \frac{9}{4}\left(\frac{5\hbar^4\lambda}{4m^2}\right)^{1/3} \approx 2.424 \left(\frac{\hbar^4\lambda}{m^2}\right)^{1/3}. \tag{14.3.20}$$

A somewhat better result could follow from using a multi-parameter trial wavefunction of the form $\psi(a,b,c,\ldots) = (x + bx^3 + cx^5 + \cdots)e^{-x^2/a^2}$. Note that it is important to be careful to only include odd-parity terms in the trial wavefunction, to avoid contamination from the even-parity ground state. The actual first excited state energy eigenvalue $E_1$ turns out to be

$$E_1 \quad = \quad 2.394\ldots \left(\frac{\hbar^4\lambda}{m^2}\right)^{1/3}, \tag{14.3.21}$$

as can be obtained from numerical minimization of, say, a 6-parameter trial wavefunction.

Next, consider as a three-dimensional example the hydrogen atom Hamiltonian, written in the spherical coordinate position representation as

$$H = T + V, \qquad T = -\frac{\hbar^2\nabla^2}{2m_e}, \qquad V = -\frac{e^2}{r}. \tag{14.3.22}$$

Although we know the exact wavefunction, let us ignore the temptation to use it, and instead try a Gaussian trial wavefunction with a variational parameter $k$,

$$\psi(k) = e^{-k^2 r^2}. \tag{14.3.23}$$

To find the expectation value of the Hamiltonian, we first compute the kinetic contribution,

$$\langle\psi|T|\psi\rangle \;=\; -\frac{\hbar^2}{2m_e}4\pi\int_0^\infty dr\, r^2\, e^{-k^2r^2}\left(\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}\right)e^{-k^2r^2} = 4\pi\left(\frac{\hbar^2}{m_e k}\sqrt{\frac{\pi}{2}}\frac{3}{16}\right), \tag{14.3.24}$$

then the potential contribution,

$$\langle\psi|V|\psi\rangle \;=\; 4\pi\left(-\frac{e^2}{4k^2}\right), \tag{14.3.25}$$

and finally the normalization factor,

$$\langle\psi|\psi\rangle \;=\; 4\pi\left(\sqrt{\frac{\pi}{2}}\frac{1}{8k^3}\right). \tag{14.3.26}$$

Putting these together, we have

$$E(k) \;=\; \frac{3\hbar^2 k^2}{2m_e} - \sqrt{\frac{8}{\pi}}e^2 k. \tag{14.3.27}$$

The minimum is found from $\partial E(k)/\partial k = 0$ to be at

$$k_{\text{min}} = \sqrt{\frac{8}{\pi}}\frac{e^2 m_e}{3\hbar^2} = \sqrt{\frac{8}{\pi}}\frac{1}{3a_0}. \tag{14.3.28}$$

Plugging this in gives the estimate of the ground state energy

$$E(k_{\text{min}}) \;=\; -\left(\frac{8}{3\pi}\right)\frac{e^2}{2a_0} \approx -0.849\,\text{Rydberg}. \tag{14.3.29}$$

Of course, the exact answer is $-1$ Rydberg. Even though our choice of one-parameter trial wavefunctions was not a very good one, because the harmonic oscillator is a poor approximation to the Coulomb potential, the result was within about 15% of the correct answer.

## 14.4 Helium atom ground state from the variational method

We conclude this chapter with a more practical example in which the variational method shines: the helium atom consisting of two electrons with charge $-e$ and a nucleus with charge $+2e$. The Hamiltonian for this system was already given in eqs. (13.3.1)–(13.3.3) of section 13.3, where we treated it using first-order perturbation theory.

To use the variational method in the most simple way, let us choose a factorized trial wavefunction of the form

$$\Psi(\tilde{Z},\vec{r}_1,\vec{r}_2) \;=\; \psi(\tilde{Z},\vec{r}_1)\,\psi(\tilde{Z},\vec{r}_2), \tag{14.4.1}$$

where [compare eq. (10.1.42)]

$$\psi(\tilde{Z}, \vec{r}) \;=\; \sqrt{\frac{\tilde{Z}^3}{\pi a_0^3}}\, e^{-\tilde{Z}r/a_0} \tag{14.4.2}$$

is the normalized ground state wavefunction for a hydrogen-like atom with general atomic number $\tilde{Z}$. However, here we do not set $\tilde{Z}$ equal to $Z = 2$ in the trial wavefunction, but instead treat it as a continuous variational parameter, adjusting it to find the minimum expectation value of the total $H$. In contrast, $Z$ in the Hamiltonian is fixed to be equal to the actual number of protons in the nucleus. For helium, $Z = 2$, but we will leave it general for the following discussion, so that our results apply to any ion with 2 electrons.

We therefore need to compute, and then minimize:

$$E(\tilde{Z}) \;=\; E_1(\tilde{Z}) + E_2(\tilde{Z}) + E_{12}(\tilde{Z}), \tag{14.4.3}$$

where $E_1(\tilde{Z}) = E_2(\tilde{Z}) = \langle\Psi|H_1|\Psi\rangle$, and $E_{12}(\tilde{Z}) = \langle\Psi|H_{12}|\Psi\rangle$, since we have arranged for $\langle\Psi|\Psi\rangle = 1$ in this example. Because each electron partly screens the charge $+Ze$ nucleus from the other electron, we anticipate that the value $\tilde{Z}_{\rm min}$ that minimizes $E(\tilde{Z})$ will be in the range $Z - 1 < \tilde{Z}_{\rm min} < Z$, so that for helium, $1 < \tilde{Z}_{\rm min} < 2$.

To find the non-interaction contributions, we first compute

$$E_1(\tilde{Z}) \;=\; \int d^3\vec{r}_1\, \psi(\vec{r}_1)^* \left(-\frac{\hbar^2}{2m_e}\nabla_1^2 - \frac{Ze^2}{r_1}\right) \psi(\vec{r}_1) \int d^3\vec{r}_2\, |\psi(\vec{r}_2)|^2. \tag{14.4.4}$$

Note that the potential energy contribution is $-Ze^2/r_1$, not $-\tilde{Z}e^2/r_1$, because $\tilde{Z}$ is our variational parameter, not the actual nuclear charge. In this expression, $\int d^3\vec{r}_2\, |\psi(\vec{r}_2)|^2 = 1$, and the remaining integral factor evaluates simply, to give

$$E_1(\tilde{Z}) \;=\; \frac{e^2}{2a_0}\left(\tilde{Z}^2 - 2\tilde{Z}Z\right). \tag{14.4.5}$$

Of course, $E_2(\tilde{Z}) = E_1(\tilde{Z})$ follows from an identical calculation with the labels 1 and 2 interchanged. The more difficult part is the interaction contribution

$$E_{12}(\tilde{Z}) \;=\; e^2 \int d^3\vec{r}_1 \int d^3\vec{r}_2\, |\psi(\vec{r}_1)|^2\, |\psi(\vec{r}_2)|^2\, \frac{1}{|\vec{r}_1 - \vec{r}_2|}. \tag{14.4.6}$$

Fortunately, we have already evaluated this integral in eqs. (13.3.8)–(13.3.19), with the result

$$E_{12}(\tilde{Z}) \;=\; \frac{5\tilde{Z}e^2}{8a_0}. \tag{14.4.7}$$

Plugging this into eq. (14.4.3), along with $E_1(\tilde{Z}) = E_2(\tilde{Z})$ from eq. (14.4.5), gives

$$E(\tilde{Z}) \;=\; \tilde{Z}(\tilde{Z} - 2Z + 5/8)\frac{e^2}{a_0}. \tag{14.4.8}$$

Minimizing this with respect to our variational parameter $\tilde{Z}$ gives

$$\tilde{Z}_{\rm min} \;=\; Z - 5/16 \;=\; 27/16. \tag{14.4.9}$$

Note that indeed $Z - 1 < \tilde{Z}_{\rm min} < Z$, as we had anticipated.

Our variational method estimate for the ground state energy of the helium atom is therefore

$$E(\tilde{Z}_{\rm min}) \;=\; -\tilde{Z}^2_{\rm min}\frac{e^2}{a_0} \;=\; -\left(\frac{27}{16}\right)^2 \frac{e^2}{a_0} \;=\; -5.695\,\text{Rydberg} \;=\; -77.49\,\text{eV}. \tag{14.4.10}$$

This can be compared to the experimental value (the minimum ionization energy needed to completely remove both electrons from the He atom in its ground state), which is

$$E_{\rm exp} \;=\; -5.807\,\text{Rydberg} \;=\; -79.005\,\text{eV}. \tag{14.4.11}$$

As expected, our variational method estimate is higher than the true value, but by less than 2%. Recall that the first-order perturbation theory method in section 13.3 gave instead

$$E_{\rm pert.} \;=\; -5.5\,\text{Rydberg} \;=\; -74.83\,\text{eV}, \tag{14.4.12}$$

which is about 5.3% higher than the experimental value. The variational method, even with the simplest trial wavefunction, performs better than first-order perturbation theory. If one uses a more sophisticated non-factorized trial wavefunction with more parameters, and asks a computer to do the integrations and minimization numerically, then the variational method will win even more convincingly for a given amount of human calculation effort expended. The variational method also scales nicely to more complex problems in atomic and molecular physics, where perturbation theory often tends to have more difficulty.

# 15 Fine, hyperfine, and magnetic effects for the hydrogen atom

## 15.1 Relativistic kinetic, spin-orbit, and Darwin corrections

The hydrogen atom Hamiltonian was treated in chapter 10 in a non-relativistic approximation. There, it was argued that relativistic effects should affect the energy eigenvalues by a factor suppressed by the square of the fine-structure constant $\alpha = 1/137.036\ldots$. Now it is time to find those corrections, using stationary-state perturbation theory.

By an expansion of the Dirac equation, it is possible to show that a more accurate approximation for the electron (with mass called $m_e$ in this section) moving in a general spherically symmetric electrostatic potential[†] $\Phi(R)$ is given by

$$H \;=\; H_0 + H_{\rm rel} + H_{\rm SO} + H_{\rm Darwin}, \tag{15.1.1}$$

where

$$H_0 \;=\; \frac{P^2}{2m_e} - e\Phi(R) \tag{15.1.2}$$

is the non-relativistic approximation we have already used, with $P^2 = \vec{P} \cdot \vec{P}$, and

$$H_{\rm rel} \;=\; -\frac{(P^2)^2}{8m_e^3c^2}, \tag{15.1.3}$$

$$H_{\rm SO} \;=\; -\frac{e}{2m_e^2c^2}\frac{1}{R}\frac{d\Phi}{dR}\vec{L}\cdot\vec{S}, \tag{15.1.4}$$

$$H_{\rm Darwin} \;=\; -\frac{\hbar^2 e}{8m_e^2c^2}\nabla^2\Phi \tag{15.1.5}$$

are called the relativistic kinematic, spin-orbit, and Darwin terms, respectively. (The last is named for Charles G. Darwin, the grandson of the Charles Darwin famous for evolution.) These three terms give rise to numerical effects that are of the same order, collectively called the **fine structure** corrections to the hydrogen atom energy levels. Each of them is a relativistic effect, as suggested by the $1/c^2$ factors, which would make them formally vanish if the speed of light were infinite. In this section, we will calculate their effects using perturbation theory. Before doing so, let us briefly review the physical reasons behind the fine structure terms.

First, $H_{\rm rel}$ can be understood as coming from the binomial expansion of the special relativistic expression for the energy of a massive particle with classical momentum $p$,

$$E \;=\; \sqrt{m_e^2c^4 + p^2c^2} \;=\; m_ec^2 + \frac{p^2}{2m_e} - \frac{(p^2)^2}{8m_e^3c^2} + \cdots\,. \tag{15.1.6}$$

---

[†] For hydrogen-like atoms, $\Phi(R) = Ze/R$, but in eqs. (15.1.2)–(15.1.5) we choose to be more general.

The first term is just a constant, Einstein's famous equivalence of rest mass and energy. We can ignore it, since it causes all time-dependent states to be multiplied by the same phase. The second term is the kinetic term in $H_0$, and the third gives $H_{\rm rel}$ when $p$ is promoted from classical momentum to the quantum momentum operator.

The $H_{\rm SO}$ contribution can be understood as the energy due to the interaction of the intrinsic magnetic moment of the electron with the magnetic field $\vec{B}$ that it experiences because it is moving. Naively, this gives a contribution to the Hamiltonian $-\vec{\mu}_e \cdot \vec{B}$, where the magnetic field as seen in the electron's rest frame moving with velocity $\vec{v}$ is, in gaussian cgs units,

$$\vec{B} \;=\; -\frac{\vec{v}}{c} \times \vec{E}. \tag{15.1.7}$$

Here, we can use $\vec{v} = \vec{p}/m_e$ and, for a spherically symmetric potential,

$$\vec{E} \;=\; -\vec{\nabla}\Phi = -\frac{\vec{r}}{r}\frac{d\Phi}{dr}, \tag{15.1.8}$$

so that, promoting $\vec{r}$ and $\vec{p}$ to quantum operators,

$$\vec{B} \;=\; \frac{1}{m_e c}\frac{1}{R}\frac{d\Phi(R)}{dR}\,(\vec{P} \times \vec{R}) \;=\; -\frac{1}{m_e c}\frac{1}{R}\frac{d\Phi(R)}{dR}\,\vec{L}. \tag{15.1.9}$$

There is no operator ordering issue in writing the cross-product $\vec{L} = \vec{R} \times \vec{P} = -\vec{P} \times \vec{R}$, because different rectangular coordinate components of $\vec{P}$ and $\vec{R}$ commute. The magnetic moment of the electron is opposite the direction of the spin, because the charge $-e$ is negative,

$$\vec{\mu}_e \;=\; -\frac{e}{m_e c}\vec{S}, \tag{15.1.10}$$

where we have used eqs. (4.3.18) and (4.3.19) with the (very good) approximation $g_e \approx 2$. So, naively we would have

$$H_{\text{SO, naive}} \;=\; -\vec{\mu}_e \cdot \vec{B} \;=\; -\frac{e}{m_e^2 c^2}\frac{1}{R}\frac{d\Phi}{dR}\,\vec{L} \cdot \vec{S}. \tag{15.1.11}$$

The extra factor of $1/2$ in the true $H_{\rm SO}$ in eq. (15.1.4) is known as the Thomas precession effect. Llewellyn Thomas explained it in 1925 as a classical relativistic effect of the electron being at rest in a rotating (not inertial) reference frame. Alternatively, it can be derived directly from the Dirac equation.

Finally, the physical origin of the Darwin term $H_{\rm Darwin}$ is related to the fact that, as we noted following eq. (10.1.71), the behavior of the electron is modified on a distance scale comparable to its reduced Compton wavelength, $\lambdabar_e = \hbar/m_e c$, due to the effects of virtual electron-positron production, which are not captured by the non-relativistic Hamiltonian. A proper derivation from the Dirac equation shows that this ends up making the replacement

$$\Phi(R) \;\to\; \Phi(R) - \frac{\lambdabar_e^2}{8}\nabla^2\Phi, \tag{15.1.12}$$

resulting in the correction $H_{\text{Darwin}}$ in eq. (15.1.5).

Now we apply first-order perturbation theory to evaluate the $H_{\text{rel}}$ contribution to the energies of the unperturbed states $|n, l, m\rangle$. These unperturbed states have degeneracies for $n > 1$, so we are really doing degenerate perturbation theory, but $W = H_{\text{rel}}$ is already diagonal in this basis, so we can directly apply eq. (13.6.4). This gives

$$E^{(1),\,\text{rel}}_{n,l,m} \;=\; \langle H_{\text{rel}}\rangle_{n,l,m} \;=\; -\frac{1}{8m_e^3c^2}\langle (P^2)^2\rangle_{n,l,m}, \tag{15.1.13}$$

where here, and in the following, $\langle\cdots\rangle_{n,l,m}$ denotes the expectation value $\langle n, l, m|\cdots|n, l, m\rangle$. To evaluate this efficiently, we can use a trick, by rewriting it using $P^2 = 2m_e(H_0 + e^2/R)$,

$$\begin{aligned} E^{(1),\,\text{rel}}_{n,l,m} &= -\frac{1}{2m_ec^2}\langle (H_0 + e^2/R)^2\rangle_{n,l,m} && (15.1.14)\\ &= -\frac{1}{2m_ec^2}\left(\mathcal{E}_n^2 + 2\mathcal{E}_n e^2\langle 1/R\rangle_{n,l,m} + e^4\langle 1/R^2\rangle_{n,l,m}\right), && (15.1.15)\end{aligned}$$

where $\mathcal{E}_n = -e^2/2a_0n^2$ is the eigenvalue of $H_0$ for the state $|n, l, m\rangle$. The expectation values of $1/R$ and $1/R^2$ were given in eqs. (10.1.64) and (10.1.67), respectively. Plugging those results into eq. (15.1.15) gives

$$E^{(1),\,\text{rel}}_{n,l,m} \;=\; \alpha^2\frac{e^2}{2a_0}\frac{1}{n^3}\left(\frac{3}{4n} - \frac{1}{l+1/2}\right), \tag{15.1.16}$$

where we have used $a_0 = \hbar^2/m_ee^2$ to write the result in terms of the fine structure constant $\alpha = e^2/\hbar c \approx 1/137.036$, and $e^2/2a_0 = 1$ Rydberg $\approx 13.606$ eV.

By itself, eq. (15.1.16) is not very useful, because the spin-orbit and Darwin terms contribute at the same order. To evaluate the spin-orbit contribution, we note that for the special case of the hydrogen atom, the electrostatic potential is $\Phi = e/R$, so that eq. (15.1.4) becomes

$$H_{\text{SO}} \;=\; \frac{e^2}{2m_e^2c^2}\frac{1}{R^3}\vec{L}\cdot\vec{S}. \tag{15.1.17}$$

This vanishes for all $l = 0$ states, including the ground state $n = 1$, since they have no orbital angular momentum. Because the $n > 1$ states have $H_0$ eigenvalue degeneracies, we are really doing degenerate perturbation theory, which means that we must use a basis in which $H_{\text{SO}}$ is diagonal. We use a standard trick for dot products of angular momentum operators,

$$\vec{L}\cdot\vec{S} \;=\; \frac{1}{2}(J^2 - L^2 - S^2) \;=\; \frac{\hbar^2}{2}\left[j(j+1) - l(l+1) - s(s+1)\right], \tag{15.1.18}$$

which is diagonal in the total angular momentum orthobasis $|n, l, s, j, m_j\rangle$. (In particular, we do *not* want to use the product orthobasis $|n, l, s, m_l, m_s\rangle$ here, because then $\vec{L}\cdot\vec{S}$ would not be diagonal, as required for degenerate perturbation theory.) Using $s = 1/2$, it follows that the first-order correction is

$$E^{(1),\,\mathrm{SO}}_{n,l,j} \;=\; \frac{\hbar^2 e^2}{4m_e^2 c^2}\,[j(j+1) - l(l+1) - 3/4]\;\langle 1/R^3\rangle_{n,l,m}, \qquad (l \neq 0). \tag{15.1.19}$$

The expectation value of $1/R^3$ was given in eq. (10.1.68), reproduced here for convenience,

$$\langle 1/R^3\rangle_{n,l,m} = \frac{1}{a_0^3 n^3 l(l+1)(l+1/2)}. \tag{15.1.20}$$

It follows that

$$E^{(1),\,\mathrm{SO}}_{n,l,j} \;=\; \frac{\hbar^2 e^2}{2m_e^2 c^2 a_0^3}\frac{j(j+1) - l(l+1) - 3/4}{n^3 l(l+1)(2l+1)}, \qquad (l \neq 0). \tag{15.1.21}$$

Now, using $j = l \pm 1/2$, and writing the prefactor in terms of the fine structure constant and the Rydberg energy,

$$E^{(1),\,\mathrm{SO}}_{n,l,j} \;=\; \alpha^2\left(\frac{e^2}{2a_0}\right)\frac{-1 \pm (2l+1)}{2n^3 l(l+1)(2l+1)}, \qquad (l \neq 0,\; j = l \pm 1/2), \tag{15.1.22}$$

to go along with $E^{(1),\,\mathrm{SO}}_{n,0,1/2} = 0$.

Now we turn to the Darwin term of eq. (15.1.5), which for the hydrogen atom includes, in the position representation, a factor $\nabla^2\Phi = e\nabla^2(1/r) = -4\pi e\delta^{(3)}(\vec{r})$. It is left as a short exercise to compute the first-order energy correction

$$E^{(1),\mathrm{Darwin}} \;=\; \delta_{l,0}\,\alpha^2\left(\frac{e^2}{2a_0}\right)\frac{1}{n^3}. \tag{15.1.23}$$

which only depends on $n, l$ and is only non-zero for $l = 0$.

We are finally ready to combine the three fine-structure contributions, $E^{\mathrm{fine}} = E^{(1),\mathrm{rel}} + E^{(1),\mathrm{SO}} + E^{(1),\mathrm{Darwin}}$, from eqs. (15.1.16), (15.1.22), and (15.1.23). This can be done in three separate cases,

$$(l = 0, j = 1/2):\quad E^{\mathrm{fine}} \;=\; \frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\left(\frac{3}{4n} - 2 + 0 + 1\right), \tag{15.1.24}$$

$$(l \neq 0, j = l + 1/2):\quad E^{\mathrm{fine}} \;=\; \frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\left(\frac{3}{4n} - \frac{2}{2l+1} + \frac{1}{(l+1)(2l+1)} + 0\right), \tag{15.1.25}$$

$$(l \neq 0, j = l - 1/2):\quad E^{\mathrm{fine}} \;=\; \frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\left(\frac{3}{4n} - \frac{2}{2l+1} - \frac{1}{l(2l+1)} + 0\right). \tag{15.1.26}$$

All three cases are seen to be encompassed by the simple formula

$$E^{\mathrm{fine}}_{n,j} \;=\; \frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\left(\frac{3}{4n} - \frac{1}{j+1/2}\right), \tag{15.1.27}$$

which notably depends only on $n$ and $j$, and not separately on $l$, as one perhaps might have guessed. (The same result can be obtained by solving the Dirac equation directly.)

For the lowest few energy levels, in terms of the quantity

$$E_f = \alpha^2 e^2/2a_0 = 7.245 \times 10^{-4} \text{ eV}, \tag{15.1.28}$$

one has fine structure shifts for spectroscopic states for $n = 1$:

$$1S_{1/2} \quad (n=1, l=0, j=1/2) \qquad E_{\text{fine}} \;=\; -E_f/4, \tag{15.1.29}$$

and for $n = 2$:

$$\left.\begin{array}{l} 2S_{1/2} \quad (n=2, l=0, j=1/2) \\ 2P_{1/2} \quad (n=2, l=1, j=1/2) \end{array}\right\} \quad E_{\text{fine}} = -5E_f/64, \tag{15.1.30}$$

$$2P_{3/2} \quad (n=2, l=1, j=3/2) \qquad E_{\text{fine}} = -E_f/64, \tag{15.1.31}$$

where the $2S_{1/2}$ and $2P_{1/2}$ states stay degenerate, and for $n = 3$:

$$\left.\begin{array}{l} 3S_{1/2} \quad (n=3, l=0, j=1/2) \\ 3P_{1/2} \quad (n=3, l=1, j=1/2) \end{array}\right\} \quad E_{\text{fine}} = -E_f/36, \tag{15.1.32}$$

$$\left.\begin{array}{l} 3P_{3/2} \quad (n=3, l=1, j=3/2) \\ 3D_{3/2} \quad (n=3, l=2, j=3/2) \end{array}\right\} \quad E_{\text{fine}} = -E_f/108, \tag{15.1.33}$$

$$3D_{5/2} \quad (n=3, l=2, j=5/2) \qquad E_{\text{fine}} = -E_f/324, \tag{15.1.34}$$

where the $3S_{1/2}$ and $3P_{1/2}$ states stay degenerate, as do the $3P_{3/2}$ and $3D_{3/2}$ states. Including electron spin, each principal quantum number level $n$ has $(2s+1)\sum_{l=0}^{n-1}(2l+1) = 2n^2$ states. The fine structure splits each of those levels into $n$ sub-levels, with $j = 1/2, 3/2, \ldots, n-1/2$. The degeneracies are further enhanced by a factor of 2 for the proton spin, but then are further split by the hyperfine structure to be discussed next.

## 15.2 Hyperfine structure of hydrogen

The hyperfine structure takes into account that an atomic nucleus also has a magnetic dipole moment (along its spin direction) and an electric quadrupole moment (if the nuclear spin is at least 1). For the $^1$H isotope of hydrogen with a nucleus consisting of one proton and no neutrons, the magnetic moment of the proton is related to its spin-1/2 operator, which we denote $\vec{S}_p$, by

$$\vec{\mu}_p \;=\; \frac{g_p e}{2m_p c}\vec{S}_p, \tag{15.2.1}$$

(see section 4.3), where $m_p$ is the proton's mass, and $g_p = 5.5856946893(16)$ is the proton's $g$-factor. The hyperfine contributions to the energy eigenvalues are numerically suppressed compared to the fine structure contributions because of the fact that $m_p \gg m_e$.

The interaction between the magnetic field created by the proton's magnetic moment and the moving electron introduces a proton spin-orbit coupling term in the Hamiltonian,

$$H_{\text{SO}}^{\text{proton}} = \frac{e}{m_e c R^3}\,\vec{\mu}_p \cdot \vec{L} = \frac{g_p e^2}{2 m_e m_p c^2}\frac{1}{R^3}\vec{S}_p \cdot \vec{L}. \tag{15.2.2}$$

In addition, there is the energy of the magnetostatic interaction of the dipole moments of the proton and the electron,

$$H_{\text{dipole-dipole}} = \frac{1}{R^3}\left[\vec{\mu}_e \cdot \vec{\mu}_p - \frac{3}{R^2}(\vec{\mu}_e \cdot \vec{R})\,(\vec{\mu}_p \cdot \vec{R})\right] - \frac{8\pi}{3}\vec{\mu}_e \cdot \vec{\mu}_p\,\delta^{(3)}(\vec{R}), \tag{15.2.3}$$

in which the final term is called a contact term. These two contributions are of the same numerical order, and so, using

$$\vec{\mu}_e = -\frac{g_e e}{2 m_e c}\vec{S} \tag{15.2.4}$$

with $g_e \approx 2$ and $\vec{S}$ the electron's spin, the full hyperfine contribution to the Hamiltonian for hydrogen is

$$\begin{aligned} H_{\text{hf}} &= H_{\text{SO}}^{\text{proton}} + H_{\text{dipole-dipole}} & (15.2.5)\\ &= \frac{g_p e^2}{2 m_e m_p c^2}\left[\frac{1}{R^3}(\vec{L} - \vec{S}) \cdot \vec{S}_p + \frac{3}{R^5}(\vec{R} \cdot \vec{S})(\vec{R} \cdot \vec{S}_p) + \frac{8\pi}{3}\vec{S} \cdot \vec{S}_p\,\delta^{(3)}(\vec{R})\right]. & (15.2.6) \end{aligned}$$

We now treat this as a perturbation of the fine-structure results found in the previous section.

Even after taking into account the fine structure, there were still energy degeneracies. Therefore, we again need to do degenerate perturbation theory by choosing a good set of unperturbed orthobasis states, so that $H_{\text{hf}}$ is diagonal in that basis. The appropriate basis is again a total angular momentum basis, but now taking into account the proton's spin in addition to the electron's spin and orbital angular momentum. We define

$$\begin{aligned} \vec{J} &= \vec{L} + \vec{S}, & (15.2.7)\\ \vec{F} &= \vec{J} + \vec{S}_p = \vec{L} + \vec{S} + \vec{S}_p, & (15.2.8) \end{aligned}$$

so that $\vec{F}$ ($\vec{J}$) is the total angular momentum including (excluding) the proton's spin. The eigenvalues for $J^2$ and $J_z$ are denoted $\hbar^2 j(j+1)$ and $\hbar m_j$, where the allowed values for $j$ are $l-1/2$ and $l+1/2$, and in each case $m_j$ takes on the $2j+1$ values $-j, -j+1, \ldots, j-1, j$. Because $\vec{F}$ arises from adding the orbital angular momentum to two spin-1/2 angular momenta, the eigenvalues for $F^2$ and $F_z$ are $\hbar^2 f(f+1)$ and $\hbar m_f$, with integer values $f = j-1/2$ and $j+1/2$, and in each case $m_f = -f, -f+1, \ldots, f-1, f$. The total angular momentum orthobasis states are constructed by first combining $\vec{L}$ and $\vec{S}$ eigenstates to obtain an orthobasis $|l, s, j, m_j\rangle$, and then taking the tensor product with the proton spin eigenstates $|s_p, m_{s_p}\rangle$ and going to the total angular momentum basis to obtain states labeled $|l, s, j, s_p, f, m_f\rangle$. Here the $s$ and $s_p$ spin labels are trivial and can be omitted, as they are both fixed to be 1/2. So, including the principal quantum number $n$, the CSCO $= \{H_0, L^2, J^2, F^2, F_z\}$ eigenstate orthobasis states can be labeled

$$|n, l, j, f, m_f\rangle. \tag{15.2.9}$$

We now want to show that $H_{\rm hf}$ is diagonal in this basis, and obtain its expectation values. It is convenient to treat separately the case $l = 0$ (including the ground state) and the case $l \neq 0$.

For $l = 0$, the first two terms in eq. (15.2.6) do not contribute. To see this, note that the spherical symmetry of $l = 0$ states guarantees that the expectation value of these two terms is proportional to the angular integral of $3(\hat{r}\cdot\vec{S})(\hat{r}\cdot\vec{S}_p) - \vec{S}\cdot\vec{S}_p$, where we have used $\vec{L} = 0$ for $l = 0$. Such an integral is a scalar, and is linear in each of the two spins, and depends on no other vectors, so it can only be proportional to their dot product. That is, we must have

$$\int d\Omega\, \Big[3(\hat{r}\cdot\vec{S})(\hat{r}\cdot\vec{S}) - \vec{S}\cdot\vec{S}_p\Big] \;=\; a\vec{S}\cdot\vec{S}_p \tag{15.2.10}$$

for some constant number $a$. The value of $a$ can now be discovered to be 0 by evaluating the left side for $\vec{S} = \vec{S}_p = \hat{z}$.

The last term in eq. (15.2.6), proportional to $\delta^{(3)}(\vec{R})$, does contribute for $l = 0$, and only in that case. To see this, note that

$$\langle n, l, m_l|\delta^{(3)}(\vec{R})|n, l, m_l\rangle \;=\; |\psi_{n,l,m_l}(0,0,0)|^2, \tag{15.2.11}$$

the square of the wavefunction at the origin. Using the radial wavefunction at $r = 0$ from eq. (10.1.41), this is non-zero only if $l = 0$, which of course implies $m_l = 0$ also, and

$$\langle n, l, m_l|\delta^{(3)}(\vec{R})|n, l, m_l\rangle \;=\; \delta_{l,0}\,\delta_{m_l,0}\,\frac{1}{\pi a_0^3 n^3}. \tag{15.2.12}$$

Now, with $l = 0$, $\vec{F} = \vec{S} + \vec{S}_p$, so

$$\vec{S}\cdot\vec{S}_p \;=\; \frac{1}{2}(F^2 - S^2 - S_p^2). \tag{15.2.13}$$

This is indeed diagonal in the total angular momentum basis $|n, l, j, f, m_f\rangle$, with $F^2 - S^2 - S_p^2 = \hbar^2[f(f+1) - 3/2]$, where the possible values of $f$ are 0 and 1. Putting together the factors,

$$\langle H_{\rm hf}\rangle \;=\; \frac{g_p e^2}{2m_e m_p c^2}\,\frac{8\pi}{3}\,\frac{\hbar^2}{2}[f(f+1) - 3/2]\,\frac{1}{\pi a_0^3 n^3} \qquad (l = 0), \tag{15.2.14}$$

so we arrive at the hyperfine energy correction for $l = 0$, $j = 1/2$ states,

$$E_{\rm hf}^{(l=0)} \;=\; \langle H_{\rm hf}\rangle \;=\; \frac{g_p m_e}{m_p}\,\frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right) \times \begin{cases} 2/3 & \text{for } f = 1, \\ -2 & \text{for } f = 0. \end{cases} \tag{15.2.15}$$

In particular, the hyperfine interaction splits the $n = 1$ ground state by an amount

$$E_\gamma \;=\; E_{\rm hf}^{(f=1)} - E_{\rm hf}^{(f=0)} \;=\; \frac{8 g_p m_e}{3 m_p} \alpha^2 \left(\frac{e^2}{2a_0}\right) \;=\; 5.878 \times 10^{-6}\,\text{eV}. \tag{15.2.16}$$

We had already discussed this in eqs. (11.2.27)–(11.2.29); here we have shown how the numerical prefactor arises. The actual experimental value $5.87433 \times 10^{-6}$ eV is very slightly less than eq. (15.2.16), due to higher order effects. The $f = 1$ state is higher in energy than the $f = 0$ state, and when the former decays to the latter it emits a photon with this energy splitting, whose frequency is $\nu = \omega/2\pi = E_\gamma/2\pi\hbar = 1.42041 \times 10^9$ Hz, corresponding to a wavelength

$$\lambda \;=\; c/\nu \;=\; 21.106\,\text{cm}. \tag{15.2.17}$$

This is the "21 centimeter line" of radio astronomy, which comes from a photon emission or absorption between states with different orientations of the electron and proton spins in the hydrogen atom ground state.

Now take the case $l \neq 0$. We will use the Landé projection formula eq. (12.2.10) derived in section 12.2 for a general vector operator, given in our present problem by the vector that appears dotted into $\vec{S}_p$ in eq. (15.2.6),

$$\vec{V} \;=\; \frac{1}{R^3}(\vec{L} - \vec{S}) + \frac{3}{R^5}(\vec{R} \cdot \vec{S})\,\vec{R}, \tag{15.2.18}$$

so that

$$H_{\rm hf}^{l\neq 0} \;=\; \frac{g_p e^2}{2 m_e m_p c^2} \vec{V} \cdot \vec{S}_p. \tag{15.2.19}$$

Here we have used the fact that the $\delta^{(3)}(\vec{R})$ term does not contribute for $l \neq 0$, as noted in the calculation of eq. (15.2.12). Because $\vec{V}$ itself does not involve $\vec{S}_p$ at all, for the purposes of evaluating its matrix elements we can temporarily ignore the $f$ and $m_f$ quantum numbers when using eqs. (12.2.10) and (12.2.11), and so take the total angular momentum operator to be just $\vec{J} = \vec{L} + \vec{S}$, the generator of rotations for the orbital and electron spin degrees of freedom.[†] Also, in our present problem, we are calculating an energy correction as an expectation value, so the degeneracy labels $\alpha$ and $\beta$ appearing in eq. (12.2.10) are the same, $\alpha = (n, l)$, and the Landé projection formula eq. (12.2.10) gives

$$\langle \alpha, j, m_j' | \vec{V} | \alpha, j, m_j \rangle \;=\; \langle j, m_j' | \vec{J}\, | j, m_j \rangle \, \frac{\langle \alpha, j \| \vec{J} \cdot \vec{V} \| \alpha, j \rangle}{\hbar^2 j(j+1)}, \tag{15.2.20}$$

where the reduced matrix element is the scalar expectation value

$$\langle \alpha, j \| \vec{J} \cdot \vec{V} \| \alpha, j \rangle \;=\; \langle \alpha, j, m_j | \vec{J} \cdot \vec{V}\, | \alpha, j, m_j \rangle\,. \tag{15.2.21}$$

---

[†] If $\vec{V}$ had contained $\vec{S}_p$, then the Landé projection formula eq. (12.2.10) would apply only if $\vec{J}$ were replaced by $\vec{F}$, the total angular momentum operator including the proton spin.

Recall that, because of Theorem 12.2.1, since $\vec{J}\cdot\vec{V}$ is a scalar operator, $\langle\alpha, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle$ does not actually depend on $m_j$ at all, despite its appearance on the right side of eq. (15.2.21).

To evaluate $\langle\alpha, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle$ we therefore need the expectation values $\langle\vec{J}\cdot(\vec{L}-\vec{S})/R^3\rangle$ and $\langle(\vec{R}\cdot\vec{J})(\vec{R}\cdot\vec{S})/R^5\rangle$ in a state $|\alpha, j, m_j\rangle$. For the first of these, we calculate:

$$\vec{J}\cdot(\vec{L}-\vec{S}) \;=\; (\vec{L}+\vec{S})\cdot(\vec{L}-\vec{S}) \;=\; L^2 - S^2 \;=\; \hbar^2[l(l+1) - 3/4]. \tag{15.2.22}$$

For the second, we use $\vec{R}\cdot\vec{L} = \vec{R}\cdot(\vec{R}\times\vec{P}) = 0$ to obtain $\vec{R}\cdot\vec{J} = \vec{R}\cdot\vec{S}$, and therefore

$$(\vec{R}\cdot\vec{J})(\vec{R}\cdot\vec{S}) \;=\; (\vec{R}\cdot\vec{S})(\vec{R}\cdot\vec{S}) \;=\; \frac{\hbar^2}{4}(\vec{R}\cdot\vec{\sigma})(\vec{R}\cdot\vec{\sigma}) \;=\; \hbar^2 R^2/4, \tag{15.2.23}$$

where we have used the Pauli matrix representation of spin, and then the identity eq. (8.2.18). Putting eqs. (15.2.22) and (15.2.23) into eq. (15.2.21) gives

$$\langle\alpha, j\|\vec{J}\cdot\vec{V}\|\alpha, j\rangle \;=\; \hbar^2 l(l+1)\,\langle n, l|\frac{1}{R^3}|n, l\rangle \tag{15.2.24}$$

for the operator $\vec{V}$ defined in eq. (15.2.18). So, we can write for eq. (15.2.19), within matrix elements with common $n, l, j$,

$$H_{\text{hf}}^{l\neq 0} \;=\; \frac{g_p e^2}{2m_e m_p c^2}\frac{l(l+1)}{j(j+1)}\,\vec{J}\cdot\vec{S}_p\,\frac{1}{R^3}. \tag{15.2.25}$$

In the orthobasis $|n, l, j, f, m_f\rangle$, the angular momentum operator factor in eq. (15.2.25) is diagonal, and evaluates to

$$\vec{J}\cdot\vec{S}_p \;=\; \frac{1}{2}(F^2 - J^2 - S_p^2) \;=\; \frac{\hbar^2}{2}(f(f+1) - j(j+1) - 3/4). \tag{15.2.26}$$

The radial part of the expectation value was found in eq. (10.1.68),

$$\langle 1/R^3\rangle \;=\; \frac{1}{a_0^3 n^3 l(l+1)(l+1/2)}. \tag{15.2.27}$$

Using eqs. (15.2.26) and (15.2.27) to find the expectation value of eq. (15.2.25), and eliminating $j$ in favor of $f$, gives

$$E_{\text{hf}} \;=\; \langle H_{\text{hf}}\rangle \;=\; \frac{g_p m_e}{m_p}\frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\frac{\pm 1}{(l+1/2)(2f+1)}, \tag{15.2.28}$$

where the $+$ sign applies if $f = j + 1/2$ and the $-$ sign if $f = j - 1/2$. Actually, this agrees with the $l = 0$, $j = 1/2$ formula in eq. (15.2.15), so that eq. (15.2.28) is an all-purpose result, correct for all states $|n, l, j, f, m_f\rangle$ in the total angular momentum basis.

Note that the hyperfine energy contribution is independent of the quantum number $m_f$. This had to be true, because the energy eigenvalues for the hydrogen atom cannot possibly depend on the orientation of the total angular momentum with respect to an arbitrarily chosen $\hat{z}$ direction. Therefore, each state with total angular momentum quantum number $f$ has a remaining exact degeneracy of $2f+1$. The only way to break this degeneracy is to introduce some external effect that picks out a special direction in space, for example an external electric or magnetic field.

The average hyperfine energy contribution for any given $n, l, j$ level is 0. Of the $2(2j+1)$ states before the hyperfine splitting is taken into account, the $2j+2$ states with $f = j+1/2$ are raised by an amount multiplied by $1/(2j+2)$, while the $2j$ states with $f = j-1/2$ are lowered by the same amount multiplied by $1/2j$.

In terms of the total $n = 1$ hyperfine splitting $E_\gamma$ defined in eq. (15.2.16), the hyperfine splittings of the $n = 2$ states are

$$2S_{1/2}(l=0, j=1/2): \qquad E_{\rm hf} = \begin{cases} E_\gamma/32 & (f=1) \\ -3E_\gamma/32 & (f=0) \end{cases} \quad\rightarrow\quad \Delta E_{\rm hf}^{2S_{1/2}} = E_\gamma/8, \tag{15.2.29}$$

$$2P_{1/2}(l=1, j=1/2): \qquad E_{\rm hf} = \begin{cases} E_\gamma/96 & (f=1) \\ -E_\gamma/32 & (f=0) \end{cases} \quad\rightarrow\quad \Delta E_{\rm hf}^{2P_{1/2}} = E_\gamma/24, \tag{15.2.30}$$

$$2P_{3/2}(l=1, j=3/2): \qquad E_{\rm hf} = \begin{cases} E_\gamma/160 & (f=2) \\ -E_\gamma/96 & (f=1) \end{cases} \quad\rightarrow\quad \Delta E_{\rm hf}^{2P_{3/2}} = E_\gamma/60. \tag{15.2.31}$$

So far, we have evaluated the fine and hyperfine contributions, by working at first order in degenerate perturbation theory. However, there is a second-order effect, called the **Lamb shift** after Willis E. Lamb Jr., which for $l = 0$ states is numerically larger than the hyperfine effect. The Lamb shift arises due to the electron emitting and reabsorbing virtual photons, quanta of the electromagnetic field. To calculate it in a systematic way requires the toolbox of quantum electrodynamics, including renormalization of ultraviolet divergent contributions and a careful treatment of infrared effects. Understanding these issues played a critical role in the development of relativistic quantum field theory, but here we can only summarize the results numerically. The Lamb shift for atomic hydrogen states $|n, l, j\rangle$ has been calculated to be, approximately,

$$\Delta E_{\rm Lamb} \;=\; \frac{\alpha^3}{n^3}\left(\frac{e^2}{2a_0}\right) \times \begin{cases} 6.50 & (\text{for } l=0,\ j=1/2), \\ 0.026 \pm \dfrac{1}{\pi(j+1/2)(2l+1)} & (\text{for } l\neq 0,\ j=l\pm 1/2), \end{cases} \tag{15.2.32}$$

where the decimal coefficients actually have a slight dependence on $n$. Note that the Lamb shift is much larger for $s$-wave ($l = 0$) states than for $l \neq 0$ states. The extra factor of $\alpha$ here (compared to the $\alpha^2$ factor for fine and hyperfine energy corrections) is indicative of an effect of second order in perturbation theory. Effects that are even higher order in $\alpha = 1/137.036$ give slight modifications to each of the fine, hyperfine, and Lamb shift contributions, so each of the energy splittings that we have calculated here cannot be trusted beyond about the per cent level. More heroic calculations, not reviewed here, have been done to improve beyond that accuracy.

The fine, hyperfine, and Lamb shift splitting effects on the $n = 2$ states of atomic hydrogen are summarized and illustrated in Figure 15.2.1.

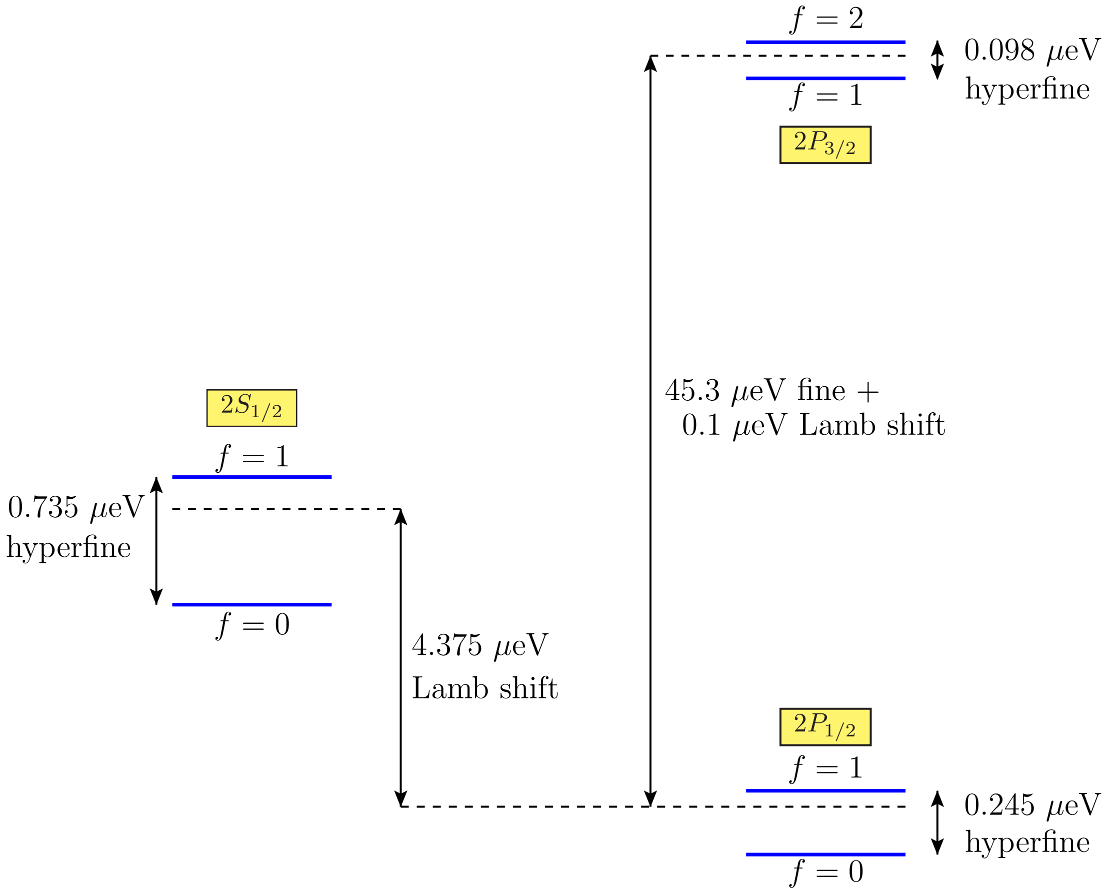

Figure 15.2.1: The fine, hyperfine, and Lamb shift energy splittings in $\mu$eV for the $n = 2$ energy levels of the hydrogen atom, with $nl_j$ spectroscopic notation where $l$ is the orbital angular momentum replaced by its (capitalized) letter code as in eq. (10.1.20), $j$ is the total angular momentum quantum number excluding the proton spin, and $f$ is the total angular momentum quantum number including the proton spin. Relative energy spacings are indicated qualitatively, but not to scale. The Lamb-shift splitting between the $2S_{1/2}$ and $2P_{1/2}$ levels is the experimental one, which differs slightly from the calculation quoted in eq. (15.2.32) due to higher order effects. Each energy level shown has a degeneracy $2f + 1$.

## 15.3 Hydrogen atom in external magnetic field (Zeeman and Paschen–Back effects)

Atomic and molecular state properties can be probed using an external magnetic field. In this section, we consider the energy levels of the hydrogen atom in the presence of a uniform constant magnetic field, which is taken without loss of generality to point in the $\hat{z}$ direction,

$$\vec{B} = B\hat{z}. \tag{15.3.1}$$

This is associated with the Coulomb-gauge vector potential in spherical coordinates,

$$\vec{A} = \frac{1}{2}Br\sin\theta\,\hat{\phi}. \tag{15.3.2}$$

We wish to apply this in the Hamiltonian for a charged particle moving in an electromagnetic field, eq. (4.3.35), which includes $\vec{A}\cdot\vec{P}$. Working in the position representation with $\vec{P} = -i\hbar\vec{\nabla}$,

$$\vec{A}\cdot\vec{P} = \frac{1}{2}B\Big(-i\hbar\frac{\partial}{\partial\phi}\Big) = \frac{1}{2}BL_z, \tag{15.3.3}$$

where the last equality uses the differential representation of $L_z$ of eq. (8.6.7). Therefore, from eq. (4.3.35), the external magnetic field contribution to the electron's Hamiltonian in the hydrogen atom is

$$\Delta H = \frac{eB}{2m_e c}\left(L_z + g_e S_z\right) + \frac{e^2B^2}{8m_e c^2}r^2\sin^2\theta. \tag{15.3.4}$$

Let us first get a rough idea of the order of magnitude of these contributions, for typical laboratory magnetic fields. Putting in the numerical values of the constants gives

$$\Delta H = (5.8\times10^{-5}\,\text{eV})\frac{1}{\hbar}\left(L_z + g_e S_z\right)\frac{B}{\text{Tesla}} + (6.2\times10^{-11}\,\text{eV})\frac{r^2}{a_0^2}\sin^2\theta\left(\frac{B}{\text{Tesla}}\right)^2. \tag{15.3.5}$$

From this, we estimate that the term linear in $B$ will contribute comparably to the fine-structure effects in atomic hydrogen if $B$ is of order 1 Tesla. For fields typically accessible in laboratories (of order 10 Tesla or less), the term quadratic in $B$ is much smaller, so we will neglect it. Taking $g_e = 2$, the new Hamiltonian contribution to be considered in this section is therefore

$$H_B = \mu_B B\left(L_z + 2S_z\right)/\hbar, \tag{15.3.6}$$

where the Bohr magneton $\mu_B$ was defined in eq. (4.3.26). Our goal is to understand the energy splitting of hydrogen atomic states due to this Hamiltonian perturbation, as a function of the magnetic field strength.

Let us first consider the weak-field limit, in which $\mu_B B$ can be treated as a perturbation that is small compared to the fine-structure contributions, but still large compared to the hyperfine energy splittings, which we will neglect. This is called the **Zeeman effect**, after Pieter Zeeman. We apply degenerate first-order perturbation theory to the fine-structure eigenstates $|n, l, j, m\rangle$ with unperturbed energies

$$E = \frac{e^2}{2a_0}\left[-\frac{1}{n^2} + \frac{\alpha^2}{n^3}\left(\frac{3}{4n} - \frac{1}{j+1/2}\right)\right]. \tag{15.3.7}$$

Following the general discussion of degenerate perturbation theory in section 13.6, the energy shifts due to $B$ will be the eigenvalues of the $(2j+1) \times (2j+1)$ perturbation matrix

$$(H_B)_{m',m} \;=\; \frac{\mu_B B}{\hbar} \langle n, l, j, m'|(L_z + 2S_z)|n, l, j, m\rangle\,, \tag{15.3.8}$$

for each energy level with fixed $n, l, j$. (There is no need to consider matrix elements between different values of $l$, because the perturbation $L_z + 2S_z$ cannot connect states with different $l$. We will also suppress the common label $n$ in the following.) Now we write

$$L_z + 2S_z \;=\; J_z + S_z, \tag{15.3.9}$$

and replace $J_z$ by its eigenvalue, using $\langle l, j, m'|J_z|l, j, m\rangle = \hbar m \delta_{m,m'}$. To finish the calculation, we need the matrix elements $\langle l, j, m'|S_z|l, j, m\rangle$. Here, the Landé projection formula of eqs. (12.2.10) and (12.2.11) comes to our aid, with $\vec{S}$ in the role of $\vec{V}$. It tells us that the matrix elements of $S_z$ are proportional to those of $J_z$, with

$$\langle l, j, m'|S_z|l, j, m\rangle \;=\; \langle l, j, m'|J_z|l, j, m\rangle\, \frac{\langle l, j, m|\vec{J}\cdot\vec{S}|l, j, m\rangle}{\hbar^2 j(j+1)}. \tag{15.3.10}$$

Now, we again evaluate $\langle l, j, m'|J_z|l, j, m\rangle = \hbar m \delta_{m,m'}$, and are delighted to see that the whole $H_B$ is proportional to $\delta_{m,m'}$, and so is actually a diagonal matrix on the degenerate subspace of total angular momentum basis states. It remains to calculate the matrix element $\langle l, j, m|\vec{J}\cdot\vec{S}|l, j, m\rangle$. This can be done using the standard trick for dot products of angular momenta,

$$\vec{J}\cdot\vec{S} \;=\; \frac{1}{2}\left[J^2 + S^2 - (\vec{J}-\vec{S})^2\right] \;=\; \frac{1}{2}\left[J^2 + S^2 - L^2\right], \tag{15.3.11}$$

which can be replaced by its eigenvalue, $\hbar^2[j(j+1) + s(s+1) - l(l+1)]/2$, when acting on the state $|l, j, m\rangle$.

Putting the ingredients of eq. (15.3.8) together, the energy shifts are

$$\Delta E_B \;=\; g \mu_B B m, \tag{15.3.12}$$

where

$$g \;=\; \frac{3}{2} + \frac{s(s+1) - l(l+1)}{2j(j+1)} \tag{15.3.13}$$

is called the **Landé $g$-factor**. These energy splittings are directly proportional to the magnetic quantum number $m$ for the total angular momentum. Using the facts that $s = 1/2$ and $j = l \pm 1/2$, we can rewrite the Landé $g$-factor for the hydrogen atom as

$$g \;=\; \begin{cases} 1 + \frac{1}{2j} & (\text{for } j = l + 1/2), \\ 1 - \frac{1}{2j+2} & (\text{for } j = l - 1/2). \end{cases} \tag{15.3.14}$$

The weak-field Zeeman energy shifts for the lowest few values of $j$ are therefore

$$\begin{aligned}
\Delta E_{S_{1/2}} &= 2\mu_B B m, \qquad (m = \pm 1/2), && (15.3.15)\\
\Delta E_{P_{1/2}} &= \frac{2}{3}\mu_B B m, \qquad (m = \pm 1/2), && (15.3.16)\\
\Delta E_{P_{3/2}} &= \frac{4}{3}\mu_B B m, \qquad (m = \pm 3/2,\, \pm 1/2), && (15.3.17)\\
\Delta E_{D_{3/2}} &= \frac{4}{5}\mu_B B m, \qquad (m = \pm 3/2,\, \pm 1/2), && (15.3.18)\\
\Delta E_{D_{5/2}} &= \frac{6}{5}\mu_B B m, \qquad (m = \pm 5/2,\, \pm 3/2,\, \pm 1/2), && (15.3.19)
\end{aligned}$$

etc. By applying a weak magnetic field, the energy degeneracies for each principal quantum number $n$ are eliminated, and the observed splittings can be used to count the (previously degenerate) states and deduce their angular momentum quantum numbers. This phenomenon is the historical reason for calling $m$ the "magnetic quantum number".

Next, consider the opposite limit, called the **Paschen–Back** regime after Friedrich Paschen and Ernst Back, in which the effect of the external magnetic field is taken to be much stronger than the fine-structure effect, which will be treated as a perturbation. The Hamiltonian $H_0 = P^2/2m_e - e^2/R$ commutes with the magnetic field Hamiltonian $H_B$ in eq.( 15.3.6), so the exact energy eigenstates of $H_0 + H_B$ are the product angular momentum states $|n, l, m_l, m_s\rangle$, with energy eigenvalues $E = -e^2/2a_0 n^2 + \Delta E_B$, where

$$\Delta E_B = \mu_B B(m_l + 2m_s). \qquad (15.3.20)$$

Now we must apply the fine-structure Hamiltonian to these states as a perturbation. A key point is that when doing so, we cannot use the states $|n, l, j, m\rangle$ as the unperturbed states, as we did in the previous treatment of the fine structure with a vanishing or weak external magnetic field. This is because the strong magnetic field has eliminated the spherical symmetry and the degeneracy associated with it. Instead, degenerate perturbation theory tells us that the fine-structure energy shifts are obtained as the expectation values of the fine-structure Hamiltonian in the product basis states $|n, l, m_l, m_s\rangle$, with their degeneracies for fixed $m_l + 2m_s$.

First, consider the easier case that $l = 0$. Then the spin-orbit Hamiltonian vanishes, and the fine-structure contribution to the energies are the same as found in eq. (15.1.24) from the $H_{\text{rel}}$ and $H_{\text{Darwin}}$ terms. Since $m_l = 0$ in this case, the combined energy shift due to the strong external $B$ field and the fine structure is

$$\Delta E^{l=0}_{B,\,\text{fine}} = 2\mu_B B m_s + \frac{\alpha^2}{n^3}\left(\frac{e^2}{2a_0}\right)\left(\frac{3}{4n} - 1\right). \qquad (15.3.21)$$

Now consider the case $l \neq 0$. From the spin-orbit Hamiltonian in eq. (15.1.17), the energy correction from first-order perturbation theory is

$$\langle n, l, m_l, m_s | H_{\rm SO} | n, l, m_l, m_s \rangle \;=\; \frac{e^2}{2m_e^2 c^2} \langle 1/R^3 \rangle \, \langle l, m_l, m_s | \vec{L} \cdot \vec{S} | l, m_l, m_s \rangle \,. \qquad (15.3.22)$$

The expectation value $\langle 1/R^3 \rangle$ was already given in eq. (15.1.20). To evaluate the expectation value of $\vec{L} \cdot \vec{S}$, it is convenient to use

$$\vec{L} \cdot \vec{S} \;=\; \frac{1}{2}(L_+ S_- + L_- S_+) + L_z S_z, \qquad (15.3.23)$$

because in the product basis states the expectation value of $L_z S_z$ evaluates to $\hbar^2 m_l m_s$, and the expectation values of $S_\pm$ and $L_\pm$ vanish. (Note the difference from the previously treated spin-orbit case with a weak magnetic field, where one must instead evaluate the expectation value of $\vec{L} \cdot \vec{S}$ in the total angular momentum basis states.) Combining the spin-orbit result with the $H_{\rm rel}$ contribution from eq. (15.1.16), and recalling from eq. (15.1.23) that the Darwin contribution vanishes for $l \neq 0$, we obtain the total first-order perturbation contribution

$$\Delta E^{l\neq 0}_{B,\rm fine} \;=\; \mu_B B(m_l + 2m_s) + \alpha^2 \frac{e^2}{2a_0} \frac{1}{n^3} \left( \frac{m_l m_s}{l(l+1/2)(l+1)} + \frac{3}{4n} - \frac{1}{l+1/2} \right) \qquad (15.3.24)$$

from the strong external $B$ field and the fine-structure effects.

Finally, let us consider the intermediate case in which the effects of the fine structure $H_{\rm fine} = H_{\rm rel} + H_{\rm SO} + H_{\rm Darwin}$ and external magnetic field $H_B$ Hamiltonians are comparable to each other, so that $H_{\rm fine} + H_B$ must be treated together as a single perturbation to $H_0$. The preceding results for weak $B$ and for strong $B$ can be obtained as special-case limits of the following analysis. As before, we will ignore the smaller hyperfine effects. For the unperturbed states, it is not a bad idea to choose[†] the total angular momentum basis elements $|n, l, j, m\rangle$, so that the matrix for the fine-structure Hamiltonian $H_{\rm fine}$ is diagonal and we can make use of results already found in section 15.1.

For $l = 0$ states, there is no distinction between $S_z$ and $J_z$, so $H_B = 2\mu_B B m_s$ is also diagonal in the total angular momentum basis. This means that the energy shift for the $l = 0$ states is the same as we already found in eq. (15.3.21),

For $l \neq 0$ states, things are more complicated. Let us consider only the $n = 2$ level with $l = 1$, where the six unperturbed states are

$$|j, m\rangle \;=\; |\tfrac{3}{2}, \tfrac{3}{2}\rangle\,,\; |\tfrac{3}{2}, \tfrac{1}{2}\rangle\,,\; |\tfrac{3}{2}, -\tfrac{1}{2}\rangle\,,\; |\tfrac{3}{2}, -\tfrac{3}{2}\rangle\,, \qquad (2P_{3/2}), \qquad (15.3.25)$$

$$|\tfrac{1}{2}, \tfrac{1}{2}\rangle\,,\; |\tfrac{1}{2}, -\tfrac{1}{2}\rangle\,, \qquad (2P_{1/2}). \qquad (15.3.26)$$

The matrix elements of $L_z + 2S_z$ between any pair of these states can be quickly evaluated using eqs. (11.3.16)–(11.3.21), which give these $|j, m\rangle$ states in terms of the product orthobasis states

---

[†] It is also possible to choose the product basis eigenstates of $H_B$ as the unperturbed states. Then the $H_{\rm fine}$ matrix elements will be off-diagonal, but the final results for the energy corrections will be the same.

$|m_l, m_s\rangle$, allowing the operator $L_z + 2S_z$ to be replaced by $\hbar(m_l + 2m_s)$. It is useful to note that the only non-zero matrix elements in the $|j, m\rangle$ basis are between states with the same $m$, because $L_z + 2S_z$ commutes with $J_z$. Therefore, most of the elements of the $6 \times 6$ matrix of perturbations for the $2P_{3/2}$ and $2P_{1/2}$ states vanish. For $H_B$ in eq. (15.3.6), the only relevant non-zero matrix elements in the $|j, m\rangle$ basis with positive $m$ are

$$\begin{aligned}
\langle \tfrac{3}{2}, \tfrac{3}{2}|(L_z + 2S_z)/\hbar|\tfrac{3}{2}, \tfrac{3}{2}\rangle &= 2, && (15.3.27)\\
\langle \tfrac{3}{2}, \tfrac{1}{2}|(L_z + 2S_z)/\hbar|\tfrac{3}{2}, \tfrac{1}{2}\rangle &= 2/3, && (15.3.28)\\
\langle \tfrac{3}{2}, \tfrac{1}{2}|(L_z + 2S_z)/\hbar|\tfrac{1}{2}, \tfrac{1}{2}\rangle &= \langle \tfrac{1}{2}, \tfrac{1}{2}|(L_z + 2S_z)/\hbar|\tfrac{3}{2}, \tfrac{1}{2}\rangle \;=\; -\sqrt{2}/3, && (15.3.29)\\
\langle \tfrac{1}{2}, \tfrac{1}{2}|(L_z + 2S_z)/\hbar|\tfrac{1}{2}, \tfrac{1}{2}\rangle &= 1/3, && (15.3.30)
\end{aligned}$$

while the corresponding matrix elements with negative $m$ each have the opposite sign. From eq. (15.1.27), $H_{\text{fine}}$ is diagonal, with non-zero matrix elements

$$\begin{aligned}
\langle \tfrac{3}{2}, m|H_{\text{fine}}|\tfrac{3}{2}, m\rangle &= -\frac{1}{64}\alpha^2 \frac{e^2}{2a_0}, \qquad (m = \pm 3/2, \pm 1/2), && (15.3.31)\\
\langle \tfrac{1}{2}, m|H_{\text{fine}}|\tfrac{1}{2}, m\rangle &= -\frac{5}{64}\alpha^2 \frac{e^2}{2a_0}. \qquad (m = \pm 1/2). && (15.3.32)
\end{aligned}$$

We can now put together these results to find the matrix elements of $W = H_{\text{fine}} + H_B$. Using the notations

$$a = \frac{\alpha^2}{64}\frac{e^2}{2a_0}, \qquad b = \mu_B B \tag{15.3.33}$$

for simplicity, and ordering the basis elements as $|\tfrac{3}{2}, \tfrac{3}{2}\rangle$, $|\tfrac{3}{2}, -\tfrac{3}{2}\rangle$, $|\tfrac{3}{2}, \tfrac{1}{2}\rangle$, $|\tfrac{1}{2}, \tfrac{1}{2}\rangle$, $|\tfrac{3}{2}, -\tfrac{1}{2}\rangle$, $|\tfrac{1}{2}, -\tfrac{1}{2}\rangle$, the matrix representation for the perturbation is

$$W = \begin{pmatrix}
-a+2b & 0 & 0 & 0 & 0 & 0\\
0 & -a-2b & 0 & 0 & 0 & 0\\
0 & 0 & -a+\frac{2}{3}b & -\frac{\sqrt{2}}{3}b & 0 & 0\\
0 & 0 & -\frac{\sqrt{2}}{3}b & -5a+\frac{1}{3}b & 0 & 0\\
0 & 0 & 0 & 0 & -a-\frac{2}{3}b & \frac{\sqrt{2}}{3}b\\
0 & 0 & 0 & 0 & \frac{\sqrt{2}}{3}b & -5a-\frac{1}{3}b
\end{pmatrix}. \tag{15.3.34}$$

This contains only $1 \times 1$ and $2 \times 2$ non-zero blocks, with eigenvalues

$$\begin{aligned}
\Delta E^{n=2,\, l=1}_{B,\,\text{fine}} \;=\;& -a \pm 2b, && (15.3.35)\\
& -3a + b/2 \pm \sqrt{4a^2 + 2ab/3 + b^2/4}, && (15.3.36)\\
& -3a - b/2 \pm \sqrt{4a^2 - 2ab/3 + b^2/4}. && (15.3.37)
\end{aligned}$$

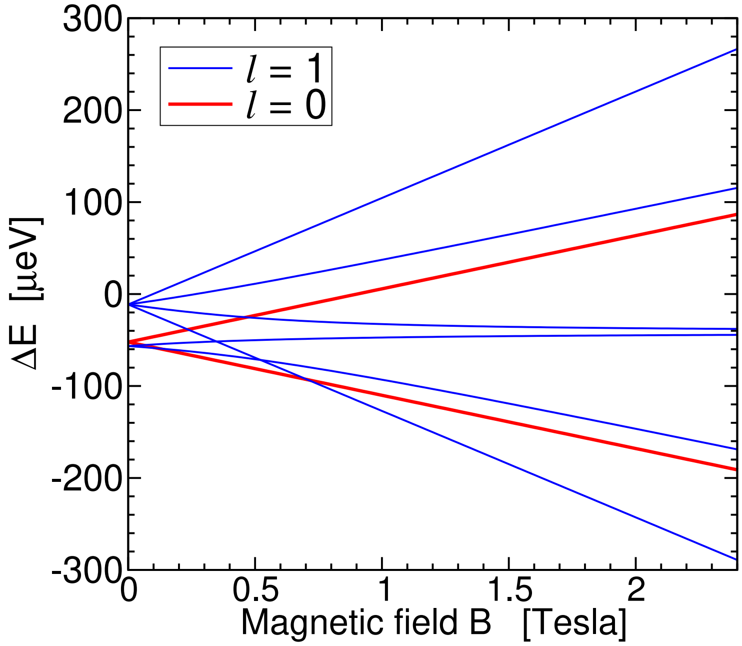

Figure 15.3.1: Energy corrections (in $\mu$eV) to the $n = 2$ states of atomic hydrogen, due to the fine structure Hamiltonian and an external magnetic field $B$ (in Tesla), as found in eqs. (15.3.35)–(15.3.38). At small $B$, the weak-field Zeeman effect energy eigenstates approach total angular momentum eigenstates $|l, s, j, m\rangle$; the group of upper four states are $2P_{3/2}$, and the lower four states are $2P_{1/2}$ and $2S_{1/2}$. The small Lamb shift of 4.375 $\mu$eV has also been included for the $2S_{1/2}$ states, and is barely visible. At large $B$, the strong-field (Paschen–Back regime) energy eigenstates approach product angular momentum eigenstates $|l, s, m_l, m_s\rangle$, with $m_l + 2m_s = 2$, 1, 0, $-1$, and $-2$, from top to bottom.

In the same notation, the $l = 0$ energy corrections from eq. (15.3.21) are

$$\Delta E^{n=2,\,l=0}_{B,\,\text{fine}} \;=\; -5a \pm b, \qquad\qquad . \tag{15.3.38}$$

These eight energy corrections to the $n = 2$ states are graphed in Figure 15.3.1 as a function of the external magnetic field $B$, also including the small positive Lamb shift of 4.375 $\mu$eV for the $l = 0$ eigenstates. It is left as an exercise for the reader to confirm that expanding eqs. (15.3.35)–(15.3.38) in small $b/a$ recovers the weak-field limit of eqs. (15.1.27) and (15.3.15)-(15.3.17), and expanding in small $a/b$ recovers the strong-field limit found in eqs. (15.3.21) and (15.3.24).

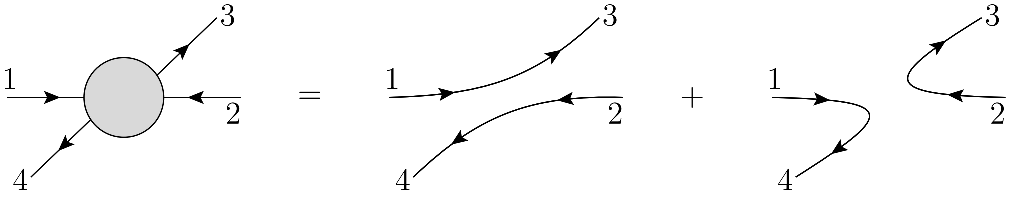

Figure 16.1.1: In scattering of indistinguishable particles (for example electrons) $1, 2 \to 3, 4$, each of the final-state particles labeled 3 and 4 have an equal claim to be the same as each of the initial-state particles 1 and 2 of the same type. It is allowed and useful to label the identical particles, but physical results cannot depend on our labeling choice.

# 16 Identical particles

## 16.1 Intrinsic indistinguishability of identical particles

Particles are said to be **identical** if there is no way of distinguishing them, even in principle. In particular, they must have the same mass, spin magnitude, charge, and any other intrinsic properties. Classically, one can keep track of individual particles, even if they are identical, by following their trajectories, which are determined by the equations of motion. However, in quantum mechanics, this is impossible.

For example, if we scatter two electrons off of each other, as illustrated in Figure 16.1.1, we may choose to label the electrons in the initial state by $1, 2$, and the electrons in the final state by $3, 4$. However, it has no meaning to say that final-state electron 3 is uniquely the same as the initial-state electron 1. Both of the outgoing electrons 3 and 4 have an equal claim to being the same as the incoming electron 1 (or 2). Although we may choose a labeling scheme, and it is generally quite useful to do so, the physics results cannot depend at all on that choice.

One way to describe a state with $N$ identical particles is to take a tensor product of $N$ identical individual state spaces. Suppose that we have an orthobasis of single-particle kets $|\alpha_n\rangle$, where each $\alpha_n$ represents a set of CSCO labels for one particle. An orthobasis of kets for the $N$-particle Hilbert space is, mathematically,[†]

$$|\alpha_1\rangle \otimes |\alpha_2\rangle \otimes \cdots \otimes |\alpha_N\rangle. \tag{16.1.1}$$

For example, we could take $\alpha_n = (\vec{r}_n, m_n)$, and the orthobasis elements would be

$$|\vec{r}_1, m_1\rangle \otimes |\vec{r}_2, m_2\rangle \otimes \cdots \otimes |\vec{r}_N, m_N\rangle, \tag{16.1.2}$$

where the interpretation of $|\vec{r}_n, m_n\rangle$ is that we are certain to find the particle at position $\vec{r}_n$ with spin component $S_z = \hbar m_n$. Alternatively, $\alpha_n$ could instead include the momentum $\vec{p}_n$

---

[†] We say "mathematically" here because the physical $N$-identical-particle Hilbert space is a proper subset of the mathematical $N$-identical-particle Hilbert space, as we will soon see.

rather than the position, or it could include the energy, $L^2$, and $L_z$ quantum numbers of the $n$th particle, or some other choice of CSCO eigenvalues. In any case, we can dispense with the tensor product notation and write eq. (16.1.1) as a single ket for a given orthobasis element,

$$|\alpha_1, \alpha_2, \dots \alpha_N\rangle. \tag{16.1.3}$$

The fact that the particles are identical means that although we may choose to label them by integers $n = 1, 2, \dots, N$, we cannot distinguish them, and so physical results cannot depend on our arbitrary choice of labels. In particular, this means that any observable operator must be unchanged if we exchange any pair of labels.

For example, the Hamiltonian for two identical particles of mass $m$ can be of the form

$$H = \frac{P_1^2}{2m} + \frac{P_2^2}{2m} + V(\vec{R}_1) + V(\vec{R}_2) + V_{\rm int}(|\vec{R}_1 - \vec{R}_2|), \tag{16.1.4}$$

since this is invariant under the exchange of labels 1 ↔ 2. It is important here that the potentials $V(\vec{r})$ for the individual identical particles 1 and 2 are the same function. The total momentum operator for $N$ identical particles,

$$\vec{P} = \vec{P}_1 + \vec{P}_2 + \cdots + \vec{P}_N, \tag{16.1.5}$$

is likewise an observable, but the individual operators $\vec{P}_n$ appearing in it are not observables. Similarly, we can define the observable operator for the density at a fixed position $\vec{r}$ as

$$\rho(\vec{r}) = \sum_{n=1}^{N} \delta^{(3)}(\vec{r} - \vec{R}_n). \tag{16.1.6}$$

Here $\vec{r}$ is an ordinary 3-vector that labels which operator we are talking about, and $\vec{R}_n$ are the individual position operators for the identical particles.

**Pair-exchange operators** exchange the labels of two identical particles, and therefore are defined by

$$P_{ij}|\dots, \alpha_i, \dots, \alpha_j, \dots\rangle = |\dots, \alpha_j, \dots, \alpha_i, \dots\rangle. \tag{16.1.7}$$

So, for example, $P_{13}|\alpha_1, \alpha_2, \alpha_3\rangle = |\alpha_3, \alpha_2, \alpha_1\rangle$ in the special case of a system of three identical particles. Note that $P_{ij}$ is both unitary and Hermitian (like the parity operator), and $P_{ij} = P_{ji}$, and $(P_{ij})^2 = I$. It follows from the last property that if $P_{ij}\,|\psi\rangle = \lambda\,|\psi\rangle$, then $\lambda^2 = 1$ unless $|\psi\rangle$ is the null ket, so the only possible eigenvalues of $P_{ij}$ are $\lambda = \pm 1$. Despite being Hermitian, $P_{ij}$ is not itself an observable, since, for example, $P_{12}$ is not left unchanged if we exchange identical particles 1 and 3.

Now, since observables must be symmetric under the pair exchange of labels, we have

**Theorem 16.1.1** *For any observable $A$, and any pair-exchange operator $P_{ij}$ for identical particles,*

$$[P_{ij}, A] \;=\; 0. \tag{16.1.8}$$

Proof: write $A = A(a_1, a_2, \ldots, a_N)$, where the $a_n$ are operators that act non-trivially only on orthobasis kets $|\alpha_n\rangle$ for the $n$th particle. Since $A$ must be symmetric under exchange of labels,

$$A(\ldots, a_i, \ldots, a_j, \ldots) = A(\ldots, a_j, \ldots, a_i, \ldots). \tag{16.1.9}$$

It follows that

$$P_{ij}A|\ldots, \alpha_i, \ldots, \alpha_j, \ldots\rangle \;=\; A|\ldots, \alpha_j, \ldots, \alpha_i, \ldots\rangle \;=\; AP_{ij}|\ldots, \alpha_i, \ldots, \alpha_j, \ldots\rangle, \tag{16.1.10}$$

where the first equality made used of eq. (16.1.9). Since the states $|\ldots, \alpha_i, \ldots, \alpha_j, \ldots\rangle$ are an orthobasis, the theorem follows by linearity since $P_{ij}A$ must equal $AP_{ij}$ acting on every state.

In particular, the Hamiltonian must commute with every $P_{ij}$,

$$[H, P_{ij}] \;=\; 0. \tag{16.1.11}$$

Now suppose that $|E\rangle$ is an eigenket of $H$, with energy eigenvalue $E$. Then,

$$H\left(P_{ij}|E\rangle\right) = P_{ij}H|E\rangle = EP_{ij}|E\rangle, \tag{16.1.12}$$

so $P_{ij}|E\rangle$ is also an eigenket of $H$ with the same energy eigenvalue $E$. This fact sometimes goes by the name of **exchange degeneracy**.

However, while mathematically useful, the exchange degeneracy is really a fake physically, because it only applies to mathematical kets, and not to physical states. The reason is the **spin-statistics principle**: physical states are not just any kets, but are required to be eigenstates of all exchange operators $P_{ij}$ for pairs of identical particles, with eigenvalue $+1$ if the identical particles being exchanged are bosons (have integer spin), and eigenvalue $-1$ if they are fermions (have spin 1/2, 3/2, ...). These are the only two possible eigenvalues, as noted after eq. (16.1.7). Fermions are said to obey **Fermi–Dirac statistics**, and bosons are said to obey **Bose–Einstein statistics**.

The spin-statistics principle is sometimes taken as a postulate, but it can proved as a theorem in 4-d relativistic quantum field theory, the fundamental quantum mechanical framework that underlies our best understanding of the universe at small distance scales. Since it can be proved in that context, rather than assumed, it was not listed among the postulates of quantum mechanics in section 3.1. On the other hand, if a quantum system is not assumed to obey special relativity, then the spin-statistics principle cannot be proved, but can be adopted (or not) as an assumption. Alternatively, one can view the spin-statistics principle simply as an experimentally observed fact, subject at any time to potential falsification.

The spin-statistics principle applies not just for the particles that we currently believe to be fundamental, including bosons (photons, gluons, $W$ and $Z$ bosons, and Higgs bosons) and fermions (electrons, muons, tau leptons, neutrinos, quarks), but even for composite bosons (for example, H atoms, deuterons, $^4$He nuclei, pions) and composite fermions (protons, neutrons, $^3$He nuclei). Whether a composite particle is a boson or a fermion is determined simply by whether its constituents include an even or odd number of fermions, respectively. For example, a $^4$He nucleus consists of two protons and two neutrons. Doing a pair exchange of two $^4$He nuclei therefore involves the simultaneous exchange of four fermions, resulting in a factor of $(-1)^4$. The proton can be viewed as consisting of $3+n$ quarks and $n$ antiquarks, where naively $n=0$, but the presence of virtual particles inside the proton means that $n$ is actually indeterminate. Nevertheless, $n$ takes on integer values, so the pair exchange of two protons always results in factor of $(-1)^{3+n}(-1)^n = -1$, making the proton a fermion.

The exchange degeneracy described by eq. (16.1.12) is completely eliminated from physical states by the spin-statistics principle. For example, if there are two identical particles, then $|\alpha_1, \alpha_2\rangle$ and $|\alpha_2, \alpha_1\rangle$ are not eigenstates of $P_{12}$ unless $\alpha_1 = \alpha_2$. If $\alpha_1 \neq \alpha_2$, then using

$$P_{12}|\alpha_1, \alpha_2\rangle \;=\; |\alpha_2, \alpha_1\rangle, \qquad P_{12}|\alpha_2, \alpha_1\rangle \;=\; |\alpha_1, \alpha_2\rangle, \tag{16.1.13}$$

one finds that the linear combinations

$$|\alpha_1, \alpha_2\rangle_S \;=\; \frac{1}{\sqrt{2}}\left(|\alpha_1, \alpha_2\rangle + |\alpha_2, \alpha_1\rangle\right), \tag{16.1.14}$$

$$|\alpha_1, \alpha_2\rangle_A \;=\; \frac{1}{\sqrt{2}}\left(|\alpha_1, \alpha_2\rangle - |\alpha_2, \alpha_1\rangle\right), \tag{16.1.15}$$

are normalized eigenkets of $P_{12}$, with eigenvalues $+1$ and $-1$, respectively. Therefore, according to the spin-statistics principle, only the linear combination $|\alpha_1, \alpha_2\rangle_S$ is a physical state if the two particles are identical bosons, and only $|\alpha_1, \alpha_2\rangle_A$ is a physical state if the two particles are identical fermions. In either case, if $|\alpha_1, \alpha_2\rangle$ is an energy eigenstate, then exactly one linear combination is physical, and there is no true energy degeneracy associated with the exchange of identical particles. (There may, however, be energy degeneracies for other reasons.)

If instead $\alpha_1 = \alpha_2$, then $|\alpha_1, \alpha_1\rangle_S = |\alpha_1, \alpha_1\rangle$ is a physical state if the particles are bosons; note that only the normalization has changed compared to eq. (16.1.14). But if the particles are fermions, then $|\alpha_1, \alpha_1\rangle_A$ is the null ket, so there is no such physical state. This is the **Pauli exclusion principle**; two identical fermions are not allowed to be in the same state. In particular, they are not allowed to have the same eigenvalues for a CSCO.

To construct physical states that satisfy the spin-statistics principle for $N > 2$ identical particles, we first construct general **permutation operators**, which are arbitrary products of pair-exchange operators, for example $P = P_{12}P_{13}P_{45}$. (Note that the order of the $P_{ij}$ matters, because they do not always commute.) There are many different ways of writing a given permutation operator as a product in this way, because of identities like $P_{12}P_{23}P_{12} = P_{13}$, but the total number of distinct permutation operators is exactly $N!$. They can all be classified as odd or even, according to whether the number of pair-exchange operators in the product is even or odd. If $P$ is an even permutation, we define $(-1)^P = +1$, and if $P$ is an odd permutation, we write $(-1)^P = -1$.

We can then define the **symmetrization and antisymmetrization operators** as sums over all $N!$ distinct permutation operators for $N$ identical particles,

$$\mathcal{S} = \frac{1}{N!}\sum_P P, \tag{16.1.16}$$

$$\mathcal{A} = \frac{1}{N!}\sum_P (-1)^P P. \tag{16.1.17}$$

These are projection operators onto the subspaces of completely symmetric and antisymmetric kets, respectively, and satisfy the properties $\mathcal{S}^2 = \mathcal{S}$ and $\mathcal{A}^2 = \mathcal{A}$ and $\mathcal{S}\mathcal{A} = \mathcal{A}\mathcal{S} = 0$. Furthermore, we have $P_{ij}\mathcal{S} = \mathcal{S}$ and $P_{ij}\mathcal{A} = -\mathcal{A}$ for all $i, j$, so that any ket acted on by $\mathcal{S}$ will be an eigenstate of every $P_{ij}$ with eigenvalue 1, and any ket acted on by $\mathcal{A}$ will be an eigenstate of every $P_{ij}$ with eigenvalue $-1$.

It follows that the physical orthobasis states for $N$ identical fermions are proportional to $\mathcal{A}|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle$,

$$|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_A = \frac{1}{\sqrt{N!}}\sum_P (-1)^P P|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle. \tag{16.1.18}$$

These are the simultaneous eigenstates for every pair-exchange operator $P_{ij}$ with eigenvalue $-1$, unique up to normalization, that one can build out of the 1-particle orthobasis states $|\alpha_i\rangle$. The prefactor has been chosen to make the ket have unit norm if the $|\alpha_i\rangle$ are orthonormal and all distinct from each other. An equivalent way to write this is the **Slater determinant**, named after John C. Slater,

$$|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_A = \frac{1}{\sqrt{N!}}\begin{vmatrix} |\alpha_1\rangle_1 & |\alpha_1\rangle_2 & \ldots & |\alpha_1\rangle_N \\ |\alpha_2\rangle_1 & |\alpha_2\rangle_2 & \ldots & |\alpha_2\rangle_N \\ \vdots & \vdots & \ddots & \vdots \\ |\alpha_N\rangle_1 & |\alpha_N\rangle_2 & \ldots & |\alpha_N\rangle_N \end{vmatrix}, \tag{16.1.19}$$

where the products obtained from the determinant are to be understood as

$$|\alpha_i\rangle_1\,|\alpha_j\rangle_2 \ldots |\alpha_k\rangle_N = |\alpha_i, \alpha_j, \ldots, \alpha_k\rangle. \tag{16.1.20}$$

If any two of the $\alpha_i$ are the same, then the determinant vanishes, and there is no corresponding physical state at all. This is the more general version of the Pauli exclusion principle.

Similarly, the physical orthobasis states for $N$ identical bosons are proportional to completely symmetrized kets $\mathcal{S}|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle$, because these are the simultaneous eigenstates for all of the pair-exchange operators $P_{ij}$ with eigenvalue $+1$, unique up to normalization, that one can build out of the 1-particle orthobasis states. Here, the normalization factor is complicated slightly by the fact that for bosons the $\alpha_i$ can be the same. If we write $n_i$ for the number of times each distinct $\alpha_i$ is repeated in the list, and take the $|\alpha_i\rangle$ to be orthonormal, then the normalized physical orthobasis states are

$$|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_S \;=\; \sqrt{\frac{N!}{n_1! n_2! \cdots n_N!}}\; \mathcal{S}|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle. \tag{16.1.21}$$

This can also be written as

$$|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_S \;=\; \sqrt{\frac{n_1! n_2! \cdots n_N!}{N!}}\; \sum_P{}' \, P|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle, \tag{16.1.22}$$

where $\sum_P'$ means that the $n_1! n_2! \cdots n_N!$ redundancies in the sum are eliminated, by including each distinct ket $P|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle$ only once.

For example, consider $N = 3$ identical particles. Let us construct the physical states that can be made using a finite number $n$ of orthobasis states for the individual particles, $|\alpha_i\rangle$ with $i = 1, 2, \ldots, n$. If the three identical particles are bosons, then we have $n(n+1)(n+2)/6$ physical orthobasis states, consisting of $n(n-1)(n-2)/6$ states with distinct $i$, $j$, $k$,

$$\begin{aligned} |\alpha_i, \alpha_j, \alpha_k\rangle_S \;=\; & \frac{1}{\sqrt{6}}\Big(|\alpha_i, \alpha_j, \alpha_k\rangle + |\alpha_i, \alpha_k, \alpha_j\rangle + |\alpha_j, \alpha_k, \alpha_i\rangle + |\alpha_j, \alpha_i, \alpha_k\rangle \\ & + |\alpha_k, \alpha_i, \alpha_j\rangle + |\alpha_k, \alpha_j, \alpha_i\rangle\Big), \end{aligned} \tag{16.1.23}$$

and $n(n-1)$ orthobasis states with distinct $i$ and $j$,

$$|\alpha_i, \alpha_i, \alpha_j\rangle_S \;=\; \frac{1}{\sqrt{3}}\Big(|\alpha_i, \alpha_i, \alpha_j\rangle + |\alpha_i, \alpha_j, \alpha_i\rangle + |\alpha_j, \alpha_i, \alpha_i\rangle\Big), \tag{16.1.24}$$

and $n$ orthobasis states,

$$|\alpha_i, \alpha_i, \alpha_i\rangle_S \;=\; |\alpha_i, \alpha_i, \alpha_i\rangle. \tag{16.1.25}$$

Each of these is an eigenstate of each of $P_{12}$, $P_{13}$, and $P_{23}$, with eigenvalue $+1$. For fermions, there are only the $n(n-1)(n-2)/6$ orthobasis states,

$$\begin{aligned} |\alpha_i, \alpha_j, \alpha_k\rangle_A \;=\; & \frac{1}{\sqrt{6}}\Big(|\alpha_i, \alpha_j, \alpha_k\rangle - |\alpha_i, \alpha_k, \alpha_j\rangle + |\alpha_j, \alpha_k, \alpha_i\rangle - |\alpha_j, \alpha_i, \alpha_k\rangle \\ & + |\alpha_k, \alpha_i, \alpha_j\rangle - |\alpha_k, \alpha_j, \alpha_i\rangle\Big) \end{aligned} \tag{16.1.26}$$

for distinct $i, j, k$.

An alternative way of specifying the orthobasis states for $N$ identical particles is by giving the **occupation number** for each of the 1-particle orthobasis states $|\alpha_i\rangle$. The occupation number is simply the number of times $n_i$ that each $\alpha_i$ appears in the symmetrized or antisymmetrized state ket. For bosons, the occupation numbers are just the same as the numbers $n_1, n_2, \ldots$ appearing in eq. (16.1.22). For fermions, each occupation number is always 0 or 1, due to the Pauli exclusion principle. Then, in both cases, the orthobasis states can be written using the occupation-number notation $|n_1, n_2, n_3, \ldots\rangle$, where some ordering has been chosen for the $\alpha_i$. The individual occupation numbers can change with time, due to interactions. However, we will usually treat the total of the occupation numbers as fixed at $\sum_i n_i = N$. (Quantum field theory is what happens to you when you allow $N$ to be variable.)

So far, we have considered orthobasis states for $N$ identical particles, built out of the orthobasis states $|\alpha_i\rangle$ for the individual particles. It is crucial that the resulting states are indeed an orthobasis for the full Hilbert state space of $N$ identical particles, whether they are bosons or fermions. This means that if we define any $N$ linear combinations of 1-particle states

$$|\phi_j\rangle = \sum_i c_{ji}|\alpha_i\rangle, \qquad (j = 1, \ldots, N), \tag{16.1.27}$$

which may be taken to have unit norm but are not necessarily the elements of an orthobasis, then the allowed physical states for identical fermions can be written as

$$|\phi_1, \phi_2, \ldots, \phi_N\rangle_A = \mathcal{A}|\phi_1, \phi_2, \ldots, \phi_N\rangle, \tag{16.1.28}$$

and the allowed physical states for identical bosons are

$$|\phi_1, \phi_2, \ldots, \phi_N\rangle_S = \mathcal{S}|\phi_1, \phi_2, \ldots, \phi_N\rangle, \tag{16.1.29}$$

up to normalization. For fermions, a general Slater determinant ket

$$|\phi_1, \phi_2, \ldots, \phi_N\rangle_A = \begin{vmatrix} |\phi_1\rangle_1 & |\phi_1\rangle_2 & \cdots & |\phi_1\rangle_N \\ |\phi_2\rangle_1 & |\phi_2\rangle_2 & \cdots & |\phi_2\rangle_N \\ \vdots & \vdots & \ddots & \vdots \\ |\phi_N\rangle_1 & |\phi_N\rangle_2 & \cdots & |\phi_N\rangle_N \end{vmatrix} \tag{16.1.30}$$

is physical (provided that it is not null), and can be shown to be a unique linear combination of the orthobasis states $|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_A$ defined by eq. (16.1.19). Here, the states $|\phi_i\rangle$ must be linearly independent, because otherwise the Slater determinant will vanish. Similarly, if the identical particles are bosons, then the state

$$|\phi_1, \phi_2, \ldots, \phi_N\rangle_S = \sum_P P|\phi_1, \phi_2, \ldots, \phi_N\rangle, \tag{16.1.31}$$

is an allowed physical state, even if the $|\phi_i\rangle$ are not orthobasis states or even linearly independent, and any state of this form can be written in a unique way as a linear combination of the orthobasis states $|\alpha_1, \alpha_2, \ldots, \alpha_N\rangle_S$ defined by eq. (16.1.22).

Consider $N$ identical particles, each of which has a single-particle Hamiltonian of the form

$$H_i = \frac{P_i^2}{2m} + V(\vec{R}_i). \tag{16.1.32}$$

For now, we are neglecting any kind of Hamiltonian interaction between the particles [like $V_{\rm int}$ in eq. (16.1.4)]. Suppose that the single-particle energy eigenstate position wavefunctions (neglecting spin) are

$$\phi_0(\vec{r}),\ \phi_1(\vec{r}),\ \phi_2(\vec{r}),\ \phi_3(\vec{r}),\ \ldots, \tag{16.1.33}$$

with individual $H_i$ energy eigenvalues

$$\mathcal{E}_0 < \mathcal{E}_1 \leq \mathcal{E}_2 \leq \mathcal{E}_3 \leq \cdots . \tag{16.1.34}$$

We would like to now consider the eigenstates of the total unperturbed Hamiltonian $H_0 = \sum_i H_i$. Interactions between the particles can be treated later as a perturbation.

If $N = 2$, then the ground state has $E = 2\mathcal{E}_0$, with wavefunction

$$\psi(\vec{r}_1, \vec{r}_2) \;=\; \phi_0(\vec{r}_1)\phi_0(\vec{r}_2). \tag{16.1.35}$$

Since this position-wavefunction part of the state is symmetric under exchange of the labels $1, 2$, the spin state must also be symmetric if the particles are bosons, and the spin state must be antisymmetric if the particles are fermions. Call the spins $\vec{S}_1$ and $\vec{S}_2$, so that the total spin operator is $\vec{S} = \vec{S}_1 + \vec{S}_2$, and denote the eigenvalue of the operator $S^2$ as $\hbar^2 s(s+1)$ where $s$ is the total-spin quantum number. For spin-0 bosons, the total-spin quantum number is of course $s = 0$, and for spin-1 bosons it must be either $s = 0$ or $s = 2$, because of the addition of angular momentum rule

$$1 \otimes 1 \;=\; 0_S \oplus 1_A \oplus 2_S. \tag{16.1.36}$$

For spin-1/2 fermions, the spin state must be $s = 0$, the antisymmetric combination in

$$\frac{1}{2} \otimes \frac{1}{2} \;=\; 0_A \oplus 1_S. \tag{16.1.37}$$

We will explore this in more detail in the next two sections.

For any number $N \geq 2$ bosons, one can have a completely symmetric position wavefunction

$$\psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) \;=\; \phi_0(\vec{r}_1)\phi_0(\vec{r}_2)\phi_0(\vec{r}_3)\cdots\phi_0(\vec{r}_N), \tag{16.1.38}$$

together with a symmetric spin state, with a total spin quantum number that is an even integer, due to eq. (11.4.35). To minimize the energy, it is therefore favorable to simply maximize the occupation number of the single-particle ground state, forming a **Bose–Einstein condensate** with ground-state energy eigenvalue

$$E = N\mathcal{E}_0. \tag{16.1.39}$$

The ground state in the occupation number notation is $|N, 0, 0, \ldots\rangle$, where the ordering of 1-particle state labels is the same as for the energy eigenvalues.

However, for $N \geq 2$ spin-1/2 fermions, at most two can be put into each spatial wavefunction, due to the Pauli exclusion principle. So, one can assign two fermions to $\phi_0$ (with $s = 0$ for their antisymmetric combined spin state), two more to $\phi_1$ (again with $s = 0$ for their spin state) and so on. If $N$ is even, then the last two fermions in an $s = 0$ state will occupy the $\phi_{\frac{N}{2}-1}$ wavefunction, while if $N$ is odd then the last electron will have wavefunction $\phi_{\frac{N-1}{2}}$. The total unperturbed energy for $N$ fermions is therefore

$$E = \begin{cases} 2\mathcal{E}_0 + 2\mathcal{E}_1 + \cdots + 2\mathcal{E}_{\frac{N}{2}-1} & \text{(even } N\text{)}, \\ 2\mathcal{E}_0 + 2\mathcal{E}_1 + \cdots + 2\mathcal{E}_{\frac{N-3}{2}} + \mathcal{E}_{\frac{N-1}{2}} & \text{(odd } N\text{)}. \end{cases} \tag{16.1.40}$$

One application of this counting is to the shell model for electrons in atoms. If one ignores the electron-electron Coulomb repulsion, fine, and hyperfine interactions, and just considers electrons as moving in a central potential, then each single particle state with eigenvalue $E_{n,l,m}$ for $m = -l, \ldots, l$ can be occupied by 2 electrons in an $s = 0$ state. So, in the ground state of the multi-electron atom, the unperturbed energy level $E_{n,l}$ can contain up to $2(2l+1)$ electrons. We will explore this in more detail in section 16.4.

## 16.2 Wavefunctions and spin for two identical particles

In practice, the spin degrees of freedom often play a special role in specifying the orthobasis of states for identical particles. As an example, consider a system of two electrons. We can write the orthobasis kets as

$$|\alpha_1, \alpha_2\rangle \;\;=\;\; |\vec{r}_1, m_1, \vec{r}_2, m_2\rangle, \tag{16.2.1}$$

with possible values $\pm 1/2$ for each of the spin magnetic quantum numbers $m_1$ and $m_2$. The total spin is $\vec{S} = \vec{S}_1 + \vec{S}_2$, and the operator $S^2$ has eigenvalues $\hbar^2 s(s+1)$ where $s = 0, 1$, while the $S_z$ eigenvalue $\hbar m_S$ can have $m_S = 0$ (if $s = 0$) and $m_S = 1, 0, -1$ (if $s = 1$). The state can be specified in a hybrid form, with wavefunctions for the spatial degrees of freedom and kets for the spin degrees of freedom, as

$$\psi_{0,0}(\vec{r}_1, \vec{r}_2)\, \chi_{0,0} + \sum_{m_S=-1}^{1} \psi_{1,m_S}(\vec{r}_1, \vec{r}_2)\, \chi_{1,m_S}. \tag{16.2.2}$$

Here $\chi_{s,m_S}$ is the orthobasis ket for a state with total spin quantum numbers $(s, m_S)$, so that in terms of the product orthobasis kets,

$$\chi_{0,0} = \frac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle - |{\downarrow\uparrow}\rangle\right), \tag{16.2.3}$$

$$\chi_{1,1} = |{\uparrow\uparrow}\rangle, \tag{16.2.4}$$

$$\chi_{1,0} = \frac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle + |{\downarrow\uparrow}\rangle\right), \tag{16.2.5}$$

$$\chi_{1,-1} = |{\downarrow\downarrow}\rangle. \tag{16.2.6}$$

The pair-exchange operator separates into spatial and spin parts, as

$$P_{12} = P_{12}^{\text{spatial}} P_{12}^{\text{spin}}. \tag{16.2.7}$$

The total spin singlet ($s = 0$) state is antisymmetric (eigenvalue $-1$ for $P_{12}^{\text{spin}}$), while the total spin triplet ($s = 1$) states are each symmetric (eigenvalue $+1$ for $P_{12}^{\text{spin}}$). Since the total state must be antisymmetric to describe identical fermions, we must have

$$\psi_{0,0}(\vec{r}_1, \vec{r}_2) = \psi_{0,0}(\vec{r}_2, \vec{r}_1), \tag{16.2.8}$$

$$\psi_{1,m_S}(\vec{r}_1, \vec{r}_2) = -\psi_{1,m_S}(\vec{r}_2, \vec{r}_1). \tag{16.2.9}$$

Thus $\psi_{0,0}$ is a symmetric spatial wavefunction for the total spin singlet, and $\psi_{1,m_S}$ are three distinct antisymmetric spatial wavefunctions for the total spin triplets. In the context of the spatial wavefunctions, "symmetric" (or "antisymmetric") means eigenvalue $+1$ (or $-1$) for $P_{12}^{\text{spatial}}$. If the operator $S^2$ commutes with the Hamiltonian $H$, then they have common eigenstates, so one can label the energy levels as either $s = 1$ or $s = 0$.

Suppose the two electrons are in a potential $V(\vec{r})$ that is large in magnitude compared to their Coulomb interaction with each other and any spin interactions, which we therefore neglect. Then, in that approximation,

$$H = \frac{1}{2m_e}(P_1^2 + P_2^2) + V(\vec{R}_1) + V(\vec{R}_2), \tag{16.2.10}$$

which could be the unperturbed Hamiltonian in a perturbative approach, and the Schrödinger equation for the wavefunction $\psi(\vec{r}_1, \vec{r}_2)$ is

$$\left[-\frac{\hbar^2}{2m_e}(\nabla_1^2 + \nabla_2^2) + V(\vec{r}_1) + V(\vec{r}_2) - E\right]\psi = 0. \tag{16.2.11}$$

Let $\phi_a(\vec{r})$ be the wavefunction solutions to the single-particle eigenvalue differential equation

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V(\vec{r}) - \mathcal{E}_a\right]\phi_a(\vec{r}) = 0. \tag{16.2.12}$$

Then, there are spin singlets (but not spin triplets) with energies $E = 2\mathcal{E}_a$, and total spatial wavefunctions

$$\psi_{0,0}(\vec{r}_1, \vec{r}_2) = \phi_a(\vec{r}_1)\phi_a(\vec{r}_2). \tag{16.2.13}$$

This will apply, for example, if $\mathcal{E}_a$ is the non-degenerate lowest eigenvalue for eq. (16.2.12), so that, as already noted in the previous section, the ground state of the two-electron system is a spin singlet state. Also, for each pair of distinct eigenfunctions $\phi_a(\vec{r})$ and $\phi_b(\vec{r})$ with eigenvalues $\mathcal{E}_a$ and $\mathcal{E}_b$, we have four energy eigenstates with $E = \mathcal{E}_a + \mathcal{E}_b$, with spatial wavefunctions

$$\psi_{0,0}(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}\left[\phi_a(\vec{r}_1)\phi_b(\vec{r}_2) + \phi_b(\vec{r}_1)\phi_a(\vec{r}_2)\right], \qquad (s = 0), \tag{16.2.14}$$

$$\psi_{1,m_S}(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}\left[\phi_a(\vec{r}_1)\phi_b(\vec{r}_2) - \phi_b(\vec{r}_1)\phi_a(\vec{r}_2)\right], \qquad (s = 1,\ m_S = 1,\ 0,\ -1), \tag{16.2.15}$$

which are total spin singlets and triplets, respectively. The eigenvalues $\mathcal{E}_a$ may have degeneracies, so that in the notation we are using here, $\mathcal{E}_a$ and $\mathcal{E}_b$ could be equal even if $\phi_a(\vec{r})$ and $\phi_b(\vec{r})$ are distinct eigenfunctions.

For each of these possibilities, consider the probability to find one electron within a volume $d^3\vec{r}_1$ near $\vec{r}_1$ and the other within a volume $d^3\vec{r}_2$ near $\vec{r}_2$. From Postulate 4 in section 3.1,

$$d\mathcal{P}(\vec{r}_1, \vec{r}_2) = d^3\vec{r}_1\, d^3\vec{r}_2\, |\psi(\vec{r}_1, \vec{r}_2)|^2, \tag{16.2.16}$$

where

$$|\psi(\vec{r}_1, \vec{r}_2)|^2 = \frac{1}{2}\left(|\phi_a(\vec{r}_1)|^2|\phi_b(\vec{r}_2)|^2 + |\phi_b(\vec{r}_1)|^2|\phi_a(\vec{r}_2)|^2 \pm 2\mathrm{Re}\left[\phi_a(\vec{r}_1)\phi_b^*(\vec{r}_1)\phi_b(\vec{r}_2)\phi_a^*(\vec{r}_2)\right]\right), \tag{16.2.17}$$

with the + sign for spin-singlet states and the − sign for spin-triplet states. Note that for either sign, eq. (16.2.17) is invariant under the exchange $\vec{r}_1 \leftrightarrow \vec{r}_2$.

The ± term in eq. (16.2.17) is known as the **exchange density**. As a consequence of it, when the electrons are in the spin-triplet state, they avoid each other. To see this, note that the probability density vanishes when $\vec{r}_1 = \vec{r}_2$, and therefore is small (by continuity) when $\vec{r}_1 \approx \vec{r}_2$. Conversely, in the spin-singlet state, the two electrons "attract"; there is an enhanced probability density for $\vec{r}_1 \approx \vec{r}_2$ due to constructive interference. This attraction and repulsion is not due to any electromagnetic interaction or other Hamiltonian interaction between the identical fermions, but rather due to the Fermi–Dirac statistics.

Electrons in Argonne are identical to those in Batavia. So, it is natural to wonder why do we not need to worry about significant effects due to correlation effects from antisymmetrization of wavefunctions belonging to very distant electrons. Suppose that we define normalized wavefunctions that are peaked in Argonne and Batavia, with very little overlap, so that $\phi_A(\vec{r})$ is insignificant outside of a small volume somewhere in Argonne, and $\phi_B(\vec{r})$ is insignificant outside of a small volume in Batavia, and therefore

$$\phi_A(\vec{r})\phi_B(\vec{r}) \approx 0 \tag{16.2.18}$$

everywhere. The total wavefunction is

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}\big[\phi_A(\vec{r}_1)\phi_B(\vec{r}_2) \pm \phi_B(\vec{r}_1)\phi_A(\vec{r}_2)\big], \tag{16.2.19}$$

where the + sign applies for spin singlets and the − sign for spin triplet states. Note that there is no sense in which the electron labeled 1 is nearer to Argonne, or nearer to Batavia; it is always equally likely to be found in either place.

Now we can ask: what is the probability of seeing an electron within an infinitesimal volume $d^3\vec{r}$ near $\vec{r}$? The answer is

$$d\mathcal{P}(\vec{r}) = d^3\vec{r}\int d^3\vec{r}_2\, |\psi(\vec{r}, \vec{r}_2)|^2 + d^3\vec{r}\int d^3\vec{r}_1\, |\psi(\vec{r}_1, \vec{r})|^2. \tag{16.2.20}$$

The first term is the probability that the electron labeled 1 is near $\vec{r}$, summed over all possible electron 2 positions, and the second term is the same with $1 \leftrightarrow 2$. This can be rewritten as

$$\begin{aligned} d\mathcal{P}(\vec{r}) = d^3\vec{r}\,\bigg(&|\phi_A(\vec{r})|^2 \int d^3\vec{x}\,|\phi_B(\vec{x})|^2 + |\phi_B(\vec{r})|^2 \int d^3\vec{x}\,|\phi_A(\vec{x})|^2 \\ &\pm 2\mathrm{Re}\left[\phi_A(\vec{r})\phi_B(\vec{r})^* \int d^3\vec{x}\,\phi_A(\vec{x})^*\phi_B(\vec{x})\right]\bigg). \end{aligned} \tag{16.2.21}$$

The important point is that the last term with ± is (doubly!) negligible, because of eq. (16.2.18). In each of the first two terms, the integral is 1 by the assumed normalization, so

$$d\mathcal{P}(\vec{r}) = d^3\vec{r}\,\big(|\phi_A(\vec{r})|^2 + |\phi_B(\vec{r})|^2\big) \tag{16.2.22}$$

to a very good approximation. This shows that, despite the form of the wavefunction in eq. (16.2.19), and despite the fact that our labeling $1, 2$ of the electrons did not distinguish either one as being "the electron in Batavia", the probability density decouples into two separate contributions with negligible interference, for sufficiently separated identical particle wavefunctions. It is the lack of wavefunction overlap that enforces the decoupling, not the electron labels. (The same argument works if the particles are bosons.) When in Batavia, we can usually ignore electrons in Argonne, and vice versa.

Suppose that two identical particles are bosons; let us consider the cases of spin 0 and spin 1. For spin 0, the spin degrees of freedom are trivial, and in place of eq. (16.2.2) we have simply a symmetric wavefunction

$$\psi_{0,0}(\vec{r}_1, \vec{r}_2) = \psi_{0,0}(\vec{r}_2, \vec{r}_1). \tag{16.2.23}$$

The case of spin 1 bosons is more interesting, because the two individual spins can combine to form nine different total spin states $\chi_{s,m_S}$, with $s = 0, m_S = 0$ and $s = 1, m_S = 1, 0, -1$ and $s = 2, m_S = 2, 1, 0, -1, -2$. By the addition of angular momentum method discussed in section 11.4, one can write these in terms of the product orthobasis kets $|m_{s_1} : m_{s_2}\rangle$, where the eigenvalues of the individual spin operators $S_{1z}$ and $S_{2z}$ are $\hbar m_{s_1}$ and $\hbar m_{s_2}$, with $m_{s_1}, m_{s_2} = -1, 0, 1$. The total spin $S^2, S_z$ eigenstates are, from eqs. (11.4.25)–(11.4.33),

$$\chi_{2,\pm 2} = |\pm 1 : \pm 1\rangle, \tag{16.2.24}$$

$$\chi_{2,\pm 1} = \frac{1}{\sqrt{2}}\big(|\pm 1 : 0\rangle + |0 : \pm 1\rangle\big), \tag{16.2.25}$$

$$\chi_{2,0} = \frac{1}{\sqrt{6}}\big(|1 : -1\rangle + 2|0 : 0\rangle + |-1 : 1\rangle\big), \tag{16.2.26}$$

$$\chi_{1,\pm 1} = \frac{1}{\sqrt{2}}\big(|\pm 1 : 0\rangle - |0 : \pm 1\rangle\big), \tag{16.2.27}$$

$$\chi_{1,0} = \frac{1}{\sqrt{2}}\big(|1 : -1\rangle - |-1 : 1\rangle\big), \tag{16.2.28}$$

$$\chi_{0,0} = \frac{1}{\sqrt{3}}\big(|1 : -1\rangle - |0 : 0\rangle + |-1 : 1\rangle\big). \tag{16.2.29}$$

These expressions show that $\chi_{s,m_S}$ has $P_{12}^{\text{spin}}$ eigenvalue equal to $(-1)^s$. This same fact can be expressed in the shorthand form for addition of angular momentum

$$1 \otimes 1 = 2_S \oplus 1_A \oplus 0_S. \tag{16.2.30}$$

where the $S, A$ indicate symmetry or antisymmetry under exchange of the two spins. Therefore, the total state for two identical spin-1 bosons is, again in a hybrid notation with wavefunctions for spatial degrees of freedom and kets for spin degrees of freedom,

$$\Psi = \psi_{0,0}(\vec{r}_1, \vec{r}_2)\chi_{0,0} + \sum_{m_S=-1}^{1} \psi_{1,m_S}(\vec{r}_1, \vec{r}_2)\chi_{1,m_S} + \sum_{m_S=-2}^{2} \psi_{2,m_S}(\vec{r}_1, \vec{r}_2)\chi_{2,m_S}. \tag{16.2.31}$$

This contains nine component wavefunctions satisfying

$$\psi_{s,m_S}(\vec{r}_1, \vec{r}_2) = (-1)^s \psi_{s,m_S}(\vec{r}_2, \vec{r}_1). \tag{16.2.32}$$

The position wavefunction for the total spin 0 and 2 states are symmetric, and the position wavefunctions for total spin 1 are antisymmetric, under exchange of the boson labels.

## 16.3 Excited states of the helium atom

The binding energy of the ground state of the two-electron helium atom was studied in sections 13.3 and 14.4. In doing so, we ignored the spin degrees of freedom, as well as the fact that the electrons are identical particles. This is now seen to be justified by the fact that the electrons can arrange themselves into a total spin $s = 0$ state. Then the total spatial wavefunction is the symmetric $\psi_{0,0}(\vec{r}_1, \vec{r}_2) = \psi_{0,0}(\vec{r}_2, \vec{r}_1)$ appearing in eqs. (16.2.2), (16.2.8), and (16.2.13), just as we had assumed (without bothering to justify it in detail) in sections 13.3 and 14.4. As we found in section 14.4, the helium ground state spatial wavefunction is approximately just that of the product of two identical $1s$ hydrogen-like states with $Z_{\text{eff}} \approx 1.7$. The ground state of helium does not have a counterpart with total spin $s = 1$, because that would force the spatial wavefunction to be antisymmetric under exchange of the two electron levels, which is inconsistent with it being the product of two identical lowest-energy single particle states.

However, for the excited states of helium, we must explicitly take into account both spin and the Fermi–Dirac statistics of the electrons. The Hamiltonian is once again given by eqs. (13.3.1)–(13.3.3). We can start by neglecting the $H_{12} = e^2/|\vec{r}_1 - \vec{r}_2|$ interaction Hamiltonian, and later include it as a perturbation. Consider the spatial wavefunctions that can be formed from the individual $Z = 2$ hydrogen-like states, with one electron in the ground state ($1s$) and the other in a higher level, $2s$ or $2p$ or $3s$ or $3p$ or $3d$ or $4s$, etc.:

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}} \left[ \phi_{1,0,0}(\vec{r}_1)\phi_{n,l,m}(\vec{r}_2) \pm \phi_{1,0,0}(\vec{r}_2)\phi_{n,l,m}(\vec{r}_1) \right]. \tag{16.3.1}$$

Here, if the + sign applies, then the spatial wavefunction is symmetric, and so the total-spin state must have $s = 0$; these are called para-helium states. (The ground state is also a para-helium state.) Conversely, if the − sign applies, then the spatial wavefunction is antisymmetric, and so the total-spin state must be symmetric with $s = 1$; these are called ortho-helium states. For both of these, the unperturbed energy eigenvalues of $H_0 = H_1 + H_2$ are

$$\mathcal{E}_{n,l,m} \;=\; -\frac{Z^2 e^2}{2a_0}\left(1 + 1/n^2\right). \tag{16.3.2}$$

Now, treating $H_{12}$ as a perturbation, we obtain the first-order energy correction

$$E^{(1)}_{n,l,m} \;=\; \langle\psi| H_{12} |\psi\rangle \;=\; e^2 \left( I_{n,l,m} \pm J_{n,l,m} \right), \tag{16.3.3}$$

where again the + sign applies to para-helium ($s = 0$) and the − sign to ortho-helium ($s = 1$), and

$$I_{n,l,m} \;=\; \int d^3\vec{r}_1 \int d^3\vec{r}_2 \; |\phi_{1,0,0}(\vec{r}_1)|^2 \, |\phi_{n,l,m}(\vec{r}_2)|^2 \, \frac{1}{|\vec{r}_1 - \vec{r}_2|}, \tag{16.3.4}$$

$$J_{n,l,m} \;=\; \int d^3\vec{r}_1 \int d^3\vec{r}_2 \; \phi_{1,0,0}(\vec{r}_1)\phi_{1,0,0}(\vec{r}_2)\, \phi_{n,l,m}(\vec{r}_1)\phi_{n,l,m}(\vec{r}_2)^* \, \frac{1}{|\vec{r}_1 - \vec{r}_2|}. \tag{16.3.5}$$

Here $I_{n,l,m}$, which is manifestly real and positive, is called the "direct integral", while $J_{n,l,m}$ is called the "exchange integral". Although we will not bother to evaluate them explicitly, it is worth knowing that $J_{n,l,m}$ is also real and positive, but smaller in magnitude than $I_{n,l,m}$. Therefore, each ortho-helium ($s = 1$) state is always lower in energy than the corresponding para-helium ($s = 0$) state. A way of understanding this qualitatively is to note that in the $s = 0$ combination with antisymmetric spins, the symmetric spatial wavefunction results in a higher electrostatic repulsion between the two electrons, due to their greater wavefunction overlap, than for the $s = 1$ combination with antisymmetric spatial wavefunction.

For $s = 0$, the total angular momentum quantum number $j$ is the same as the orbital angular momentum of the excited electron state, $l$. For $l = 0$, $j$ is the same as $s$. For $s = 1$ and $l \geq 1$, one can instead have $j$ values $l - 1$, $l$, or $l + 1$. A sketch of the energy levels for the lowest few states of the neutral helium atom is shown in Figure 16.3.1, classified by their unperturbed quantum numbers on the left, and by their ${}^{2s+1}l_j$ angular momentum labels on the right. The excited states are each split into $s = 0$, para-helium (upper) and $s = 1$, ortho-helium (lower) levels. Transitions between para-helium states and ortho-helium states are highly suppressed, because they require spin flips, so to a good approximation they form two distinct sets of spectral lines. For this reason it was thought by early investigators that they might actually be two separate kinds of helium.

Each of the ortho-helium levels with $l > 1$ is split by small fine-structure effects into its separate $j = l - 1$, $l$ and $l + 1$ states. These splittings are not shown in Figure 16.3.1, but move the states with larger $j$ slightly lower in energy. We will discuss this in the next section in the more general context of multi-electron atoms.

Regarding hyperfine effects, the helium nucleus exists in two stable isotopes. Almost all naturally occurring helium nuclei are ${}^4$He, a boson with spin 0. A spin-0 particle has no special direction and therefore cannot have a magnetic moment or electric quadrupole moment, so ${}^4$He atoms have no hyperfine splitting. A small fraction of naturally occurring helium nuclei are ${}^3$He, a fermion with spin 1/2. The ${}^3$He atomic states with a given $j$ are therefore further split by the hyperfine contribution.

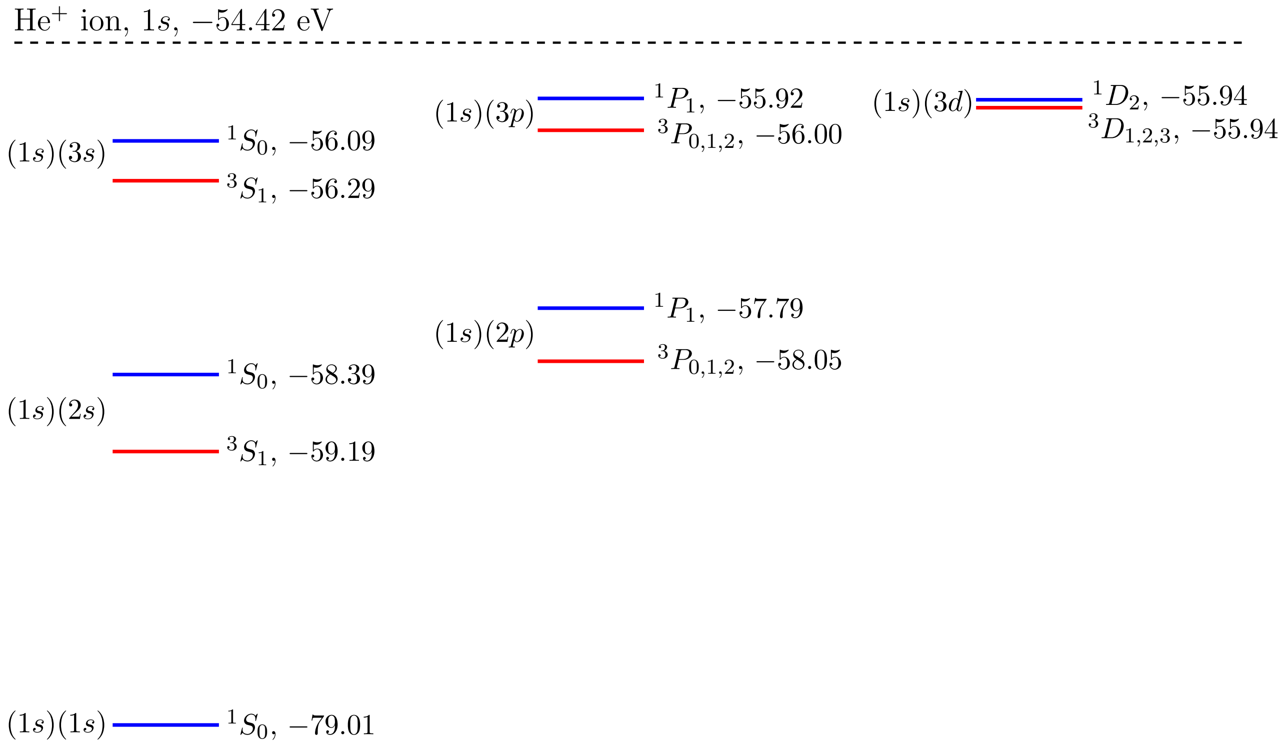

Figure 16.3.1: The lowest few energy levels of the neutral helium atom, with electron configurations, $^{2s+1}l_j$ spectroscopic notations, and energies in eV. Relative energies spacings are indicated qualitatively, but not to scale. The ground state has both electrons in the lowest energy level $n = 1$ and opposite spins ($s = 0$, para-helium). The remaining states have one electron in the lowest level $1s$ and one in an excited energy level $n > 1$, and are split into para-helium (upper, blue lines, $s = 0$) and ortho-helium (lower, red lines, $s = 1$) states. Would-be states with both electrons in an excited level are above the ground-state energy of the He$^+$ ion, shown as the dashed line, for which one electron is in the ground state and the other has been completely ionized away. Not shown are other states with $(1s)(4s)$ and $(1s)(4p)$ etc., which fall below the dashed line and asymptotically approach it for large $n$ for one electron. Small fine structure effects, also not shown, split the ortho-helium levels $^3P_{0,1,2}$, $^3D_{1,2,3}$, etc., into their different $j$ components, with larger $j$ slightly lower in energy.

## 16.4 Multi-electron atoms

Consider the problem of finding the energy eigenstates for multi-electron atoms. To a good approximation, the Hamiltonian depends on the non-relativistic kinetic energy of the electrons, their electrostatic attraction to the nucleus, their pairwise electrostatic repulsion, and the couplings between the electron spins and their angular momenta. These effects alone are enough to provide a challenge that we cannot hope to solve exactly, and approximation methods must be used. In the following, we will not consider other, usually smaller, effects, including the contributions due to the relativistic correction to the electron kinetic energies, Darwin terms, and hyperfine contributions due to the magnetic dipole and electric quadrupole moments of the nucleus. These can be added as further perturbations.

One way to make progress is the **central-field method**. We invent a central potential $U(r)$, a rotationally invariant function (of the radial coordinate only) that is chosen to approximate the electrostatic potential seen by an electron at a given point due to the nucleus and the suitably averaged effects of the other electrons. Although it is not easy to make a rigorous definition of "suitably averaged", there are motivated techniques for estimating $U(r)$ by iteration (notably the Hartree and Hartree–Fock approximations, due to Douglas Hartree and improved by Vladimir Fock) that will not be discussed here. In any case, the choice of $U(r)$ is arbitrary, in principle, because we can add it to one part of the Hamiltonian (the unperturbed part, or the part used to define trial states for use with the variational principle) and subtract it from another (the perturbation part), so that the total Hamiltonian does not depend on it.

More specifically, we can write the approximate Hamiltonian as

$$\begin{aligned} H &= H_0 + W + H_{\rm SO}, && (16.4.1)\\ H_0 &= \sum_{i=1}^{N_e} \left( \frac{P_i^2}{2m_e} + U(R_i) \right), && (16.4.2)\\ W &= \sum_{i=1}^{N_e} \left( -U(R_i) - \frac{Ze^2}{R_i} + \sum_{j=1}^{i-1} \frac{e^2}{|\vec{R}_i - \vec{R}_j|} \right), && (16.4.3)\\ H_{\rm SO} &= \sum_{i=1}^{N_e} \vec{\Omega}_i \cdot \vec{S}_i. && (16.4.4) \end{aligned}$$

This Hamiltonian is invariant under the exchange of any two electrons, as it must be. In the spin-orbit term eq. (16.4.4),

$$\vec{\Omega}_i = -\frac{e}{2m_e^2c^2} \big(\vec{\nabla}\Phi_i\big) \times \vec{P}_i \qquad (16.4.5)$$

where the electric potential seen by the $i$th electron is

$$\Phi_i(\vec{R}_i) = \frac{Ze}{R_i} - \sum_{j\neq i} \frac{e}{|\vec{R}_i - \vec{R}_j|}. \qquad (16.4.6)$$

The contribution to the spin-orbit interaction $H_{\rm SO}$ from each electron generalizes the case of a spherically symmetric potential in eq. (15.1.4), and for hydrogen in eq. (15.1.17), which would be recovered for $\Phi = e/R$. Note that there is no operator ordering problem associated with the placement of $\vec{P}_i$ in eq. (16.4.5), because of the vector calculus identity $\vec{\nabla} \times (\vec{\nabla}\Phi) = 0$. We have distinguished the number of protons in the nucleus $Z$ from the number of electrons $N_e$, so that the formulas apply to charged ions as well as neutral atoms.

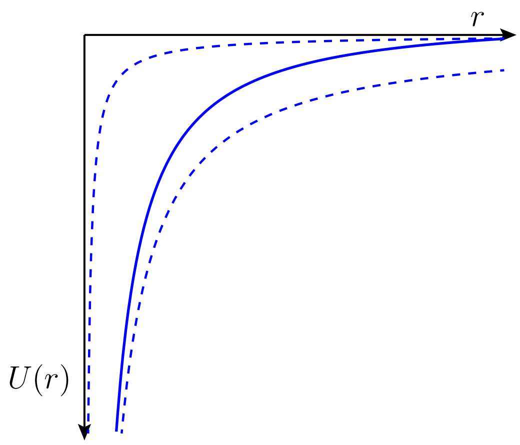

Figure 16.4.1: The solid line shows the qualitative shape of a choice for the central-field potential energy function $U(r)$ for the electrons in a neutral atom with $N_e = Z = 10$. The upper dashed and lower dashed lines show the potentials $-e^2/r$ (maximally screened by 9 electrons closer to the nucleus) and $-10e^2/r$ (no screening, all electrons farther from the nucleus), respectively.

One should aim to choose $U(R)$ in such a way that the eigenstates of $H_0$ are as close as possible to those of the full Hamiltonian. A reasonable choice might interpolate between

$$U(r) \quad \approx \quad \begin{cases} -Ze^2/r & \text{(small } r\text{)}, \\ -(Z - N_e + 1)e^2/r & \text{(large } r\text{)}. \end{cases} \tag{16.4.7}$$

The idea behind this choice is the principle from classical electrostatics, following from Gauss' Law, that the electric field and electric potential at a point $\vec{r}$ in a spherically symmetric charge distribution are both determined only by the total charge contained within $r$. Thus, at small $r$ there is no screening of the charge $+Ze$ nuclear potential due to other electrons, because they are likely to be found farther from the nucleus, while at large $r$ the other $N_e - 1$ electrons are all likely to be closer to the nucleus and so the effective number of charges seen by an outermost very distant electron is only $Z - N_e + 1$. A sketch of a plausible choice of $U(r)$ is shown in Figure 16.4.1, compared to the asymptotic limits of no screening at small $r$ and maximum screening at large $r$. A choice of $U(r)$ should be justified *a posteriori* by the successful convergence of perturbation theory or the variational method.

The single-electron eigenstates of the central-field Hamiltonian $H_0$, called **orbital states** or **orbitals**, will have wavefunctions of the form

$$\phi_{n,l,m_l,m_s}(r, \theta, \phi) \quad = \quad R^U_{n,l}(r) Y^{m_l}_l(\theta, \phi)\, \chi_{m_s}, \tag{16.4.8}$$

where $\chi_{m_s} = |{\uparrow}\rangle$ or $|{\downarrow}\rangle$ for the spin states. The radial wavefunctions $R^U_{n,l}(r)$ will not be the same as for a hydrogen-like atom, because $U(r)$ is not proportional to $1/r$. However, they will be qualitatively similar. In particular, the label $n$ is a positive integer, and then $0 \leq l \leq n-1$, with $n-l-1$ equal to the number of radial zeros for $r > 0$. In general, the orbital energy eigenvalues of $H_0$ will increase with $n$, but will also depend on $l$, unlike hydrogen-like atom energy eigenstates. The unperturbed eigenstates of $H_0$ will be Slater determinants formed out of $N_e$ states of the form of eq. (16.4.8). Then the effects of $W$ and $H_{\text{SO}}$ can be added as perturbations.

The set of orbitals with the same values of $n$ and $l$ are called a **shell**, and consist of

$$g_{n,l}^{\text{shell}} = 2(2l+1) \tag{16.4.9}$$

states, due to the $m_s$ and $m_l$ quantum numbers. (Here we are neglecting any further degeneracy due to nuclear spins, which are different for distinct isotopes of the same element.) The orbital states and shells are labeled $nl$, but with $l = 0, 1, 2, 3, \ldots$ replaced by the spectroscopic code letter $s, p, d, f, \ldots$, just as for the hydrogen atom states. The state of a given atom can be specified by giving the occupation numbers of each shell, called the **electron configuration**, subject to the Pauli exclusion principle, enforced by the Slater determinant, that we cannot put two electrons in the same state due to their Fermi–Dirac statistics. Using eq. (16.4.9), the maximum number of electrons in an $s, p, d, f$ shell is $2, 6, 10, 14$, respectively. For a given $l$, smaller $n$ tends to have lower energy, but this does not always mean that the shells fill up from smaller to larger $n$, as we will soon see. For a given $n$, smaller $l$ states tend to have lower energies. The reason is that electrons with smaller $l$ have support closer to the nucleus, and that part of the wavefunction is less screened from the attractive nucleus by the other electrons in the already-filled shells.

The $1s$ shell can fit up to 2 electrons, and so is the only shell necessary for H and He, but is insufficient to accommodate larger $Z$ atoms. The next to fill is the $2s$ shell, which again fits up to 2 electrons, and so is the last shell needed for Li and Be. The $2p$ shell fills next, and can fit up to 6 electrons, so it is enough to accommodate the neutral atoms with $Z$ up to 10, namely B, C, N, O, Fe, and Ne. Likewise, the $3s$ shell accommodates the last electron for Na and Mg, while the $3p$ shell is used for Al, Si, P, S, Cl, Ar. However, the $4s$ shell tends to fill before the $3d$ shell, in K, Ca, Sc, Ti, V, Mn, Fe, Co, Ni, and Zn. This tendency has two exceptions; within this sequence from $Z = 19$ to $Z = 30$, Cr and Cu break the trend, by having only one electron in the $4s$ shell, with the $3d$ shell containing the other five and ten electrons, respectively.

The shell filling order for multi-electron atoms is, empirically,

$$1s,\ 2s,\ 2p,\ 3s,\ 3p,\ \begin{pmatrix} 4s \\ 3d \end{pmatrix},\ 4p,\ \begin{pmatrix} 5s \\ 4d \end{pmatrix},\ 5p,\ \begin{pmatrix} 6s \\ 4f \\ 5d \end{pmatrix},\ 6p,\ \begin{pmatrix} 7s \\ 5f \\ 6d \end{pmatrix}, \tag{16.4.10}$$

where the cases in parentheses correspond to ambiguities due to close energies, with the most common (but not universally followed) filling order from top to bottom. The fact noted earlier that smaller $l$ tends to give lower energy for a given $n$ explains why $2s$ fills before $2p$, and why the $4s$ can compete with the $3d$, and the $5s$ can compete with the $4d$, etc.

Some example electron configurations are

$$
\begin{gathered}
\text{H:}\quad 1s^1, \qquad \text{He:}\quad 1s^2, \qquad \text{Li:}\quad 1s^2\,2s^1, \qquad \text{C:}\quad 1s^2\,2s^2\,2p^2,\\
\text{N:}\quad 1s^2\,2s^2\,2p^3, \qquad \text{Na:}\quad 1s^2\,2s^2\,2p^6\,3s^1,\\
\text{Hg:}\quad 1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,4s^2\,3d^{10}\,4p^6\,5s^2\,4d^{10}\,5p^6\,6s^2\,4f^{14}\,5d^{10}.
\end{gathered}
$$

The chemical properties of a given element are mostly dependent on the electron configuration, and the similarities between elements in a given column of the periodic table are due to having similar configurations of electrons in the outermost (partially filled) shells.

The inert (noble) gases He, Ne, Ar, Kr, Xe, Rn are those that have a full $p$ shell, and all previous shells filled, leading to a very low tendency to form chemical bonds. It is common to give the electron configuration for any element by just listing the filled orbitals in excess of that for the previous inert gas. For example, for the inert gases themselves, one can write the electron configurations as

$$
\begin{gathered}
\text{He:}\quad 1s^2, \qquad \text{Ne:}\quad [\text{He}]2s^2\,2p^6, \qquad \text{Ar:}\quad [\text{Ne}]3s^2\,3p^6, \qquad \text{Kr:}\quad [\text{Ar}]4s^2\,3d^{10}\,4p^6,\\
\text{Xe:}\quad [\text{Kr}]5s^2\,4d^{10}\,5p^6, \qquad \text{Rn:}\quad [\text{Xe}]6s^2\,4f^{14}\,5d^{10}\,6p^6.
\end{gathered}
$$

For the alkali metals, there is an $s$ shell with only one electron,

$$
\begin{gathered}
\text{Li:}\quad [\text{He}]2s^1, \qquad \text{Na:}\quad [\text{Ne}]3s^1, \qquad \text{K:}\quad [\text{Ar}]4s^1,\\
\text{Rb:}\quad [\text{Kr}]5s^1, \qquad \text{Cs:}\quad [\text{Xe}]6s^1, \qquad \text{Fr:}\quad [\text{Rn}]7s^1,
\end{gathered}
$$

while the halogens are all missing 1 electron in a $p$ shell,

$$
\begin{gathered}
\text{F:}\quad [\text{He}]2s^2\,2p^5, \qquad \text{Cl:}\quad [\text{Ne}]3s^2\,3p^5, \qquad \text{Br:}\quad [\text{Ar}]3d^{10}\,4s^2\,4p^5,\\
\text{I:}\quad [\text{Kr}]4d^{10}\,5s^2\,5p^5, \qquad \text{At:}\quad [\text{Xe}]4f^{14}\,5d^{10}\,6s^2\,6p^5.
\end{gathered}
$$

A good periodic table of the elements will give the electron configuration for the atomic ground state of each element.

The energy needed to remove a single electron from the ground state of a neutral atom is called the **first-ionization energy**. Experimental results for the first-ionization energy as a function of the atomic number $Z$ are shown in Figure 16.4.2, including as the first two data points 13.6 eV for H and 24.6 eV for He. (Note that the result for He can be obtained from numbers given in Figure 16.3.1.) Because of the shielding of the nuclear charge by the innermost electrons, the results do not depend very strongly on $Z$; the outermost electrons are attracted to the nucleus as if they "see" a reduced net charge of order $e$ rather than the full nuclear charge $Ze$. The first-ionization energies are largest for the inert gases, which each have a complete

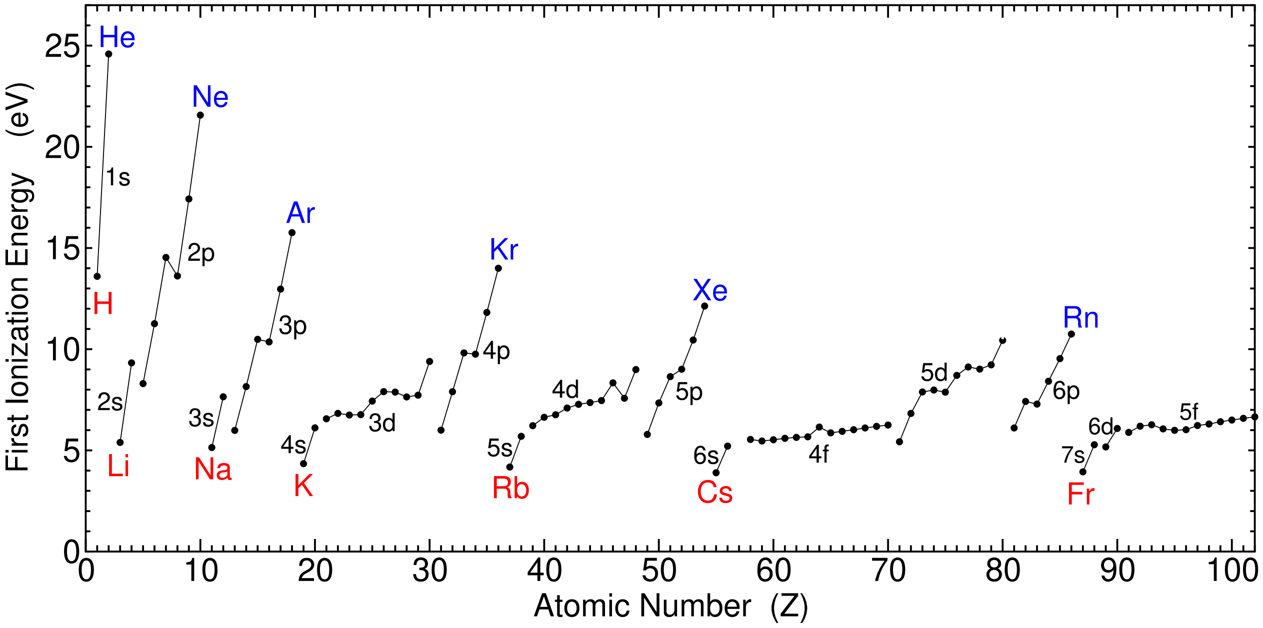

Figure 16.4.2: The energy (in eV) needed to remove a single electron from the ground state of a neutral atom, as a function of the atomic number $Z$. Atoms with the same highest-energy shells ($1s$, $2s$, $2p$, $3s$, …) are connected by lines to guide the eye, as labeled. The inert gas elements (He, Ne, Ar, Kr, Xe, Rn) each have completely filled $s$- and $p$-shells and thus a large first-ionization energy, while the alkali metal atoms (Li, Na, K, Rb, Cs, Fr) each have a lone $s$-shell electron and a small first ionization energy.

$p$ shell, and are smallest for alkali metals, which each have a single, relatively weakly bound, $s$-shell electron available for ionization.

A more complete specification of the structure of the atoms will include information about the angular momentum quantum numbers. Suppose that we can treat the spin-orbit contribution to the Hamiltonian $H_{\text{SO}}$ as small compared to $H_0 + W$ in eqs. (16.4.1)–(16.4.4). In practice, this is especially likely to be a good approximation if $Z$ is not too large. Now, all three of the summed orbital, summed spin, and total angular momentum operators

$$\vec{L} = \sum_{i=1}^{N_e} \vec{L}_i, \qquad \vec{S} = \sum_{i=1}^{N_e} \vec{S}_i, \qquad \vec{J} = \vec{L} + \vec{S} \tag{16.4.11}$$

commute with $H_0$ (as well as $W$). One can therefore choose unperturbed orthobasis states that are not only eigenstates of $H_0$, but also eigenstates of the operators $L^2$, $S^2$, $L_z$ and $S_z$ with eigenvalues $\hbar^2 L(L+1)$, $\hbar^2 S(S+1)$, $\hbar m_L$, and $\hbar m_S$ respectively. These states also carry other quantum number labels (including the unperturbed $H_0$ energy eigenvalue) which we will symbolically denote as $N$, so that the states are labeled

$$|N L S m_L m_S\rangle\,. \tag{16.4.12}$$

The $H_0$ energy eigenvalues have energy degeneracy $g_N = (2S+1)(2L+1)$, due to the magnetic quantum numbers $m_L, m_S$. Alternatively, one can choose as an orthobasis the linear combinations of the same unperturbed states that are eigenstates of $H_0$, $S^2$, $L^2$, and $J^2$ [with eigenvalue $\hbar^2 J(J+1)$], and $J_z$ (with eigenvalue $\hbar m_J$),

$$|NLSJm_J\rangle = \sum_{m_L, m_S} C^{LSJ}_{m_L m_S m_J} \, |NLSm_L m_S\rangle \,, \tag{16.4.13}$$

where $C^{LSJ}_{m_L m_S m_J}$ are the Clebsch–Gordan coefficients. The total number of such states for fixed $N, L, S$ is again $(2S+1)(2L+1)$. The approximation in which the states constructed in this way from eigenstates of $H_0$ (or perhaps $H_0 + W$) are close to being eigenstates of the full Hamiltonian is called the **LS coupling** (or **Russell–Saunders coupling**, after Henry N. Russell and Frederick Saunders) approximation. Note that this approximation is only good to the extent that $H_{\text{SO}}$ can be treated as small, since it does not commute with $S^2$ or $L^2$ if there is more than one electron.

We now investigate the spin-orbit fine-structure energy corrections, using degenerate perturbation theory. To do so, we must find the matrix elements of $H_{\text{SO}}$ on the degenerate subspaces of $H_0$, and diagonalize it. We start by working with the product orthobasis kets, which decompose into linear combinations of tensor products of orbital and spin angular momentum eigenkets,

$$|NLSm_L m_S\rangle = \sum_{\alpha,\beta} c_{\alpha,\beta} \, |\alpha N L m_L\rangle \otimes |\beta S m_S\rangle \,. \tag{16.4.14}$$

Here $\alpha$ is a degeneracy label whose presence reflects the fact that there can be more than one state with eigenvalues $L, m_L$, and similarly $\beta$ labels the different spin states with eigenvalues $S, m_S$. The whole state $|NLSm_L m_S\rangle$ is a Slater determinant incorporating the antisymmetry under interchange of any two electrons. This puts constraints, which we will not explicitly identify or make use of, on the coefficients $c_{\alpha,\beta}$.

The matrix elements of $H_{\text{SO}}$ from eq. (16.4.4) on the degenerate subspace of states are

$$\begin{aligned}\langle NLSm'_L m'_S|H_{\text{SO}}|NLSm_L m_S\rangle \;\; &= \\ &\sum_{\alpha,\beta}\sum_{\alpha',\beta'} c_{\alpha,\beta} c^*_{\alpha',\beta'} \sum_{i=1}^{N_e} \langle \alpha' N L m'_L|\vec{\Omega}_i|\alpha N L m_L\rangle \cdot \langle \beta' S m'_S|\vec{S}_i|\beta S m_S\rangle \,.\end{aligned} \tag{16.4.15}$$

Electrons in complete shells can be omitted, because pairs of opposite spin expectation values with the same orbital quantum numbers will always cancel in the sum over $i$. Now we will invoke the Landé projection formula, eq. (12.2.10), twice. First, note that each $\vec{\Omega}_i$ is a vector operator with respect to the total orbital angular momentum operator $\vec{L}$, without involving spin. Therefore, for $L \neq 0$, we have

$$\langle \alpha' N L m'_L|\vec{\Omega}_i|\alpha N L m_L\rangle \;\; = \;\; \langle L m'_L|\vec{L}|L m_L\rangle \, \frac{\langle \alpha' N L \| \vec{L}\cdot\vec{\Omega}_i \| \alpha N L\rangle}{\hbar^2 L(L+1)}. \tag{16.4.16}$$

As noted in Theorem 12.2.2, for the case $L = 0$, the matrix element simply vanishes, and there is no spin-orbit splitting. Similarly, each of the individual electron spin operators $\vec{S}_i$ is a vector with respect to rotations generated by the total spin angular momentum operator $\vec{S}$, without involving orbital angular momentum, so we can write, for $S \neq 0$,

$$\langle \beta' S m'_S | \vec{S}_i | \beta S m_S \rangle \;=\; \langle S m'_S | \vec{S} | S m_S \rangle \, \frac{\langle \beta' S \| \vec{S} \cdot \vec{S}_i \| \beta S \rangle}{\hbar^2 S(S+1)}. \qquad (16.4.17)$$

while for $S = 0$ the matrix element again vanishes. Therefore, putting things together,

$$\langle N L S m'_L m'_S | H_{\text{SO}} | N L S m_L m_S \rangle \;=\; \zeta(N, L, S) \, \langle L S m'_L m'_S | \vec{L} \cdot \vec{S} | L S m_L m_S \rangle \,, \qquad (16.4.18)$$

where the factor

$$\zeta(N, L, S) \;=\; \sum_{\alpha,\beta} \sum_{\alpha',\beta'} c_{\alpha,\beta} c^*_{\alpha',\beta'} \sum_{i=1}^{N_e} \frac{\langle \alpha' N L \| \vec{L} \cdot \vec{\Omega}_i \| \alpha N L \rangle}{\hbar^2 L(L+1)} \frac{\langle \beta' S \| \vec{S} \cdot \vec{S}_i \| \beta S \rangle}{\hbar^2 S(S+1)} \qquad (16.4.19)$$

may be quite difficult to evaluate exactly, but has the important property that it does not depend on the magnetic quantum numbers $m_L, m_S, m'_L, m'_S$ at all, because the reduced matrix elements $\langle \alpha' N L \| \vec{L} \cdot \vec{\Omega}_i \| \alpha N L \rangle$ and $\langle \beta' S \| \vec{S} \cdot \vec{S}_i \| \beta S \rangle$ do not. For a given atom, $\zeta(N, L, S)$ can, in principle, be evaluated, perhaps in approximations such as replacing the electric potential $\Phi_i(\vec{R}_i)$ in eq. (16.4.5) by a spherically symmetric approximation. This is beyond our scope here; we will merely consider it to be an empirical function of $N, L, S$.

Equation (16.4.18) shows that the matrix elements of $H_{\text{SO}}$ are proportional to those of $\vec{L} \cdot \vec{S} = \hbar^2 (J^2 - L^2 - S^2)/2$, where $J^2$, $L^2$ and $S^2$ are understood to be operators. To properly conduct degenerate perturbation theory, it is therefore clear that we should switch to the orthobasis $|N L S J m_J\rangle$, in which this operator is diagonal. We therefore arrive at a formula for the spin-orbit correction to the energies, at first order in degenerate perturbation theory,

$$\Delta E(N, L, S, J) \;=\; \zeta(N, L, S) \hbar^2 \big[ J(J+1) - L(L+1) - S(S+1) \big]/2, \qquad (16.4.20)$$

where $J$, $L$, and $S$ are now numbers. This vanishes in the special cases $L = 0$ (which then implies $J = S$) and $S = 0$ (which then implies $J = L$), consistent with observations already made in the previous paragraph. Equation (16.4.20) shows that the degeneracy $(2S+1)(2L+1)$ is partially broken by $H_{\text{SO}}$, but states with a given $J$ are not split, with remaining degeneracies $g_J = 2J + 1$, since $\Delta E(N, L, S, J)$ does not depend on $m_J$.

We can now use eq. (16.4.20) to obtain a simple formula for the energy differences between successive eigenvalues $J - 1$ and $J$, for fixed $N$, $L$, and $S$,

$$\Delta E(N, L, S, J) - \Delta E(N, L, S, J-1) \;=\; \hbar^2 \zeta(N, L, S) J. \qquad (16.4.21)$$

This result, that the spin-orbit fine-structure energy splitting between states $J$ and $J-1$ with fixed $N, L, S$ is proportional to $J$, is called the **Landé interval rule**, as observed in spectral lines from transitions between atomic states. It can be used as a check of the validity of the $LS$ coupling scheme approximation.

The preceding discussion shows how, within the $LS$ coupling approximation, energy eigenstates can be built by first adding all of the spins, then adding all of the orbital angular momenta, and then adding the two to get the total angular momentum. The angular momentum quantum numbers of these states are then traditionally specified with the **term symbol**

$$^{2S+1}L_J \quad = \quad \text{Russell–Saunders } (LS) \text{ spectroscopic notation,} \tag{16.4.22}$$

where the eigenvalues of operators $J^2$, $S^2$, and $L^2$ are numbers $\hbar^2 J(J+1)$, $\hbar^2 S(S+1)$, and $\hbar^2 L(L+1)$, respectively, and then the value of $L$ is replaced by its spectroscopic code $S, P, D$, $F, G, H, \ldots$ for $L = 0, 1, 2, 3, 4, 5, \ldots$. Thus, for example, the hydrogen and helium atom ground states are given as $^2S_{1/2}$ and $^1S_0$, respectively. Many editions of the periodic table of elements list the Russell–Saunders notation for the atomic ground states along with the electron configuration.

The list of possible term symbols $^{2S+1}L_J$ of an atom can be obtained from looking only at the **valence electrons**, the ones in the incompletely filled shells. This is because each electron in a filled shell with a given $m_s$, $m_l$, and $m_j$ always has a partner with the opposite values of those quantum numbers, giving no net contribution to $S$, $L$, and $J$. The parity of the state is useful for selection rules for transitions. From eq. (8.7.7), it is $\pi = (-1)^{\sum_i l_i}$, where $l_i$ are the individual orbital angular momenta, specified by the electronic configuration. [Note that for orbitals the parity is defined with respect to the fixed nucleus as the origin, not the center of mass of the valence electrons, so even for two electrons $\pi$ is not necessarily equal to $(-1)^L$.]

For an atom with a partly filled shell, there may be several candidates for the ground state term symbol. Three empirical guidelines, known as **Hund's Rules** after Friedrich Hund, can be used to predict which the ground state will be, given the electron configuration. They are:

Hund's Rule 1: The $L, S, J$ multiplet with the lowest energy will have the largest $S$ from among the candidates.
(Qualitative justification: for larger $S$, the spin wavefunction is more symmetric, which means the spatial wavefunction is more antisymmetric, which means the electrons are less often near each other, which gives less repulsive forces between them, and so lower energy. We have already seen this effect in action; it is why each ortho-helium $S = 1$ state is lower in energy than its para-helium $S = 0$ counterpart.)

Hund's Rule 2: If there are two or more candidates selected by the first rule that have the same largest $S$, then the lowest energy will have the largest $L$ among them.

(Qualitative justification: larger $L$ means that the orbital wavefunctions can be arranged to have less overlap, again leading to smaller repulsive forces and thus lower energy.)

Hund's Rule 3: If there are two or more candidates selected by the first two rules with the same $S$ and $L$, then the spin-orbit coupling splits apart the different levels of $J = |L-S|, \ldots, L+S$. If an incomplete shell is not more than half-filled, then the ground state has $J = |L-S|$, otherwise the ground state has $J = L+S$.

(Qualitative justification: the form of eq. (16.4.20) immediately tells us that for fixed $L, S$, the minimum energy will occur for one of the two extremes for $J$, either $|L-S|$ or $L+S$. Hund's Rule 3 is equivalent to the empirical fact that $\zeta(N, L, S)$ in eq. (16.4.20) turns out to be positive if the shell is not more than half-filled, and negative otherwise.)

We now illustrate the prediction of term symbols, using Hund's rules where necessary, with examples from the beginning of the periodic table. For the first four elements, Hund's rules are not needed.

Hydrogen (H, $Z = 1$) has electron configuration $1s^1$. Because there is one electron with $S = 1/2$ and $L = 0$, the only possibility is $J = 1/2$, and the term symbol is ${}^2S_{1/2}$.

Helium (He, $Z = 2$) has electron configuration $1s^2$. As always when there are no partially filled shells, the ground state is ${}^1S_0$. The same result is obtained for all inert (noble) gases.

Lithium (Li, $Z = 3$) has electron configuration $1s^2\,2s^1$. Just as for hydrogen, there is one electron that is not in a filled shell, with $S = 1/2$ and $L = 0$, so the term symbol is ${}^2S_{1/2}$. The same applies for all alkali metals, the elements in the same periodic table column as Li, which have a single $s$ shell electron.

Beryllium (Be, $Z = 4$) has electron configuration $1s^2\,2s^2$, with only completely filled shells, so the ground state is ${}^1S_0$. The same applies for all elements in the Be periodic table column.

Boron (B, $Z = 5$) has electron configuration $1s^2\,2s^2\,2p^1$. Here, the $1s$ and $2s$ shells are irrelevant, since they are completely filled. In the incomplete $2p$ shell, we have one electron with $L = 1$ and $S = 1/2$. Since this is the only choice, we do not need Hund's rules 1 or 2. The possible candidates have $J = 1/2$ and $J = 3/2$, with term symbols ${}^2P_{1/2}$ and ${}^2P_{3/2}$. Because the last incomplete shell is less than half filled (1 out of 6), Hund's rule 3 tells us that the ground state has $J = |L - S| = 1/2$ and is ${}^2P_{1/2}$. The same result is obtained for all elements in the B column of the periodic table, which have a single $p$ electron.

Carbon (C, $Z = 6$) has electron configuration $1s^2\,2s^2\,2p^2$. We start by computing the total number of allowed states associated with the incomplete $2p$ shell. The 2 electrons have 6 orbital states from which to choose, so there are $(6{\cdot}5)/(2{\cdot}1) = 15$ antisymmetric states. Now, let us see how these 15 states are divided into ${}^{2S+1}L_J$ multiplets. Combining the two spin-1/2's gives $S = 0$ (the antisymmetric combination) and $S = 1$ (the symmetric combination). Combining the two orbital angular momenta, each with $l = 1$, gives $L = 0$ (symmetric), and $L = 1$ (antisymmetric), and $L = 2$ (symmetric). Therefore, requiring that the full state be antisymmetric under exchange of the two electrons, we have the following allowed combinations. First, there is $L = 2, S = 0$, which implies $J = 2$. This possibility ${}^1D_2$ consists of $2J + 1 = 5$ states. Second, there is $L = 1, S = 1$, which allows $J = 0, 1, 2$, so the terms are ${}^3P_0$, ${}^3P_1$, and ${}^3P_2$, with 1, 3, and 5 states respectively. Third, there is $L = 0, S = 0$, which of course allows only one state, ${}^1S_0$ with $J = 0$. As a check, the total of these possibilities is indeed 15. Now we are ready to use Hund's rules. Hund's rule 1 selects the largest possible value of $S$, which is $S = 1$. Hund's rule 2 tells us nothing, because the remaining competitors all have $L = 1$. Finally, for Hund's rule 3, we note that the $2p$ shell is not more than half-filled (2 out of 6), so that $J = |L - S| = 0$ in the ground state. We therefore have $S = 1$, $L = 1$, $J = 0$, and the ground state of carbon is ${}^3P_0$. The same logic and result applies to all elements in the C column of the periodic table, which have two $p$ electrons.

Nitrogen (N, $Z = 7$) has electron configuration $1s^2\,2s^2\,2p^3$. The total number of possible electron states associated with the three electrons in the $2p$ shell is $(6 \cdot 5 \cdot 4)/(3 \cdot 2 \cdot 1) = 20$ antisymmetric states. To see which $L, S, J$ combinations these correspond to, first consider the combination of the orbital angular momenta, which gives $1 \otimes 1 \otimes 1 = 0 \oplus 1 \oplus 1 \oplus 1 \oplus 2 \oplus 2 \oplus 3$ for the possible values of $L$. If $L = 3$, one can check that the spatial wavefunction constructed by addition of angular momentum would be totally symmetric under exchange of the electrons, but there is no totally antisymmetric spin state for $\geq 3$ electrons; we conclude from this that $L$ cannot be 3. If $L = 0$, then the spatial wavefunction is instead totally antisymmetric, from which we conclude that the spin state must be totally symmetric, so $S = 3/2$. Conversely, if $S = 3/2$, then only the totally antisymmetric $L = 0$ combination is allowed. We conclude that the only $(L, S)$ combinations that are allowed are $L = 0, S = 3/2$ with $J = 3/2$ (term ${}^4S_{3/2}$, 4 states), and $L = 1, S = 1/2$ with $J = 1/2$ and $3/2$ (terms ${}^2P_{1/2}$ and ${}^2P_{3/2}$, 6 states), and $L = 2, S = 1/2$ with $J = 3/2$ and $J = 5/2$ (terms ${}^2D_{3/2}$ and ${}^2D_{5/2}$, 10 states). As a check, these indeed add up to 20 states. Hund's rule 1 tells us that the ground state should have maximal $S$, which selects the $L = 0$, $S = 3/2$, $J = 3/2$ combination. This is the only one that realizes that maximal value of $S$, so the other two Hund's rules are not needed, and the ground state term is ${}^4S_{3/2}$. The same logic and result applies to all elements in the N column of the periodic table, which have three $p$ electrons.

Oxygen (O, $Z = 8$) has electron configuration $1s^2\,2s^2\,2p^4$. As in the previous examples, we start by doing the counting of the total number of physical antisymmetric electron states for the incomplete shell. Since there are 4 electrons and 6 available states, that number is $(6 \cdot 5 \cdot 4 \cdot 3)/(4 \cdot 3 \cdot 2 \cdot 1) = 15$. However, as a useful trick it is better to consider the 2 "holes", by which we mean the electrons omitted from the completely filled $2p$ shell. Treating the holes as equivalent spin-1/2 particles, we again obtain $(6 \cdot 5)/(2 \cdot 1) = 15$ states. A complete $2p$ shell would have $L = S = 0$, so we treat the 2 holes as carrying the same $S$ and $L$ as the 4 electrons. So, the allowed candidate multiplets are the same as we found for carbon with $2p^2$, namely $(L, S, J) = (2, 0, 2)$, $(1, 1, 0)$, $(1, 1, 1)$, $(1, 1, 2)$, and $(0, 0, 0)$, with term symbols ${}^1D_2$, ${}^3P_0$, ${}^3P_1$, ${}^3P_2$, and ${}^1S_0$. Hund's rule 1 says the ground state has $S = 1$, and Hund's rule 2 does not differentiate between the remaining candidates which all have $L = 1$. To apply Hund's rule 3, we note that the $2p$ shell is now more than half-filled with electrons (4 out of 6), so $J = L+S = 2$ for the ground state. (Note that this is the difference between carbon and oxygen; although the enumeration of the candidate terms is easier in terms of the holes, Hund's rule 3 for the selection of the ground state still applies to the actual electrons, not the holes.) So, the ground state term for oxygen is ${}^3P_2$. The same logic and result applies to all elements in the O column of the periodic table, with four $p$ electrons.

Fluorine (F, $Z = 9$) has electron configuration $1s^2\, 2s^2\, 2p^5$. Using the trick introduced in the oxygen example, we get the allowed $(L, S, J)$ multiplets by considering the $2p$ shell to consist of just one hole, rather than 5 electrons. This means that, just as for boron, the only combinations are $L = 1, S = 1/2$ with $J = 1/2$ or $3/2$. Since these have the same $L$ and $S$, Hund's rules 1 and 2 provide no information. To apply Hund's rule 3, we note that the $2p$ shell is more than half full (5 out of 6), and so the ground state has $J = L + S = 3/2$, with term symbol ${}^2P_{3/2}$. The same applies to all halogens, the elements in the F column of the periodic table, with five $p$ electrons.

Hund's rules correctly predict the ground state of all neutral atoms for which the ground state term symbol is unambigously measured. They also often give the correct result for charged ions, and for neutral atoms with an electron configuration different from the ground state. For example, the singly ionized carbon ion $\mathrm{C}^+$ has electron configuration $1s^2\ 2s^2\, 2p$, which is the same as boron, and has the same term symbol ${}^2P_{1/2}$, as correctly predicted by Hund's rules. Also, the neutral carbon atom has excited states with electron configuration $1s^2\ 2s\, 2p^3$. Among those excited states, Hund's rules correctly predict that the lowest energy term is ${}^5S_2$.

An important experimental tool for exposing the properties of atomic states is to observe the behavior of energy levels and transition spectra as a function of an applied magnetic field, called the Zeeman effect. In section 15.3, we studied the Zeeman effect for atomic hydrogen, and much of that discussion can be adapted to the multi-electron case. Adding the contributions to the Hamiltonian linear in a constant, uniform magnetic field $\vec{B} = B\hat{z}$, one obtains a result of the same form as eq. (15.3.6),

$$H_B \;=\; \mu_B B(L_z + 2S_z)/\hbar, \tag{16.4.23}$$

but now with $L_z$ and $S_z$ respectively equal to the total orbital and spin angular momentum operators, obtained as the sums of the operators for the individual electrons. Within the $LS$ coupling scheme approximation, the analysis for the weak-field (small-$B$) Zeeman effect in multi-electron atoms just mimics what we already did for hydrogen from eqs. (15.3.8)–(15.3.13), so one finds that atomic states characterized by $N, L, S, J$ obtain energy splittings proportional to the total angular momentum quantum number $m_J$,

$$\Delta E_B \;=\; g\,\mu_B B\, m_J, \qquad (m_J = -J,\, -J+1, \ldots,\, J-1,\, J), \tag{16.4.24}$$

where the Landé $g$-factor is

$$g \;=\; \frac{3}{2} + \frac{S(S+1) - L(L+1)}{2J(J+1)}. \tag{16.4.25}$$

(Of course, unlike the atomic hydrogen case, in general $S$ need not be equal to $1/2$, and $J$ need not be equal to $L \pm 1/2$.) For weak external magnetic fields, the magnetic moment of the atom is therefore related to its total angular momentum by $-g\mu_B \vec{J}/\hbar$, and the splitting exposes the degeneracies and angular momentum quantum numbers for the multi-electron atomic states.

The strong magnetic field (Paschen–Back) limit for multi-electron atoms is also quite similar to the example of the hydrogen atom, as given in eqs. (15.3.20)–(15.3.24). This includes the use of $|NLSm_Lm_S\rangle$, in which $H_B$ is diagonal, as the degenerate unperturbed states to which $H_{\text{SO}}$ is applied as a perturbation. For multi-electron atoms, the counterpart of eq. (15.3.22) is

$$\langle NLSm_Lm_S|H_{\text{SO}}|NLSm_Lm_S\rangle \;=\; \zeta(N,L,S)\,\langle NLSm_Lm_S|\vec{L}\cdot\vec{S}|NLSm_Lm_S\rangle\,, \tag{16.4.26}$$

where $\zeta(N,L,S)$ is the quantity appearing in eqs. (16.4.19)–(16.4.21) and $\vec{L}\cdot\vec{S} = \frac{1}{2}(L_+S_- + L_-S_+) + L_zS_z$ evaluates to $\hbar^2 m_L m_S$. Therefore, we have in the strong-field limit

$$\Delta E_{B,\,\text{fine}} \;=\; \mu_B B(m_L + 2m_S) + \hbar^2\zeta(N,L,S)m_Lm_S, \tag{16.4.27}$$

where we have kept only the spin-orbit part of the fine-structure effect, which contains the dependence on $m_L$ and $m_S$. For the intermediate case that the fine-structure and external magnetic fields are comparable, things are more complicated and one should find the energy shift by treating $H_B + H_{\text{SO}}$ together as a single perturbation, to be diagonalized using either the basis of states $|NLSJm_J\rangle$ or the basis $|NLSm_Lm_S\rangle$.

It is important to keep in mind that the $LS$ coupling scheme is just an approximation. The assumption that $H_{\text{SO}}$ can be treated as a small effect becomes less appropriate for atoms with larger $Z$, and so the $LS$ coupling scheme approximation is not as good for heavier atoms. To understand this, consider an approximation in which the potential in the spin-orbit interaction is spherically symmetric, so that we have a sum of terms like eq. (15.1.4) for each electron,

$$H_{\rm SO} \;=\; \sum_{i=1}^{N_e} \xi(R_i)\vec{L}_i \cdot \vec{S}_i, \qquad\qquad \xi(r) = -\frac{e}{2m_e^2 c^2}\frac{1}{r}\frac{\partial \Phi}{\partial r}. \tag{16.4.28}$$

Now, as a qualitative approximation, suppose that the electrons are in hydrogen-like orbitals arising from a potential $\Phi(r) = -Z_{\rm eff} e^2/r$, where $Z_{\rm eff} < Z$ roughly takes into account the effects of partial screening of the nuclear charge by the other electrons. Then $H_{\rm SO}$ scales like $Z_{\rm eff}/R^3$, and the characteristic size of orbitals scales like $1/Z_{\rm eff}$, as can be seen from eq. (10.1.65), for example. This implies that the effects of $H_{\rm SO}$ scale like $Z_{\rm eff}^4$.

In the limit that the effects of $H_{\rm SO}$ are too large to be treated as a perturbation, one can instead use the **jj coupling approximation**. Since eq.(16.4.28) involves the individual $\vec{L}_i \cdot \vec{S}_i$, it is useful to first perform addition of angular momentum on each electron, $\vec{J}_i = \vec{L}_i + \vec{S}_i$, and then construct $\vec{J} = \sum_{i=1}^{N_e} \vec{J}_i$. Then $H_0 + H_{\rm SO}$ is diagonal in an orthobasis labeled by the quantum numbers $l_i$, $s_i$, and $j_i$ associated with the operators $L_i^2$, $S_i^2$, and $J_i^2$. In particular, the dominant spin-orbit coupling term is diagonal in that basis and can be evaluated according to

$$H_{\rm SO} \;=\; \frac{\hbar^2}{2}\sum_{i=1}^{N_e} \xi(R_i)[j_i(j_i+1) - l_i(l_i+1) - s_i(s_i+1)], \tag{16.4.29}$$

where $s_i = 1/2$, and $l_i$ is specified by the electron configuration. The $W$ part of the Hamiltonian in eq. (16.4.3) does not commute with the operators $L_i^2$ and $J_i^2$, so it is treated as a perturbation, which then splits the states into different values of the total angular momentum quantum number $J$, with a remaining degeneracy $2J+1$ (neglecting small hyperfine effects associated with electronic interactions with the nucleus). In the $jj$ coupling scheme, the term symbols specifying the angular momentum can thus be written as

$$(j_1, j_2, \ldots, j_n)_J, \tag{16.4.30}$$

where $n$ is the number of electrons in the unfilled shell, and the remaining degeneracy is $2J+1$.

The standard example of the $jj$ coupling scheme is provided by the lead atom (Pb, $Z = 82$). The electron configuration is $[\text{Xe}]4f^{14}\,5d^{10}\,6s^2\,6p^2$, so what matters is the two $l = 1$ electrons in the unfilled $6p$ shell. Just as for carbon, the number of antisymmetric states that we can form from two $p$ electrons in the same shell is $6 \cdot 5/2 = 15$. Each electron has total angular momentum $j = 1/2$ or $3/2$. Now we can use

$$\frac{1}{2} \otimes \frac{1}{2} \;=\; 0_A \oplus 1_S \tag{16.4.31}$$

$$\frac{3}{2} \otimes \frac{3}{2} \;=\; 0_A \oplus 1_S \oplus 2_A \oplus 3_S, \tag{16.4.32}$$

$$\frac{1}{2} \otimes \frac{3}{2} \;=\; 1 \oplus 2 \tag{16.4.33}$$

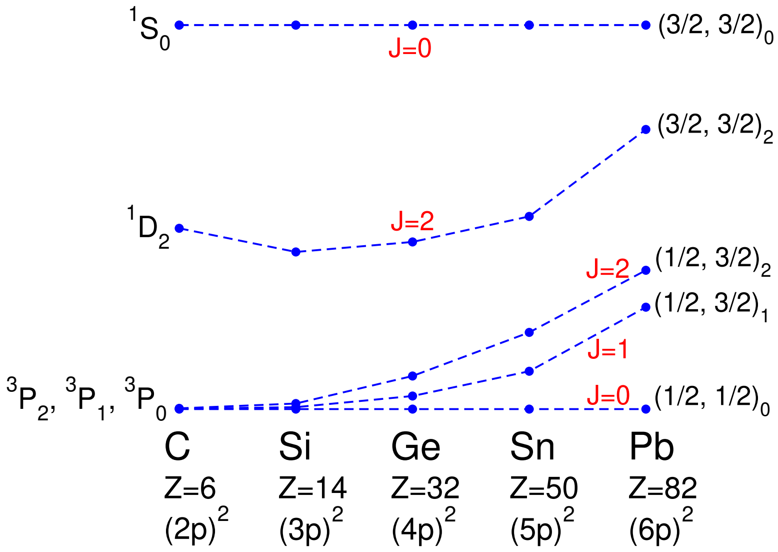

Figure 16.4.3: Experimental results for energy levels of neutral atoms with two $p$ electrons in an incomplete shell. The vertical energy scales are adjusted independently for each element (C, Si, Ge, Sn, Pb) so that the highest and lowest $J = 0$ energies are at the same level. On the left side, the $LS$-coupling terms $^{2S+1}L_J$ are given, and on the right the $jj$-coupling terms $(j_1, j_2)_J$ are given. The dashed lines are to guide the eye, connecting atomic states with the same $J$. In the limit of the perfect $LS$-coupling approximation, the states $^3P_{0,1,2}$ would be degenerate, while in the limit of perfect $jj$-coupling the states $(3/2, 3/2)_{0,2}$ would be degenerate and the states $(1/2, 3/2)_{1,2}$ would be degenerate. The $jj$-coupling approximation is seen to be approached (quite imperfectly) for the high-$Z$ case of Pb.

to find that the possible $jj$ coupling terms for two $p$ electrons are

$$(j_1, j_2)_J \;=\; (\tfrac{1}{2}, \tfrac{1}{2})_0, \quad (\tfrac{1}{2}, \tfrac{3}{2})_1, \quad (\tfrac{1}{2}, \tfrac{3}{2})_2, \quad (\tfrac{3}{2}, \tfrac{3}{2})_2, \quad (\tfrac{3}{2}, \tfrac{3}{2})_0, \tag{16.4.34}$$

with degeneracies 1, 3, 5, 5, 1, respectively (which indeed add up to 15 as a check). Experimentally, it turns out that the order given in eq. (16.4.34) is in increasing energy, so the ground state has $J = 0$. By way of comparison, the Hund's rules prediction for the $LS$-coupling term for Pb is $^3P_0$ (by an argument identical to the one for carbon), which also has $J = 0$, so the two schemes agree in that respect. However, in some respects the $jj$ coupling scheme provides a better numerical approximation to the energy levels and other properties of the lowest few eigenstates of Pb than the $LS$ coupling scheme does. Figure 16.4.3 shows the experimental energy levels for neutral atoms with two $p$ electrons in an incomplete shell, comparing the low-$Z$ case of C and Si, which nearly realize the perfect $LS$-coupling approximation, to the high-$Z$ case of Pb where the $jj$-coupling approximation does better, and the intermediate cases of Ge and Sn. In general, for atoms with high and medium $Z$, neither the $LS$ scheme nor the $jj$ scheme is very accurate, and more complicated descriptions are appropriate.

# 17 Heisenberg and interaction representations

## 17.1 The Heisenberg picture and equations of motion

Everything we have done so far uses a formulation that is sometimes called the **Schrödinger picture** of quantum mechanics. This just means that the state of the system evolves according to Postulate 6 of section 3.1,

$$i\hbar \frac{d}{dt}|\psi(t)\rangle \;=\; H|\psi(t)\rangle. \tag{17.1.1}$$

However, there are other, equivalent, formulations of quantum mechanics, called "pictures", which are related to the Schrödinger picture by doing a time-dependent unitary transformation on the state of the system and a corresponding transformation to all operators. Done consistently, this cannot affect the predictions for physically measurable quantities, but the intermediate steps for solving problems will differ. The main reason for considering other pictures is to make some calculations easier to do and/or understand.

Recall from the discussion in section 3.4 that the solution to eq. (17.1.1), starting from an initial state $|\psi(t_0)\rangle$, is

$$|\psi(t)\rangle \;=\; U(t,t_0)|\psi(t_0)\rangle, \tag{17.1.2}$$

where $U(t,t_0)$ is a unitary operator satisfying the differential equation

$$i\hbar \frac{d}{dt}U(t,t_0) \;=\; HU(t,t_0), \tag{17.1.3}$$

with $U(t,t_0)^{-1} = U(t,t_0)^\dagger = U(t_0,t)$. If $H$ has no explicit time dependence, then the solution to eq. (17.1.3) is relatively easy, just $U(t,t_0) = e^{-i(t-t_0)H/\hbar}$. In that case, there is no particular advantage to the Heisenberg or interaction pictures described in the following.

In the **Heisenberg picture**, all time dependence is transferred from the state of the system to the operators. To accomplish this, define

$$|\psi^{(H)}\rangle \;=\; U(t,t_0)^\dagger|\psi(t)\rangle, \tag{17.1.4}$$

$$A^{(H)}(t) \;=\; U(t,t_0)^\dagger A(t) U(t,t_0), \tag{17.1.5}$$

where $|\psi^{(H)}\rangle$ is the state ket in the Heisenberg picture, and for each Schrödinger picture operator $A(t)$, the corresponding Heisenberg picture operator is denoted $A^{(H)}(t)$. Combining eqs. (17.1.2) and (17.1.4) gives

$$|\psi^{(H)}\rangle = |\psi(t_0)\rangle. \tag{17.1.6}$$

Thus the Heisenberg picture state ket $|\psi^{(H)}\rangle$ is independent of the time $t$; it only depends on the initial condition at $t_0$. Note that there are really infinitely many Heisenberg pictures, one for each choice of the reference initial time $t_0$. We include a superscript $^{(H)}$ to denote the state of the system or an operator as being in the Heisenberg picture. If an object is not written with such a superscript, it can be assumed to be in the Schrödinger picture. For convenience and simplicity, we will often write $U = U(t, t_0)$ in the rest of this section.

A possible point of confusion is that it is only the state of the system that we are defining differently in the Heisenberg and Schrödinger pictures via the unitary transformation in eq. (17.1.4). Both the Schrödinger picture state ket $|\psi(t)\rangle$ and the Heisenberg picture state ket $|\psi^{(H)}\rangle$ can be expressed as (different!) linear combinations of a set of *fixed* orthobasis kets, which are defined independently of the choice of picture. A convenient way to choose the fixed orthobasis kets is as time-independent eigenstates of some appropriate time-independent CSCO in the Schrödinger picture. Once so chosen, the orthobasis kets do not depend on $t$, by definition.

Inner products and matrix elements are the same in both pictures, since they are related by a unitary transformation. To see this, consider a matrix element of an arbitrary observable $A$ between two possible states of the system $|\psi\rangle$ and $|\chi\rangle$ that both obey Schrödinger's equation in the Schrödinger picture. We have

$$\langle\chi(t)|A(t)|\psi(t)\rangle \;=\; \langle\chi(t)|UU^\dagger A(t) UU^\dagger|\psi(t)\rangle \;=\; \langle\chi^{(H)}|A^{(H)}(t)|\psi^{(H)}\rangle. \tag{17.1.7}$$

This shows the physical equivalence of the two pictures.

Even though many important observable operators (for example, $A =$ position, momentum, or spin) are time-independent in the Schrödinger picture, the corresponding Heisenberg picture operators $A^{(H)}(t)$ will generally depend on $t$, and often in a complicated way, depending on the Hamiltonian. This is the price to be paid in the Heisenberg picture for the convenience that the state of the system has no $t$ dependence.

However, there is an important special case in which $A^{(H)}(t)$ does not depend on $t$. Suppose that in the Schrödinger picture, $H$ and $A$ do not depend on $t$, and $[A, H] = 0$. Then $A$ commutes with $U(t, t_0)$ as well, since the latter is built out of $H$. In that case,

$$A^{(H)} \;=\; U^\dagger A U \;=\; U^\dagger U A \;=\; A, \tag{17.1.8}$$

and we say that $A$ is a **constant of motion**, following classical physics terminology. Note that this corresponds to a conserved quantity, as defined in section 5.1 in the Schrödinger picture. In particular, if $H$ does not depend explicitly on $t$, then $H^{(H)} = H$, and the Schrödinger and Heisenberg picture Hamiltonians are equal.

More generally, we can ask how $A^{(H)}(t)$ evolves in time. To find out, we calculate, using the product rule for derivatives,

$$\frac{dA^{(H)}}{dt} \;=\; \left[\frac{d}{dt}U(t,t_0)^\dagger\right] AU + U^\dagger A \left[\frac{d}{dt}U(t,t_0)\right] + U^\dagger \frac{\partial A}{\partial t} U. \tag{17.1.9}$$

Evaluating the derivatives in square brackets using the differential equation (17.1.3) gives

$$\frac{dA^{(H)}}{dt} = \frac{i}{\hbar}U^\dagger HAU - \frac{i}{\hbar}U^\dagger AHU + U^\dagger \frac{\partial A}{\partial t}U, \tag{17.1.10}$$

which can be rewritten as

$$\frac{dA^{(H)}}{dt} = \frac{i}{\hbar}[H^{(H)}, A^{(H)}] + \left(\frac{\partial A}{\partial t}\right)^{(H)}. \tag{17.1.11}$$

The meaning of the last term is that we take the derivative of $A$ with respect to its explicit dependence on $t$ in the Schrödinger picture, and then convert the result to the Heisenberg picture. Equation (17.1.11) is called the **Heisenberg equation of motion** for the operator $A^{(H)}$. Since the state of the system does not change in the Heisenberg picture in the absence of a measurement, this equation captures all physical effects of unitary time evolution. It is a direct analog in quantum mechanics of the classical equations of motion.

The form of **equal-time commutation relations** does not change if one switches pictures. If one has a commutation relation

$$[A, B] = C, \tag{17.1.12}$$

in the Schrödinger picture, then

$$[A^{(H)}(t), B^{(H)}(t)] = [U^\dagger AU, U^\dagger BU] = U^\dagger[A,B]U = U^\dagger CU = C^{(H)}(t). \tag{17.1.13}$$

So, for example,

$$[X^{(H)}(t), P^{(H)}(t)] = [X, P] = i\hbar, \tag{17.1.14}$$

and if $\vec{J}$ is an angular momentum operator, then

$$[J_a^{(H)}(t), J_b^{(H)}(t)] = i\hbar\epsilon_{abc}J_c^{(H)}(t), \qquad (a,b,c = x,y,z). \tag{17.1.15}$$

It is crucial that we specified equal-time commutation relations, because in general the commutator $[A^{(H)}(t), B^{(H)}(t')]$ does not have a simple relationship to $[A, B]$ unless $t' = t$. For example,

$$[X^{(H)}(t), P^{(H)}(t')] \neq i\hbar \tag{17.1.16}$$

in general, unless $t' = t$.

To illustrate how the Heisenberg picture works in the simple case that the Hamiltonian does not explicitly depend on time, suppose that in the Schrödinger picture we have the usual Hamiltonian for a spin-less particle of mass $m$ moving in a potential,

$$H = \frac{P^2}{2m} + V(X), \tag{17.1.17}$$

where $V(x) = \sum_n v_n x^n$, with $v_n$ a set of fixed real numbers. Then

$$\left(P^{(H)}\right)^2 \;=\; U^\dagger P U U^\dagger P U \;=\; U^\dagger P^2 U, \tag{17.1.18}$$

and similarly

$$\left(X^{(H)}\right)^n \;=\; U^\dagger X^n U. \tag{17.1.19}$$

Therefore, the Heisenberg picture Hamiltonian is

$$H^{(H)} \;=\; U^\dagger H U \;=\; \frac{(P^{(H)})^2}{2m} + V(X^{(H)}). \tag{17.1.20}$$

The Schrödinger picture position and momentum operators have no explicit time dependence,

$$\frac{\partial X}{\partial t} \;=\; 0, \qquad\qquad \frac{\partial P}{\partial t} \;=\; 0, \tag{17.1.21}$$

and so we obtain from eq. (17.1.11), by using eq. (17.1.14),

$$\frac{dX^{(H)}}{dt} \;=\; \frac{i}{\hbar}[H^{(H)},\, X^{(H)}] \;=\; \frac{P^{(H)}}{m}, \tag{17.1.22}$$

$$\frac{dP^{(H)}}{dt} \;=\; \frac{i}{\hbar}[H^{(H)}, P^{(H)}] \;=\; -V'(X^{(H)}). \tag{17.1.23}$$

The Heisenberg equations of motion (17.1.22) and (17.1.23) have the same form as the classical equations of motion, but with the classical position and momentum promoted to the corresponding Heisenberg-picture operators. Using them, one can check that

$$\frac{d}{dt}H^{(H)} \;=\; 0, \tag{17.1.24}$$

which verifies that $H^{(H)} = H$ is indeed a constant of motion.

## 17.2 The interaction picture and transition amplitudes

Suppose that we have a Hamiltonian (in the Schrödinger picture) of the form

$$H(t) \;=\; H_0 + W(t), \tag{17.2.1}$$

where $W(t)$ may depend explicitly on time, but $H_0$ does not. Then one can define another picture called the **interaction picture**, also known as the **Dirac picture**.

The idea of the interaction picture is to transfer from the state ket to the operators only the part of the time dependence that $H_0$ is responsible for. This is done by defining, for the ket describing the Schrödinger picture state of the system $|\psi(t)\rangle$ and for all observables $A$,

$$|\psi^{(I)}(t)\rangle \;=\; e^{i(t-t_0)H_0/\hbar}|\psi(t)\rangle, \tag{17.2.2}$$

$$A^{(I)}(t) \;=\; e^{i(t-t_0)H_0/\hbar} A\, e^{-i(t-t_0)H_0/\hbar}. \tag{17.2.3}$$

The superscript ${}^{(I)}$ is used to denote the interaction picture state and operators; when it is not present, the state or operator should be assumed to be in the Schrödinger picture. Note that $H_0^{(I)} = e^{i(t-t_0)H_0/\hbar} H_0 e^{-i(t-t_0)H_0/\hbar} = H_0$, so a distinguishing superscript for $H_0$ is not needed. However, in general, $W^{(I)}(t) \neq W(t)$, so we must be careful to distinguish them from now on. As in the case of the Heisenberg picture, the interaction picture depends on a choice of reference time $t_0$, which is sometimes taken to be 0. Note that eqs. (17.2.2) and (17.2.3) can be considered a change of basis, since $e^{-i(t-t_0)H_0/\hbar}$ is a unitary operator.

Clearly, if $W(t) = 0$, then the interaction and Heisenberg pictures are the same, while if $H_0 = 0$, then the interaction and Schrödinger pictures are the same. More generally, the interaction picture is intermediate between the Schrödinger and Heisenberg pictures. At first, it might seem that the interaction picture combines the worst features of both, since both the state and observable operators evolve non-trivially in time. However, it turns out to be useful for an elegant formulation of time-dependent perturbation theory, when the effects of $H_0$ are large but easily calculable, and the effects of $W(t)$ can be treated as the perturbation.

By an argument exactly analogous to the derivation of eq. (17.1.11), one can find the equation giving the time evolution of operators in the interaction picture,

$$\frac{dA^{(I)}}{dt} = \frac{i}{\hbar}[H_0, A^{(I)}] + \left(\frac{\partial A}{\partial t}\right)^{(I)}. \tag{17.2.4}$$

Here, the meaning of the last term is that one should take the derivative of $A$ with respect to its explicit time dependence in the Schrödinger picture, and then convert the result to the interaction picture using eq. (17.2.3).

Let us now define the unitary operator $\mathcal{U}_I(t, t_0)$ which gives the time evolution of $|\psi^{(I)}(t)\rangle$ in terms of its initial condition at $t = t_0$, according to

$$|\psi^{(I)}(t)\rangle = \mathcal{U}_I(t, t_0)|\psi^{(I)}(t_0)\rangle. \tag{17.2.5}$$

To avoid confusion, it is important to note that $\mathcal{U}_I(t, t_0)$ is not the same as $U^{(I)}(t, t_0)$, which is defined to be the interaction-picture version of the Schrödinger picture unitary time-evolution operator $U(t, t_0)$. In fact, the relation between these three unitary operators is

$$\mathcal{U}_I(t, t_0) = U^{(I)}(t, t_0) e^{i(t-t_0)H_0/\hbar} = e^{i(t-t_0)H_0/\hbar} U(t, t_0). \tag{17.2.6}$$

The last equality follows directly from the definition eq. (17.2.3), while the first equality then follows from eqs. (17.1.2), (17.2.2), and (17.2.5).

Our goal now is to solve for $\mathcal{U}_I(t, t_0)$. To accomplish this, we calculate

$$i\hbar \frac{d}{dt}|\psi^{(I)}(t)\rangle = i\hbar \frac{d}{dt}\left(e^{i(t-t_0)H_0/\hbar}|\psi(t)\rangle\right) \tag{17.2.7}$$

$$= -e^{i(t-t_0)H_0/\hbar} H_0 |\psi(t)\rangle + i\hbar e^{i(t-t_0)H_0/\hbar} \frac{d}{dt}|\psi(t)\rangle. \tag{17.2.8}$$

Using the Schrödinger equation $i\hbar\frac{d}{dt}|\psi\rangle = (H_0 + W)|\psi\rangle$, this becomes

$$i\hbar\frac{d}{dt}|\psi^{(I)}(t)\rangle = e^{i(t-t_0)H_0/\hbar} W|\psi\rangle, \tag{17.2.9}$$

which we can rewrite in a form involving only interaction picture objects, as

$$i\hbar\frac{d}{dt}|\psi^{(I)}(t)\rangle = W^{(I)}(t)|\psi^{(I)}(t)\rangle. \tag{17.2.10}$$

Plugging the definition eq. (17.2.5) into eq. (17.2.10), we get the operator equation

$$i\hbar\frac{d}{dt}\mathcal{U}_I(t,t_0) = W^{(I)}(t)\,\mathcal{U}_I(t,t_0). \tag{17.2.11}$$

Integrating both sides gives

$$i\hbar\big[\mathcal{U}_I(t,t_0) - \mathcal{U}_I(t_0,t_0)\big] = \int_{t_0}^{t} dt'\, W^{(I)}(t')\,\mathcal{U}_I(t',t_0), \tag{17.2.12}$$

or, rewriting using $\mathcal{U}_I(t_0,t_0) = I$,

$$\mathcal{U}_I(t,t_0) = I - \frac{i}{\hbar}\int_{t_0}^{t} dt'\, W^{(I)}(t')\,\mathcal{U}_I(t',t_0). \tag{17.2.13}$$

Finally, this integral equation can be solved by iteration,

$$\mathcal{U}_I(t,t_0) = I - \frac{i}{\hbar}\int_{t_0}^{t} dt'\, W^{(I)}(t') + \left(-\frac{i}{\hbar}\right)^2 \int_{t_0}^{t} dt' \int_{t_0}^{t'} dt''\, W^{(I)}(t')W^{(I)}(t'') + \cdots, \tag{17.2.14}$$

or, rewriting more systematically,

$$\mathcal{U}_I(t,t_0) = \sum_{N=0}^{\infty}\left(-\frac{i}{\hbar}\right)^N \int_{t_0}^{t} dt_N \int_{t_0}^{t_N} dt_{N-1} \cdots \int_{t_0}^{t_2} dt_1\, W^{(I)}(t_N)W^{(I)}(t_{N-1})\cdots W^{(I)}(t_1). \tag{17.2.15}$$

This solution can be verified by plugging it directly into eq. (17.2.11).

Equation (17.2.15) is the **Dyson series** solution for the time evolution of the state in the interaction picture; compare to the similar structure of eq. (3.4.26). In general, $W^{(I)}(t)$ may not commute with $W^{(I)}(t')$ when $t \neq t'$. Therefore, we are careful to put the operators in increasing time order, from right to left, as specified by the nested integrations with $t_0 < t_1 < t_2 < \cdots < t_{N-1} < t_N < t$. In practice, eq. (17.2.15) is usually approximated by truncating it to the first few terms, treating $W^{(I)}$ as small. A more compact (but less directly useful) way of writing eq. (17.2.15) is

$$\mathcal{U}_I(t,t_0) = T\,\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} dt'\, W^{(I)}(t')\right], \tag{17.2.16}$$

where $T$ is the time-ordering symbol introduced in the discussion surrounding eq. (3.4.25), to which eq. (17.2.16) can be compared. The symbol $T$ is an instruction to re-order the operators following it, so that after expanding the exponential, $W^{(I)}(t_j)$ appears to the right of $W^{(I)}(t_k)$ whenever $t_j < t_k$. The time-ordering rule is that earlier operators act first on kets.

Given our solution for $\mathcal{U}_I(t, t_0)$ in eq. (17.2.15), one can express the time dependence of the state back in the Schrödinger picture. Using eqs. (17.1.2) and (17.2.6),

$$|\psi(t)\rangle \;=\; U(t,t_0)|\psi(t_0)\rangle \;=\; e^{-i(t-t_0)H_0/\hbar}\,\mathcal{U}_I(t,t_0)|\psi(t_0)\rangle. \tag{17.2.17}$$

It is convenient to express this in terms of an orthobasis of eigenstates $|n\rangle$ of $H_0$ with eigenvalues $\mathcal{E}_n$. Using completeness twice, we have

$$\begin{aligned}|\psi(t)\rangle &= \sum_m\sum_n |m\rangle\langle m|\, e^{-i(t-t_0)\,H_0/\hbar}\,\mathcal{U}_I(t,t_0)\,|n\rangle\langle n|\psi(t_0)\rangle & (17.2.18)\\ &= \sum_m\sum_n e^{-i(t-t_0)\mathcal{E}_m/\hbar}\,|m\rangle\,\langle m|\mathcal{U}_I(t,t_0)|n\rangle\,\langle n|\psi(t_0)\rangle. & (17.2.19)\end{aligned}$$

Now, define the **transition amplitudes** for the eigenstates of $H_0$ as

$$a_{n\to m}(t,t_0) \;\equiv\; \langle m|\mathcal{U}_I(t,t_0)|n\rangle. \tag{17.2.20}$$

Then, given the initial state coefficients,

$$c_n(t_0) \;=\; \langle n|\psi(t_0)\rangle, \tag{17.2.21}$$

it follows that the Schrödinger picture state of the system at time $t$ is

$$|\psi(t)\rangle \;=\; \sum_m c_m(t)|m\rangle, \tag{17.2.22}$$

where, from eq. (17.2.19),

$$c_m(t) \;=\; \sum_n e^{-i(t-t_0)\mathcal{E}_m/\hbar}\,a_{n\to m}(t,t_0)\,c_n(t_0). \tag{17.2.23}$$

Note that $a_{n\to m}(t_0,t_0) = \langle m|n\rangle = \delta_{n,m}$, so that eq. (17.2.23) is consistent for $t = t_0$. The non-trivial part of the time evolution of the system due to $W(t)$ is encoded in the transition amplitudes $a_{n\to m}(t,t_0)$, which in turn rely on our solution for $\mathcal{U}_I(t,t_0)$ in eq. (17.2.15).

As a special case, suppose we are only interested in transitions between eigenstates of $H_0$, but with time evolution governed by the full Hamiltonian $H = H_0 + W(t)$. The probability to start at time $t_0$ in one of the $H_0$ eigenstates $|n\rangle$ and find the system at time $t$ in another $H_0$ eigenstate $|m\rangle$ is

$$\mathcal{P}_{n\to m} \;=\; |a_{n\to m}|^2. \tag{17.2.24}$$

More generally, the transition amplitudes $a_{n\to m}$ defined in eq. (17.2.20) can be used to find the time evolution of an arbitrary state $|\psi(t)\rangle$, through eqs. (17.2.22) and (17.2.23).

# 18 Time-dependent perturbation theory

## 18.1 The short-time and sudden approximations

As a warm-up for a more general treatment of time-dependent perturbation theory, suppose that the Hamiltonian for a system is

$$H(t) \;=\; \begin{cases} H_0 & (t<0), \\ H_{\rm int} & (0<t<\delta), \\ H_0' & (t>\delta). \end{cases} \tag{18.1.1}$$

The starting and ending Hamiltonians are assumed to have no time dependence, and to have known eigenstates and eigenvalues, $|n\rangle$ and $\mathcal{E}_n$ for $H_0$, and $|m'\rangle$ and $\mathcal{E}_{m'}$ for $H_0'$. In the interregnum $0 < t < \delta$, the Hamiltonian is $H_{\rm int}$. Consider the **short-time approximation**, in which $\delta$ is small compared to the time scales set by $\hbar$ divided by characteristic energy differences. The idea is to take advantage of the smallness of $\delta$ to develop an approximation for the state of the system at late times, despite the possibly complicated form of $H_{\rm int}$, and in particular even if we do not know its eigenstates and eigenvalues. We assume that $\delta$ is so small that any explicit time dependence of $H_{\rm int}$ over that time interval can be ignored.

For each of $t \leq 0$ and $t \geq \delta$, we can use the Schrödinger equation to write the form of the state as

$$|\psi(t)\rangle \;=\; \sum_n b_n e^{-it\mathcal{E}_n/\hbar}|n\rangle \qquad (t\leq 0), \tag{18.1.2}$$

$$|\psi(t)\rangle \;=\; \sum_{m'} c_{m'} e^{-i(t-\delta)\mathcal{E}_{m'}/\hbar}|m'\rangle \qquad (t\geq \delta), \tag{18.1.3}$$

for some constants $b_n$ and $c_{m'}$. For later convenience, a constant phase $e^{i\mathcal{E}_{m'}\delta/\hbar}$ has been included in eq. (18.1.3), by choice of the definition of $c_{m'}$. For $0 < t < \delta$, the Schrödinger equation is

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = H_{\rm int}|\psi(t)\rangle. \tag{18.1.4}$$

So, to first order in $\delta$,

$$|\psi(\delta)\rangle \;=\; \left(I - i\frac{\delta}{\hbar}H_{\rm int}\right)|\psi(0)\rangle. \tag{18.1.5}$$

Comparing this to eq. (18.1.2) with $t = 0$, and eq. (18.1.3) with $t = \delta$, we find

$$\sum_{m'} c_{m'}|m'\rangle \;=\; \left(I - i\frac{\delta}{\hbar}H_{\rm int}\right)\sum_n b_n|n\rangle. \tag{18.1.6}$$

Now taking the inner product with another $H_0'$ eigenstate $\langle k'|$, and then relabeling $k' \to m'$, we get the solution

$$c_{m'} \;=\; \sum_n b_n a_{n\to m'}, \tag{18.1.7}$$

where the transition amplitude is

$$a_{n\to m'} \;=\; \langle m'|n\rangle - i\frac{\delta}{\hbar}\langle m'|H_{\rm int}|n\rangle. \tag{18.1.8}$$

Given the initial state, specified by the coefficients $b_n$ of the $H_0$ eigenstates, the final state is determined by the coefficients $c_{m'}$ of the $H_0'$ eigenstates.

One application of eq. (18.1.8) is to the special case $H_0' = H_0$, so that $H_{\rm int} - H_0$ acts as a perturbation for a short time interval on an otherwise constant Hamiltonian. Then $\langle m|n\rangle = \delta_{n,m}$, and

$$a_{n\to m} \;=\; \delta_{n,m} - i\frac{\delta}{\hbar}\sum_n \langle m|H_{\rm int}|n\rangle. \tag{18.1.9}$$

For example, if the initial state is an eigenstate $|n\rangle$ of $H_0$, so that only one of the $b_n$ is non-zero, then the probability of finding the system in a different $H_0$ eigenstate $|m\rangle$ after $t > \delta$ is

$$\mathcal{P}_{n\to m} \;=\; \frac{\delta^2}{\hbar^2}|\langle m|H_{\rm int}|n\rangle|^2. \tag{18.1.10}$$

In the next section, we will see that this result can be obtained as a special case application of a general treatment that we will develop for time-dependent perturbation theory; see eqs. (18.2.8) and (18.2.9), with $W = H_{\rm int} - H_0$ for $0 < t < \delta$.

The **sudden approximation** is obtained if we assume that $\delta$ is so small that $H_{\rm int}$ is irrelevant, but $H_0' \neq H_0$. Then, if the system was originally in an $H_0$ eigenstate $|n\rangle$ for $t < 0$, the probability of finding the system in an $H_0'$ eigenstate $|m'\rangle$ for $t > 0$ is

$$\mathcal{P}_{n\to m'} \;=\; |c_{m'}|^2 \;=\; |\langle m'|n\rangle|^2. \tag{18.1.11}$$

This follows from the $\delta \to 0$ limit of eq. (18.1.8).

As an example, suppose that a one-dimensional harmonic oscillator is in its ground state, when the angular frequency in the Hamiltonian suddenly doubles, due to some outside agent. Let us compute the probability that the oscillator will be found in the ground state of the new Hamiltonian. Working in the position representation, the normalized wavefunctions of the ground states are

$$\langle x|0\rangle_\omega \;=\; \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp(-m\omega x^2/2\hbar) \qquad \text{(before)}, \tag{18.1.12}$$

$$\langle x|0\rangle_{2\omega} \;=\; \left(\frac{2m\omega}{\pi\hbar}\right)^{1/4}\exp(-m\omega x^2/\hbar) \qquad \text{(after)}. \tag{18.1.13}$$

Then, in the sudden approximation, we compute

$${}_{2\omega}\langle 0|0\rangle_\omega = \int_{-\infty}^{\infty} dx\; {}_{2\omega}\langle 0|x\rangle\langle x|0\rangle_\omega = \sqrt{\frac{\sqrt{2}m\omega}{\pi\hbar}}\int_{-\infty}^{\infty} dx\, \exp(-3m\omega x^2/2\hbar) \;=\; \frac{2^{3/4}}{3^{1/2}}, \tag{18.1.14}$$

which results in $\mathcal{P}_{0_\omega \rightarrow 0_{2\omega}} = 2\sqrt{2}/3 \approx 0.9428$.

As a more practical example, consider the effect of the decay of the tritium ($^3$H) nucleus on the atomic electron state. The tritium nucleus consists of 1 proton and 2 neutrons, so $Z = 1$. It undergoes a $\beta$ decay to a helium isotope and an electron and antineutrino,

$$^3\text{H} \;\; \rightarrow \;\; ^3\text{He} + e^- + \overline{\nu}_e. \tag{18.1.15}$$

The antineutrino has only weak interactions, and so can be ignored for the present purposes. The $\beta$-decay electron is emitted with average (maximum) kinetic energy 0.0057 MeV (0.0186 MeV), so that its average (maximum) speed can be computed to be $0.15c$ ($0.26c$). These are much larger than the typical semi-classical speeds of electrons in a $Z = 1$ atom, $v = \alpha c \approx c/137$. Therefore, it is a good approximation to also ignore the $\beta$-decay electron, as it leaves the scene very quickly.

The remaining $^3$He atom consists of the nucleus with 2 protons and 1 neutron, so $Z = 2$, together with the left-over atomic electron from the $^3$H atom. Therefore, in the sudden approximation the probability of the atomic electron starting in the ground state of $^3$H and ending up in the ground state of the $^3\text{He}^+$ ion is $\mathcal{P}_{1,0,0\rightarrow 1,0,0} = |a_{1,0,0\rightarrow 1,0,0}|^2$, where

$$a_{1,0,0\rightarrow 1,0,0} \;=\; \int d^3\vec{r}\; \left[\psi^{Z=2}_{1,0,0}(\vec{r}\,)\right]^* \psi^{Z=1}_{1,0,0}(\vec{r}\,). \tag{18.1.16}$$

Recalling that

$$\psi^{Z}_{1,0,0}(\vec{r}\,) \;=\; \sqrt{\frac{Z^3}{\pi a_0^3}}\, e^{-Zr/a_0}, \tag{18.1.17}$$

we have

$$a_{1,0,0\rightarrow 1,0,0} \;=\; 4\pi \int_0^\infty dr\, r^2\, \frac{2^{3/2}}{\pi a_0^3} e^{-3r/a_0} \;=\; 2^{9/2}/3^3, \tag{18.1.18}$$

and the probability of the electron remaining in the atomic ground state after the decay is

$$\mathcal{P}_{1,0,0\rightarrow 1,0,0} \;=\; 2^9/3^6 \;=\; 512/729 \;\approx\; 0.702. \tag{18.1.19}$$

## 18.2 Transition amplitudes and probabilities in perturbation theory

In section 3.4, we learned how to solve for the time-dependence of quantum state. If the Hamiltonian does not depend on time, and one can solve its eigenvalue problem, then the time evolution can always be found by writing the unitary operator $U(t, t_0)$ in terms of the stationary states with appropriate phases factors that depend on the energy; see eqs. (3.4.1) and (3.4.7). However, if the Hamiltonian does depend on time, then as we saw in section 17.2, the time evolution is given in terms of a Dyson series that typically cannot be obtained exactly in closed form. The approximation that follows from truncating the Dyson series in the interaction picture is known as time-dependent perturbation theory.

Just as in our discussion of the interaction picture, we suppose that the full Hamiltonian is

$$H(t) = H_0 + W(t), \tag{18.2.1}$$

where $H_0$ does not depend explicitly on time, and we assume that we know (at least approximately) the solutions to its eigenvalue problem,

$$H_0|n\rangle \;=\; \mathcal{E}_n|n\rangle. \tag{18.2.2}$$

Since $H_0$ is an observable, the kets $|n\rangle$ form a complete orthobasis. The remaining part of the full Hamiltonian, $W(t)$, contains the explicit time dependence. Then, given the state of the system at time $t_0$, the solution for the state at a later time $t$ is found in terms of the transition amplitudes $a_{n\to m}$ defined by eq. (17.2.20), which relies on eq. (17.2.15), where

$$W^{(I)}(t) \;=\; e^{i(t-t_0)H_0/\hbar}\, W(t)\, e^{-i(t-t_0)H_0/\hbar} \tag{18.2.3}$$

is the perturbation part of the Hamiltonian, translated into the interaction picture.

The time-dependent perturbative expansion simply amounts to assuming that $W(t)$ is small, so that we can expand the transition amplitude in eq. (17.2.20) as

$$a_{n\to m}(t, t_0) \;=\; a^{(0)}_{n\to m} + a^{(1)}_{n\to m} + a^{(2)}_{n\to m} + a^{(3)}_{n\to m} + \cdots\,, \tag{18.2.4}$$

where each term $a^{(N)}_{n\to m}$ contains $N$ factors of $W$, following from the corresponding terms in eq. (17.2.15). In practice, eq. (18.2.4) is then truncated at some finite $N$, usually at $N = 1$ or $N = 2$. Note that the zeroth order contribution is just

$$a^{(0)}_{n\to m} \;=\; \langle m|n\rangle \;=\; \delta_{n,m}, \tag{18.2.5}$$

as follows from the identity operator ($N = 0$) part of $\mathcal{U}_I$.

The preceding tells us that the first-order contribution to the transition amplitude is

$$a^{(1)}_{n\to m} \;=\; -\frac{i}{\hbar}\int_{t_0}^{t} dt_1\, \langle m|e^{i(t_1-t_0)H_0/\hbar}\, W(t_1)\, e^{-i(t_1-t_0)H_0/\hbar}|n\rangle. \tag{18.2.6}$$

Fortunately, each $H_0$ can be evaluated acting on a neighboring eigenket or eigenbra. So, defining

$$\omega_{mn} = (\mathcal{E}_m - \mathcal{E}_n)/\hbar, \tag{18.2.7}$$

we obtain

$$a^{(1)}_{n\to m} \;=\; -\frac{i}{\hbar}\int_{t_0}^{t} dt_1\, e^{i(t_1-t_0)\omega_{mn}}\, \langle m|\, W(t_1)\, |n\rangle. \tag{18.2.8}$$

This result provides the time evolution of a general state from $t_0$ to $t$, at the leading order in time-dependent perturbation theory, using eqs. (17.2.22)–(17.2.23) with $a_{n\to m} = \delta_{n,m} + a^{(1)}_{n\to m}$.

Before moving on to consider the higher-order contributions, consider the special case that the system starts at time $t_0$ in an eigenstate $|n\rangle$ of $H_0$, and we then measure $H_0$ at time $t$. The transition probability to transfer from the stationary state $|n\rangle$ to a *different* eigenstate $|m\rangle$ of $H_0$, at the leading non-trivial order in the perturbation, is simply

$$\mathcal{P}_{n\to m} \;=\; |a^{(1)}_{n\to m}|^2, \qquad\qquad \text{(for } m \neq n\text{)}. \tag{18.2.9}$$

The validity of the approximation requires that we find $|a^{(1)}_{n\to m}| \leq 1$, so that the probability does not exceed 1. This is a necessary but not sufficient condition, because the sum of $|a^{(1)}_{n\to m}|^2$ for multiple final states $|m\rangle$ could still exceed 1. And, even if that does not happen, the series expansion in eq. (18.2.4) could still fail to converge when higher-order terms are included. The perturbation expansion is often, but not always, more reliable if $|a^{(1)}_{n\to m}| \ll 1$.

It is also instructive to consider what happens if we try to use the first-order approximation to directly compute the probability to remain in the initial stationary state $|n\rangle$. Since $a^{(0)}_{n\to n} = 1$, we get, instead of eq. (18.2.9):

$$\mathcal{P}_{n\to n} \;=\; |1 + a^{(1)}_{n\to n} + a^{(2)}_{n\to n} + \cdots|^2. \tag{18.2.10}$$

Now, as part of the Hamiltonian, $W(t_1)$ must be Hermitian, so its expectation values are always real, and

$$a^{(1)}_{n\to n} \;=\; -\frac{i}{\hbar}\int_{t_0}^{t} dt_1\, \langle n|W(t_1)|n\rangle \tag{18.2.11}$$

is a pure imaginary number. Therefore, although one might naively neglect $a^{(2)}_{n\to n}$ as being of higher order, eq. (18.2.10) would then give $1 + |a^{(1)}_{n\to n}|^2$, which is clearly unacceptable for a probability unless the first-order correction vanishes. To get a meaningful result from this direct calculation, we must work consistently to at least second order, by also including the contribution from $a^{(2)}_{n\to n}$ in the transition amplitude, with the result [dropping contributions of third order in $W(t)$]

$$\mathcal{P}_{n\to n} \;=\; 1 + |a^{(1)}_{n\to n}|^2 + 2\text{Re}[a^{(2)}_{n\to n}]. \tag{18.2.12}$$

Alternatively, the probability to remain in the initial eigenstate of $H_0$ can be obtained indirectly, from the complementary probability,

$$\mathcal{P}_{n\to n} \;=\; 1 - \sum_{m\neq n} \mathcal{P}_{n\to m} \;=\; 1 - \sum_{m\neq n} |a^{(1)}_{n\to m}|^2, \tag{18.2.13}$$

which is also second order in the perturbation. Comparing these, we learn that

$$2\mathrm{Re}[a^{(2)}_{n\to n}] \;=\; -\sum_m |a^{(1)}_{n\to m}|^2, \tag{18.2.14}$$

with the sum over $m$ now including $n$.

Now consider the second-order contribution to the transition amplitude, obtained from eqs. (17.2.15), (18.2.3), and eq. (17.2.20),

$$a^{(2)}_{n\to m} \;=\; \left(-\frac{i}{\hbar}\right)^2 \int_{t_0}^{t} dt_2 \int_{t_0}^{t_2} dt_1\, \langle m|e^{i(t_2-t_0)H_0/\hbar}W(t_2)e^{i(t_1-t_2)H_0/\hbar}W(t_1)e^{-i(t_1-t_0)H_0/\hbar}|n\rangle. \tag{18.2.15}$$

To evaluate this, we use completeness, by inserting a sum over all eigenstates of $H_0$,

$$I = \sum_{k_1} |k_1\rangle\langle k_1|, \tag{18.2.16}$$

immediately before $W(t_1)$. This puts every $H_0$ adjacent to one of its eigenstate bras or kets, allowing it to be replaced by $\mathcal{E}_n$, $\mathcal{E}_m$, or $\mathcal{E}_{k_1}$, with the result

$$a^{(2)}_{n\to m} \;=\; \left(-\frac{i}{\hbar}\right)^2 \int_{t_0}^{t} dt_2 \int_{t_0}^{t_2} dt_1 \sum_{k_1} e^{i(t_2-t_0)\omega_{mk_1}} \langle m|W(t_2)|k_1\rangle\, e^{i(t_1-t_0)\omega_{k_1 n}}\, \langle k_1|W(t_1)|n\rangle. \tag{18.2.17}$$

It is left as an exercise to perform the check of taking the special case $|m\rangle = |n\rangle$ and manipulating the result to verify that eq. (18.2.14) indeed holds.

In the same way, the third-order contribution is found to be

$$\begin{aligned} a^{(3)}_{n\to m} \;=\; & \left(-\frac{i}{\hbar}\right)^3 \int_{t_0}^{t} dt_3 \int_{t_0}^{t_3} dt_2 \int_{t_0}^{t_2} dt_1 \sum_{k_1}\sum_{k_2} e^{i(t_3-t_0)\omega_{mk_2}} \langle m|W(t_3)|k_2\rangle \\ & e^{i(t_2-t_0)\omega_{k_2k_1}}\, \langle k_2|W(t_2)|k_1\rangle\, e^{i(t_1-t_0)\omega_{k_1 n}}\, \langle k_1|W(t_1)|n\rangle. \end{aligned} \tag{18.2.18}$$

A diagrammatic representation of the contributions $a^{(N)}_{n\to m}$ is shown in Figure 18.2.1. Each line with an arrow represents an eigenstate of $H_0$, as labeled. The vertices represent insertions of the perturbation Hamiltonian, and can be interpreted in two equivalent ways. For a vertex with incoming state line $k$, outgoing state line $k'$, and time label $t_j$, one can assign either

$$-\frac{i}{\hbar}\langle k'|W^{(I)}(t_j)|k\rangle \tag{18.2.19}$$

in the interaction picture, or

$$-\frac{i}{\hbar}e^{i(t_j-t_0)\omega_{k'k}}\, \langle k'|W(t_j)|k\rangle \tag{18.2.20}$$

in the Schrödinger picture. Then one integrates $\int_{t_0}^{t_{j+1}} dt_j$ for each of the intermediate times $t_1, \ldots, t_N$ in the diagram for $a^{(N)}_{n\to m}$, with the identification $t_{N+1} = t$. All internal state labels $k_j$ are summed over. This diagrammatic mnemonic allows us to quickly write down the formula for $a^{(N)}_{n\to m}$, corresponding to each sub-diagram starting with $n$ and ending with $m$.

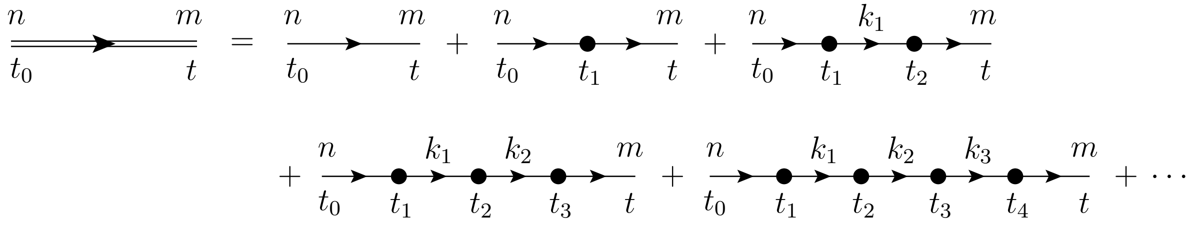

Figure 18.2.1: A diagrammatic representation of the time-dependent perturbation theory expansion for the transition amplitude $a_{n\to m}(t, t_0)$, represented by the double line with arrow. In the sum on the right side of the equality, each sub-diagram with $N$ vertices represents $a^{(N)}_{n\to m}(t, t_0)$. Each single line with arrow represents an $H_0$ eigenstate, as labeled. Each vertex labeled by a time $t_j$ and incoming line $k$ and outgoing line $k'$ represents a factor of $-\frac{i}{\hbar}e^{i(t_j-t_0)\omega_{k'k}}\langle k'|W(t_j)|k\rangle$, with an integration $\int_{t_0}^{t_{j+1}} dt_j$. This is done for each of the intermediate times $t_1, \ldots, t_N$, with the identification $t_{N+1} = t$. All internal $H_0$ eigenstate labels $k_j$ are summed over.

## 18.3 Applying first-order time-dependent perturbation theory, and Fermi's golden rule

In this section and the next, we consider the application of time-dependent perturbation theory in the approximation of keeping only the first-order contribution to the transition amplitude, eq. (18.2.8).

As a first example, consider a one-dimensional harmonic oscillator with the usual unperturbed Hamiltonian $H_0 = \frac{P^2}{2m} + \frac{1}{2}m\omega^2X^2$, taken to be in its ground state $|0\rangle$ in the far past, at time $t = -\infty$. A perturbation

$$W(t) \;\;=\;\; -qE_0e^{-t^2/\tau^2}X \tag{18.3.1}$$

is applied, corresponding to a continuous turning on and then off of a uniform electric field $E_0$ that couples to the particle's charge $q$. The constant $\tau$ controls the time scale of the turn-on and turn-off. Our goal is to find the probability that the particle is found in a given excited state $|n\rangle$ when the perturbation has become negligible in the far future as $t \to \infty$.

Applying eq. (18.2.8) with $\mathcal{E}_n = \hbar\omega(n + 1/2)$, the amplitude for the transition in the far future, when the perturbation has effectively turned off, is

$$a^{(1)}_{0\to n}(t = \infty, t_0) \;\;=\;\; -\frac{i}{\hbar}e^{-in\omega t_0}\int_{t_0}^{\infty} dt_1\, e^{in\omega t_1}(-qE_0e^{-t_1^2/\tau^2})\langle n|X|0\rangle. \tag{18.3.2}$$

The necessary matrix element is

$$\langle n|X|0\rangle \;\;=\;\; \sqrt{\frac{\hbar}{2m\omega}}\delta_{n,1}, \tag{18.3.3}$$

so that within the first-order approximation, there is a selection rule that only the first excited state can be reached. We now drop the constant phase $e^{-i\omega t_0}$, since we will be taking the squared magnitude to find the probability. Then, taking the limit $t_0 \to -\infty$,

$$a^{(1)}_{0\to 1} \;=\; \frac{iqE_0}{\sqrt{2m\hbar\omega}} \int_{-\infty}^{\infty} dt_1 \, \exp\left(i\omega t_1 - t_1^2/\tau^2\right). \tag{18.3.4}$$

The integral can be done by completing the square, using the change of variables $u = t_1/\tau - i\omega\tau/2$, resulting in a transition probability

$$\mathcal{P}(0\to 1) \;=\; \frac{\pi q^2 E_0^2 \tau^2}{2m\hbar\omega} e^{-\omega^2\tau^2/2}, \tag{18.3.5}$$

while $a^{(1)}_{0\to n}$ and $\mathcal{P}(0 \to n)$ vanish for all $n > 1$. Since $\mathcal{P}$ cannot exceed 1, the approximation leading to eq. (18.3.5) must fail if the applied field is too large, that is if $q^2E_0^2\tau^2$ is too large in comparison to $m\hbar\omega e^{\omega^2\tau^2/2}$.

As a second example with many applications, we consider a generic unspecified $H_0$, and a perturbation that turns on at $t = 0$ and is constant thereafter,

$$W(t) \;=\; \begin{cases} 0 & (\text{for } t < 0), \\ V & (\text{for } t > 0). \end{cases} \tag{18.3.6}$$

To be as general as possible, let $V$ be an arbitrary Hermitian operator, with the only restriction that it has no time dependence. Now consider the probability that the perturbation will induce a transition from an initial $H_0$ eigenstate $|i\rangle$ to a different final $H_0$ eigenstate $|f\rangle$. From eq. (18.2.8), the amplitude for the transition is

$$a^{(1)}_{i\to f} \;=\; -\frac{i}{\hbar}\langle f|V|i\rangle \int_0^t dt_1 \, e^{i\omega_{fi}t_1}, \tag{18.3.7}$$

where we have exploited the fact that the matrix element of $V$ does not depend on time to pull it out of the integral, and defined $\omega_{fi} = (\mathcal{E}_f - \mathcal{E}_i)/\hbar$. Evaluating the integral, and squaring the magnitude of the result, we obtain the transition probability for $t > 0$,

$$\mathcal{P}_{i\to f}(t) \;=\; |a^{(1)}_{i\to f}|^2 \;=\; \frac{4\sin^2(t\omega_{fi}/2)}{\hbar^2\omega_{fi}^2} \, |\langle f|V|i\rangle|^2, \tag{18.3.8}$$

valid at first order in perturbation theory. For future convenience, we define the function

$$F(\mathcal{E}_f - \mathcal{E}_i, \, t) \;\equiv\; \frac{4}{(\mathcal{E}_f - \mathcal{E}_i)^2} \sin^2\left(\frac{t(\mathcal{E}_f - \mathcal{E}_i)}{2\hbar}\right), \tag{18.3.9}$$

so that

$$\mathcal{P}_{i\to f}(t) \;=\; F(\mathcal{E}_f - \mathcal{E}_i, \, t)\, |\langle f|V|i\rangle|^2. \tag{18.3.10}$$

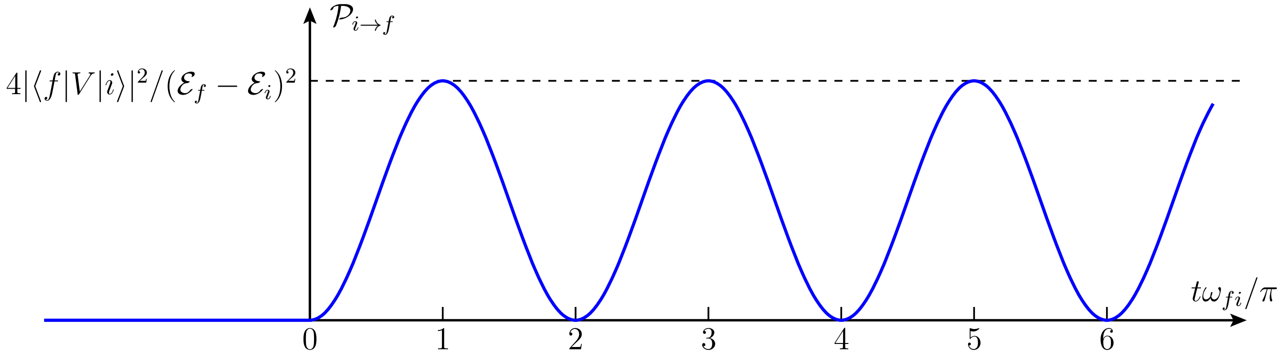

Figure 18.3.1: The transition probability at first-order in time-dependent perturbation theory, $\mathcal{P}_{i\to f} = |a^{(1)}_{i\to f}|^2$, as a function of the time $t$, for a constant perturbation $V$ that turns on at time $t = 0$. For short times, the probability grows quadratically like $t^2|\langle f|V|i\rangle|^2/\hbar^2$, and for longer times oscillates between 0 and a maximum value $4|\langle f|V|i\rangle|^2/(\mathcal{E}_f-\mathcal{E}_i)^2$.

The transitions at first order in perturbation theory are limited to those that satisfy

$$\langle f|V|i\rangle \neq 0, \tag{18.3.11}$$

an example of a matrix element selection rule.

When applying eq. (18.3.10) in practice, it is important to distinguish the cases for which $|f\rangle$ is a discrete state with an isolated $H_0$ eigenvalue $\mathcal{E}_f$ (for example, a bound state of an electron in an atom) from those in which $|f\rangle$ is a state in a continuum (for example, a free-particle plane wave, or an unbound state of an electron ionized from an atom but still under the influence of the Coulomb potential of the nucleus).

First suppose that $|f\rangle$ is discrete. In that case, eq. (18.3.10) can be directly applied. For very small $t$, the transition probability grows quadratically, because for small $t$,

$$F(\mathcal{E}_f-\mathcal{E}_i, t) \approx t^2/\hbar^2, \tag{18.3.12}$$

which is independent of the unperturbed energies. Therefore, eq.(18.3.10) becomes

$$\mathcal{P}_{i\to f} \approx \frac{t^2}{\hbar^2}|\langle f|V|i\rangle|^2 \qquad \text{(small } t\text{)}. \tag{18.3.13}$$

Of course, this must fail for sufficiently large $t$, as the probability cannot exceed 1. Equation (18.3.13) agrees with the result we had already found in the short-time approximation in eq. (18.1.10). For longer time scales, the transition probability for a given discrete final state oscillates between 0 and a maximum value $4|\langle f|V|i\rangle|^2/(\mathcal{E}_f-\mathcal{E}_i)^2$, as illustrated in Figure 18.3.1.

Due to the denominator, the final states $|f\rangle$ that have $\mathcal{E}_f \approx \mathcal{E}_i$ will tend to have the largest probabilities, provided that the matrix element $\langle f|V|i\rangle$ is not too small in magnitude. Also, the probability for a given final state $|f\rangle$ vanishes whenever $t = 2\pi n/\omega_{fi} = 2\pi\hbar n/(\mathcal{E}_f - \mathcal{E}_i)$ for integer $n$. These features are illustrated by the graph of $F(\mathcal{E}_f - \mathcal{E}_i, t)$ as a function of $\mathcal{E}_f - \mathcal{E}_i$, shown in Figure 18.3.2. The final states $|f\rangle$ most likely to be reached at a given time $t$ are those with energies that lie under the main peak around $\mathcal{E}_i$, which is bounded by

$$|\mathcal{E}_f - \mathcal{E}_i| \;<\; 2\pi\hbar/t. \tag{18.3.14}$$

The main peak gets narrower and taller for larger $t$, so that in the limit of long times, the most likely transitions are those for which the difference in unperturbed energies is constrained, increasingly strongly, by eq. (18.3.14).

Now consider the case that $|f\rangle$ is not discrete, so that the unperturbed energy eigenvalues $\mathcal{E}_f$ and the matrix elements $\langle f|V|i\rangle$ each form a continuum. In that case, we write

$$|f\rangle \;=\; |\mathcal{E}_f, u_{\mathcal{E}_f}\rangle, \tag{18.3.15}$$

where $u_{\mathcal{E}_f}$ represents the degeneracy label (or labels) for the unperturbed energy. (For example, the degeneracy labels could include the angular direction of a final-state momentum vector, and/or an unobserved spin.) We also define the density of states $\rho(\mathcal{E}_f, u_{\mathcal{E}_f})$ according to

$$\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f} \;=\; \begin{pmatrix}\text{number of states between } \mathcal{E}_f \text{ and } \mathcal{E}_f + d\mathcal{E}_f, \\ \text{and between } u_{\mathcal{E}_f} \text{ and } u_{\mathcal{E}_f} + du_{\mathcal{E}_f}\end{pmatrix}, \tag{18.3.16}$$

where, for now, the degeneracy labels are also assumed continuous. The corresponding contri-

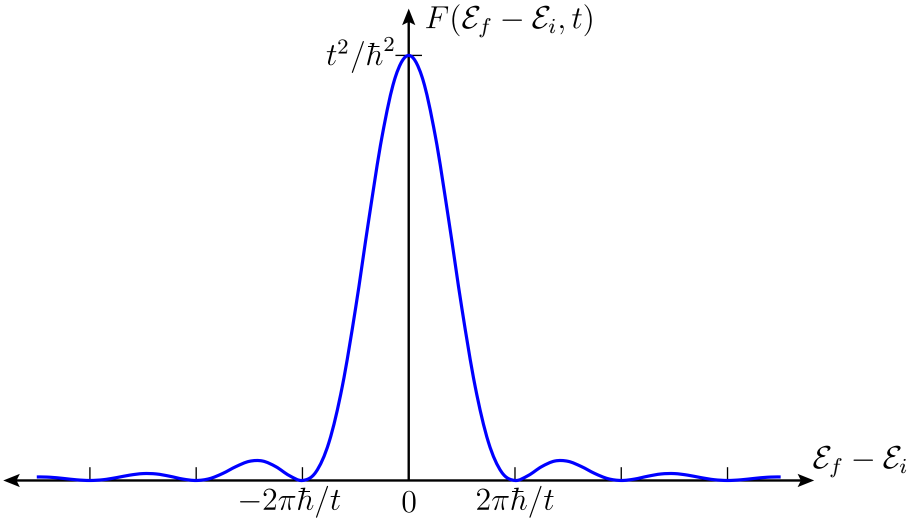

Figure 18.3.2: The function $F(\mathcal{E}_f - \mathcal{E}_i, t)$ defined by eq. (18.3.9) and appearing in the transition probability for first-order time-dependent perturbation theory in eq. (18.3.10), as a function of $\mathcal{E}_f - \mathcal{E}_i$. The main peak height $t^2/\hbar^2$ grows quadratically with time, while its width shrinks proportionally to $1/t$. The total area under the curve is $2\pi t/\hbar$.

bution to the transition probability is, from eq. (18.3.10),

$$d\mathcal{P}_{i\to f} \;=\; |\langle f|V|i\rangle|^2\, F(\mathcal{E}_f - \mathcal{E}_i, t)\, \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f}, \tag{18.3.17}$$

in terms of the function $F$ defined in eq.(18.3.9). In practice, eq. (18.3.17) must always be integrated over some ranges of $\mathcal{E}_f$ and $u_{\mathcal{E}_f}$ in order to give the probability of transition to that range of states,

$$\Delta\mathcal{P} \;=\; \int d\mathcal{P}_{i\to f}. \tag{18.3.18}$$

If we want only the transition probability for some of the final states (for example, those with energy in some range, or momentum direction inside some cone), this can be enforced either by integrating only over the appropriate ranges of energy and degeneracy labels, or by putting a projection operator before $V$ within the matrix element $\langle f|V|i\rangle$.

In the limit of large time $t$, we can take advantage of the fact that $F(\mathcal{E}_f - \mathcal{E}_i, t)$ as a function of $\mathcal{E}_f$ becomes increasingly sharply peaked near $\mathcal{E}_i$, and so approaches proportionality to a delta function. To understand the proportionality, we note that

$$\int_{-\infty}^{\infty} d\mathcal{E}_f\, F(\mathcal{E}_f - \mathcal{E}_i, t) \;=\; 2\pi t/\hbar, \tag{18.3.19}$$

so that we can make the replacement

$$F(\mathcal{E}_f - \mathcal{E}_i, t) \;\to\; \frac{2\pi t}{\hbar}\delta(\mathcal{E}_f - \mathcal{E}_i) \tag{18.3.20}$$

for large $t$. The transition probability eq. (18.3.17) is then directly proportional to $t$,

$$d\mathcal{P}_{i\to f} \;=\; t\, dR_{i\to f}, \tag{18.3.21}$$

where the proportionality can therefore be interpreted as a transition rate per unit time,

$$dR_{i\to f} \;=\; \frac{2\pi}{\hbar}\, |\langle f|V|i\rangle|^2\, \delta(\mathcal{E}_f - \mathcal{E}_i)\, \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f}. \tag{18.3.22}$$

For a group of states with $\mathcal{E}_f \approx \mathcal{E}_i$, the total transition rate to those states is obtained by integration over $\mathcal{E}_f$, and integrating (or summing) over the degeneracy label(s) $u_{\mathcal{E}_f}$ for it. Doing the energy integration, and using the notation of eq. (18.3.15), we get the transition rate

$$R \;=\; \frac{2\pi}{\hbar} \int du_{\mathcal{E}_f}\, \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, \left|\langle \mathcal{E}_f, u_{\mathcal{E}_f}|V|i\rangle\right|^2, \tag{18.3.23}$$

where it is now understood that only final states with $\mathcal{E}_f = \mathcal{E}_i$ are included, and again projection operators can be used inside the matrix element to restrict to particular classes of final states with some desired properties. Either eq. (18.3.23) [or its equivalent (18.3.22)] is one version of

**Fermi's golden rule**. (It was actually originally obtained by Dirac, but Fermi popularized it.) If the energy degeneracy labels $u_{\mathcal{E}_f}$ are discrete rather than continuous, then one makes the replacement $\int du_{\mathcal{E}_f} \rightarrow \sum_{u_{\mathcal{E}_f}}$.

Although the derivation of the golden rule turned $F$ into a delta function by assuming that $t$ is large, we have to be careful if it is *too* large. This is simply because interpretation of eq. (18.3.23) as a constant rate clearly requires that $t < 1/R$, so that the probability does not exceed 1. More generally, for times that are not short compared to $1/R$, the transition rate decreases as the initial state $|i\rangle$ is depleted. If we write $\mathcal{P}_{i\rightarrow i}(t) = e^{-R_{\text{tot}}t}$ for the probability of remaining in the initial state, including the effects of changing to *all* possible final states, then the transition rate to a particular class of final states approaches $Re^{-R_{\text{tot}}t}$ at late times.

The golden rule can also be applied as an approximation in the case of a group of many discrete but closely spaced final states with approximately the same matrix elements $\langle \mathcal{E}_f, u_{\mathcal{E}_f}|V|i\rangle$. This works if the unperturbed energy differences $\Delta\mathcal{E}_f$ between neighboring discrete final states are small enough that there are many states under the main peak in the function $F(\mathcal{E}_f - \mathcal{E}_i, t)$. This provision requires that

$$t\,\Delta\mathcal{E}_f \quad \ll \quad 2\pi\hbar, \tag{18.3.24}$$

so that the elapsed time is short compared to the inverse energy spacing between adjacent states.

## 18.4 Harmonic time-dependent perturbations

An even more common application of time-dependent perturbation theory is the case of a **harmonic time-dependent perturbation**, in which $W(t)$ is proportional to a sine or cosine function of time. For example, this includes the case of electromagnetic waves interacting with a charged particle. Suppose that the Hamiltonian perturbation turns on at time $t = 0$, so

$$W(t) \quad = \quad \begin{cases} 0 & \text{(for } t < 0\text{)}, \\ Ve^{-i\omega t} + V^\dagger e^{i\omega t} & \text{(for } t > 0\text{)}, \end{cases} \tag{18.4.1}$$

where $V$ is an operator that does not depend on time. Note that $W(t)$ is Hermitian, but $V$ need not be. This perturbation can also be written as $(V + V^\dagger)\cos(\omega t) + i(V^\dagger - V)\sin(\omega t)$ for $t > 0$, with $(V + V^\dagger)$ and $i(V^\dagger - V)$ both Hermitian operators.

Applying first-order time-dependent perturbation theory in the form of eq. (18.2.8), the transition amplitude is

$$a^{(1),\omega}_{i\rightarrow f}(t) \quad = \quad -\frac{i}{\hbar}\int_0^t dt' \left(e^{i(\omega_f-\omega_i-\omega)t'}\langle f|V|i\rangle + e^{i(\omega_f-\omega_i+\omega)t'}\langle f|V^\dagger|i\rangle\right) \tag{18.4.2}$$

$$= \quad \frac{1 - e^{i(\omega_f-\omega_i-\omega)t}}{\hbar(\omega_f - \omega_i - \omega)}V_{fi} \; + \; \frac{1 - e^{i(\omega_f-\omega_i+\omega)t}}{\hbar(\omega_f - \omega_i + \omega)}V^\dagger_{fi}, \tag{18.4.3}$$

where we have used a short-hand notation for the time-independent matrix elements,

$$V_{fi} = \langle f|V|i\rangle, \qquad V^\dagger_{fi} = \langle f|V^\dagger|i\rangle. \tag{18.4.4}$$

Requiring that these do not vanish provides a selection rule for first-order time-dependent perturbation theory.

The first term in eq. (18.4.3) is resonant only if $\omega_f \approx \omega_i + \omega$, and the second term only if $\omega_f \approx \omega_i - \omega$. For $\omega \neq 0$, at most one of these conditions can be satisfied for any particular $\mathcal{E}_f$, so only one term will be important within the resonance approximation. Therefore, we can make the approximation of ignoring the cross terms in $|a^{(1),\omega}_{i\to f}(t)|^2$. Then the probability of transition between eigenstates of $H_0$ is

$$\mathcal{P}_{i\to f} = F(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega,\, t)\,|V_{fi}|^2 + F(\mathcal{E}_f - \mathcal{E}_i + \hbar\omega,\, t)\,|V^\dagger_{fi}|^2, \tag{18.4.5}$$

where the same function $F$ defined by eq. (18.3.9) appears again. The application of eq. (18.4.5) is slightly different depending on whether the final state is discrete or part of a continuum, although the final results will have a similar appearance.

First, consider the continuum case, for which we write

$$|f\rangle = |\mathcal{E}_f, u_{\mathcal{E}_f}\rangle \tag{18.4.6}$$

for the group of final states with $H_0$ eigenvalue $\mathcal{E}_f$ and degeneracy label $u_{\mathcal{E}_f}$. For large $t$, the same argument that led to eq. (18.3.22) now gives

$$d\mathcal{P}_{i\to f} = t\, dR_{i\to f}, \tag{18.4.7}$$

where the differential contribution to the transition rate is

$$dR_{i\to f} = \frac{2\pi}{\hbar}\Big[\delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\,|V_{fi}|^2 + \delta(\mathcal{E}_f - \mathcal{E}_i + \hbar\omega)\,|V^\dagger_{fi}|^2\Big]\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f}. \tag{18.4.8}$$

The effect of the first term can be referred to as **absorption**; since $\mathcal{E}_f = \mathcal{E}_i + \hbar\omega$, the Hamiltonian perturbation $V$ causes the initial state to gain energy $\hbar\omega$ and jump up to the final state. The second term is **emission**, as the perturbation $V^\dagger$ cause the initial state to lose energy $\hbar\omega$ and jump down to the final state. Since we ignored the cross-terms, and the final state energies must be different in the two cases anyway, it makes sense to separate these, and write

$$dR^{\text{abs}}_{i+\hbar\omega\to f} = \frac{2\pi}{\hbar}|V_{fi}|^2\,\delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f}, \tag{18.4.9}$$

$$dR^{\text{em}}_{i\to f+\hbar\omega} = \frac{2\pi}{\hbar}|V^\dagger_{fi}|^2\,\delta(\mathcal{E}_f - \mathcal{E}_i + \hbar\omega)\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, d\mathcal{E}_f\, du_{\mathcal{E}_f}. \tag{18.4.10}$$

After integration over $\mathcal{E}_f$,

$$\begin{aligned} dR^{\rm abs}_{i+\hbar\omega\to f} &= \frac{2\pi}{\hbar}\,|\langle\mathcal{E}_f, u_{\mathcal{E}_f}|V|i\rangle|^2\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\,du_{\mathcal{E}_f}, && (18.4.11)\\ dR^{\rm em}_{i\to f+\hbar\omega} &= \frac{2\pi}{\hbar}\,|\langle\mathcal{E}_f, u_{\mathcal{E}_f}|V^\dagger|i\rangle|^2\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\,du_{\mathcal{E}_f}, && (18.4.12)\end{aligned}$$

where it is now understood that the energy conservation conditions

$$\begin{aligned} \mathcal{E}_f &= \mathcal{E}_i + \hbar\omega, \qquad &&\text{(absorption)}, && (18.4.13)\\ \mathcal{E}_f &= \mathcal{E}_i - \hbar\omega, \qquad &&\text{(emission)}, && (18.4.14)\end{aligned}$$

are enforced, due to the delta functions. Equations (18.4.11) and (18.4.12), or their equivalents (18.4.9) and (18.4.10), are **Fermi's golden rule for harmonic perturbations**. If the energy degeneracy labels $u_{\mathcal{E}_f}$ are continuous, then they should be integrated over, and if they are discrete then the differential $du_{\mathcal{E}_f}$ is removed and they can be summed over.

If the transition from $|i\rangle$ to $|f\rangle$ by absorbing energy $\hbar\omega$ from $V$ is allowed, then it follows that for the same $\omega$ one can also have the reverse process of $|f\rangle$ to $|i\rangle$ by emitting energy $\hbar\omega$ to $V^\dagger$; both the energy conservation and matrix element selection rules work. For both absorption and emission, we can obtain the differential rate for the opposite process $dR_{f\to i}$ by interchanging $\mathcal{E}_f \leftrightarrow \mathcal{E}_i$ and $|V_{fi}|^2 \leftrightarrow |V^\dagger_{if}|^2$. Note that the relevant matrix elements are simply related by

$$V^\dagger_{if} \;=\; \langle i|V^\dagger|f\rangle \;=\; (\langle f|V|i\rangle)^* = (V_{fi})^*, \qquad (18.4.15)$$

so $|V^\dagger_{if}|^2 = |V_{fi}|^2$. Let us now change the $i$ and $f$ labels to a more neutral notation $a$ and $b$, since each can play the role of initial and final state. If we multiply eq. (18.4.12) with $i = a$ and $f = b$ by $\rho(\mathcal{E}_a, u_{\mathcal{E}_a})du_{\mathcal{E}_a}$, the result is the same as multiplying eq. (18.4.11) with $i = b$ and $f = a$ by $\rho(\mathcal{E}_b, u_{\mathcal{E}_b})du_{\mathcal{E}_b}$, after integrating over the same intervals in the degeneracy labels in each case. This shows that

$$\rho(\mathcal{E}_a, u_{\mathcal{E}_a})\,\frac{dR^{\rm em}_{a\to b+\hbar\omega}}{du_{\mathcal{E}_b}} \;=\; \rho(\mathcal{E}_b, u_{\mathcal{E}_b})\,\frac{dR^{\rm abs}_{b+\hbar\omega\to a}}{du_{\mathcal{E}_a}}. \qquad (18.4.16)$$

This is known as the **detailed balance** relation between emission and absorption rates and the densities of states.

Let us now reconsider eq. (18.4.5), this time assuming that the final state $|f\rangle$ is discrete. We now suppose that the perturbations occur as an incoherent distribution with a range of angular frequencies. By "incoherent", it is implied that interference contributions will be negligible, so that to a good approximation the probabilities, not the amplitudes, add for different driving frequencies. To see how this works, consider instead of eq. (18.4.1), a perturbation that is a superposition of many such terms with infinitesimally spaced $\omega$'s with smoothly varying real amplitudes $c_\omega$ and, crucially, phases $\varphi_\omega$ that are assumed to be random, with no relation (coherence) whatsoever between neighboring values of $\omega$,

$$W(t) \;=\; \sum_\omega c_\omega \left(e^{i\varphi_\omega} V e^{-i\omega t} + e^{-i\varphi_\omega} V^\dagger e^{i\omega t}\right) \qquad \text{(for } t > 0\text{)}. \tag{18.4.17}$$

After isolating the emission and absorption resonant pieces as before, the square of the transition amplitude $a^{(1)}_{i\to f}$ will contains terms of the form

$$\left|a^{(1)}_{i\to f}\right|^2 \;=\; \sum_\omega \sum_{\omega'} c_\omega c_{\omega'} e^{i(\varphi_\omega - \varphi_{\omega'})} \cdots, \tag{18.4.18}$$

where the $\cdots$ represents quantities that are relatively slowly varying functions of $\omega$ and $\omega'$. Because of the random phases and the slow variation of the summand, all terms with $\omega' \neq \omega$ now cancel in the limit of infinitesimal spacing of the sums, leaving behind a single sum of the form $\sum_\omega c_\omega^2 \cdots$, coming from the subset of terms with $\omega' = \omega$, which add with constructive interference. This sum can now be converted to an integral

$$\sum_\omega c_\omega^2 \;\to\; \int d\omega\, \rho(\omega), \tag{18.4.19}$$

described by some density per unit frequency $\rho(\omega)$. This means that the transition probabilities for continuous $\omega$ are assumed[†] to be weighted by a function $\rho(\omega)$ that characterizes the perturbation. For example, if the perturbations take the form of electromagnetic radiation, one can relate $\rho(\omega)$ to the intensity spectrum of the incoherent light, as we will do in section 19.2. In the presence of the integral over continuous $\omega$, even though $\mathcal{E}_f$ is discrete, we can again replace

$$F(\mathcal{E}_f - \mathcal{E}_i \pm \hbar\omega,\, t) \;\to\; \frac{2\pi t}{\hbar}\delta(\mathcal{E}_f - \mathcal{E}_i \pm \hbar\omega), \tag{18.4.20}$$

and so obtain the differential transition rates

$$dR^{\text{abs}}_{i+\hbar\omega\to f} \;=\; \frac{2\pi}{\hbar}|V_{fi}|^2\, \delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\, \rho(\omega) d\omega, \tag{18.4.21}$$

$$dR^{\text{em}}_{i\to f+\hbar\omega} \;=\; \frac{2\pi}{\hbar}|V^\dagger_{fi}|^2\, \delta(\mathcal{E}_f - \mathcal{E}_i + \hbar\omega)\, \rho(\omega) d\omega. \tag{18.4.22}$$

After integrating over $\omega$ according to $\displaystyle\int d\omega\, \delta(\mathcal{E}_f - \mathcal{E}_i \pm \hbar\omega) = 1/\hbar$, we have

$$R^{\text{abs}}_{i+\hbar\omega\to f} \;=\; \frac{2\pi}{\hbar^2}\, |\langle f|V|i\rangle|^2\, \rho(\omega), \tag{18.4.23}$$

$$R^{\text{em}}_{i\to f+\hbar\omega} \;=\; \frac{2\pi}{\hbar^2}\, |\langle f|V^\dagger|i\rangle|^2\, \rho(\omega), \tag{18.4.24}$$

---

[†] It is also possible to take the distribution of perturbations to be coherent, which means that the phases are slowly varying functions of $\omega$. In that case one cannot neglect interference cross-terms in the transition amplitude $a^{(1)}_{i\to f}$ for given initial and final states. We will not consider that case here.

where it is understood now that $\omega$ is fixed in terms of $\mathcal{E}_f$ and $\mathcal{E}_i$ by eqs. (18.4.14) and (18.4.13). The energy degeneracies of the final state $|f\rangle$ are assumed here to be also discrete, and so are not indicated explicitly, but can be summed over.

Although eqs. (18.4.11)–(18.4.12) and (18.4.23)–(18.4.24) look similar, it bears emphasis that they apply in slightly different circumstances. In eqs. (18.4.23)–(18.4.24), the final states are discrete, and the driving angular frequency is selected (by energy conservation) from an incoherent continuum of perturbations to the Hamiltonian with some assumed density function with respect to angular frequency, $\rho(\omega)$. In eqs. (18.4.11)–(18.4.12), it is the final state that is selected by energy conservation from among the continuous final states with density with respect to energy, $\rho(\mathcal{E}_f, u_{\mathcal{E}_f})$, for a driving angular frequency that can be treated as a single monochromatic Hamiltonian perturbation. In the latter case, the results could also be summed over an incoherent continuous distribution. Both of these situations arise in the interaction of matter with electromagnetic waves, depending on whether the final state is a bound or unbound electronic state. This is the subject of the next chapter.

# 19 Absorption and emission of light

## 19.1 Electrons in the presence of electromagnetic waves

In this chapter, we will be discussing the interactions of electrons with a classical electromagnetic wave. The Hamiltonian is obtained from the Coulomb gauge result in eq. (4.3.35), with $q = -e$ and $\gamma = -g_e e/2m_e c \approx -e/m_e c$ for electrons,

$$H = H_0 + \frac{e}{m_e c}\vec{A}\cdot\vec{P} + \frac{e^2}{2m_e c^2}A^2 - e\Phi + \frac{e}{m_e c}\vec{S}\cdot\vec{B}, \tag{19.1.1}$$

where the electron's Hamiltonian in the absence of the electromagnetic wave is

$$H_0 = \frac{P^2}{2m_e} + U(\vec{R}). \tag{19.1.2}$$

We have included here an additional potential $U$, which includes all effects not associated with the wave, for example the Coulomb attraction of the electron to a fixed heavy nucleus, or repulsion by other electrons. Let the wave have unit polarization vector $\hat{\varepsilon}$, and wavevector

$$\vec{k} = \hat{k}\omega/c, \tag{19.1.3}$$

so that the classical electric and magnetic fields are

$$\begin{aligned}\vec{E}(\vec{r},t) &= \hat{\varepsilon}E_0\sin(\vec{k}\cdot\vec{r}-\omega t), && (19.1.4)\\ \vec{B}(\vec{r},t) &= (\hat{k}\times\hat{\varepsilon})E_0\sin(\vec{k}\cdot\vec{r}-\omega t), && (19.1.5)\end{aligned}$$

which satisfy Maxwell's equations provided that $\hat{\varepsilon}\cdot\hat{k} = 0$. In Coulomb gauge, the potentials describing this classical wave are

$$\begin{aligned}\Phi(\vec{r},t) &= 0, && (19.1.6)\\ \vec{A}(\vec{r},t) &= -\hat{\varepsilon}\frac{c}{\omega}E_0\cos(\vec{k}\cdot\vec{r}-\omega t), && (19.1.7)\end{aligned}$$

as one can check using eqs. (4.3.8)–(4.3.9) and (4.3.33).

If the wave is not too intense, it is a good approximation to drop the non-linear term proportional to $A^2$, so that our Hamiltonian simplifies to $H = H_0 + W$, where

$$\begin{aligned}W(t) &= \frac{e}{m_e c}\left(\vec{A}\cdot\vec{P}+\vec{B}\cdot\vec{S}\right) && (19.1.8)\\ &= \frac{eE_0}{m_e\omega}\Big[-\cos(\vec{k}\cdot\vec{R}-\omega t)\,\hat{\varepsilon}\cdot\vec{P}+\sin(\vec{k}\cdot\vec{R}-\omega t)\,(\vec{k}\times\hat{\varepsilon})\cdot\vec{S}\Big]. && (19.1.9)\end{aligned}$$

This can be treated as a perturbation using the formalism of section 18.4, specifically by writing

$$W(t) = Ve^{-i\omega t} + V^\dagger e^{i\omega t}, \tag{19.1.10}$$

as in eq. (18.4.1), with

$$V = -\frac{eE_0}{2m_e\omega} e^{i\vec{k}\cdot\vec{R}} \left[\hat{\varepsilon}\cdot\vec{P} + i(\vec{k}\times\hat{\varepsilon})\cdot\vec{S}\right]. \tag{19.1.11}$$

To put this into a more compact form, we use the identity $(\vec{k}\times\hat{\varepsilon})\cdot\vec{S} = -\hat{\varepsilon}\cdot(\vec{k}\times\vec{S})$ to obtain

$$V = -\frac{eE_0}{2m_e\omega}\,\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}, \tag{19.1.12}$$

where we have defined a vector operator

$$\vec{\mathcal{P}}_{\vec{k}} \equiv e^{i\vec{k}\cdot\vec{R}}\big(\vec{P} - i\vec{k}\times\vec{S}\big). \tag{19.1.13}$$

Note that $\vec{\mathcal{P}}_{\vec{k}}$ has the same units as momentum, but it is not Hermitian. Its matrix elements will be used in our discussion of absorption and emission of electromagnetic waves in the next few sections. The polarization vector $\hat{\varepsilon}$ can now be taken to be complex, for example to describe circularly polarized waves, but must satisfy $\hat{\varepsilon}\cdot\hat{k} = 0$ and $\hat{\varepsilon}^*\cdot\hat{\varepsilon} = 1$.

The Hamiltonian for a system of $N_e$ electrons in an external classical electromagnetic wave can be similarly treated using an appropriate $H_0$ (which will include the repulsive potentials between each pair of electrons, their attractive potentials due to the heavy nucleus, and spin-orbit and spin-spin couplings) together with a harmonic perturbation

$$V = -\frac{eE_0}{2m_e\omega}\sum_{i=1}^{N_e}\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k},i}, \tag{19.1.14}$$

Here, each $\vec{\mathcal{P}}_{\vec{k},i}$ is defined in terms of the position, momentum, and spin operators $\vec{R}_i$, $\vec{P}_i$, and $\vec{S}_i$ for the individual electrons, as in eq. (19.1.13). This is the linearized approximation of a Hamiltonian perturbation describing the interaction with an external electromagnetic wave.

## 19.2 Absorption of electromagnetic waves

Consider an electron described by some Hamiltonian $H_0$, which then encounters a classical electromagnetic wave. As described in the previous section, we can now apply time-dependent perturbation theory as discussed in section 18.4, with

$$V = -\frac{eE_0}{2m_e\omega}\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}} \tag{19.2.1}$$

in eq. (18.4.1), where $E_0$ and $\hat{\varepsilon}$ are the electric field amplitude and unit polarization vector of the wave, and the operator $\vec{\mathcal{P}}_{\vec{k}}$ was defined in eq. (19.1.13) in terms of the wavevector. We will compute results in terms of matrix elements of $\vec{\mathcal{P}}_{\vec{k}}$, and then work out the simpler electric dipole and higher multipole approximations in sections 19.4 and 19.5, using an expansion in the angular frequency $\omega$.

Consider the rate for an initial $H_0$ eigenstate $|i\rangle$ to absorb energy from the wave and end up in a group of final $H_0$ eigenstates

$$|f\rangle \;=\; |\mathcal{E}_f, u_{\mathcal{E}_f}\rangle. \tag{19.2.2}$$

For now, we take these to have continuous eigenvalues $\mathcal{E}_f$ and degeneracy labels $u_{\mathcal{E}_f}$ and density of states $\rho(\mathcal{E}_f, u_{\mathcal{E}_f})$ as defined in eq. (18.3.16). Fermi's golden rule, eq. (18.4.9), then gives the rate for absorption transitions,

$$dR^{\text{abs}}_{i+\hbar\omega\rightarrow f} \;=\; \frac{2\pi}{\hbar}\left(\frac{eE_0}{2m_e\omega}\right)^2 \left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2 \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, du_{\mathcal{E}_f}\; \delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\, d\mathcal{E}_f. \tag{19.2.3}$$

The time-averaged flux of energy in the incident wave that is causing these transitions is

$$\text{energy flux in wave} \;=\; \frac{\text{energy}}{(\text{time})(\text{area})} \;=\; \frac{cE_0^2}{8\pi}. \tag{19.2.4}$$

This can be obtained by time-averaging either $cu_{\text{EM}}$ in eq. (4.3.6), or the magnitude of the Poynting vector $|\vec{S}|$ in eq. (4.3.7). Since both eq. (19.2.3) and eq. (19.2.4) are proportional to $E_0^2$, it make sense to define a measurable quantity proportional to their ratio, in which the incident wave intensity cancels out. The appropriate ratio is called the **absorption cross-section**:

$$\sigma^{\text{abs}} \;=\; \frac{\text{absorbed energy/time}}{\text{energy flux in incident wave}}. \tag{19.2.5}$$

Note that $\sigma^{\text{abs}}$ has units of area. The energy absorbed in each $i \rightarrow f$ transition is $\hbar\omega$, so

$$d\sigma^{\text{abs}}_{i+\hbar\omega\rightarrow f} \;=\; \hbar\omega\, dR^{\text{abs}}_{i+\hbar\omega\rightarrow f}/(cE_0^2/8\pi), \tag{19.2.6}$$

or, in terms of the fine-structure constant $\alpha = e^2/\hbar c$,

$$d\sigma^{\text{abs}}_{i+\hbar\omega\rightarrow f} \;=\; \frac{4\pi^2\alpha\hbar}{m_e^2\omega}\left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2 \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, du_{\mathcal{E}_f}\; \delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\, d\mathcal{E}_f. \tag{19.2.7}$$

If the final state energies $\mathcal{E}_f$ are continuous near $\mathcal{E}_i + \hbar\omega$, one can immediately integrate over them, with the result

$$d\sigma^{\text{abs}}_{i+\hbar\omega\rightarrow f} \;=\; \frac{4\pi^2\alpha\hbar}{m_e^2\omega}\left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2 \rho(\mathcal{E}_f, u_{\mathcal{E}_f})\, du_{\mathcal{E}_f}, \tag{19.2.8}$$

where it is now implicitly required that $\mathcal{E}_f = \mathcal{E}_i + \hbar\omega$. Equation (19.2.8) also applies as an approximation if the relevant final states are discrete but very closely spaced. If some of the degeneracy labels are discrete rather than continuous, then one should remove $du_{\mathcal{E}_f}$ and sum over them rather than integrating.

Now suppose instead that the final states selected by energy conservation have discrete $H_0$ eigenvalues $\mathcal{E}_f$. We will discuss this situation in three different ways, appropriate for different circumstances.

First, suppose that the electromagnetic fields are plane waves with fixed propagation direction $\hat{k}$ and polarization $\hat{\varepsilon}$, but with incoherent phases and a range of angular frequencies wide enough to cover the final state in question. Then, as discussed in section 18.4, the incoherent phases mean that we can sum the probabilities (not the amplitudes) to obtain a transition rate. From using eq. (19.2.1) in eq. (18.4.21),

$$R^{\rm abs}_{i+\hbar\omega\rightarrow f} \;=\; \int_0^\infty d\omega\, \rho(\omega)\, \frac{2\pi}{\hbar} \delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega) \left(\frac{eE_0}{2m_e\omega}\right)^2 \left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2, \tag{19.2.9}$$

where $\rho(\omega)$ is the density per unit angular frequency of incoherent electromagnetic wave perturbations of the form eq. (18.4.1). Let $I$ be the total intensity (power per unit area) in the electromagnetic radiation, so that

$$dI \;=\; \frac{cE_0^2}{8\pi}\rho(\omega) d\omega. \tag{19.2.10}$$

We thus obtain, in terms of the intensity per unit angular frequency, $dI/d\omega$,

$$R^{\rm abs}_{i+\hbar\omega\rightarrow f} \;=\; \int_0^\infty d\omega\, \frac{dI}{d\omega}\, \delta(\mathcal{E}_f - \mathcal{E}_i - \hbar\omega)\, \frac{4\pi^2\alpha}{m_e^2\omega^2} \left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2. \tag{19.2.11}$$

Doing the integration over $\omega$, we find[†] an absorption rate

$$R^{\rm abs}_{i+\hbar\omega\rightarrow f} \;=\; \frac{4\pi^2\alpha}{m_e^2\hbar\omega^2}\,\frac{dI}{d\omega} \left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2, \tag{19.2.12}$$

where $\omega = (\mathcal{E}_f - \mathcal{E}_i)/\hbar$ has now been implicitly fixed.

As a second scenario, suppose that the system is bathed in electromagnetic radiation that is not only phase-incoherent and broad-band in $\omega$, but may be coming from all directions and with all possible polarization vectors. Although we have been discussing the electromagnetic field as a classical wave, in this case we choose to frame our discussion in terms of the numbers of photon quanta with energy $\hbar\omega$ and particular wavenumbers and polarizations. To do so, we take the system to be in a large cubic box of side $L$, with periodic boundary conditions giving allowed wavevectors

$$\vec{k} \;=\; \hat{k}\omega/c \;=\; \frac{2\pi}{L}(\hat{x}n_x + \hat{y}n_y + \hat{z}n_z) \tag{19.2.13}$$

where $(n_x, n_y, n_z)$ are integers. In the large-volume continuum limit, a discrete sum over these wavevectors will correspond to the integration

$$\sum_{\vec{k}} \;\rightarrow\; \left(\frac{L}{2\pi}\right)^3 \int d^3\vec{k} \;=\; \left(\frac{L}{2\pi c}\right)^3 \int_0^\infty d\omega\,\omega^2 \int d\Omega_{\vec{k}}. \tag{19.2.14}$$

---

[†] This could also be obtained directly from eq. (18.4.23).

Denote by $N_{\vec{k},\hat{\varepsilon}}$ the occupation number of photons in a mode with wavevector $\vec{k}$ and polarization vector $\hat{\varepsilon}$. Then the energy associated with that mode is

$$\hbar\omega\, N_{\vec{k},\hat{\varepsilon}} \;=\; (\text{volume})\left(\frac{\text{energy}}{\text{volume}}\right) \;=\; L^3\big(E_0^{\vec{k},\hat{\varepsilon}}\big)^2/8\pi, \tag{19.2.15}$$

where $E_0^{\vec{k},\hat{\varepsilon}}$ is the corresponding classical electric field amplitude.

Going back to eq. (18.4.5), and using the assumption of incoherence to sum the probabilities rather than amplitudes, the total probability for absorption is

$$\mathcal{P}^{\text{abs}}_{i+\hbar\omega\to f} \;=\; \sum_{\vec{k}}\sum_{\hat{\varepsilon}}\left(\frac{eE_0^{\vec{k},\hat{\varepsilon}}}{2m_e\omega}\right)^2 \big|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\big|^2\, F(\mathcal{E}_f-\mathcal{E}_i-\hbar\omega,\, t). \tag{19.2.16}$$

Now, using eqs. (19.2.14) and (19.2.15), this becomes

$$\mathcal{P}^{\text{abs}}_{i+\hbar\omega\to f} \;=\; \int_0^\infty d\omega \int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} \frac{e^2\hbar\omega}{4\pi^2 m_e^2 c^3} N_{\vec{k},\hat{\varepsilon}}\big|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\big|^2\, F(\mathcal{E}_f-\mathcal{E}_i-\hbar\omega,\, t). \tag{19.2.17}$$

Note that the polarizations $\hat{\varepsilon}$ that are being summed over depend on the direction of $\vec{k}$, because of the transversality constraint $\hat{\varepsilon}\cdot\vec{k} = 0$ for electromagnetic waves. As usual, we now apply $F(\mathcal{E}_f-\mathcal{E}_i-\hbar\omega,\, t) = (2\pi t/\hbar)\delta(\mathcal{E}_f-\mathcal{E}_i-\hbar\omega)$, and do the integral over $\omega$ to obtain $\mathcal{P}^{\text{abs}}_{i+\hbar\omega\to f} = t R^{\text{abs}}_{i+\hbar\omega\to f}$, where the absorption rate is

$$R^{\text{abs}}_{i+\hbar\omega\to f} \;=\; \frac{\alpha\omega}{2\pi m_e^2 c^2}\int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} N_{\vec{k},\hat{\varepsilon}}\,\big|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\big|^2, \tag{19.2.18}$$

in terms of the photon occupation numbers $N_{\vec{k},\hat{\varepsilon}}$ for modes with $\omega = c|\vec{k}| = (\mathcal{E}_f-\mathcal{E}_i)/\hbar$. As one might expect, the contribution to the rate from each mode is proportional to the number of photons available in that mode. Comparing eq. (19.2.12) to eq. (19.2.18), we see that

$$\frac{dI}{d\omega} \;\to\; \frac{\hbar\omega^3}{8\pi^3c^2}\int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} N_{\vec{k},\hat{\varepsilon}} \tag{19.2.19}$$

is the effective intensity per unit angular frequency in the bath of photons.

The third way of treating the absorption to a discrete final state is to compute a cross-section, as we have already done for the case of absorption by a continuous group of final states, in eq. (19.2.8). To do so, it is necessary to realize that there will always be at least a small effective broadening of the discrete energy level of the final-state, for two reasons. First, $|f\rangle$ must be able to decay, at least back to the lower-energy initial state $|i\rangle$ if not others, leading to a finite lifetime. The second reason, which is often more important in practice, is that there are environmental effects on the $H_0$ system from perturbations due to nearby atoms with a thermal distribution of velocities, and from fluctuations in the electromagnetic fields.

Both types of line-broadening effects imply that the energy level of the final state is not perfectly sharp, but is described by a Breit–Wigner resonance, with a small finite width that may need to be obtained empirically. Therefore, when using eq. (19.2.7), the density of final states, which naively would be a delta function in the case of discrete $|f\rangle$, should be

$$\rho(\mathcal{E}_f, u_{\mathcal{E}_f}) \;=\; f_{\rm BW}(\mathcal{E}_f - \mathcal{E}_{f0}, \hbar\gamma) \;\equiv\; \frac{\hbar\gamma}{2\pi} \frac{1}{(\mathcal{E}_f - \mathcal{E}_{f0})^2 + \hbar^2\gamma^2/4}. \tag{19.2.20}$$

Here $\mathcal{E}_{f0}$ is the nominal discrete $H_0$ eigenvalue before the line-broadening is taken into account, and $\hbar\gamma$ is the full width at half maximum (FWHM) of the **Breit–Wigner line-shape** (named after Gregory Breit and Wigner, and also known as the **Lorentzian line-shape**) distribution for the energy $\mathcal{E}_f$, which is now continuous. Note that the normalization is

$$\int_{-\infty}^{\infty} d\mathcal{E}_f \, f_{\rm BW}(\mathcal{E}_f - \mathcal{E}_{f0}, \hbar\gamma) \;=\; 1, \tag{19.2.21}$$

so that in the narrow-width approximation,

$$\lim_{\gamma \to 0} f_{\rm BW}(\mathcal{E}_f - \mathcal{E}_{f0}, \, \hbar\gamma) \;=\; \delta(\mathcal{E}_f - \mathcal{E}_{f0}), \tag{19.2.22}$$

as expected. The absorption cross-section for a nominally discrete final state $|f\rangle$ becomes, after using eq. (19.2.20) in eq. (19.2.7) and integrating over $d\mathcal{E}_f$,

$$\sigma^{\rm abs}_{i+\hbar\omega\to f} \;=\; \frac{2\pi\alpha\hbar^2\gamma}{m_e^2\omega} \frac{\left|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\right|^2}{(\mathcal{E}_{f0} - \mathcal{E}_i - \hbar\omega)^2 + \hbar^2\gamma^2/4}. \tag{19.2.23}$$

Here we have removed the $du_{\mathcal{E}_f}$ under the assumption that the degeneracy labels for the final state, such as angular momentum quantum numbers, are also discrete and can be summed over. This cross-section is a strongly peaked function of $\omega$, becoming sharper in the limit $\gamma \to 0$.

## 19.3 Induced and spontaneous emission of light

We now turn to the case of transitions between electronic states involving emission of energy in the form of electromagnetic radiation. The presence of a classical electromagnetic wave with angular frequency $\omega$ will cause an initial $H_0$ eigenstate $|i\rangle$ to fall to a state $|f\rangle$ of lower energy, provided that the $H_0$ eigenvalues satisfy $\mathcal{E}_i - \mathcal{E}_f = \hbar\omega$. If $dI/d\omega$ is the intensity of incoherent electromagnetic waves per unit $\omega$, then we can apply exactly the same reasoning we used to arrive at eq. (19.2.12), except that we use eq. (18.4.22) for emission rather than eq. (18.4.21) for absorption. The result is the rate for **induced emission** (or **stimulated emission**),

$$R^{\rm ind.\ em.}_{i\to f+\hbar\omega} \;=\; \frac{4\pi^2\alpha}{m_e^2\hbar\omega^2}\frac{dI}{d\omega}\left|\langle f|\hat{\varepsilon}^*\cdot\vec{\mathcal{P}}^\dagger_{\vec{k}}\,|i\rangle\right|^2, \tag{19.3.1}$$

where now $\omega = (\mathcal{E}_i - \mathcal{E}_f)/\hbar$ is implicit. This assumes that the wavevector direction $\hat{k}$ is fixed. For the case of light incoming from all directions, the transition rate can be expressed in terms of the photon occupation numbers using eq. (19.2.19), again just as we did for the absorption case in arriving at eq. (19.2.18),

$$R^{\text{ind. em.}}_{i\rightarrow f+\hbar\omega} \;=\; \frac{\alpha\omega}{2\pi m_e^2 c^2} \int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} N_{\vec{k},\hat{\varepsilon}} \left|\langle f|\hat{\varepsilon}^* \cdot \vec{\mathcal{P}}^\dagger_{\vec{k}}\,|i\rangle\right|^2. \qquad (19.3.2)$$

The reason for referring to this as "induced" is that it is proportional to the intensity of incident light, or equivalently to the numbers of photons already present.

Despite the fact that our calculations so far do not show it, the transition from $|i\rangle$ to $|f\rangle$ with the emission of a photon can also occur even if there are no incident classical waves or photons present. That process is called **spontaneous emission**, and we will now show that it yields a result equal to eq. (19.3.2) but with the factor $N_{\vec{k},\hat{\varepsilon}}$ replaced by 1. The phenomenon of spontaneous emission, in which an electronic state decays to a lower energy state by emitting a photon, is the quantum mechanical analog of the classical Larmor radiation of an accelerating charge. However, it can only occur if there is a quantum state with lower energy, while classical Larmor radiation would always occur; compare to the discussion in section 1.1.

To obtain the rate for spontaneous emission in the most direct way, the electromagnetic fields should be treated quantum mechanically, rather than classically as we have done. Recall that our method so far was to treat the potentials $\Phi$ and $\vec{A}$ and their derivatives $\vec{E}$ and $\vec{B}$ as classical quantities, which are then evaluated in terms of the position and momentum operators for the electron that witnesses them. In a more logical and complete approach, both the electronic and radiation degrees of freedom should be treated as a unified quantum system. Then $\Phi$ and $\vec{A}$ will be operators in their own right, rather than just functions of the position operator, and will act on a Hilbert space that describes not just the electronic state, but the photons as well. This allows a direct computation of spontaneous emission (and more complicated phenomena). In this section, instead of pursuing that direct method, we will derive the results for spontaneous emission using an indirect argument, based on inference from requiring the consistency of statistical mechanics, called the **Einstein A and B coefficients** method.

Consider the thought experiment of a cavity kept in thermal equilibrium at temperature $T$ with the electromagnetic radiation inside of it. The walls of the cavity are made of a very large number of atoms that include two states $|a\rangle$ and $|b\rangle$ with energy difference

$$\mathcal{E}_b - \mathcal{E}_a \;=\; \hbar\omega, \qquad (19.3.3)$$

so that the photon absorption process $a + \hbar\omega \rightarrow b$ and the emission process $b \rightarrow a + \hbar\omega$ are both constantly occurring to ensure the equilibrium. According to statistical mechanics, and our discussion in section 3.5, the relative probability for a state with energy $E$ is proportional to the Boltzmann factor $e^{-E/k_BT}$, where $k_B$ is Boltzmann's constant. Therefore, since $\hbar\omega N_{\vec{k},\hat{\varepsilon}}$ is the energy of a state with $N_{\vec{k},\hat{\varepsilon}}$ photons in the mode with wavevector $\vec{k}$ (with $k = \omega/c$) and polarization $\hat{\varepsilon}$, the average number of photons in that mode is

$$\overline{N}_{\vec{k},\hat{\varepsilon}} \;=\; \left(\sum_{n=0}^{\infty} n e^{-n\hbar\omega/k_BT}\right)\Big/\sum_{n=0}^{\infty} e^{-n\hbar\omega/k_BT} \;=\; \frac{1}{e^{\hbar\omega/k_BT}-1}. \tag{19.3.4}$$

In equilibrium, each $N_{\vec{k},\hat{\varepsilon}}$, which we will refer to as $N$ for short in the following, is constant and equal to the mean number given by eq. (19.3.4). Therefore, we must have

$$\left(\frac{dN}{dt}\right)_{\text{spon. em.}} + \left(\frac{dN}{dt}\right)_{\text{ind. em.}} + \left(\frac{dN}{dt}\right)_{\text{absorption}} \;=\; 0, \tag{19.3.5}$$

with contributions from spontaneous emission, induced emission, and absorption rates that are respectively of the forms

$$\left(\frac{dN}{dt}\right)_{\text{spon. em.}} \;=\; n_b A, \tag{19.3.6}$$

$$\left(\frac{dN}{dt}\right)_{\text{ind. em.}} \;=\; N n_b B_{ba}, \tag{19.3.7}$$

$$\left(\frac{dN}{dt}\right)_{\text{absorption}} \;=\; -N n_a B_{ab}. \tag{19.3.8}$$

Here, we have used the fact that the spontaneous and induced emission contributions to the rate for creation of photons must both be proportional to the number of atoms $n_b$ in the state $|b\rangle$, while the absorption rate contribution to the rate for subtraction of photons is proportional to the number of atoms $n_a$ in the state $|a\rangle$. Also, the induced emission and absorption rates are both proportional to the number of photons already present $N$, but by definition the spontaneous emission rate has no such factor. Our goal is to determine the ratios of the coefficients[†] of proportionality, $A$, $B_{ba}$, and $B_{ab}$. These are in turn proportional to the transition rates $R^{\text{spon. em.}}_{b\to a+\hbar\omega}$, $R^{\text{ind. em.}}_{b\to a+\hbar\omega}$, and $R^{\text{abs}}_{a+\hbar\omega\to b}$, respectively, but with the corresponding factors of $N$ extracted from the last two, since we have put them explicitly in eqs. (19.3.7) and (19.3.8).

The numbers $n_a$ and $n_b$ are unknown, but since the system is in thermal equilibrium, we know their ratio, as they are each proportional to the corresponding Boltzmann factor. So,

$$n_a/n_b \;=\; e^{-(\mathcal{E}_a-\mathcal{E}_b)/k_BT} \;=\; e^{\hbar\omega/k_BT}. \tag{19.3.9}$$

Using eqs. (19.3.6)–(19.3.9) in eq. (19.3.5), we can solve for the photon occupation number,

$$N \;=\; \frac{A}{B_{ab}e^{\hbar\omega/k_BT} - B_{ba}}. \tag{19.3.10}$$

---

[†] Einstein's original coefficients were normalized differently.

Now the key point is that $A$, $B_{ab}$, and $B_{ba}$ should depend only on the properties of the states $|a\rangle$ and $|b\rangle$, but not on the temperature. Therefore, the only way to reconcile the functional dependences on $T$ of eqs. (19.3.4) and (19.3.10) is

$$A = B_{ab} = B_{ba}, \tag{19.3.11}$$

and, reassuringly, this indeed does give eq. (19.3.4).

The equality $B_{ab} = B_{ba}$ is just a check of the detailed balance equality between $R^{\text{abs}}_{a+\hbar\omega\to b}$ and $R^{\text{ind. em.}}_{b\to a+\hbar\omega}$, which we had already found by direct computation in eqs. (19.2.18) and (19.3.2). The new requirement that $A = B_{ba}$ says that, because $N$ was already factored out of $B_{ba}$, the rate we calculated for induced emission alone is related to the total emission rate by

$$\begin{aligned} \text{induced} &\;\to\; \text{induced} + \text{spontaneous} \qquad (19.3.12)\\ N_{\vec{k},\hat{\varepsilon}} &\;\to\; N_{\vec{k},\hat{\varepsilon}} + 1. \qquad (19.3.13)\end{aligned}$$

Thus, the rate for spontaneous emission is obtained by simply removing the factor of $N_{\vec{k},\hat{\varepsilon}}$ from the formula eq. (19.3.2) for induced emission,

$$R^{\text{spon. em.}}_{i\to f+\hbar\omega} = \frac{\alpha\omega}{2\pi m_e^2 c^2}\int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} \left|\langle f|\hat{\varepsilon}^* \cdot \vec{\mathcal{P}}^\dagger_{\vec{k}}\,|i\rangle\right|^2. \tag{19.3.14}$$

A more direct calculation that treats the electromagnetic fields quantum mechanically gives the same result.

## 19.4 Electric dipole approximation

The case in which the wavelength of the electromagnetic wave is much larger than the spatial extent of the initial-state wavefunction is called the **electric dipole approximation**. To see why this might apply as a valid approximation, suppose that the electron is in a hydrogen-like atomic orbital state with principal quantum number $n$ and atomic number $Z$. Then the minimum photon energy needed to ionize the state obeys $\hbar\omega = Z^2e^2/2a_0n^2$ from eq. (10.1.16), while the characteristic spatial extent of the initial wavefunction is $\langle R\rangle = 3n^2a_0/2Z$, using eq. (10.1.65) with $l = 0$. For larger $l$ in the initial state, the characteristic size $\langle R\rangle$ is even smaller (for fixed $n$), according to eq. (10.1.65). To compare the atomic size scale to the inverse wavenumber of the light (with minimum ionizing energy),

$$k\langle R\rangle = \frac{\omega\langle R\rangle}{c} \sim \frac{3\alpha Z}{4}. \tag{19.4.1}$$

This is smaller than 1, even when $Z$ is large, since $\alpha = e^2/\hbar c \approx 1/137$. For transitions between bound states, the relevant wavelength is even longer because the change in energy is smaller.

As an extreme case, the wavelength is macroscopic for transitions between states with the same principal quantum number, which have very nearly the same energy. This means that in transitions between bound states we can make the long-wavelength limit approximation

$$e^{i\vec{k}\cdot\vec{R}} \approx 1 \tag{19.4.2}$$

over the relevant spatial extent of the atom. In the same limit, one can can also neglect the $\vec{k} \times \vec{S}$ contribution compared to the momentum operator $\vec{P}$ in $\vec{\mathcal{P}}_{\vec{k}}$ of eq. (19.1.13). This is made plausible by the rough estimate

$$\frac{|\vec{k} \times \vec{S}|}{|\vec{P}|} \sim \frac{(\omega/c)\hbar}{\hbar/\langle R\rangle} \tag{19.4.3}$$

(using the uncertainty relation for position and momentum in the denominator), which is parametrically the same as eq. (19.4.1), and so again suppressed by a factor $\alpha Z$.

Therefore, the electric dipole (long-wavelength) approximation is defined by taking

$$\langle f|\vec{\mathcal{P}}_{\vec{k}}|i\rangle \approx \langle f|\vec{P}|i\rangle \tag{19.4.4}$$

in the absorption and emission rates. Now we can use a trick, by noticing that if the only momentum dependence of $H_0$ is in the kinetic term, then

$$\big[H_0, \vec{R}\big] = \frac{1}{2m_e}\big[P^2, \vec{R}\big] = -i\frac{\hbar}{m_e}\vec{P}, \tag{19.4.5}$$

so that in the case of absorption,

$$\langle f|\hat{\varepsilon}\cdot\vec{P}|i\rangle = \frac{im_e}{\hbar}\hat{\varepsilon}\cdot\langle f|[H_0,\vec{R}]|i\rangle = \frac{im_e}{\hbar}(\mathcal{E}_f - \mathcal{E}_i)\,\langle f|\hat{\varepsilon}\cdot\vec{R}\,|i\rangle = im_e\omega\langle f|\hat{\varepsilon}\cdot\vec{R}\,|i\rangle. \tag{19.4.6}$$

For the emission case, the only difference is that $\hat{\varepsilon} \to \hat{\varepsilon}^*$ and $\omega \to -\omega$, so

$$\langle f|\hat{\varepsilon}^*\cdot\vec{P}|i\rangle = -im_e\omega\langle f|\hat{\varepsilon}^*\cdot\vec{R}\,|i\rangle. \tag{19.4.7}$$

The sign change makes no difference, and the complex conjugation of the polarization vector has no effect for linear polarizations, but makes a difference for circular polarization vectors [such as $\hat{\varepsilon} = (\hat{x} \pm i\hat{y})/\sqrt{2}$ if $\hat{k} = \hat{z}$].

For example, using eq. (19.4.6) in eq. (19.2.8), we have for the absorption cross-section in the case of continuous final state energies

$$d\sigma_{i\to f}^{\text{abs, dipole}} = 4\pi^2\alpha\hbar\omega\,\big|\langle f|\hat{\varepsilon}\cdot\vec{R}\,|i\rangle\big|^2\,\rho(\mathcal{E}_f, u_{\mathcal{E}_f})\,du_{\mathcal{E}_f}. \tag{19.4.8}$$

Recalling that $\hat{\varepsilon}$ is the direction of the polarization of the absorbed electromagnetic wave, we see that the relevant matrix element in the electric dipole approximation is proportional to the position operator along the polarization direction. More generally, the electric dipole approximation amounts to making the substitutions

$$|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle|^2 \;\to\; m_e^2\omega^2\,|\langle f|\hat{\varepsilon}\cdot\vec{R}\,|i\rangle|^2 \qquad \text{(electric dipole absorption)} \tag{19.4.9}$$

in the absorption rates for discrete final states with incoherent plane-wave light in eq. (19.2.12), incoherent light from all directions in (19.2.18), and in the cross-section for a line-broadened discrete final state in eq. (19.2.23). Similarly, one makes the replacement

$$|\langle f|\hat{\varepsilon}^* \cdot \vec{\mathcal{P}}^\dagger_{\vec{k}}\,|i\rangle|^2 \;\;\rightarrow\;\; m_e^2\omega^2\,|\langle f|\,\hat{\varepsilon}^* \cdot \vec{R}\,|i\rangle|^2 \qquad \text{(electric dipole emission)} \qquad (19.4.10)$$

in the induced emission rates eq. (19.3.1) for plane waves with a fixed direction of propagation and eq.( 19.3.2) for light from all directions in terms of the photon occupation numbers, and in eq. (19.3.14) for spontaneous emission.

In the case of spontaneous emission in eq. (19.3.14), the matrix element $\langle f|\hat{\varepsilon}^* \cdot \vec{R}|i\rangle$ in the electric dipole approximation does not depend directly on the wavevector magnitude, but does depend indirectly on the wavevector direction because of the requirement $\hat{\varepsilon}^* \cdot \hat{k} = 0$. Temporarily fixing $\hat{k} = \hat{z}$, the two independent polarization directions to be summed over can be taken along $\hat{x}$ and $\hat{y}$, so that

$$\sum_{\hat{\varepsilon}} \big|\langle f|\hat{\varepsilon}^* \cdot \vec{R}\,|i\rangle\big|^2 \;=\; \big|\langle f|X|i\rangle\big|^2 + \big|\langle f|Y|i\rangle\big|^2. \qquad (19.4.11)$$

Then, integrating over all possible directions for $\hat{k}$ will effectively give an average of the two terms over three possible directions, so

$$\int d\Omega_{\vec{k}} \sum_{\hat{\varepsilon}} \big|\langle f|\hat{\varepsilon}^* \cdot \vec{R}\,|i\rangle\big|^2 \;=\; (4\pi)\frac{2}{3}\Big(\big|\langle f|X|i\rangle\big|^2 + \big|\langle f|Y|i\rangle\big|^2 + \big|\langle f|Z|i\rangle\big|^2\Big) \qquad (19.4.12)$$

$$= \;\frac{8\pi}{3}\langle f|\vec{R}\,|i\rangle \cdot \langle i|\vec{R}\,|f\rangle. \qquad (19.4.13)$$

This reduces eq. (19.3.14) to the simple formula

$$R^{\text{spon. em.}}_{i\rightarrow f+\hbar\omega} \;=\; \frac{4\alpha\omega^3}{3c^2}\,\big|\langle f|\vec{R}\,|i\rangle\big|^2 \qquad (19.4.14)$$

for the rate for an initial state $|i\rangle$ to decay to a final state $|f\rangle$ by the spontaneous emission of a photon. Note that the right side involves

$$\big|\langle f|\vec{R}\,|i\rangle\big|^2 \;\equiv\; \langle f|\vec{R}\,|i\rangle \cdot \langle i|\vec{R}\,|f\rangle, \qquad (19.4.15)$$

the complex square of the magnitude of a vector matrix element.

Suppose that there is no background electromagnetic radiation to induce emission or absorption, and let the number of atoms in the state $|i\rangle$ be $N_i(t)$ at time $t$. Then at time $t + \Delta t$,

$$N_i(t + \Delta t) \;=\; N_i(t) - \Delta t\, R_i N_i(t), \qquad (19.4.16)$$

where the total decay rate of the initial state is

$$R_i \;=\; \sum_f R^{\text{spon. em.}}_{i\rightarrow f+\hbar\omega}. \qquad (19.4.17)$$

Then

$$\frac{N_i(t + \Delta t) - N_i(t)}{\Delta t} = -R_i N_i(t), \tag{19.4.18}$$

which in the limit $\Delta t \to 0$ becomes the differential equation

$$\frac{dN_i}{dt} = -R_i N_i, \tag{19.4.19}$$

with the solution

$$N_i(t) = N_i(t_0) e^{-(t-t_0)R_i}. \tag{19.4.20}$$

Therefore $\tau_i = 1/R_i$ is the lifetime of the state $|i\rangle$, defined as the amount of time needed for the number of atoms in that state to decrease by a factor of $1/e \approx 0.367879$ in the absence of other perturbations. The half-life, defined as the amount of time required for the number to decrease by half, is related to this by $t_{1/2} = \tau_i \ln 2 = 0.693147/R_i$.

The preceding results imply matrix element selection rules that govern both absorption and emission in the electric dipole approximation when $|i\rangle$ and $|f\rangle$ are angular momentum eigenstates. These follow immediately from eqs. (12.2.15)–(12.2.18). First, eq. (12.2.15) tells us that the change in the the orbital angular momentum quantum number $l$ must be exactly 1 unit, $\Delta l = l_f - l_i = \pm 1$. As a corollary, the parities of the initial and final states must be opposite, since the parity of a state with orbital angular momentum $l$ is $(-1)^l$ from eq. (8.7.5); this can also be seen directly from the parity selection rule, since the operator $\vec{R}$ has odd parity. Also, eqs. (12.2.16)–(12.2.18) tell us that the $z$-component of the orbital angular momentum changes by 1 unit or less: $\Delta m_l = m_{l,f} - m_{l,i} = 0, \pm 1$. As special cases, if the polarization of the wave is along $\hat{\varepsilon} = \hat{z}$, then eq. (12.2.16) says $\Delta m_l = 0$, and if it is in the $x, y$ plane, then eqs. (12.2.17)–(12.2.18) demand $\Delta m_l = \pm 1$. The spin is not affected by the position operator, so that $m_{s,f} = m_{s,i}$.

We therefore have the **electric dipole selection rules** for single-electron transitions,

$$\Delta l = \pm 1, \quad \Delta m_l = 0, \pm 1 \quad \Delta m_s = 0, \qquad \text{(electric dipole)}. \tag{19.4.21}$$

These can be generalized to the case of multi-electron transitions (where more than one electron changes orbitals in going from the initial to the final state), for which the electric dipole selection rules are

$$\pi_i \pi_f = -1, \quad \Delta J = 0, \pm 1, \quad \Delta m_J = 0, \pm 1, \quad J_i + J_f > 0, \quad \text{(electric dipole)}. \tag{19.4.22}$$

The first three of these follow from requiring the non-vanishing of the matrix element of the parity-odd electric dipole operator $\hat{\varepsilon} \cdot \vec{R}$. The $\Delta J$ rule follows from the fact that $\vec{R}$ is a vector operator, and therefore carries total angular momentum 1 in the sense explained in Chapter 12. In terms of addition of angular momentum, $J \otimes 1 = (J-1) \oplus J \oplus (J+1)$ for $J \geq 1$.

The last selection rule in eq. (19.4.22) is completely general, even if we do not make the electric dipole approximation. It says just this: a total angular momentum $J = 0$ state can never go to another $J = 0$ state by absorbing or emitting a single photon. The reason is that a single photon can always be written as a linear combination of transverse circular polarizations, carrying intrinsic angular momentum component along the propagation direction with eigenvalue either $+\hbar$ or $-\hbar$, which cannot be matched by the initial and final electronic states if they both have no angular momentum. Therefore, we have a general rule

$$J_i + J_f > 0, \qquad \text{(all single-photon transitions).} \tag{19.4.23}$$

However, this rule (and all other selection rules listed here) can be violated if one considers emission or absorption of multiple photons, either as distinct sequential events involving intermediate states, or by going beyond first order in time-dependent perturbation theory or through perturbations caused by the environment of the atom. Multi-photon transition rates at higher order in perturbation theory are suppressed by additional powers of the fine-structure constant, however, and so can often be neglected when they compete with single-photon transitions.

In the approximation of the $LS$ coupling scheme, the total spin, total orbital, and total angular momentum orthobasis quantum numbers $L, S, J, m_J$ are good quantum numbers for the initial and final stationary states. Then, because the total spin is not involved at all in the electric dipole moment operator, one has the additional rules

$$\Delta S = 0, \quad \Delta L = 0, \pm 1, \qquad L_i + L_f > 0, \qquad \text{(electric dipole, } LS \text{ coupling)} \tag{19.4.24}$$

Recall, from the discussion in section 16.4, that the $LS$ coupling scheme holds to a good approximation for atoms with atomic number $Z$ not too large.

## 19.5 Magnetic dipole, electric quadrupole, and higher orders

Transitions that violate the electric dipole selection rules of eqs. (19.4.21) and (19.4.22) have rates that are typically suppressed by extra powers of $Z\alpha$, or by extra inverse powers of the wavelength of the photon emitted or absorbed. For this reason, they are often referred to as "forbidden". Traditionally, this terminology does not mean that they are literally forbidden, but rather that they take place at highly suppressed rates. By the same terminology, transitions that take place through the electric dipole operator are often called "allowed".

To go beyond the electric dipole approximation and its selection rules for absorption and emission of electromagnetic radiation, consider the expansion in small $k$ of eq. (19.1.13). To make things a little simpler, let us temporarily choose the wavevector of the light as $\vec{k} = k\hat{z}$ and the polarization as $\hat{\varepsilon} = \hat{x}$, so

$$\hat{\varepsilon} \cdot \vec{\mathcal{P}}_{\vec{k}} \;=\; e^{i\vec{k}\cdot\vec{R}}\hat{\varepsilon}\cdot(\vec{P} - i\vec{k}\times\vec{S}) \;=\; e^{ikZ}(P_x + ikS_y). \tag{19.5.1}$$

Now, expanding $e^{ikZ} \approx 1 + ikZ + \cdots$, and using $ZP_x = (ZP_x - XP_z)/2 + (ZP_x + XP_z)/2 = L_y/2 + (ZP_x + P_zX)/2$, we have

$$\hat{\varepsilon} \cdot \vec{\mathcal{P}}_{\vec{k}} \;=\; P_x + i\frac{\omega}{2c}\left[L_y + 2S_y + (ZP_x + P_zX)\right] + \cdots . \tag{19.5.2}$$

Recasting this back into a form valid for general orthogonal $\hat{k}$ and $\hat{\varepsilon}$ gives

$$\hat{\varepsilon} \cdot \vec{\mathcal{P}}_{\vec{k}} \;=\; \hat{\varepsilon}\cdot\vec{P} + i\frac{\omega}{2c}\Big[(\hat{k}\times\hat{\varepsilon})\cdot(\vec{L} + 2\vec{S}) + (\hat{k}\cdot\vec{R}\,\hat{\varepsilon}\cdot\vec{P} + \hat{k}\cdot\vec{P}\,\hat{\varepsilon}\cdot\vec{R})\Big] + \cdots , \tag{19.5.3}$$

which is the result to next-to-leading order in the long-wavelength expansion. In the last term, the ordering of operators $\vec{R}$ and $\vec{P}$ does not matter; they commute because of the transversality condition $\hat{\varepsilon}\cdot\hat{k} = 0$

Therefore, for the harmonic perturbation to be used in eq. (18.4.1) and subsequent equations, we can write

$$V \;=\; V_{\text{E1}} + V_{\text{M1}} + V_{\text{E2}} + \cdots , \tag{19.5.4}$$

where the separated contributions,

$$V_{\text{E1}} \;=\; -\frac{eE_0}{2m_e\omega}\hat{\varepsilon}\cdot\vec{P}, \tag{19.5.5}$$

$$V_{\text{M1}} \;=\; -i\frac{eE_0}{4m_ec}(\hat{k}\times\hat{\varepsilon})\cdot(\vec{L} + 2\vec{S}), \tag{19.5.6}$$

$$V_{\text{E2}} \;=\; -i\frac{eE_0}{4m_ec}(\hat{k}\cdot\vec{R}\,\hat{\varepsilon}\cdot\vec{P} + \hat{k}\cdot\vec{P}\,\hat{\varepsilon}\cdot\vec{R}) \tag{19.5.7}$$

are called electric dipole (E1), magnetic dipole (M1), and electric quadrupole (E2), respectively. Note that in the magnetic dipole operator, the vector $(\hat{k}\times\hat{\varepsilon})$ is the unit vector along the magnetic field direction in the wave; the magnetic field couples to the combination $\vec{L} + 2\vec{S}$.

Single-photon transitions that are dominantly magnetic dipole in character can now be evaluated from the general formulas in sections 19.2-19.3, by using

$$\big|\langle f|\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}|i\rangle\big|^2 \;\to\; \frac{\omega^2}{4c^2}\big|\langle f|(\hat{k}\times\hat{\varepsilon})\cdot(\vec{L} + 2\vec{S})|i\rangle\big|^2 \qquad \text{(M1 absorption)}, \tag{19.5.8}$$

$$\big|\langle f|\hat{\varepsilon}^*\cdot\vec{\mathcal{P}}^\dagger_{\vec{k}}|i\rangle\big|^2 \;\to\; \frac{\omega^2}{4c^2}\big|\langle f|(\hat{k}\times\hat{\varepsilon}^*)\cdot(\vec{L} + 2\vec{S})|i\rangle\big|^2 \qquad \text{(M1 emission)}. \tag{19.5.9}$$

For the electric quadrupole term, we can do a trick similar to the one we used in the electric dipole case to get eqs. (19.4.9) and (19.4.10). Assuming that $H_0$ does not depend on the momentum operator except through the kinetic term $P^2/2m_e$, we find from eq. (19.4.5) that

$$\big[H_0,\, \hat{k}\cdot\vec{R}\,\hat{\varepsilon}\cdot\vec{R}\big] \;=\; -i\frac{\hbar}{m_e}\left(\hat{k}\cdot\vec{R}\,\hat{\varepsilon}\cdot\vec{P} + \hat{k}\cdot\vec{P}\,\hat{\varepsilon}\cdot\vec{R}\right), \tag{19.5.10}$$

where we have made use of $\hat{\varepsilon} \cdot \hat{k} = 0$. Therefore,

$$V_{\rm E2} \;=\; \frac{eE_0}{4\hbar c}\big[H_0,\, \hat{k}\cdot\vec{R}\,\hat{\varepsilon}\cdot\vec{R}\big], \tag{19.5.11}$$

and in the matrix elements, one can evaluate $H_0$ on the adjacent ket $|i\rangle$ or bra $|f\rangle$, giving a factor of $\mathcal{E}_f - \mathcal{E}_i = \hbar\omega$ for absorption and $-\hbar\omega$ for emission. It follows that in matrix elements for single-photon absorption, one can make the replacement

$$V_{\rm E2} \;\to\; \frac{eE_0\omega}{4c}\,\hat{k}\cdot\overleftrightarrow{\mathcal{Q}}\cdot\hat{\varepsilon}, \tag{19.5.12}$$

where[†] the symmetric, traceless rank-2 tensor quadrupole moment operator is

$$\mathcal{Q}_{ab} = R_a R_b - \delta_{ab} R^2/3, \tag{19.5.13}$$

for $a, b = x, y, z$. This uses the fact that the $\delta_{ab}$ part of $\mathcal{Q}_{ab}$ does not contribute to $V_{\rm E2}$ because of $\hat{\varepsilon} \cdot \hat{k} = 0$. Within matrix elements for single-photon emission,

$$V^\dagger_{\rm E2} \;\to\; -\frac{eE_0\omega}{4c}\,\hat{k}\cdot\overleftrightarrow{\mathcal{Q}}\cdot\hat{\varepsilon}^* \tag{19.5.14}$$

similarly applies. Comparing eqs. (19.5.12) and (19.5.14) to eq. (19.2.1), we see that for dominantly electric quadrupole transitions, the general formulas in sections 19.2-19.3 apply with

$$\big|\langle f|\,\hat{\varepsilon}\cdot\vec{\mathcal{P}}_{\vec{k}}\,|i\rangle\big|^2 \;\to\; \frac{m_e^2\omega^4}{4c^2}\big|\langle f|\,\hat{k}\cdot\overleftrightarrow{\mathcal{Q}}\cdot\hat{\varepsilon}\,|i\rangle\big|^2 \qquad \text{(E2 absorption)}, \tag{19.5.15}$$

$$\big|\langle f|\,\hat{\varepsilon}^*\cdot\vec{\mathcal{P}}^\dagger_{\vec{k}}\,|i\rangle\big|^2 \;\to\; \frac{m_e^2\omega^4}{4c^2}\big|\langle f|\,\hat{k}\cdot\overleftrightarrow{\mathcal{Q}}\cdot\hat{\varepsilon}^*\,|i\rangle\big|^2 \qquad \text{(E2 emission)}, \tag{19.5.16}$$

for single-photon absorption and emission, respectively.

We now turn to the selection rules for M1 and E2 transitions. First, let us use eqs. (19.5.8) and (19.5.9) to find the requirements on the initial and final states in order for a magnetic dipole transition to occur. In transitions involving a single electron, neither $\vec{L}$ nor $\vec{S}$ changes the orbital quantum number $l$, and the spin quantum number $s$ is of course fixed at 1/2. However, the components of $\vec{L}$ are linear combinations of $L_z, L_\pm$, and the components of $\vec{S}$ are linear combinations of $S_z, S_\pm$, so they can change $m_l$ and $m_s$ by zero or one unit. Furthermore, magnetic dipole transitions cannot occur between states with different principal (radial) quantum numbers, because $\vec{L}$ and $\vec{S}$ do not operate on the radial wavefunctions and do not change $l$, so the matrix element calculation in the position representation will be proportional to

$$\int_0^\infty dr\, r^2\, R_{nl}(r) R_{n'l}(r) \;=\; \delta_{n,n'}. \tag{19.5.17}$$

---

[†] For two vectors $\vec{v}$ and $\vec{w}$ with components $v_a$ and $w_a$, and a rank-2 tensor $\overleftrightarrow{T}$ with components $T_{ab}$, the notation $\vec{v}\cdot\overleftrightarrow{T}\cdot\vec{w}$ means $\sum_{a=x,y,z}\sum_{b=x,y,z} v_a T_{ab} w_b$. We have already encountered the quadrupole moment operator $\mathcal{Q}_{ab}$ in eqs. (12.1.32) and (12.1.33).

Therefore, the **magnetic dipole selection rules** are, for single-electron transitions,

$$\Delta n = 0, \qquad \Delta l = 0, \qquad \Delta m_l = 0, \pm 1, \qquad \Delta m_s = 0, \pm 1, \qquad \text{(magnetic dipole)}. \qquad (19.5.18)$$

For multi-electron transitions, the selection rules from eqs. (19.5.8) and (19.5.9) are

$$\pi_i \pi_f = +1, \quad \Delta J = 0, \pm 1, \quad \Delta m_J = 0, \pm 1, \quad J_i + J_f > 0, \qquad \text{(magnetic dipole)}. \qquad (19.5.19)$$

The parity selection rule follows from the fact that the operator $\vec{L} + 2\vec{S}$ has even parity, and the $\Delta J$ and $\Delta m_J$ selection rules follow from the fact that it is a vector, and therefore as explained in Chapter 12 it carries total angular momentum 1. The rule $J_i + J_f > 0$ is completely general, as we already noted in our discussion of the electric dipole case. In the $LS$ coupling scheme approximation, where initial and final stationary states can be assigned total angular momentum basis quantum numbers $L, S, J, m_J$, there are additional rules

$$\Delta L = 0, \qquad \Delta S = 0 \qquad \text{(magnetic dipole, } LS \text{ coupling)}. \qquad (19.5.20)$$

Note that the electric dipole and magnetic dipole contributions cannot both be relevant for a given initial and final state. Magnetic dipole transitions never change the parity, while electric dipole transitions always do.

The famous 21 centimeter hyperfine transition line within the ground state of hydrogen is a magnetic dipole transition, as it involved a spin flip but does not change the orbital angular momentum or principal quantum number, which are $l = 0$ and $n = 1$ in both the initial and final state. Since the resulting photon has an extremely long wavelength compared to the atomic size, the hyperfine transition occurs at a highly suppressed rate (compared to what one might expect if it were an electric dipole transition), because $k = \omega/c \ll 1/a_0$.

Now consider the matrix element selection rules that apply to the electric quadrupole term. Depending on the choices for $\hat{k}$ and $\hat{\varepsilon}$, the matrix elements in eqs. (19.5.15) and (19.5.16) can involve linear combinations of the operators $XY$, $XZ$, $YZ$, $X^2 - Y^2$, and $X^2 + Y^2 - 2Z^2$, which are the components of the quadrupole tensor $\overleftrightarrow{Q}$. These operators all have parity $\pi_Q = +1$, so electric quadrupole transitions can only connect states with the same parity. Acting on single-electron orbital angular momentum eigenstates, these operators can change $l$ by 0 or 2 units (but not 1 unit, as that would not satisfy the parity selection rule), and $m_l$ by 0, 1, or 2 units. Unlike the magnetic dipole interaction, the electric quadrupole operator does act non-trivially on the radial wavefunction, so there is no restriction on changes in the principal quantum number. However, it does not change the spin state. So, we have the **electric quadrupole selection rules** for single-electron transitions,

$$\Delta l = 0, \pm 2, \quad \Delta m_l = 0, \pm 1, \pm 2, \quad \Delta m_s = 0, \quad \text{(electric quadrupole)}. \qquad (19.5.21)$$

For transitions involving multiple electrons, the selection rules are

$$\pi_i\pi_f = +1, \qquad |\Delta J| \leq 2, \qquad |\Delta m_J| \leq 2, \qquad J_i + J_f > 1, \quad \text{(electric quadrupole)}, \qquad (19.5.22)$$

and in the $LS$ coupling scheme approximation in which the initial and final multi-electron stationary states have good quantum numbers $L, S, J, m_J$, there are additional selection rules

$$\Delta S = 0, \quad |\Delta L| \leq 2, \quad L_i + L_f > 1 \quad \text{(electric quadrupole, } LS \text{ coupling)}. \qquad (19.5.23)$$

The general constraint that angular momentum can change by at most two units in electric quadrupole transitions follows from the Wigner–Eckart Theorem 12.3.1 and the fact that the quadrupole operator is an irreducible tensor of order $j = 2$, as we noted in section 12.1. The electric quadrupole transitions can only occur between states that have the same parity, so they can never occur if electric dipole transitions are allowed. If the initial and final states in question have $\Delta l = 0$ or $\Delta L = 0$, then the transition can be both magnetic dipole and electric quadrupole in character, but if $\Delta L = 2$ or $\Delta J = 2$ or it changes $m_l$ or $m_J$ by two units, then it will be electric quadrupole at leading order in the long-wavelength approximation.

If one continues the expansion in $\vec{k} \cdot \vec{R}$ of $\vec{\mathcal{P}}_{\vec{k}}$, the resulting electric and magnetic multipole contributions are called E$q$ and M$q$ with $q = 1, 2, 3, \ldots$, where M2 is magnetic quadrupole, E3 is electric octopole, etc. The general selection rules for these contributions are

$$\pi_i\pi_f = (-1)^q, \qquad \text{(E}q\text{)}, \qquad (19.5.24)$$

$$\pi_i\pi_f = (-1)^{q-1}, \qquad \text{(M}q\text{)}, \qquad (19.5.25)$$

$$|\Delta J| \leq q \leq J_i + J_f, \qquad |\Delta m_J| \leq q, \qquad \text{(E}q\text{) and (M}q\text{)}. \qquad (19.5.26)$$

In the long wavelength limit, the leading E$q$ or M$q$ contributions for a transition between two states are those with the smallest $q$ that satisfies these selection rules. For a possible transition between an odd-parity state with $J = 2$ and an even-parity state with $J = 0$, the single-photon transition would be uniquely determined to be M2, magnetic quadrupole. For two states with the same parity and both with $J = 1$, the transition could be either M1 or E2. For two states both with $J = 0$, there is no allowed single-photon transition as we have already discussed, and there will instead be a highly suppressed double-photon emission.

## 19.6 Photo-electric effect for atoms

In this section, we will compute the photo-electric effect cross-section for atoms, as an application of the results obtained in section 19.2. Instead of considering the photo-electric effect for a metal surface, we will take an initial hydrogen-like atomic bound state $|i\rangle$ for an electron. The light to be absorbed has angular frequency $\omega$, polarization $\hat{\varepsilon}$, and wavenumber $\vec{k} = \hat{k}\omega/c$. The final state $|f\rangle$ is taken to be a continuum momentum eigenstate with momentum eigenvalue $\vec{p}_f = \hbar\vec{K}$. This final state is an approximation, as we are neglecting the influence on its wavefunction of the nuclear electric field; this is to make the problem more tractable, and works best in the limit that the final-state electron energy $\hbar^2K^2/2m_e$ is large compared to the binding energy of the initial state. Since we are not assuming that the wavelength of the electromagnetic wave is large compared to the de Broglie wavelength $2\pi/K$ of the final state, we will not use the multipole expansion of the previous two sections in the following derivation.

To be specific, let the initial state of the electron be a hydrogen-like atomic ground-state wavefunction

$$\langle\vec{r}|i\rangle \;=\; \psi_{1,0,0}(\vec{r}) \;=\; \sqrt{\frac{Z^3}{\pi a_0^3}}\, e^{-Zr/a_0}, \tag{19.6.1}$$

which also applies approximately for $1s$ electrons in multi-electron atoms, with

$$\mathcal{E}_i \;=\; -Z^2e^2/2a_0. \tag{19.6.2}$$

The final state electronic wavefunction is approximated by

$$\langle\vec{r}|f\rangle \;=\; \psi_{\vec{K}}(\vec{r}) \;=\; Ae^{i\vec{K}\cdot\vec{r}}, \tag{19.6.3}$$

where $A$ is a normalization constant and the momentum and energy are

$$\vec{p}_f = \hbar\vec{K}, \quad \text{and} \quad \mathcal{E}_f = \frac{\hbar^2K^2}{2m_e}. \tag{19.6.4}$$

The delta function in eq. (19.2.7) then reveals that the photon energy is

$$\hbar\omega \;=\; \hbar ck \;=\; \frac{\hbar^2K^2}{2m_e} + \frac{Z^2e^2}{2a_0}. \tag{19.6.5}$$

This shows that there is a minimum possible value of $\hbar\omega_{\text{min}} = Z^2e^2/2a_0$ in order for the process to take place, achieved when the kinetic energy of the final-state electron is small. This is in agreement with the observation that the photo-electric effect requires a minimum incident photon energy quantum $\hbar\omega$, regardless of the intensity of the light, as noted in section 1.3.

However, our calculation will work best in the opposite limit of large $K$ compared to the binding energy, so that the plane-wave approximation of eq. (19.6.3) is more nearly valid. In that limit, we have $\hbar^2K^2/2m_e \gg Z^2e^2/2a_0$, which can be rewritten as $K \gg Z/a_0$ by using $a_0 = \hbar^2/m_ee^2$. This also implies $\hbar\omega \gg Z^2e^2/2a_0$, so $\omega/c \gg Z^2e^2/2a_0\hbar c$, or $\omega/c \gg Z^2\alpha/2a_0$. The velocity of the ejected electron is then

$$v_f = \frac{p_f}{m_e} = \frac{\hbar K}{m_e} \approx \frac{2\omega}{K}, \tag{19.6.6}$$

where we used eq. (19.6.5) with the approximation that the binding energy is small compared to the photon energy. Therefore, since we will be treating the electron as non-relativistic, we need $\omega \ll cK$ so that $v_f \ll c$. So, the following results will be most reliable when

$$K \gg \omega/c \gg Z^2\alpha/2a_0, \qquad \text{and} \qquad K \gg Z/a_0. \tag{19.6.7}$$

The first condition allows us to neglect the spin contribution $\vec{k}\times\vec{S} = (\omega/c)\hat{k}\times\vec{S}$ in eq. (19.1.13) compared to the $\vec{P}$ term.

In order to have a well-defined normalization factor $A$ in eq. (19.6.3), the final state is taken to be confined to a very large cubic box with side $L \gg a_0/Z$; as a check, we will show that the dependence on $L$ cancels in the final result. Then the normalization condition is

$$1 = \int_{\text{box}} d^3\vec{r}\; |\psi_{\vec{K}}(\vec{r})|^2 \;=\; |A|^2 \int_{\text{box}} d^3\vec{r} \;=\; |A|^2 L^3, \tag{19.6.8}$$

so $A = L^{-3/2}$. Imposing periodic boundary conditions on the wavefunction implies that

$$\psi_{\vec{K}}(\vec{r}) \;=\; \psi_{\vec{K}}(\vec{r} + L[n_x\hat{x} + n_y\hat{y} + n_y\hat{y}]), \tag{19.6.9}$$

for integers $n_x, n_y, n_z$, so we obtain the allowed wavenumber components

$$K_x = \frac{2\pi}{L}n_x, \qquad K_y = \frac{2\pi}{L}n_y, \qquad K_z = \frac{2\pi}{L}n_z. \tag{19.6.10}$$

We will need the density of final states, $\rho(\mathcal{E}_f, \cos\theta_f, \phi_f)$, as defined by eq. (18.3.16), where $(\theta_f, \phi_f)$ are the spherical coordinate angles of the final-state electron's momentum vector $\vec{K}$, and serve as the degeneracy labels for the energy. For a volume of $\vec{K}$-space $dK_x dK_y dK_z$, the number of allowed final states is

$$\left(\frac{L}{2\pi}\right)^3 dK_x dK_y dK_z \;=\; \left(\frac{L}{2\pi}\right)^3 d^3\vec{K} \;=\; \left(\frac{L}{2\pi}\right)^3 K^2 dK\, d(\cos\theta_f)\, d\phi_f. \tag{19.6.11}$$

Now, since

$$d\mathcal{E}_f = \frac{\hbar^2}{m_e} K\, dK, \tag{19.6.12}$$

we have

$$\begin{pmatrix}\text{number of states in the}\\ \text{range } d\mathcal{E}_f\, d\phi_f\, d(\cos\theta_f)\end{pmatrix} \;=\; \left(\frac{L}{2\pi}\right)^3 \frac{m_e}{\hbar^2} K\, d\mathcal{E}_f\, d\phi_f\, d(\cos\theta_f), \tag{19.6.13}$$

from which we obtain, by comparison with eq. (18.3.16), the density of final electron states

$$\rho(\mathcal{E}_f, \cos\theta_f, \phi_f) \;=\; \left(\frac{L}{2\pi}\right)^3 \frac{m_e K}{\hbar^2} \tag{19.6.14}$$

in the plane-wave approximation.

Because we are assuming that the spin can be neglected due to eq. (19.6.7), eq. (19.1.13) becomes $\vec{\mathcal{P}}_{\vec{k}} = e^{i\vec{k}\cdot\vec{R}}\hat{\varepsilon}\cdot\vec{P}$. Applying eq. (19.2.8), we now obtain the differential cross-section for absorption,

$$\frac{d\sigma}{d\Omega_f} \;=\; \frac{4\pi^2\alpha K}{m_e\hbar\omega}\left(\frac{L}{2\pi}\right)^3\left|\langle f|e^{i\vec{k}\cdot\vec{R}}\hat{\varepsilon}\cdot\vec{P}|i\rangle\right|^2, \tag{19.6.15}$$

where $d\Omega_f = d(\cos\theta_f)\,d\phi_f$. It remains to evaluate the matrix element. Because of the transverse nature of electromagnetic waves $\hat{\varepsilon}\cdot\vec{k} = 0$, we have the commutator $\left[\vec{k}\cdot\vec{R},\,\hat{\varepsilon}\cdot\vec{P}\right] = 0$, so

$$\langle f|e^{i\vec{k}\cdot\vec{R}}\hat{\varepsilon}\cdot\vec{P}|i\rangle \;=\; \langle f|\hat{\varepsilon}\cdot\vec{P}e^{i\vec{k}\cdot\vec{R}}|i\rangle \;=\; \hbar\,\hat{\varepsilon}\cdot\vec{K}\,\langle f|e^{i\vec{k}\cdot\vec{R}}|i\rangle\,, \tag{19.6.16}$$

where we used $\vec{p}_f = \hbar\vec{K}$. The last matrix element is evaluated in the position representation as

$$\langle f|e^{i\vec{k}\cdot\vec{R}}|i\rangle \;=\; \int d^3\vec{r}\,\left(\frac{1}{\sqrt{L^3}}e^{-i\vec{K}\cdot\vec{r}}\right)e^{i\vec{k}\cdot\vec{r}}\left(\sqrt{\frac{Z^3}{\pi a_0^3}}e^{-Zr/a_0}\right) \;=\; \sqrt{\frac{Z^3}{\pi a_0^3 L^3}}\;I(q), \tag{19.6.17}$$

where we have defined the integral

$$I(q) \;=\; \int d^3\vec{r}\;e^{-Zr/a_0}e^{i\vec{q}\cdot\vec{r}}. \tag{19.6.18}$$

in terms of

$$\vec{q} \;\equiv\; \vec{k} - \vec{K}, \tag{19.6.19}$$

so that $\hbar\vec{q}$ is the momentum transferred from the photon to the atom.

Due to the spherical symmetry of the initial-state wavefunction, we see that $I(q)$ only depends on the magnitude of $\vec{q}$, not its direction. So, for the purpose of evaluating $I(q)$, we can temporarily choose $\vec{q} = q\hat{z}$ so that $\vec{q}\cdot\vec{r} = qr\cos\theta$, giving

$$I(q) \;=\; \int_0^\infty dr\,r^2\int_{-1}^{1}d(\cos\theta)\int_0^{2\pi}d\phi\;e^{-r(Z/a_0 - iq\cos\theta)}. \tag{19.6.20}$$

The $\phi$ integration trivially gives a factor of $2\pi$. Doing the $\cos\theta$ integration next, and the $r$ integration last,

$$I(q) \;=\; -\frac{2\pi i}{q}\int_0^\infty dr\,r\left(e^{-r(Z/a_0 - iq)} - e^{-r(Z/a_0 + iq)}\right) \;=\; \frac{8\pi a_0^3}{Z^3(1+q^2a_0^2/Z^2)^2}. \tag{19.6.21}$$

Putting together eqs. (19.6.15)-(19.6.17) and (19.6.21) we arrive at

$$\frac{d\sigma}{d\Omega_f} \;=\; \frac{32\alpha\hbar K}{m_e\omega}\left|\hat{\varepsilon}\cdot\vec{K}\right|^2\frac{a_0^3}{Z^3(1+q^2a_0^2/Z^2)^4}, \tag{19.6.22}$$

in which the dependences on the size of the box $L$ (from the density of states, and from the matrix element) have canceled, as promised. Note that the differential cross-section is largest when the electron is ejected in the direction of the polarization of the electric field of the wave.

It is now convenient to choose a coordinate system in which (without loss of generality, since the initial electron wavefunction is spherically symmetric) the incident photon momentum is in the $\hat{z}$ direction, and has polarization along the $\hat{x}$ direction, so

$$\vec{k} = \hat{z}k = \hat{z}\omega/c, \qquad \hat{\varepsilon} = \hat{x}. \tag{19.6.23}$$

Writing, from here on, $\theta = \theta_f$ and $\phi = \phi_f$ for the ejected electron's momentum spherical coordinate direction angles, with $\vec{K} = \hat{r}K$, we have

$$\begin{aligned} \hat{\varepsilon} \cdot \vec{K} &= K \sin\theta \cos\phi, && (19.6.24) \\ q^2 &= (\vec{k} - \vec{K})^2 = K^2 + (\omega/c)^2 - 2K(\omega/c)\cos\theta, && (19.6.25) \end{aligned}$$

so that eq. (19.6.22) becomes

$$\frac{d\sigma}{d\Omega} = \frac{32\alpha\hbar}{m_e\omega} \frac{a_0^3 K^3}{Z^3(N_1 - N_2\cos\theta)^4} \sin^2\theta \cos^2\phi, \tag{19.6.26}$$

in which we have introduced dimensionless quantities

$$\begin{aligned} N_1 &= 1 + a_0^2(K^2 + \omega^2/c^2)/Z^2, && (19.6.27) \\ N_2 &= 2a_0^2 K\omega/cZ^2. && (19.6.28) \end{aligned}$$

To find the total cross-section, we use the integrals

$$\begin{aligned} \int_0^{2\pi} d\phi \cos^2\phi &= \pi, && (19.6.29) \\ \int_{-1}^{1} d(\cos\theta) \frac{\sin^2\theta}{(N_1 - N_2\cos\theta)^4} &= \frac{4}{3(N_1^2 - N_2^2)^2}, && (19.6.30) \end{aligned}$$

to obtain

$$\sigma = \int d\Omega \frac{d\sigma}{d\Omega} = \frac{128\pi\alpha\hbar}{3m_e\omega} \frac{a_0^3 K^3}{Z^3(N_1^2 - N_2^2)^2}. \tag{19.6.31}$$

If the light is polarized along $\hat{y}$ rather than $\hat{x}$, then $\cos^2\phi$ should be replaced by $\sin^2\phi$ in the differential cross-section eq. (19.6.26). If the light is unpolarized we should instead make the replacement based on the average of the two transverse polarizations, $\cos^2\phi \to (\cos^2\phi + \sin^2\phi)/2 = 1/2$, which gives, after integrating over $\phi$,

$$\frac{d\sigma}{d(\cos\theta)} = \frac{32\pi\alpha\hbar}{m_e\omega} \frac{a_0^3 K^3}{Z^3(N_1 - N_2\cos\theta)^4} \sin^2\theta. \tag{19.6.32}$$

The total cross-section is independent of the polarization, because the angle $\phi$ is integrated over.

Now consider the high-energy limit $K \gg \omega/c$ and $K \gg Z/a_0$ of eq. (19.6.7), which we showed was appropriate for our plane-wave approximation for the final state. Expanding to next-to-leading order in large $K$, we have

$$\frac{1}{(N_1 - N_2\cos\theta)^4} = \left(\frac{Z}{Ka_0}\right)^8 \left(1 + \frac{8\omega}{cK}\cos\theta + \cdots\right), \tag{19.6.33}$$

so that eqs. (19.6.26) and (19.6.31) become

$$\frac{d\sigma}{d\Omega} = \frac{32\alpha\hbar}{m_e\omega}\left(\frac{Z}{Ka_0}\right)^5 \left(1 + \frac{8\omega}{cK}\cos\theta\right)\sin^2\theta\cos^2\phi, \tag{19.6.34}$$

$$\sigma = \frac{128\pi\alpha\hbar}{3m_e\omega}\left(\frac{Z}{Ka_0}\right)^5. \tag{19.6.35}$$

This can be rewritten to eliminate $K$ in favor of $\omega$ using the energy conservation condition $K = \sqrt{2m_e\omega/\hbar}$, which follows from eq. (19.6.5) in the limit of small binding energy compared to the photon energy. The result for the total cross-section is

$$\sigma = \frac{256\pi}{3}\alpha\left(\frac{a_0}{Z}\right)^2\left(\frac{|E_i|}{\hbar\omega}\right)^{7/2}, \tag{19.6.36}$$

where $a_0/Z$ is the characteristic length scale of the initial state, and $|E_i| = Z^2e^2/2a_0$ is the binding energy, and $\hbar\omega$ is the incident photon energy.

The preceding results apply not just to single-electron wavefunctions for hydrogen-like ions, but to the ejection of electrons from the $K$-shell[†] of heavier atoms. There are two $K$-shell electrons (one for each spin), so the cross-section we obtained should be multiplied by 2 to find the total $K$-shell ejection cross-section for multi-electron atoms. The most important qualitative features are that the cross-section grows with increasing $Z$ like $Z^5$, and falls sharply with increasing incident photon energy like $1/(\text{energy})^{7/2}$. The ejected electrons are preferentially emitted along the photon's electric field polarization, which is perpendicular to the photon's propagation direction, but with a slight preference for the forward direction ($\cos\theta > 0$) as a subleading effect in the high-energy expansion in eq. (19.6.34).

In the opposite limit of small $K$, eqs. (19.6.26) and (19.6.31) appear to show a threshold behavior proportional to $K^3$. However, in that regime, the calculation we did is not to be trusted, because the plane-wave approximation for the final state is quite bad.

---

[†] The execrable but traditional jargon "$K$-shell" just means the innermost shell with principal quantum number $n = 1$, called $1s$ in section 16.4. The $2s$, $2p$ electrons are called $L$-shell, and $3s$, $3p$ and $3d$ are called $M$-shell, etc. For each shell with principal quantum number $n$, there can be up to $2\sum_{l=0}^{n-1}(2l+1) = 2n^2$ electrons.

# 20 Scattering in three dimensions

## 20.1 Cross-sections and scattering amplitudes

Consider the problem of scattering of free particles from a localized target in three dimensions. Suppose that the incoming particles are in a mono-energetic beam, all moving in the same direction, which we will take to be $\hat{z}$ by a choice of coordinate system. If the beam is uniform over some area large compared to the size of the target, it can be described by a flux

$$n = \frac{\text{number of incident particles}}{(\text{area})(\text{time})}. \tag{20.1.1}$$

Here, the area is measured in the $xy$ plane, perpendicular to the beam propagation direction. The experiment then counts (or infers, perhaps from measurements of energy deposited in a detector) the number of particles scattered in a differential of solid angle

$$d\Omega = d\phi\, d(\cos\theta). \tag{20.1.2}$$

Let the number of scattered particles as a function of direction be described by

$$ds = \frac{\text{number of scattered particles in } d\Omega}{\text{time}}. \tag{20.1.3}$$

Clearly, $ds$ must be proportional to $n$, and to $d\Omega$. We therefore define the proportionality factor to be the scattering differential **cross-section**,

$$\frac{d\sigma}{d\Omega} = \frac{ds}{n\, d\Omega}. \tag{20.1.4}$$

Note that this has units of area. The total cross-section is then defined as

$$\sigma = \int d\Omega\, \frac{d\sigma}{d\Omega} = \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, \frac{d\sigma}{d\Omega}, \tag{20.1.5}$$

also with units of area. Intuitively, the cross-section is the effective area of the target for the type of scattering under consideration.

It is often true that the differential cross-section has **azimuthal symmetry**, which means that it does not depend on $\phi$. This occurs if the target and the properties of the scattering process are invariant under rotations about the initial beam propagation direction. In that case, one can give the differential cross-section in terms of

$$\frac{d\sigma}{d(\cos\theta)} = \int_0^{2\pi} d\phi\, \frac{d\sigma}{d\Omega} = 2\pi \frac{d\sigma}{d\Omega}. \tag{20.1.6}$$

In particular, this will occur if the target is spherically symmetric and the incident particles are either spin-less or have randomly oriented spins.

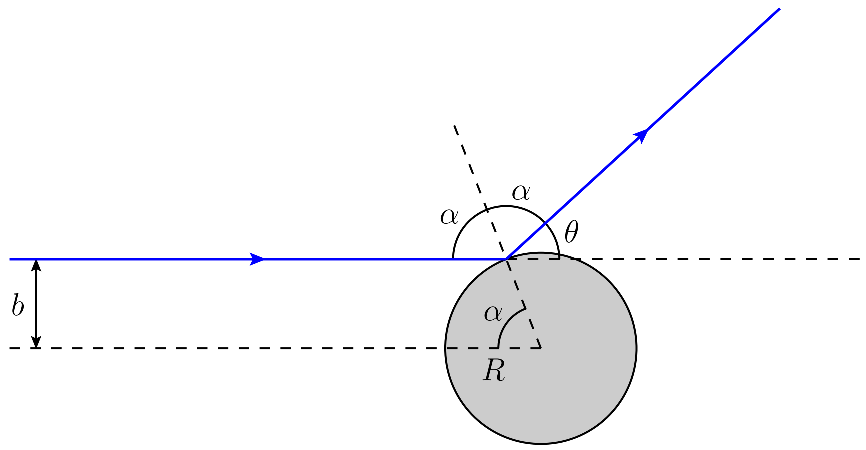

Figure 20.1.1: A classical trajectory for scattering from a hard sphere of radius $R$, with impact parameter $b$ and scattering angle $\theta$ measured from the forward direction $\hat{z}$ (the direction pointing to the right). The geometry gives the relation $b = R\cos(\theta/2)$.

For classical scattering problems, the cross-section can be computed in terms of specific trajectories followed by the incident particles. It is useful to define the **impact parameter** $b$ as the distance between a particular initial particle trajectory line (when it is still very far from the target) and the line parallel to it through a suitably chosen central point in the target, which we take to be the origin of our coordinate system.

Consider first a classical[†] example, **hard-sphere scattering**. Some (very small) particles are fired at a target ball (of radius $R$, suspended in mid-air, and heavy enough that it doesn't move when hit). Gravitational effects are neglected, so that the incident particles move in straight lines. For a particular scattering event, the impact parameter $b$ is the distance between the initial particle trajectory line and the line parallel to it through the center of the target ball, as shown in Figure 20.1.1. To find the differential cross-section, the general strategy is to relate $b$ to the scattering angle $\theta$. Using elementary geometry, the angles $\alpha$ and $\theta$ are related to each other and the impact parameter by

$$\alpha \;=\; \sin^{-1}(b/R), \qquad\qquad \theta \;=\; \pi - 2\alpha, \tag{20.1.7}$$

which implies

$$b \;=\; R\sin(\pi/2 - \theta/2) \;=\; R\cos(\theta/2). \tag{20.1.8}$$

Assuming a uniform flux $n$ of incident particles moving in the $\hat{z}$ direction, the number of them per unit time with impact parameter between $b$ and $b + db$ is equal to the product of $n$ with the area of the annulus in the $xy$ plane with inner radius $b$ and outer radius $b + db$,

$$n(2\pi b\, db) \;=\; n2\pi R\cos(\theta/2)\, R\, d(\cos(\theta/2)) \;=\; \frac{1}{2} n\pi R^2\, d(\cos\theta). \tag{20.1.9}$$

---

[†] We will later obtain the result for the corresponding quantum mechanical problem, in section 20.5.

This means that the rate of scattered particles in solid angle $d\Omega$ in eq. (20.1.3) is

$$ds \;=\; \frac{1}{2}n\pi R^2 d(\cos\theta)\,\frac{d\phi}{2\pi} \;=\; \frac{1}{4}nR^2 d\Omega. \tag{20.1.10}$$

We therefore obtain, from the definition eq. (20.1.4), the differential cross-section

$$\frac{d\sigma}{d\Omega} \;=\; R^2/4. \tag{20.1.11}$$

Because of the azimuthal symmetry, this is equivalently

$$\frac{d\sigma}{d(\cos\theta)} \;=\; \pi R^2/2, \tag{20.1.12}$$

which integrates to give the total cross-section

$$\sigma \;=\; \int_{-1}^{1} d(\cos\theta)\,\frac{d\sigma}{d(\cos\theta)} \;=\; \pi R^2. \tag{20.1.13}$$

This example has two rather special features. First, the differential cross-section is isotropic; it does not depend on $\theta$. Second, the total cross-section is simply equal to the geometrical cross-sectional area of the target; the scattering takes place if and only if the impact parameter does not exceed the radius $R$.

Another classical (and classic) example is Rutherford scattering of a light spin-less particle with charge $Z_1 e$ (for example, an $\alpha$ particle with $Z_1 = 2$) and initial energy $E$ from a static Coulomb potential

$$V(r) = Z_2 e/r, \tag{20.1.14}$$

for example, caused by a heavy nucleus containing $Z_2$ protons. From the classical hyperbolic trajectories, one can show (again by finding $\theta$ in terms of the impact parameter distance defined when the scattering particle is very far away) that

$$\frac{d\sigma}{d\Omega} \;=\; \frac{Z_1^2 Z_2^2 e^4}{16E^2 \sin^4(\theta/2)}. \tag{20.1.15}$$

This differential cross-section is certainly not constant, and in fact blows up with a non-integrable divergence for small $\theta$. To get a finite value for the total cross-section, one must modify the question slightly and count only those events in which the electron is scattered by some minimal cut-off angle $\theta_{\rm min}$. (In real-world experiments, this is justified by the finite resolution of detectors, and by the difficulties encountered if one tries to put a detector too close to the beam.) Then

$$\sigma(\theta > \theta_{\rm min}) \;=\; \int_0^{2\pi} d\phi \int_{-1}^{\cos\theta_{\rm min}} d(\cos\theta)\frac{Z_1^2 Z_2^2 e^4}{16E^2 \sin^4(\theta/2)} \;=\; \frac{\pi Z_1^2 Z_2^2 e^4}{4E^2}\left(\frac{1+\cos\theta_{\rm min}}{1-\cos\theta_{\rm min}}\right). \tag{20.1.16}$$

The divergence is recovered in the limit $\theta_{\text{min}} \to 0$. The interpretation is that in scattering from a $1/r$ potential, the initial charged particle is always scattered at least a little, with sufficient strength that the unrestricted total cross-section is infinite; we say that the Coulomb potential has infinite range.

Most scattering problems are intermediate between the extremes of hard-sphere scattering (with a sharp cut-off limited range of interaction) and Coulomb scattering (with an infinite range). Even if scattering occurs for every impact parameter or scattering angle, the integrated total cross-section $\sigma$ can still be finite, if the differential cross-section rises less quickly than $1/(1 - \cos\theta)$ for small $\theta$. If so, then the potential giving rise to it is said to have finite range.

Enough classical warm-ups; we now turn to the quantum theory of scattering. Consider a beam of particles with fixed momentum $\vec{p} = \hbar\vec{k}$ that encounter a target in the form of a potential $V(\vec{r})$ that has support within a finite volume near the origin. If the potential were completely absent, the beam could be described by a plane wavefunction

$$\psi(\vec{r}) \;=\; \langle \vec{r}|\vec{k}\rangle \;=\; \frac{1}{(2\pi)^{3/2}} e^{i\vec{k}\cdot\vec{r}} \;=\; \frac{1}{(2\pi)^{3/2}} e^{ikz}. \tag{20.1.17}$$

In writing the last equality, we chose our coordinate system so that the initial beam momentum is along the $z$ direction, so

$$\vec{k} = \hat{z}k. \tag{20.1.18}$$

The wavefunction eq. (20.1.17) cannot be normalized to unity, so we have chosen to instead normalize it so that it satisfies orthonormality and completeness relations

$$\langle \vec{k}'|\vec{k}\rangle \;=\; \delta^{(3)}(\vec{k}' - \vec{k}), \qquad\qquad \int d^3\vec{k}\, |\vec{k}\rangle\langle\vec{k}| \;=\; I, \tag{20.1.19}$$

and the probability of finding a particle per unit volume is

$$\frac{\text{probability}}{\text{volume}} \;=\; |\psi(\vec{r},t)|^2 \;=\; |e^{-iEt/\hbar}\psi(\vec{r})|^2 \;=\; \frac{1}{(2\pi)^3}, \tag{20.1.20}$$

a constant in both time and position. This normalization is rather arbitrary, but will cancel in the differential cross-section. The total flux of particles (probability for a particle in the incident beam to pass through a unit area, per unit time) described by the wavefunction is therefore

$$n \;=\; \frac{\text{probability}}{(\text{area})(\text{time})} \;=\; \left(\frac{\text{probability}}{\text{volume}}\right)(\text{speed}) \;=\; \frac{1}{(2\pi)^3}\frac{\hbar k}{m}. \tag{20.1.21}$$

Let us next anticipate the form of the wavefunction for large $r$, taking into account the scattering potential,

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}}\left[e^{ikz} + \frac{e^{ikr}}{r} f_k(\theta,\phi)\right], \qquad\qquad (r \to \infty). \tag{20.1.22}$$

We will justify this form before the end of section 20.2. For now, we just note that the first term represents the incident wave, while the quantity

$$f_k(\theta, \phi) \tag{20.1.23}$$

characterizes the outgoing flux of probability density, and is called the **scattering amplitude**. It multiplies a factor $e^{ikr}$, which corresponds to a wave traveling away from the origin, and a factor $1/r$ which indicates that the probability amplitude falls linearly with distance. This gives a probability density that falls off like $1/r^2$, which is the requisite behavior so that the probability to detect a particle scattering within a given solid angle $d\Omega$ with area $r^2 d\Omega$ is independent of the distance to the detector $r$, when $r$ is large. Due to energy conservation and the assumption that the scattering potential vanishes at large distances, the magnitude of the wavenumber is the same $k$ for both incoming and outgoing waves. The scattering amplitude $f_k(\theta, \phi)$ has units of [length] and depends only on the magnitude $k$ and the spherical coordinate angles, interpreted as those of the propagation direction of a scattered particle. In cases with azimuthal symmetry, we will write the scattering amplitude as $f_k(\theta)$.

A small detector at $(\theta, \phi)$ covering a solid angle $d\Omega$, at a large distance $r$ from the target, will see a total probability flux

$$\begin{aligned} ds &= \left(\frac{\text{probability}}{\text{volume}}\right)(\text{area})(\text{speed}) = \left|\frac{1}{(2\pi)^{3/2}} f_k(\theta, \phi) \frac{e^{ikr}}{r}\right|^2 \left(r^2 d\Omega\right) \left(\frac{\hbar k}{m}\right) && (20.1.24) \\ &= \frac{|f_k(\theta, \phi)|^2}{(2\pi)^3} \frac{\hbar k}{m} \, d\Omega. && (20.1.25) \end{aligned}$$

Here we have neglected the interference with the incident part of the wavefunction proportional to $e^{ikz}$, with the justification that in the real world the incoming particle beam has only a finite extent that will not overlap with sufficiently distant detectors measuring the scattered particles. (To summarize the assumptions made: the beam size is large compared to the target, but small compared to the distance to the detector.) Using eqs. (20.1.21) and (20.1.25) in eq. (20.1.4), we obtain the important result

$$\frac{d\sigma}{d\Omega} = |f_k(\theta, \phi)|^2. \tag{20.1.26}$$

The differential cross-section is equal to the squared magnitude of the scattering amplitude.

## 20.2 Lippmann–Schwinger equation and scattering Green function

Our goal in this section is to obtain the form of the wavefunction claimed in eq. (20.1.22), and in the process to find out how to calculate the scattering amplitude $f_k(\theta, \phi)$ for a given potential $V$. Start by writing the Hamiltonian operator as

$$H = H_0 + V(\vec{R}), \tag{20.2.1}$$

where $H_0 = P^2/2m$ is the pure kinetic term, and the scattering potential $V(\vec{R})$ is assumed to be significant only within a localized region. The eigenkets of $H_0$, with wavefunctions as in eq. (20.1.17), obey the eigenvalue equation

$$H_0|\vec{k}\rangle \;=\; E|\vec{k}\rangle, \qquad\qquad E \;=\; \hbar^2 k^2/2m, \tag{20.2.2}$$

and we will have in mind specifically $\hbar\vec{k} = \hbar k\hat{z} = \vec{p}$ for the incident beam particles.

We now look for stationary-state scattering solutions $|\psi\rangle$ to the full Schrödinger equation, with the same energy eigenvalue $E$ as the incident particles, so

$$(H_0 + V)\,|\psi\rangle \;=\; E\,|\psi\rangle\,. \tag{20.2.3}$$

Let us rewrite this as $(E - H_0)\,|\psi\rangle = V\,|\psi\rangle$, and then add $0 = (E - H_0)|\vec{k}\rangle$ to the right side:

$$(E - H_0)\,|\psi\rangle \;=\; V\,|\psi\rangle + (E - H_0)|\vec{k}\rangle. \tag{20.2.4}$$

We now act with the operator $(E - H_0)^{-1}$ on each side to get

$$|\psi\rangle \;=\; |\vec{k}\rangle + \frac{1}{E - H_0} V\,|\psi\rangle\,. \tag{20.2.5}$$

However, the inverse operator can be problematic if $H_0$ formally approaches $E$. To avoid dividing by 0, we regulate the inverse operator by deforming $E$ into the complex domain, away from the real line by an infinitesimal amount $\pm\epsilon$, with $\epsilon$ positive and to be taken to 0 later. The result is the **Lippmann–Schwinger equation**

$$|\psi^{\pm}\rangle \;=\; |\vec{k}\rangle + \frac{1}{E - H_0 \pm i\epsilon}\, V\,|\psi^{\pm}\rangle, \tag{20.2.6}$$

named after Bernard Lippmann and Julian Schwinger. We now wish to solve for the full ket $|\psi^{\pm}\rangle$, which includes the effects of scattering from the potential. It will later become apparent that the + sign is what we want, but let us keep both for now.

In the position representation, completeness with respect to position eigenstates $|\vec{r}\,'\rangle$ turns the Lippmann–Schwinger equation into its wavefunction form,

$$\langle\vec{r}\,|\psi^{\pm}\rangle \;=\; \langle\vec{r}\,|\vec{k}\rangle + \int d^3\vec{r}\,'\,\langle\vec{r}\,|\frac{1}{E - H_0 \pm i\epsilon}|\vec{r}\,'\rangle\langle\vec{r}\,'|\,V\,|\psi^{\pm}\rangle. \tag{20.2.7}$$

The last term involves a matrix element that we now write as

$$\langle\vec{r}\,|\frac{1}{E - H_0 \pm i\epsilon}|\vec{r}\,'\rangle \;=\; \frac{2m}{\hbar^2} G_{\pm}(\vec{r}, \vec{r}\,'), \tag{20.2.8}$$

which defines $G_{\pm}(\vec{r},\, \vec{r}\,')$ as a function independent of the scattering potential $V$. The normalization factor $2m/\hbar^2$ has been introduced for future convenience.

To evaluate the function $G_\pm(\vec{r},\, \vec{r}\,')$, we will now use completeness twice more, this time with respect to plane-wave states with momenta $\hbar\vec{q}$ and $\hbar\vec{q}'$, normalized just as in eq. (20.1.17), so

$$\langle \vec{r}|\vec{q}\rangle \;=\; \frac{1}{(2\pi)^{3/2}} e^{i\vec{q}\cdot\vec{r}}, \tag{20.2.9}$$

with orthonormality and completeness relations

$$\langle \vec{q}'|\vec{q}\rangle \;=\; \delta^{(3)}(\vec{q}' - \vec{q}), \qquad\qquad \int d^3\vec{q}\; |\vec{q}\rangle\langle\vec{q}| \;=\; I. \tag{20.2.10}$$

Thus eq. (20.2.8) becomes

$$G_\pm(\vec{r},\, \vec{r}\,') \;=\; \frac{\hbar^2}{2m}\int d^3\vec{q}' \int d^3\vec{q}\; \langle\vec{r}|\vec{q}'\rangle\langle\vec{q}'|\frac{1}{E - H_0 \pm i\epsilon}|\vec{q}\rangle\langle\vec{q}|\vec{r}\,'\rangle. \tag{20.2.11}$$

The reason for doing this is that now the operator $H_0$ can be evaluated as a number, using

$$H_0|\vec{q}\rangle \;=\; \frac{\hbar^2 q^2}{2m}|\vec{q}\rangle, \tag{20.2.12}$$

so that

$$\langle\vec{q}'|\frac{1}{E - H_0 \pm i\epsilon}|\vec{q}\rangle \;=\; \delta^{(3)}(\vec{q}' - \vec{q})\frac{1}{E - \hbar^2 q^2/2m \pm i\epsilon}. \tag{20.2.13}$$

Doing the $\int d^3\vec{q}'$ integration in eq. (20.2.11) using the delta function, and remembering that $E = \hbar^2 k^2/2m$ is the energy of the incident beam particles, we find an integral expression for the function defined by eq. (20.2.8),

$$G_\pm(\vec{r}, \vec{r}\,') \;=\; \int \frac{d^3\vec{q}}{(2\pi)^3}\,\frac{e^{i\vec{q}\cdot(\vec{r}-\vec{r}\,')}}{k^2 - q^2 \pm i\epsilon}. \tag{20.2.14}$$

Before doing this integral, as an aside we note that $G_\pm(\vec{r}, \vec{r}\,')$ could also be defined mathematically as the solution to a differential equation. To see this, note that

$$\nabla^2 G_\pm(\vec{r}, \vec{r}\,') \;=\; \int \frac{d^3\vec{q}}{(2\pi)^3}\,\frac{(-q^2)\, e^{i\vec{q}\cdot(\vec{r}-\vec{r}\,')}}{k^2 - q^2 \pm i\epsilon}, \tag{20.2.15}$$

where the Laplacian was taken with respect to the coordinate $\vec{r}$, taking advantage of the simple exponential dependence. Therefore,

$$(\nabla^2 + k^2) G_\pm(\vec{r}, \vec{r}\,') \;=\; \int \frac{d^3\vec{q}}{(2\pi)^3}\, e^{i\vec{q}\cdot(\vec{r}-\vec{r}\,')} \;=\; \delta^{(3)}(\vec{r} - \vec{r}\,'), \tag{20.2.16}$$

which shows that $G_\pm(\vec{r}, \vec{r}\,')$ is a **Green function** (named after mathematician George Green) for the Helmholtz differential operator $\nabla^2 + k^2$, with a delta-function source at $\vec{r}\,'$.

To evaluate the Green function, go back to eq. (20.2.14) and put the integral $\int d^3\vec{q}$ into spherical coordinates,

$$G_\pm(\vec{r},\, \vec{r}\,') \;=\; \frac{1}{(2\pi)^3}\int_0^\infty dq\, q^2 \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, \frac{e^{iq|\vec{r}-\vec{r}\,'|\cos\theta}}{k^2 - q^2 \pm i\epsilon}, \tag{20.2.17}$$

where the angles $\theta$ and $\phi$ of $\vec{q}$ are measured from the direction $\vec{r} - \vec{r}\,'$. Now do the angular integrals to get

$$G_\pm(\vec{r},\, \vec{r}\,') \;=\; \frac{i}{4\pi^2|\vec{r}-\vec{r}\,'|}\int_0^\infty dq\, \frac{q}{q^2 - k^2 \mp i\epsilon}\left(e^{iq|\vec{r}-\vec{r}\,'|} - e^{-iq|\vec{r}-\vec{r}\,'|}\right). \tag{20.2.18}$$

Let $x = |\vec{r}-\vec{r}\,'|$, and call the integral $I(x,k)$. It can be evaluated (using the method of complex variable contour integration, for example) as $I(x,k) = i\pi e^{\pm ikx}$. Therefore, the Green function is

$$G_\pm(\vec{r},\, \vec{r}\,') \;=\; -\frac{e^{\pm ik|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}. \tag{20.2.19}$$

Plugging this result into the Lippmann–Schwinger equation (20.2.7), using eq. (20.2.8), gives

$$\langle\vec{r}|\psi^\pm\rangle \;=\; \langle\vec{r}|\vec{k}\rangle - \frac{2m}{\hbar^2}\int d^3\vec{r}\,'\, \frac{e^{\pm ik|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}\,\langle\vec{r}\,'|\,V\,|\psi^\pm\rangle. \tag{20.2.20}$$

We now complete the translation into the position representation by noticing that

$$\langle\vec{r}\,'|\,V\,|\psi^\pm\rangle \;=\; \int d^3\vec{r}\,''\,\langle\vec{r}\,'|\,V\,|\vec{r}\,''\,\rangle\langle\vec{r}\,''\,|\psi^\pm\rangle \;=\; V(\vec{r}\,')\langle\vec{r}\,'|\psi^\pm\rangle, \tag{20.2.21}$$

where we have used $\langle\vec{r}\,'|\,V\,|\vec{r}\,''\,\rangle \;=\; \delta^{(3)}(\vec{r}\,' - \vec{r}\,'')V(\vec{r}\,')$. Therefore, the Lippmann–Schwinger equation in the position representation is

$$\psi^\pm(\vec{r}) \;=\; \frac{e^{ikz}}{(2\pi)^{3/2}} - \frac{2m}{\hbar^2}\int d^3\vec{r}\,'\, \frac{e^{\pm ik|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}V(\vec{r}\,')\psi^\pm(\vec{r}\,'). \tag{20.2.22}$$

Here, $\psi^\pm(\vec{r})$ is the total wavefunction, split into the part describing the incident particle beam, and the integral coming from the effect of scattering by the potential.

We now specialize to points $\vec{r}$ that are very far away from the points $\vec{r}\,'$ in the neighborhood of the origin where $V(\vec{r}\,')$ has its support. We do this by expanding for $r' \ll r$, so that

$$|\vec{r}-\vec{r}\,'| \;=\; \sqrt{r^2 - 2\vec{r}\cdot\vec{r}\,' + r'^2} \;=\; r\sqrt{1 - 2\hat{r}\cdot\vec{r}\,'/r + \cdots} \;=\; r - \hat{r}\cdot\vec{r}\,' + \cdots\,. \tag{20.2.23}$$

Let us define a wavevector $\vec{k}'$, with the same magnitude as the incoming wavevector $\vec{k} = k\hat{z}$, but directed away from the scattering region (in the radial direction),

$$\vec{k}' \;=\; k\hat{r}. \tag{20.2.24}$$

Then eq. (20.2.23) implies

$$e^{\pm ik|\vec{r}-\vec{r}\,'|} \quad \approx \quad e^{\pm ikr} e^{\mp i\vec{k}'\cdot\vec{r}\,'}. \tag{20.2.25}$$

and eq. (20.2.22) becomes, for large $r$,

$$\psi^{\pm}(\vec{r}) \quad = \quad \frac{1}{(2\pi)^{3/2}} e^{ikz} - \frac{m}{2\pi\hbar^2}\frac{e^{\pm ikr}}{r}\int d^3\vec{r}\,' \, e^{-i\vec{k}'\cdot\vec{r}\,'} V(\vec{r}\,')\, \psi^{\pm}(\vec{r}\,'). \tag{20.2.26}$$

Both of the sign choices ± that we introduced in eq. (20.2.6) are mathematically valid solutions, but we now see that we want to choose the + sign, because the factor $e^{+ikr}$ corresponds to a scattered wave moving *away* from the target potential. A factor $e^{-ikr}$ in eq. (20.2.26) would correspond to the strange situation of an incoming radial wave converging on a target from all directions and cleverly matching itself onto a plane wave; while mathematically allowed, this does not correspond to real-world scattering processes. Thus the physically relevant solution for scattering problems (dropping the + superscript from now on) is, for large $r$,

$$\psi(\vec{r}) \quad = \quad \frac{1}{(2\pi)^{3/2}}\left[e^{ikz} + \frac{e^{ikr}}{r} f_k(\theta,\phi)\right], \qquad\qquad (r \to \infty), \tag{20.2.27}$$

where we have now identified an expression for the scattering amplitude,

$$f_k(\theta,\phi) \quad = \quad -\frac{m}{2\pi\hbar^2}(2\pi)^{3/2}\int d^3\vec{r}\,' \, e^{-i\vec{k}'\cdot\vec{r}\,'} V(\vec{r}\,')\, \psi(\vec{r}\,'). \tag{20.2.28}$$

Equation (20.2.27) indeed has the form promised in eq. (20.1.22). However, eqs. (20.2.27) and (20.2.28) do not yet a constitute a solution, because eq. (20.2.27) writes $\psi$ in terms of $f_k(\theta,\phi)$, which in turn depends on $\psi$. Our remaining task in this section is to see how to solve these equations for $f_k(\theta,\phi)$ by eliminating $\psi$.

An equivalent way of writing eq. (20.2.28) is

$$f_k(\theta,\phi) \quad = \quad -\frac{4\pi^2 m}{\hbar^2}\,\langle\vec{k}'|V|\psi\rangle, \tag{20.2.29}$$

where, for general $\vec{k}'$, the ket $|\vec{k}'\rangle$ is the state with position wavefunction $e^{i\vec{k}'\cdot\vec{r}}/(2\pi)^{3/2}$. Equation (20.2.29) allows us to find at least a formal solution for the scattering amplitude, by the following maneuver. We first define the **transition operator** $T$ by

$$T|\vec{k}\rangle \quad = \quad V\,|\psi\rangle \tag{20.2.30}$$

for every plane wave ket $|\vec{k}\rangle$, where $|\psi\rangle$ is the corresponding scattering state. Since the kets $|\vec{k}\rangle$ form an orthobasis, by superposition this defines $T$ acting on any state. In terms of this operator, eq. (20.2.29) becomes a matrix element between known (plane-wave) states,

$$f_k(\theta,\phi) \quad = \quad -\frac{4\pi^2 m}{\hbar^2}\,\langle\vec{k}'|T|\vec{k}\rangle. \tag{20.2.31}$$

The angles $(\theta, \phi)$ here are defined to be the spherical coordinate angles of $\vec{k}'$ when the coordinate system has been chosen so that $\vec{k} = \hat{z}k$. The problem of finding the scattering amplitude has therefore been reduced to solving for $T$.

From the Lippmann–Schwinger equation (20.2.6) with the + sign now selected,

$$T|\vec{k}\rangle \;=\; V\left(|\vec{k}\rangle + \frac{1}{E - H_0 + i\epsilon} V\,|\psi\rangle\right) \;=\; \left(V + V\frac{1}{E - H_0 + i\epsilon}T\right)|\vec{k}\rangle. \tag{20.2.32}$$

Since this is true for every plane-wave ket $|\vec{k}\rangle$, we have an operator relation between $T$, $V$, $H_0$, and the incident particle energy $E$,

$$T \;=\; V + V\frac{1}{E - H_0 + i\epsilon}T. \tag{20.2.33}$$

This can be formally solved for $T$, with the result

$$T = \left(I - V\frac{1}{E - H_0 + i\epsilon}\right)^{-1} V. \tag{20.2.34}$$

Together with eq. (20.2.31), we can finally say that we have a solution for the scattering amplitude. However, practical evaluation of it is still a non-trivial task that typically requires approximation. Different approximation schemes amount to practical methods for evaluating the formal inverse operator on the right side of eq. (20.2.34).

## 20.3 The Optical Theorem

The **Optical Theorem** is an identity relating the total integrated cross-section to the imaginary part of the forward scattering amplitude $f_k(\theta = 0)$, as follows:

$$\sigma \;=\; \frac{4\pi}{k}\mathrm{Im}\left[f_k(0)\right]. \tag{20.3.1}$$

This is useful for at least two reasons. First, it can provide a check of the consistency of a calculation. Second, it is sometimes easier to obtain $\mathrm{Im}[f_k(0)]$ than it is to find the total cross-section directly. Note that we write $f_k(0)$ for $\theta = 0$ even though we are *not* assuming azimuthal symmetry, because the coordinate $\phi$ is meaningless and irrelevant when $\theta = 0$.

To prove the Optical Theorem, start with

$$\mathrm{Im}\left[f_k(0)\right] \;=\; -\frac{4\pi^2 m}{\hbar^2}\,\mathrm{Im}\left(\langle\vec{k}|V|\psi\rangle\right), \tag{20.3.2}$$

which is obtained from eq. (20.2.29) by setting $\langle\vec{k}'| = \langle\vec{k}|$ for forward scattering, and then taking the imaginary part. To evaluate the right side, we can apply the Hermitian adjoint of the Lippmann–Schwinger equation (20.2.6) with the + sign selected,

$$\langle\vec{k}| \;=\; \langle\psi| - \langle\psi|V\frac{1}{E - H_0 - i\epsilon}, \tag{20.3.3}$$

to obtain

$$\mathrm{Im}\left(\langle\vec{k}|V|\psi\rangle\right) \;=\; \mathrm{Im}\left(\langle\psi|V|\psi\rangle\right) - \mathrm{Im}\left(\langle\psi|V\frac{1}{E-H_0-i\epsilon}V|\psi\rangle\right). \tag{20.3.4}$$

The first term on the right side is 0, because $V$ is Hermitian, and therefore always has real expectation values. In contrast, the second term does not vanish, because the operator $1/(E - H_0 - i\epsilon)$ is not Hermitian, due to the $i\epsilon$. To make better sense of this operator, consider an analogous complex function of a real number $x$,

$$\frac{1}{x-i\epsilon} \;=\; \frac{1}{x-i\epsilon}\left(\frac{x+i\epsilon}{x+i\epsilon}\right) \;=\; \frac{x}{x^2+\epsilon^2} + i\frac{\epsilon}{x^2+\epsilon^2}. \tag{20.3.5}$$

Although it may be tempting to say that this is real in the limit of small $\epsilon$, special care must be taken when $x = 0$, where we recall that

$$\lim_{\epsilon\to 0}\frac{\epsilon}{x^2+\epsilon^2} \;=\; \pi\delta(x). \tag{20.3.6}$$

It is also useful to define the **Cauchy principal value** as

$$\mathrm{Pr}\,(1/x) = \frac{x}{x^2+\epsilon^2}, \tag{20.3.7}$$

for $\epsilon$ assumed to be infinitesimal. This is just equal to $1/x$ when $x \neq 0$, but is equal to 0 when $x = 0$. Then we can write, formally valid for all $x$ including $x = 0$,

$$\frac{1}{x-i\epsilon} \;=\; \mathrm{Pr}(1/x) + i\pi\delta(x). \tag{20.3.8}$$

The expressions on the right are distributions, which means that they are defined only inside integrations over $x$, according to the following rules, which apply for $-a < 0 < b$,

$$\int_{-a}^{b} dx\,\delta(x)f(x) \;=\; f(0), \tag{20.3.9}$$

$$\int_{-a}^{b} dx\,\mathrm{Pr}(1/x)\,f(x) \;=\; \int_{-a}^{0} dx\,\frac{f(x)}{x} + \int_{0}^{b} dx\,\frac{f(x)}{x}. \tag{20.3.10}$$

For our present purposes, the operator application of eq. (20.3.8) is

$$\frac{1}{E-H_0-i\epsilon} \;=\; \mathrm{Pr}\left(\frac{1}{E-H_0}\right) + i\pi\delta(E-H_0), \tag{20.3.11}$$

where the principal value and delta functions are Hermitian operators, just as $\mathrm{Pr}(1/x)$ and $\delta(x)$ are real functions. Thus, the first term on the right is Hermitian and has real expectation values, but the second term is anti-Hermitian and has imaginary expectation values simply because of the factor of $i$. Plugging in to eq. (20.3.4), we learn

$$\mathrm{Im}\left(\langle\vec{k}|V|\psi\rangle\right) \;=\; -\pi\,\langle\psi|V\delta(E-H_0)V|\psi\rangle \;=\; -\pi\,\langle\vec{k}|T^\dagger\delta(E-H_0)T|\vec{k}\rangle\,, \tag{20.3.12}$$

where we have used the definition of $T$ from eq. (20.2.30) to get the last equality. Next, we use completeness, by inserting $I = \int d^3\vec{k}'\, |\vec{k}'\rangle\, \langle\vec{k}'|$, and using $H_0\, |\vec{k}'\rangle = (\hbar^2 k'^2/2m)\, |\vec{k}'\rangle$, so that

$$\begin{aligned}
\text{Im}\left(\langle\vec{k}|V|\psi\rangle\right) &= -\pi \int d^3\vec{k}'\, \delta(E - \hbar^2 k'^2/2m)\, \langle\vec{k}|T^\dagger|\vec{k}'\rangle\, \langle\vec{k}'|T|\vec{k}\rangle && (20.3.13)\\
&= -\pi \int d\Omega' \int_0^\infty d(k'^2) \frac{mk'}{\hbar^2} \delta(k'^2 - k^2)\, |\langle\vec{k}|T|\vec{k}'\rangle|^2 && (20.3.14)\\
&= -\frac{\pi m k}{\hbar^2} \int d\Omega'\, |\langle\vec{k}|T|\vec{k}'\rangle|^2, && (20.3.15)
\end{aligned}$$

where we have separated the integration over $\vec{k}'$ into an angular and radial part, done a change of variable on the delta function, and then done the radial integral over $k'^2$ using the delta function. Putting this into eq. (20.3.2) gives the imaginary part of the forward scattering amplitude in terms of the matrix elements of the transition operator,

$$\text{Im}\,[f_k(0)] = \frac{4\pi^3 m^2 k}{\hbar^4} \int d\Omega'\, |\langle\vec{k}|T|\vec{k}'\rangle|^2. \tag{20.3.16}$$

Since we defined the operator $T$ acting on general plane-wave states, $|\langle\vec{k}|T|\vec{k}'\rangle|^2$ depends only on the common magnitude of $\vec{k}$ and $\vec{k}'$ and the angle between them. We can therefore view it as a function of $(\theta', \phi')$ with $\vec{k}$ held fixed, and use eqs. (20.2.31) and (20.1.26) to write

$$|\langle\vec{k}|T|\vec{k}'\rangle|^2 \;=\; \frac{\hbar^4}{16\pi^4 m^2} |f_k(\theta', \phi')|^2 \;=\; \frac{\hbar^4}{16\pi^4 m^2} \frac{d\sigma}{d\Omega'}, \tag{20.3.17}$$

where the angles $(\theta', \phi')$ are measured with respect to the $\vec{k}$ direction. Inserting this in eq. (20.3.16) and doing the $d\Omega'$ integration gives

$$\text{Im}\,[f_k(0)] = \frac{k}{4\pi}\sigma, \tag{20.3.18}$$

which is the Optical Theorem.

## 20.4 Born approximation

The **Born approximation** is a type of perturbative expansion applied to scattering problems, in which the wavefunction within the target region is assumed to be only mildly altered by the potential. In the first-order (or "leading") Born approximation, for the purposes of eq. (20.2.28) we simply take the wavefunction inside the target to be the same as the incident plane wave,

$$\psi(\vec{r}\,') \;\approx\; \frac{1}{(2\pi)^{3/2}} e^{ikz'}. \tag{20.4.1}$$

Recall that the wavevectors of the incident and scattered wavefunctions are $\vec{k} = k\hat{z}$ and $\vec{k}' = k\hat{r}$ respectively, so that $kz' = \vec{k} \cdot \vec{r}\,'$. Equation (20.2.28) then reduces to an integral that can be evaluated for any given potential,

$$f_k(\theta, \phi) \;\approx\; -\frac{m}{2\pi\hbar^2} \int d^3\vec{r}\,'\, e^{i(\vec{k}-\vec{k}')\cdot\vec{r}\,'} V(\vec{r}\,'). \tag{20.4.2}$$

Note that in this first-order Born approximation, the sign of the potential does not make any difference, since it enters linearly in the scattering amplitude, which then gets squared to give $\frac{d\sigma}{d\Omega} = |f_k(\theta, \phi)|^2$. In particular, within the leading Born approximation, an attractive potential gives the same differential cross-section as a repulsive potential with the same magnitude.

Before proceeding to use it, let us first address the question of the validity of the Born approximation. Intuitively, this relies on the potential not being too strong. Consider the wavefunction form of the Lippmann–Schwinger equation valid inside the target region, as it appeared in eq. (20.2.22) before we took the large $r$ limit. To make a rough estimate, let us call $a$ the characteristic range of the potential, meaning that $V(\vec{r})$ is significant only for $r < a$. Then we can require that in the right side of eq. (20.2.22) the correction term is much smaller in magnitude than the plane wave term,

$$\frac{2m}{\hbar^2}\left|\int d^3\vec{r}\,'\,\frac{e^{ik|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}V(\vec{r}\,')\,e^{ikz'}\right| \;\ll\; 1, \tag{20.4.3}$$

for $r < a$, where we have presumptively replaced the wavefunction $\psi$ by its putative approximation eq. (20.4.1). To make things easier, we can just consider $\vec{r}$ close to the origin, where the correction might be expected to be largest, or at least not much smaller than at other points. Then a rough condition for validity of the Born approximation is

$$\frac{m}{2\pi\hbar^2}\left|\int d^3\vec{r}\,'\,\frac{1}{r'}e^{ik(r'+z')}V(\vec{r}\,')\right| \;\ll\; 1. \tag{20.4.4}$$

Now, we make the further approximation of taking $k \approx 0$ and replacing the potential by its average over $r' < a$, a constant factor that we will call $\overline{V}$. The rest of the integral is then

$$\int d^3\vec{r}\,'\,\frac{1}{r'} \;=\; 4\pi\int_0^a dr'\,r' = 2\pi a^2. \tag{20.4.5}$$

Thus our estimate for the validity condition of the Born approximation is

$$\frac{ma^2}{\hbar^2}|\overline{V}| \;\ll\; 1, \tag{20.4.6}$$

where $a$ and $\overline{V}$ are the characteristic range and strength of the potential. This could have been guessed merely on dimensional grounds, at least in the low-energy limit of small $k$. However, this estimate for the range of validity of the Born approximation can often be too conservative, particularly for larger $k$, because then the oscillation of the integrand can produce efficient cancellation in the left side of eq. (20.4.3), not accounted for in our rough estimate. In that case, dimensional analysis suggests that the weaker condition $ma|\overline{V}|/\hbar^2 k \ll 1$ could suffice.

Let us now consider examples of the leading Born approximation. An important special case is that of a spherically symmetric potential, so that the scattering amplitude has azimuthal

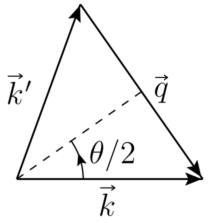

Figure 20.4.1: Geometry of vectors involved in elastic scattering from a potential. The initial and final wavevectors $\vec{k}$ and $\vec{k}'$ have equal length $k$. The vector $\vec{q} = \vec{k} - \vec{k}'$, where $\hbar\vec{q}$ is the momentum transferred from the initial particle to the scattering potential, has length $q = 2k\sin(\theta/2)$, where $\theta$ is the angle between $\vec{k}$ and $\vec{k}'$.

symmetry (no dependence on $\phi$). Define

$$\vec{q} \;=\; \vec{k} - \vec{k}', \tag{20.4.7}$$

so that $\hbar\vec{q}$ is the momentum transferred from the initial particle to the scattering potential. (See Figure 20.4.1.) Then $f_k(\theta)$ only depends on $k$ and $\theta$ through the combination

$$q = |\vec{k} - \vec{k}'| = 2k\sin(\theta/2). \tag{20.4.8}$$

Indeed, eq. (20.4.2) becomes

$$f_k(\theta) \;=\; -\frac{m}{2\pi\hbar^2}\int d^3\vec{r}'\, V(r')\, e^{iqr'\cos\theta'}, \tag{20.4.9}$$

where we have taken advantage of the spherical symmetry to choose our $\vec{r}'$ coordinate system in the integral so that $\vec{q}$ is in the $\cos\theta' = 0$ direction. It follows that

$$f_k(\theta) \;=\; -\frac{m}{2\pi\hbar^2}\int_0^\infty dr' r'^2 \int_0^{2\pi} d\phi' \int_{-1}^{1} d(\cos\theta')\, V(r')\, e^{iqr'\cos\theta'} \tag{20.4.10}$$

$$\;=\; \frac{im}{\hbar^2 q}\int_0^\infty dr' r' V(r')\big(e^{iqr'} - e^{-iqr'}\big), \tag{20.4.11}$$

which we can rewrite, without the distracting primes, as

$$f_k(\theta) \;=\; -\frac{2m}{\hbar^2 q}\int_0^\infty dr\, rV(r)\sin(qr). \tag{20.4.12}$$

This is the first-order Born approximation result for the scattering amplitude with an arbitrary spherically symmetric potential. Note that in this case, $f_k(\theta)$ is always real. In the low-energy approximation $q \to 0$, one can use $\sin(qr) \approx qr$ to find that

$$f_k(\theta) \;=\; -\frac{2m}{\hbar^2}\int_0^\infty dr\, r^2 V(r), \qquad \text{(small } q \text{ limit)}, \tag{20.4.13}$$

which is independent of $q$, and therefore independent of $\theta$. In the opposite limit of very large $q$, eq. (20.4.12) gives $f_k(\theta) \to 0$ because of the $q$ in the denominator together with the cancellation due to the very rapid oscillation of $\sin(qr)$ in the integrand.

As an example, consider the **Yukawa potential**, named after Hideki Yukawa, who used it in 1935 to model the strong interactions between nucleons due to the exchange of mesons. It is

$$V(r) \;=\; -\frac{g}{r}e^{-r/a}, \tag{20.4.14}$$

where $g$ and $a$ are constants with units of [(energy)(length)] and [length], respectively. For $r \ll a$, this has the same form as a Coulomb potential. However, for large $r \gg a$, the potential approaches 0 exponentially faster than the Coulomb potential does. For that reason it is sometimes also known as the **screened Coulomb potential**, and $a$ can be thought of as the range of the potential. Applying eq. (20.4.12) gives

$$f_k(\theta) \;=\; \frac{2mg}{\hbar^2 q}\int_0^\infty dr\, e^{-r/a}\sin(qr) \;=\; \frac{2mg}{\hbar^2}\frac{1}{q^2+1/a^2}. \tag{20.4.15}$$

Now we can use $q^2 = 4k^2\sin^2(\theta/2) = 2k^2(1-\cos\theta)$, to find

$$\frac{d\sigma}{d\Omega} \;=\; |f_k(\theta)|^2 \;=\; \left(\frac{2mg}{\hbar^2}\right)^2 \frac{1}{[2k^2(1-\cos\theta)+1/a^2]^2} \tag{20.4.16}$$

in the Born approximation. If we take the long-range limit $a \to \infty$ and let $-g = Z_1 Z_2 e^2$, then

$$\frac{d\sigma}{d\Omega} \;=\; \frac{Z_1^2 Z_2^2 e^4 m^2}{4\hbar^4 k^4 \sin^4(\theta/2)} \;=\; \frac{Z_1^2 Z_2^2 e^4}{16E^2\sin^4(\theta/2)}, \tag{20.4.17}$$

where in the last expression we used $E = \hbar^2 k^2/2m$. This is the same differential scattering cross-section as we found for the corresponding classical Rutherford scattering problem in section 20.1.

In the limit of low-energy scattering from the Yukawa potential, the rough condition for validity of the Born approximation is obtained by taking $\overline{V} = g/a$ in eq. (20.4.6), so

$$|g| \;\ll\; \hbar^2/ma. \tag{20.4.18}$$

This can be compared to the condition necessary for a bound state to exist, which can be shown (for example, using an appropriate trial wavefunction for the ground state and applying the variational principle) to be approximately $g \;>\; 0.84\hbar^2/ma$. The general lesson is that if the potential allows a bound state, then it is too strong for the Born approximation to be valid for low $E$. However, it can be valid for larger $E$, meaning $k \gg 1/a$.

Returning to the general scattering problem, let us consider how to extend the Born approximation beyond leading order. We can start with the expression eq. (20.2.34) for the transition operator, and use the expansion $(I - X)^{-1} = I + X + X^2 + X^3 + \cdots$. The result is

$$T \;=\; V + V\frac{1}{E - H_0 + i\epsilon}V + V\frac{1}{E - H_0 + i\epsilon}V\frac{1}{E - H_0 + i\epsilon}V + \cdots . \tag{20.4.19}$$

The ordering of operators matters, because $H_0$ and $V$ do not commute unless the latter is a trivial (constant) potential. Recalling, from eq. (20.2.31), that

$$f_k(\theta,\phi) \;=\; -\frac{4\pi^2 m}{\hbar^2}\langle \vec{k}'|T|\vec{k}\rangle, \tag{20.4.20}$$

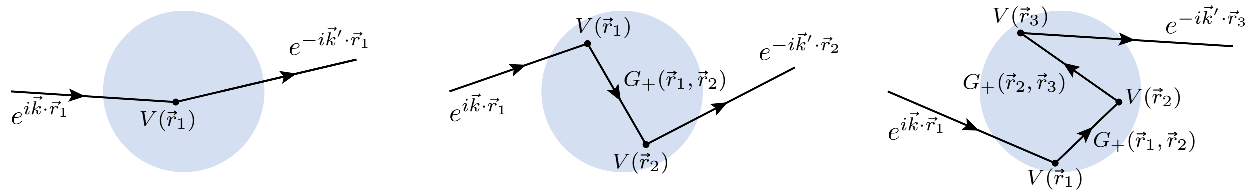

Figure 20.4.2: Diagrammatic representation of the first three terms in the Born approximation expansion for the scattering amplitude, $f_k^{(1)}$, $f_k^{(2)}$, and $f_k^{(3)}$ from eqs. (20.4.25)–(20.4.27). Each of the vertex positions $\vec{r}_n$ is integrated over, with an associated factor of $2m/\hbar^2$.

we obtain an expansion in powers of the potential,

$$f_k(\theta,\phi) \;=\; f_k^{(1)}(\theta,\phi) + f_k^{(2)}(\theta,\phi) + f_k^{(3)}(\theta,\phi) + \cdots\,, \tag{20.4.21}$$

where $f^{(N)}(\theta,\phi)$ has $N$ factors of the potential $V$. In particular,

$$f_k^{(1)}(\theta,\phi) \;=\; -\frac{4\pi^2 m}{\hbar^2}\langle\vec{k}'|V|\vec{k}\rangle \tag{20.4.22}$$

is just the leading Born approximation that we have already studied, and the second-order and third-order contributions in the Born expansion are now seen to be

$$f_k^{(2)}(\theta,\phi) \;=\; -\frac{4\pi^2 m}{\hbar^2}\langle\vec{k}'|V\frac{1}{E-H_0+i\epsilon}V|\vec{k}\rangle, \tag{20.4.23}$$

$$f_k^{(3)}(\theta,\phi) \;=\; -\frac{4\pi^2 m}{\hbar^2}\langle\vec{k}'|V\frac{1}{E-H_0+i\epsilon}V\frac{1}{E-H_0+i\epsilon}V|\vec{k}\rangle, \tag{20.4.24}$$

etc. Note that the leading Born approximation just amounts to setting $T = V$ in eq. (20.4.20).

We can now work out these contributions in the wavefunction representation, in terms of the Green function, by using completeness with respect to position and eqs. (20.2.9) and (20.2.8) for the plane wave and Green function. The results are

$$f_k^{(1)}(\theta,\phi) \;=\; -\frac{1}{4\pi}\left(\frac{2m}{\hbar^2}\right)\int d^3\vec{r}_1\, e^{i\vec{k}\cdot\vec{r}_1}V(\vec{r}_1)e^{-i\vec{k}'\cdot\vec{r}_1}, \tag{20.4.25}$$

$$f_k^{(2)}(\theta,\phi) \;=\; -\frac{1}{4\pi}\left(\frac{2m}{\hbar^2}\right)^2\int d^3\vec{r}_1\int d^3\vec{r}_2\, e^{i\vec{k}\cdot\vec{r}_1}V(\vec{r}_1)G_+(\vec{r}_1,\,\vec{r}_2)V(\vec{r}_2)e^{-i\vec{k}'\cdot\vec{r}_2}, \tag{20.4.26}$$

$$f_k^{(3)}(\theta,\phi) \;=\; -\frac{1}{4\pi}\left(\frac{2m}{\hbar^2}\right)^3\int d^3\vec{r}_1\int d^3\vec{r}_2\int d^3\vec{r}_3\, e^{i\vec{k}\cdot\vec{r}_1}V(\vec{r}_1)G_+(\vec{r}_1,\,\vec{r}_2)V(\vec{r}_2)G_+(\vec{r}_2,\,\vec{r}_3)V(\vec{r}_3)e^{-i\vec{k}'\cdot\vec{r}_3}. \tag{20.4.27}$$

A diagrammatic version of these contributions is shown in Figure 20.4.2. The formula for $f_k^{(N)}(\theta,\phi)$ contains $N$ vertices where the scattering potential acts. At each scattering vertex, labeled $n = 1,\ldots,N$, there is an integration

$$\text{vertex} \;\leftrightarrow\; \frac{2m}{\hbar^2}\int d^3\vec{r}_n\, V(\vec{r}_n), \tag{20.4.28}$$

and we assign factors

$$\text{internal lines} \quad \leftrightarrow \quad G_+(\vec{r}_n,\, \vec{r}_{n+1}) \tag{20.4.29}$$

to the lines between consecutive vertices $\vec{r}_n$ and $\vec{r}_{n+1}$. There are also two factors

$$\text{external lines} \quad \leftrightarrow \quad e^{i\vec{k}\cdot\vec{r}_1} e^{-i\vec{k}'\cdot\vec{r}_N} \tag{20.4.30}$$

associated with the initial and final momenta, and a single factor of $-1/4\pi$.

An intuitive interpretation for the Born expansion is that the particle is described by plane waves except for interactions with the potential. Between potential interactions, it propagates according to the Green function of the Helmholtz operator, and the initial and final wavefunctions are free plane-waves. All possible numbers of interactions are summed over, and then all possible interaction positions are integrated over, to give the total scattering amplitude.

## 20.5 Spherical potential scattering and the partial wave expansion

Consider scattering from a spherically symmetric potential $V(\vec{r}) = V(r)$. As usual, we take the initial particle wavenumber to be $\vec{k} = \hat{z}k$, so that we have azimuthal symmetry and the scattering amplitude can only depend on $\theta$, not $\phi$. This implies that we can write a **partial wave expansion** in the orbital angular momentum quantum number $l$,

$$f_k(\theta) \;=\; \sum_{l=0}^{\infty} (2l+1) P_l\left(\cos\theta\right) f_l(k). \tag{20.5.1}$$

The $P_l(\cos\theta)$ are the Legendre polynomials, and the factor of $2l+1$ is a convention. This equation defines the quantities $f_l(k)$, called **partial wave amplitudes**, that appear on the right side. They can equivalently be viewed as functions $f_l(E)$ of the energy $E = \hbar^2 k^2/2m$.

To better understand the partial wave expansion, recall the expansion of a plane wave in the absence of scattering, worked out at the end of section 9.2,

$$e^{ikz} \;=\; e^{ikr\cos\theta} \;=\; \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta)\, i^l j_l(kr). \tag{20.5.2}$$

The $j_l(kr)$ are spherical Bessel functions. For large $r$, we find from eq. (9.2.18) that

$$i^l j_l(kr) \;\approx\; \frac{e^{ikr} - e^{-ikr+i\pi l}}{2ikr}, \qquad\qquad (r \gg 1/k). \tag{20.5.3}$$

Plugging eqs. (20.5.1)-(20.5.3) into eq. (20.2.27) gives the scattering wavefunction

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}} \left[ e^{ikz} + f_k(\theta) \frac{e^{ikr}}{r} \right] \tag{20.5.4}$$

$$=\; \frac{1}{(2\pi)^{3/2}} \left[ \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) \left( \frac{e^{ikr} - e^{-ikr+i\pi l}}{2ikr} \right) + \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) f_l(k) \frac{e^{ikr}}{r} \right], \tag{20.5.5}$$

where the large $r$ condition is implicit. Collecting like terms, we find

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2} 2ik} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) \left( [1 + i2kf_l(k)] \frac{e^{ikr}}{r} + (-1)^{l+1} \frac{e^{-ikr}}{r} \right) \tag{20.5.6}$$

for $r \to \infty$. The terms proportional to $e^{ikr}$ and $e^{-ikr}$ are outgoing and incoming spherical waves, respectively; compare to the discussion around eqs. (9.2.25)–(9.2.28). They are called the **partial waves** of angular momentum $l$. Note that the outgoing spherical waves depend on the potential $V(r)$ through $f_l(k)$. In contrast, the incoming spherical waves do not depend on $f_l(k)$ and therefore do not depend on $V(r)$; they are just a re-writing of the incoming beam.

We are about to prove a crucial fact called **partial wave unitarity**: the factor $1 + 2ikf_l(k)$ in eq. (20.5.6) has unit norm, so we can write

$$1 + i2kf_l(k) \;=\; e^{i2\delta_l}. \tag{20.5.7}$$

This defines a set of real numbers $\delta_l$ called the **phase shifts** of the angular momentum $l$ partial waves. (The factor of 2 in the exponent is a traditional normalization convention.) Equation (20.5.7) implies that the effect of the scattering potential is just that each outgoing partial wave of orbital angular momentum $l$ acquires a phase relative to the corresponding incoming partial wave, parameterized by the phase shift $\delta_l$.

To prove that eq.(20.5.7) holds with real $\delta_l$, we use conservation of probability, based on the assumption that the potential $V(r)$ just scatters the incoming particles without absorbing them. For any wavefunction solution $\psi(\vec{r}, t)$ to the Schrödinger equation, consider the following integration over the volume of a sphere at very large constant $r = R$,

$$0 \;=\; -\frac{d}{dt} \int_{r<R} d^3\vec{r}\, |\psi(\vec{r}, t)|^2 \;=\; \int_{r<R} d^3\vec{r}\, \vec{\nabla} \cdot \vec{J} \;=\; \oint_{r=R} d\vec{a} \cdot \vec{J} \;=\; R^2 \int d\Omega\, J_r \Big|_{r=R}. \tag{20.5.8}$$

Here, the first equality expresses that the conservation of probability dictates that the norm of the state is fixed in time, the second equality uses the local conservation of probability equation (5.6.3) in terms of the probability current vector $\vec{J}$ defined in eq. (5.6.6), and the third equality uses the Divergence Theorem to express the result in terms of an integral over the surface of the sphere, with an outward-pointing area element $d\vec{a}$. Therefore, using

$$J_r = \frac{i\hbar}{2m} \left( \psi \frac{\partial \psi^*}{\partial r} - \psi^* \frac{\partial \psi}{\partial r} \right) \tag{20.5.9}$$

from eq. (5.6.6), we obtain

$$\lim_{r\to\infty} r^2 \int d\Omega\, \mathrm{Im} \left( \psi^* \frac{\partial \psi}{\partial r} \right) \;=\; 0. \tag{20.5.10}$$

Since we are considering a spherically symmetric potential, we can consider separately each common eigenstate of the Hamiltonian and the total angular momentum. So, applying to any stationary-state wavefunction of the form (for large $r$),

$$\psi_l(\vec{r}) \;=\; (2l+1)P_l(\cos\theta)\left[a_l\frac{e^{ikr}}{r} + b_l\frac{e^{-ikr}}{r}\right], \tag{20.5.11}$$

we find from eq. (20.5.10),

$$|a_l|^2 - |b_l|^2 \;=\; 0, \tag{20.5.12}$$

in which it is notable that the cross-terms containing $a_l^* b_l$ and $a_l b_l^*$ have canceled. Now, in our case of eq. (20.5.6), we have $|b_l| = 1$ for each $l$ in the superposition (not counting the overall normalization outside of the summation sign), so $|1+2ikf_l(k)| = 1$ also, establishing eq. (20.5.7). The intuitive explanation is that the magnitude of each incoming partial wave of fixed angular momentum $l$ must be equal to the magnitude of the corresponding outgoing partial wave, since the potential is not creating or destroying particles.

The general scattering problem for a spherically symmetric potential thus reduces to finding the phase shifts $\delta_l$ appearing in the wavefunction for large $r$ obtained by re-writing eq. (20.5.6),

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}}\sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta)\left[e^{i2\delta_l}e^{ikr} - e^{-i(kr-l\pi)}\right]/2ikr, \qquad (r\to\infty). \tag{20.5.13}$$

In the limit that the potential $V(r)$ vanishes, then $\delta_l = 0$, and more generally the $\delta_l$ parameterize the effect of $V(r)$ on each $l$ partial wave. We will show before the end of this section that it is a sufficient approximation to use only the lowest few partial waves $l = 0, 1, 2, \ldots$, except in the case of very high-energy scattering.

Some useful equivalent ways of writing the relation between the phase shifts and the partial wave amplitudes $f_l(k)$ are, from eq. (20.5.7),

$$f_l(k) \;=\; \frac{e^{i2\delta_l}-1}{2ik} \;=\; \frac{e^{i\delta_l}\sin\delta_l}{k} \;=\; \frac{1}{k(\cot\delta_l - i)}. \tag{20.5.14}$$

Using the next-to-last expression of eq. (20.5.14) in eq. (20.5.1), we have the scattering amplitude in terms of the phase shifts,

$$f_k(\theta) \;=\; \frac{1}{k}\sum_{l=0}^{\infty}(2l+1)e^{i\delta_l}\,\sin\delta_l\,P_l(\cos\theta). \tag{20.5.15}$$

This can be used to compute the total cross-section:

$$\sigma = 2\pi\int_{-1}^{1} d(\cos\theta)|f_k(\theta)|^2 \tag{20.5.16}$$

$$= \frac{2\pi}{k^2}\sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}(2l+1)(2l'+1)e^{i(\delta_l-\delta_{l'})}\sin(\delta_l)\sin(\delta_{l'})\int_{-1}^{1}d(\cos\theta)P_{l'}(\cos\theta)P_l(\cos\theta). \tag{20.5.17}$$

To simplify this, apply the Legendre polynomial orthogonality condition $\int_{-1}^{1} du P_{l'}(u) P_l(u) = \delta_{l,l'} 2/(2l+1)$. We arrive at

$$\sigma \;=\; \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l. \tag{20.5.18}$$

Note that even though the differential cross-section $d\sigma/d\Omega = |f_k(\theta)|^2$ certainly has interferences between partial wave amplitudes with different $l$, there is no such interference in the total cross-section. As a check, or an alternative derivation, the Optical Theorem tells us that

$$\sigma \;=\; \frac{4\pi}{k} \mathrm{Im}\big[f_k(\theta = 0)\big] \;=\; \frac{4\pi}{k} \left( \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) P_l(1) \sin(\delta_l) \mathrm{Im}[e^{i\delta_l}] \right), \tag{20.5.19}$$

which reproduces eq. (20.5.18) after using $\mathrm{Im}[e^{i\delta_l}] = \sin \delta_l$ and the Legendre polynomial normalization condition $P_l(1) = 1$.

An important consequence of eq. (20.5.14), as encoded in eq. (20.5.18), follows if we write

$$\sigma \;=\; \sum_{l=0}^{\infty} \sigma_l, \tag{20.5.20}$$

where $\sigma_l$ is the contribution from orbital angular momentum $l$. Then $\sin^2 \delta_l \leq 1$ implies

$$\sigma_l \;\leq\; \frac{4\pi}{k^2} (2l+1). \tag{20.5.21}$$

This is known as the **partial wave unitarity bound**. Equality is achieved only if $\sin^2 \delta_l = 1$, which implies $\delta_l = (n + 1/2)\pi$ for integer $n$. In that case, $\sigma_l$ is said to saturate the partial wave unitarity bound. Remarkably, this bound on the contribution $\sigma_l$ for each partial wave applies no matter how strong the scattering potential is. It becomes stronger with increasing energy, as it can be rewritten

$$\sigma_l \;\leq\; \frac{2\pi\hbar^2}{mE} (2l+1), \tag{20.5.22}$$

and it should be viewed as a requirement imposed by the conservation of probability.

Partial wave unitarity also restricts the values that $f_l(k)$ can take to a circle of radius $1/2k$ in the complex plane, as given by the first equality of eq. (20.5.14) and illustrated in Figure 20.5.1. If $\delta_l$ is small, then $f_l(k)$ is near the origin, and is almost purely real. This is realized in the leading Born approximation limit. If instead $|f_l(k)|$ is maximal, so that the partial wave unitarity bound eq. (20.5.21) is saturated, then $f_l(k)$ is at the top of the circle and is almost pure imaginary, equal to $i/k$.

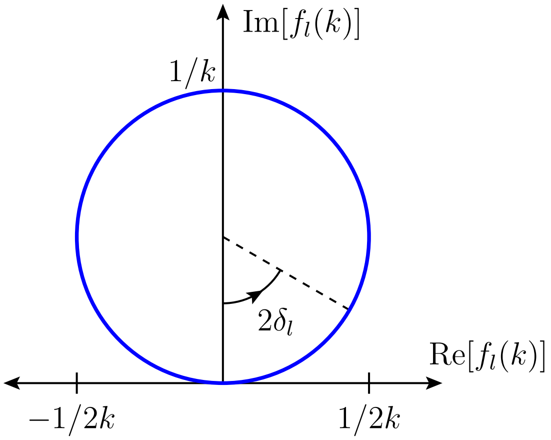

Figure 20.5.1: Partial wave unitarity illustrated: the possible values of the partial wave amplitude $f_l(k) = (e^{i2\delta_l} - 1)/2ik$ lie on a circle with radius $1/2k$ in the upper-half complex plane, and are related to the phase shift angle $\delta_l$ as shown. The Born approximation limit is realized near the bottom of the circle, for small $\sin\delta_l$, which gives small and nearly real $f_l(k)$. The partial wave unitarity bound eq. (20.5.21) is saturated for $f_l(k) \approx i/k$, near the top of the circle, with $\sin^2\delta_l = 1$, so $\delta_l = (n + 1/2)\pi$ for integer $n$.

Consider potentials that vanish completely[†] outside of a finite range, so that $V(r) = 0$ for $r > R$. Then the exact (not just the large $r$ limit) wavefunction for a stationary scattering state can be written as, taking into account the azimuthal symmetry,

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) F_l(r), \tag{20.5.23}$$

where the radial wavefunction for the angular momentum $l$ partial wave is

$$F_l(r) \;=\; \begin{cases} A_l(r) & (\text{for } r \leq R), \\ B_l(r) & (\text{for } r \geq R). \end{cases} \tag{20.5.24}$$

The exact form of the function $A_l(r)$ may be quite difficult to obtain, since it depends on the scattering potential. However, since we are taking the potential to vanish for $r \geq R$, the function $B_l(r)$ must be a superposition of the free-particle solutions that we found in section 9.2,

$$B_l(r) \;=\; c_l h_l^{(1)}(kr) + d_l h_l^{(2)}(kr), \tag{20.5.25}$$

where $c_l$ and $d_l$ are constant coefficients, and $h_l^{(1)}$ and $h_l^{(2)}$ are the spherical Hankel functions

$$h_l^{(1)}(kr) \;=\; j_l(kr) + i n_l(kr) \;\sim\; \frac{e^{i(kr - l\pi/2)}}{ikr}, \tag{20.5.26}$$

$$h_l^{(2)}(kr) \;=\; j_l(kr) - i n_l(kr) \;\sim\; -\frac{e^{-i(kr - l\pi/2)}}{ikr}, \tag{20.5.27}$$

where the large $r$ asymptotic forms are shown. Matching to our scattering wavefunction for large $r$, eq. (20.5.13), we can solve for the constants $c_l$ and $d_l$ in terms of the phase shifts $\delta_l$. The result is

$$B_l(r) \;=\; \frac{1}{2} e^{i\pi l/2} \left[ e^{i2\delta_l} h_l^{(1)}(kr) + h_l^{(2)}(kr) \right], \tag{20.5.28}$$

---

[†] This is done mostly for convenience; such potentials are not completely realistic. However, the results we are about to obtain are often a good approximation, provided that the potential $V(r)$ is instead suitably small at $r = R$ and decreases for large $r$ faster than $1/r$. This can often be realized just by choosing $R$ large enough.

or equivalently,

$$B_l(r) \;=\; e^{i\pi l/2} e^{i\delta_l} \left[\cos(\delta_l)\, j_l(kr) - \sin(\delta_l)\, n_l(kr)\right]. \tag{20.5.29}$$

To obtain scattering state solutions, one can now adopt the following strategy. First, the difficult part: solve the time-independent Schrödinger differential equation for $A_l(r)$ in the region of non-zero potential, $r \leq R$. Because this is a second-order differential equation in $r$, in general there will be two linearly independent solutions with arbitrary coefficients. To fix the coefficients, match these solutions to eq. (20.5.29) at $r = R$, using continuity of the wavefunction,

$$A_l(R) \;=\; B_l(R), \tag{20.5.30}$$

and, if the potential is finite at $r = R$, continuity of the derivative

$$A_l'(R) \;=\; B_l'(R). \tag{20.5.31}$$

Since the $B_l(R)$ are given in eq. (20.5.29) in terms of known functions, one can solve for the $\delta_l$. To write the solution as compactly as possible, it is useful to define the quantities

$$\alpha_l \equiv A_l'(R)/A_l(R), \tag{20.5.32}$$

in which the overall normalization cancels. Then, since eqs. (20.5.30) and (20.5.31) imply

$$\alpha_l = B_l'(R)/B_l(R), \tag{20.5.33}$$

we obtain the general solution for the phase shift using eq. (20.5.29),

$$\tan\delta_l \;=\; \frac{\alpha_l j_l(kR) - k j_l'(kR)}{\alpha_l n_l(kR) - k n_l'(kR)}. \tag{20.5.34}$$

Summarizing: once the $\alpha_l$ have been found using eq. (20.5.32), the phase shifts can be calculated from eq.(20.5.34), and then used in eq. (20.5.15) to obtain $f_k(\theta)$, which in turn gives the differential cross-section $|f_k(\theta)|^2$.

The expansions of the spherical Bessel and Neumann functions as given in eqs. (9.2.15) and (9.2.16) can now be used to expand eq. (20.5.34) as a series in $kR$. At leading order in the expansion, the result is

$$\tan\delta_l \;=\; \frac{(kR)^{2l+1}(l - R\alpha_l)}{(2l-1)!!\,(2l+1)!!\,(l+1+R\alpha_l)}. \tag{20.5.35}$$

For $l = 0, 1, 2, 3$, and 4, the denominator factor $(2l-1)!!\,(2l+1)!!$ is respectively $1, 3, 45, 1575$, and 99225, so the coefficient of $(kR)^{2l+1}$ gets smaller rapidly as $l$ increases. For low and intermediate energies (meaning as long as $kR$ is not very large compared to 1), it is therefore justified to keep only the first few partial wave contributions, as we had promised to show earlier.

Let us now specialize to the low energy limit, $kR \ll 1$, where it is justified to neglect all but the $s$-wave phase shift $\delta_0$ contribution, which behaves like $\tan\delta_0 \propto k$ and provides an isotropic differential cross-section. Expanding the $l = 0$ case of eq. (20.5.34) to the next-to-leading order in $k$, the result can be parameterized in the form

$$k\cot\delta_0 \;=\; -\frac{1}{a} + \frac{1}{2}r_0 k^2 + \cdots\,, \tag{20.5.36}$$

which defines length scales $a$ and $r_0$, called the **scattering length** and the **effective range** respectively. In terms of $R$ and $\alpha_0 = A_0'(R)/A_0(R)$, one finds

$$a \;=\; \frac{\alpha_0 R^2}{1+\alpha_0 R} \tag{20.5.37}$$

and the determination of $r_0$ is left as an exercise. The scattering length $a$ can be either positive or negative depending on the sign of $\alpha_0$, and can be much larger in magnitude than the length scale $R$ built into the potential, as we will see. In terms of these parameters, we can evaluate the low-energy, and therefore $s$-wave, cross-section from eq. (20.5.18),

$$\sigma \;=\; \frac{4\pi}{k^2}\frac{1}{1+\cot^2\delta_0} \;=\; \frac{4\pi a^2}{1+(1-r_0/a)a^2k^2}, \tag{20.5.38}$$

where terms of higher order than $k^2$ have been consistently neglected in the denominator. In the extreme low-energy approximation, $\sigma$ has a simple dependence on the scattering length,

$$\sigma \;\approx\; 4\pi a^2. \tag{20.5.39}$$

In principle, $a$, $r_0$, and higher order contributions to eq. (20.5.36) can be obtained experimentally from the energy dependence and interference effects in the low-energy limit.

## 20.6 Bound states, resonances, and poles in scattering amplitudes

The partial wave scattering amplitudes and cross-sections, viewed as functions of the energy $E$ analytically continued to the complex domain, have an analytic structure that conveys information about the potential. As we will explain in this section, a pole (divergence) in the partial wave scattering amplitude corresponds either to a true bound state, or to a **resonance**, also known as a **quasi-bound state**, of the potential $V(r)$.

First, consider the case that the potential has one or more discrete bound states. We assume that the constant part of the potential has been fixed so that it vanishes at large distances, $V(\infty) = 0$, so that the condition for a bound state is an isolated eigenvalue $E_{\text{bound}} < 0$ for the time-independent Schrödinger equation. To uncover the connection between such states and the scattering amplitude, consider the unbound scattering solutions for the wavefunction $\psi_E(\vec{r})$ with positive $E = \hbar^2 k^2/2m$. Here $k$ is the wavenumber far from the origin as discussed in section 20.1 and 20.2, with the asymptotic form of the wavefunction for large $r$ as given in eq. (20.5.13). Suppose we try making the replacement $k \rightarrow i\kappa$ in that wavefunction, where $\kappa$ is now taken to be a real number. It follows that the solution so obtained, if it exists, will have energy $E_{\text{bound}} = -\hbar^2\kappa^2/2m < 0$ and therefore will be a bound state.

The only obstacle to constructing bound state wavefunction solutions in this way is that they will almost always be hopelessly non-normalizable, and therefore unphysical. To see this, note that the asymptotic form for large $r$ given in eq. (20.5.13) will be, after $k \rightarrow i\kappa$,

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) \left[-e^{i2\delta_l} e^{-\kappa r} + (-1)^l e^{\kappa r}\right] /2\kappa r, \qquad (r \rightarrow \infty). \tag{20.6.1}$$

The term proportional to $e^{-\kappa r}$ is sensible, but the term $e^{\kappa r}$ evidently gives a probability density that diverges exponentially for large $r$. It appears that this wavefunction would describe a physically absurd state in which the probability to find the particle in any finite region is vanishingly small compared to the probability to find it arbitrarily far away from the origin.

However, this solution can be salvaged for special, isolated values of the energy. Since we are studying spherically symmetric potential, we can look for eigenstates of the total orbital angular momentum. Taking a term of fixed $l$ in eq. (20.6.1), and re-normalizing the wavefunction by multiplying by a constant factor $-2(2\pi)^{3/2} e^{-i2\delta_l}/(2l+1)$, we have

$$\psi_l(\vec{r}) \;=\; P_l(\cos\theta) \left[e^{-\kappa r} + \frac{(-1)^{l+1}}{e^{i2\delta_l}} e^{\kappa r}\right] /\kappa r, \qquad (r \rightarrow \infty). \tag{20.6.2}$$

In order to re-interpret this wavefunction as a sensible (normalizable) energy and angular momentum eigenstate wavefunction at large $r$, it is necessary to eliminate the offending second term by requiring

$$e^{i2\delta_l} \rightarrow \infty. \tag{20.6.3}$$

We conclude that a true bound state for the potential $V(r)$ with negative energy $E_{\text{bound}}$ must correspond to a pole in $e^{i2\delta_l}$, and thus in the scattering amplitude $f_l(k) = (e^{i2\delta_l} - 1)/(2ik)$, when viewed as a function of energy analytically continued to $E = E_{\text{bound}} < 0$. The orbital angular momentum quantum number of the bound state is $l$. So, one can relate the bound state energies, and even their angular momenta, to the poles in the partial-wave scattering amplitudes.

Since scattering solutions always have positive energy, and the bound states always have negative energy, one never actually hits the bound state pole directly in scattering experiments.

Indeed, this is clear from the fact that for positive $E$, partial wave unitarity tells us that $e^{i2\delta_l}$ has norm 1. Nevertheless, the presence of the bound states can be inferred from the functional dependence of $e^{i2\delta_l}$ (and thus the cross-section) on $E$, particularly if $-E_{\text{bound}}$ is small.

As an example, consider low-energy (small $k$) scattering as discussed in equations (20.5.36)–(20.5.39) of the previous section. Taking $k \rightarrow i\kappa$ in eq. (20.5.36), we see from eq. (20.5.14) that the $s$-wave scattering amplitude can have a pole at the solution to

$$\kappa \;=\; \frac{1}{a} + \frac{1}{2} r_0 \kappa^2, \tag{20.6.4}$$

with $\kappa$ positive to ensure that the bound-state wavefunction is well-behaved at $r \rightarrow \infty$. Thus, if the scattering length $a$ is found to be positive and not too small, one can predict a weakly bound state with $\kappa \approx 1/a$ and energy close to

$$E_{\text{bound}} \;\approx\; -\frac{\hbar^2}{2ma^2}. \tag{20.6.5}$$

Including the effect of $r_0$, this is more precisely

$$E_{\text{bound}} \;\approx\; -\frac{\hbar^2 \kappa^2}{2m}, \tag{20.6.6}$$

where $\kappa$ is the smaller solution to the quadratic equation (20.6.4), approximately

$$\kappa \;\approx\; \frac{1}{a} + \frac{r_0}{2a^2}. \tag{20.6.7}$$

The low-energy $s$-wave cross-section eq. (20.5.38) in the presence of a weakly bound state with energy $E_{\text{bound}} < 0$ can now be written in the convenient form

$$\sigma_0 \;\approx\; \frac{4\pi a^2}{1 - E/E_{\text{bound}}}, \tag{20.6.8}$$

as dictated by the $E = 0$ limit found in eq. (20.5.39), together with the requirement of a pole at $E = E_{\text{bound}}$.

Returning to the case of general (not necessarily small) $k$, the bound-state energies are not the only possible poles in the partial wave amplitude. To see this, let us look for peaks in the cross-section. The partial wave cross-section from eq. (20.5.18) can be rewritten as

$$\sigma_l \;=\; \frac{4\pi}{k^2} \frac{2l+1}{1 + \cot^2 \delta_l}, \tag{20.6.9}$$

where $\cot^2 \delta_l$ is often quite large, as for example in the Born approximation limit. However, if it happens for some $E$ that $\cot \delta_l \approx 0$, then $\sigma_l$ will peak at that energy, saturating the partial wave unitarity bound. Suppose that $E = E_{\text{res}}$ is a resonant energy that makes $\cot \delta_l = 0$. For energies that are close to this, we can expand $\cot \delta_l$ in $E - E_{\text{res}}$, so that

$$\cot(\delta_l) \;=\; \cot(\delta_l)\Big|_{E=E_{\text{res}}} + (E - E_{\text{res}}) \frac{d}{dE} \cot(\delta_l)\Big|_{E=E_{\text{res}}} + \mathcal{O}(E - E_{\text{res}})^2. \tag{20.6.10}$$

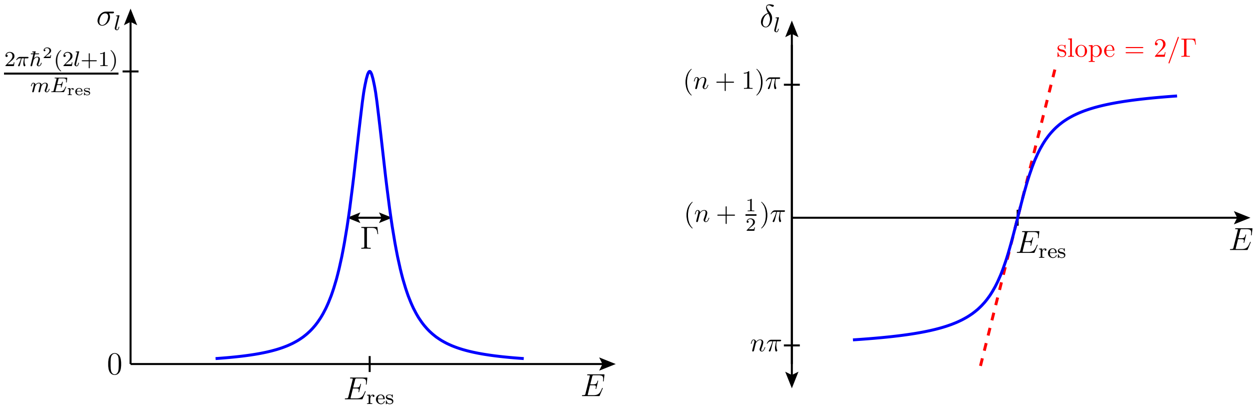

Figure 20.6.1: Behavior of the partial-wave cross-section $\sigma_l$ and the phase shift $\delta_l$ as a function of the scattering energy $E$ near a resonance pole $E_{\rm res} - i\Gamma/2$. In the left figure, the partial-wave cross-section is given by the Breit–Wigner lineshape. The phase shift $\delta_l$ increases by $\pi$ near each resonance, and for the lowest such resonance $n$ in the right figure can be taken to be 0.

The first term on the right vanishes by assumption, and we define the quantity $\Gamma$, with units of energy, according to

$$\frac{2}{\Gamma} \;=\; -\frac{d}{dE}\cot(\delta_l)\Big|_{E=E_{\rm res}}, \tag{20.6.11}$$

so that, near the resonant energy,

$$\cot(\delta_l) \;\approx\; -2(E-E_{\rm res})/\Gamma. \tag{20.6.12}$$

Using this in eq. (20.6.9) gives, for $E$ close to $E_{\rm res}$,

$$\sigma_l \;\approx\; \frac{4\pi(2l+1)}{k^2}\frac{\Gamma^2/4}{(E-E_{\rm res})^2+\Gamma^2/4}. \tag{20.6.13}$$

Excluding the energy dependence of the $1/k^2 = \hbar^2/2mE$ factor, this is the Breit–Wigner, or Lorentzian, lineshape with full width at half maximum (FWHM) $\Gamma$ for the quasi-bound state resonance with angular momentum $l$ and energy $E_{\rm res}$. It is depicted in Figure 20.6.1.

Near $E = E_{\rm res}$, eq. (20.6.12) tells us that the phase shift behaves like

$$\delta_l \;\approx\; \pi\,(n+1/2) + \tan^{-1}\left(\frac{E-E_{\rm res}}{\Gamma/2}\right), \tag{20.6.14}$$

for some integer $n$. This behavior is also sketched in Figure 20.6.1, showing that the phase shift rises through $\pi(n+1/2)$ near a resonant energy. These resonances correspond to poles in the partial wave scattering amplitude as a function of complex energy, at $E = E_{\rm res} - i\Gamma/2$, because

$$f_l(k) \;=\; \frac{1}{k\,(\cot\delta_l - i)} \approx -\frac{\Gamma/2k}{E-E_{\rm res}+i\Gamma/2}, \tag{20.6.15}$$

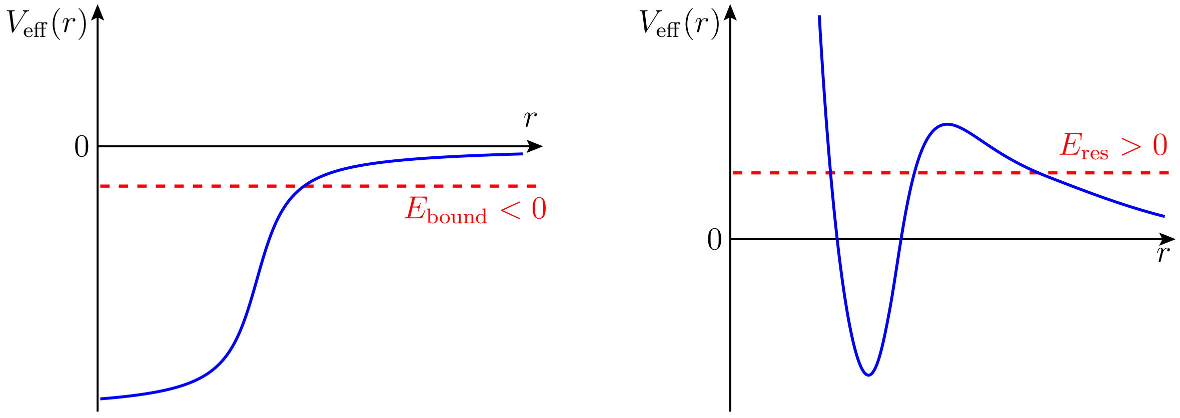

Figure 20.6.2: Sketches of example effective potentials $V_{\rm eff}(r)$ and energies for a true bound state with $l = 0$ (left) and a quasi-bound state resonance (right), which will cause the scattering amplitude and the cross-section to have poles at $E = E_{\rm bound} < 0$ or at $E = E_{\rm res} - i\Gamma/2$, respectively. In the latter case, the local minimum of the effective potential could be either negative (as shown) or positive, and could arise from the combination of an attractive potential $V(r) < 0$ and the repulsive centrifugal core $\hbar^2 l(l+1)/2mr^2$.

which follows from eq. (20.5.14). The assignment of complex energy $E_{\rm res} - i\Gamma/2$ is appropriate for an unstable state, as discussed at the end of section 5.6.

Figure 20.6.2 shows plausible potential shapes that could give rise to true bound state (left panel) and resonant quasi-bound state (right panel) poles in the scattering amplitude. Recall from our discussion at the end of section 9.4 that in three dimensions, the existence of a true bound state with energy $E_{\rm bound} < 0$ is not guaranteed unless the potential well is sufficiently deep. Unlike a true bound state, a quasi-bound state resonant energy $E_{\rm res}$ is positive. One way that such a quasi-bound state can arise is from the interplay between an attractive central potential $V(r)$ and the orbital angular momentum contribution to the effective potential for the radial wavefunction in the Schrödinger equation (recall the discussion in section 9.1),

$$V_{\rm eff}(r) \quad = \quad V(r) + \frac{\hbar^2 l(l+1)}{2mr^2}. \tag{20.6.16}$$

For $l \neq 0$, the centrifugal term is repulsive and grows stronger at small $r$. The potential shown in the right panel of Figure 20.6.2 has a finite barrier height between its minimum well and the asymptotic region at very large distances where it vanishes. Any corresponding state peaked inside the effective well with positive energy is not a true bound state; the wavefunction will always leak outside by tunneling. This is the reason for the terminology "quasi-bound state". Since a quasi-bound state resonant pole occurs only for complex $E$, it again can never be hit directly in scattering experiments, but can be inferred from the complex analytic singularity structure of the scattering amplitudes. The smaller $\Gamma$ is, the closer one can approach the

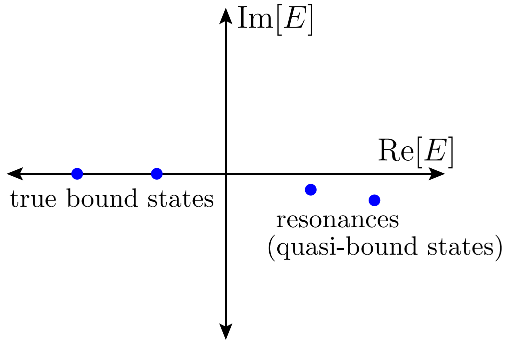

Figure 20.6.3: Positions of poles of the partial wave amplitudes and $e^{i2\delta_l}$, in the complex energy plane. Quasi-bound state resonances have poles in the analytic continuation to complex $E$ below the real $E$ axis, at $E = E_{\text{res}} - i\Gamma/2$. True bound states have poles on the negative real $E$ axis, at $E = E_{\text{bound}} < 0$. Physical scattering energies are only on the positive real $E$ axis. Therefore, these poles are never directly accessed in scattering, where $e^{i2\delta_l}$ is always finite with unit norm, but their influence can be seen in the dependence of the cross-section on energy.

resonance pole. However, $e^{i2\delta_l}$ always has unit norm for physical (real positive) values of the scattering energy $E$.

A schematic map of the two kinds of poles in the complex $E$ plane for $e^{i2\delta_l}$, and therefore for $f_l(k)$, are shown in Figure 20.6.3. They consist of true bound states on the negative real $E$ axis and quasi-bound states below the positive real $E$ axis.

## 20.7 Examples of scattering from spherical potentials

Consider the scattering of a particle from a hard (impenetrable) sphere, so that the potential is

$$V(r) \;=\; \begin{cases} \infty & (\text{for } r \leq R), \\ 0 & (\text{for } r \geq R). \end{cases} \tag{20.7.1}$$

This is the quantum mechanical version of the classical scattering of a particle from a hard sphere, discussed in section 20.1. In this particular case, it is not necessary to use eqs. (20.5.31)–(20.5.34), because the solution for $r \leq R$ is trivial,

$$\psi(\vec{r}) \;=\; 0 \qquad\qquad (r \leq R), \tag{20.7.2}$$

corresponding to $A_l(r) = 0$. Meanwhile, for $r \geq R$, we write

$$\psi(\vec{r}) \;=\; \frac{1}{(2\pi)^{3/2}} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) B_l(r), \tag{20.7.3}$$

with $B_l(r)$ given by eq. (20.5.29). To match to the vanishing solution at $r = R$, we need $B_l(R) = 0$, so

$$\tan\delta_l \;=\; j_l(kR)/n_l(kR), \qquad\qquad (l = 0, 1, 2, \ldots) \tag{20.7.4}$$

for each of the $s$-wave, $p$-wave, $d$-wave, … partial shifts. This is especially simple for the $s$-wave,

$$\tan\delta_0 \;=\; [\sin(kR)/kR]\,/\,[-\cos(kR)/kR] \;=\; -\tan(kR), \tag{20.7.5}$$

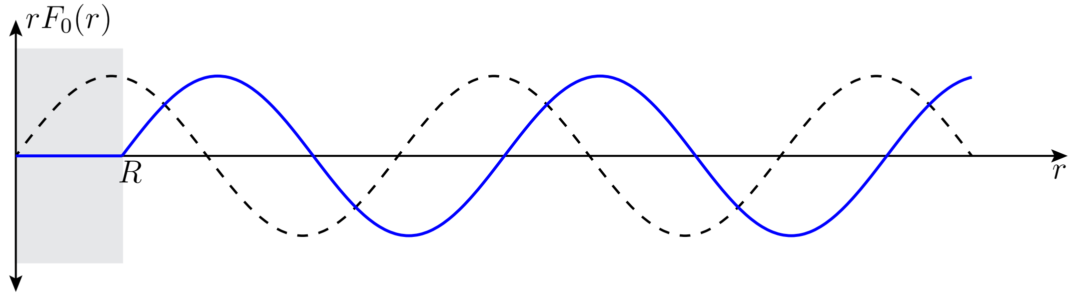

Figure 20.7.1: The solid line shows the $s$-wave radial wavefunction $rF_0(r)$ as a function of $r$, for scattering from a hard sphere with $V = \infty$ in the shaded region $r < R$. The dashed line shows what $rF_0(r)$ would be if the potential were absent, corresponding to free particle propagation. The potential pushes out the wavefunction by a distance $R$, giving a phase shift $\delta_0 = -kR$.

with the solution

$$\delta_0 \;=\; -kR. \tag{20.7.6}$$

Therefore, for $l = 0$,

$$B_0(r) \;\propto\; \cos\delta_0 \frac{\sin(kr)}{kr} + \sin\delta_0 \frac{\cos(kr)}{kr} \;=\; \frac{1}{kr}\sin(kr + \delta_0) \;=\; \frac{1}{kr}\sin(k(r - R)). \tag{20.7.7}$$

As illustrated in Figure 20.7.1, the outgoing $s$-wave is just shifted out by a distance $R$, compared to the free-particle solution.

More generally, consider all $\delta_l$ in the low-energy limit $kR \ll 1$ in which the wavenumber is much smaller than the inverse size of the hard sphere. In that case,

$$j_l(kR) \approx \frac{(kR)^l}{(2l+1)!!}, \qquad n_l(kR) \approx -\frac{(2l-1)!!}{(kR)^{l+1}}, \tag{20.7.8}$$

so, from eq. (20.7.4),

$$\tan\delta_l \;=\; -\frac{(kR)^{2l+1}}{(2l-1)!!\,(2l+1)!!}, \tag{20.7.9}$$

which quickly becomes very small as $l$ is increased. [Note that this result for $\tan\delta_l$ could also be obtained simply as the $\alpha_l \to \infty$ limit of eq. (20.5.35).]

In the extreme low energy limit $kR \ll 1$, only the $s$-wave contributes, resulting in

$$\frac{d\sigma}{d\Omega} \approx |f_0(k)|^2 = \frac{\sin^2\delta_0}{k^2} = \frac{\sin^2(kR)}{k^2} \approx R^2, \qquad (kR \ll 1). \tag{20.7.10}$$

Since this is isotropic (constant in both $\phi$ and $\cos\theta$), the total cross-section is

$$\sigma \;=\; \int d\Omega \frac{d\sigma}{d\Omega} \;=\; 4\pi R^2, \qquad (kR \ll 1), \tag{20.7.11}$$

which is 4 times the classical result $\sigma_{\text{classical}} = \pi R^2$ that we found in section eq (20.1.13).

In the opposite limit of very high scattering energy, many $\delta_l$ can contribute. For $kR \gg 1$, one can show that $\sigma \approx 2\pi R^2$, which is still twice the classical result. This can be understood from the fact that since $V(r)$ is discontinuous at $r = R$, no matter how big $E$ is, the wavefunction varies sharply on length scales $\ll 1/k$. To get $\sigma = \pi R^2 = \sigma_{\text{classical}}$, it is necessary to take a smoothed potential (a "softened hard sphere"), in which the the potential decreases to $V = 0$ over a distance scale much smaller than $R$ but much larger than $1/k$.

As another example, consider a finite well or barrier, with potential

$$V(r) \;=\; \begin{cases} V_0 & (\text{for } r \leq R), \\ 0 & (\text{for } r \geq R), \end{cases} \tag{20.7.12}$$

where the constant $V_0$ can be either negative (an attractive well) or positive (a repulsive core potential). Let us only consider $s$-wave scattering, valid at low energy. Outside the potential's range, the radial wavefunction is proportional to

$$F_0(r) \;=\; e^{i\delta_0}\left[\cos\delta_0 j_0(kr) - \sin\delta_0 n_0(kr)\right] \;=\; e^{i\delta_0}\frac{\sin(kr+\delta_0)}{kr}, \qquad (r \geq R), \tag{20.7.13}$$

with the normalization chosen arbitrarily, and

$$\frac{\hbar^2 k^2}{2m} \;=\; E. \tag{20.7.14}$$

Inside the potential's range, assuming $E > V_0$, the radial wavefunction is

$$F_0(r) \;=\; C j_0(k'r) \;=\; C\sin(k'r)/k'r, \qquad (r \leq R), \tag{20.7.15}$$

with a relative normalization constant $C$, and

$$\frac{\hbar^2 k'^2}{2m} = E - V_0. \tag{20.7.16}$$

The solution $n_0(k'r)$ is rejected here, because it is not normalizable at $r = 0$, as discussed in section 9.2. Now we match the solutions at $r = R$, to obtain an equation that determines $C$,

$$C\sin(k'R)/k'R \;=\; e^{i\delta_0}\sin(kR+\delta_0)/kR. \tag{20.7.17}$$

We also match $d\psi/dr$ at $r = R$, by specializing the convenient general results of eqs. (20.5.32) and (20.5.34),

$$\tan\delta_0 \;=\; \frac{(k/k')\sin(k'R)\cos(kR) - \cos(k'R)\sin(kR)}{(k/k')\sin(k'R)\sin(kR) + \cos(k'R)\cos(kR)}. \tag{20.7.18}$$

The preceding assumed $E > V_0$. If instead $E < V_0$, then $k'$ will be imaginary, but eqs. (20.7.17) and (20.7.18) are valid after making the replacements $\sin(k'R)/k' \to \sinh(\kappa R)/\kappa$ and $\cos(k'R) \to \cosh(\kappa R)$, where $\hbar^2\kappa^2/2m = V_0 - E$.

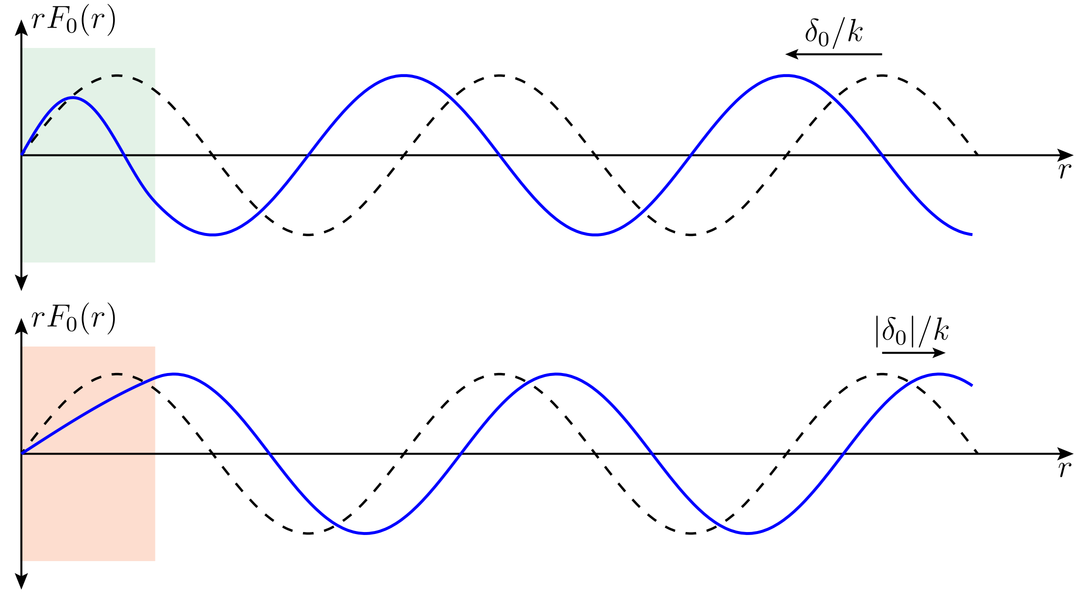

Figure 20.7.2: The solid lines show examples of the $s$-wave radial wavefunction $rF_0(r)$ as a function of $r$, for scattering from a spherical-well potential in the shaded region, if the potential is attractive (top) or repulsive (bottom). The dashed line shows the corresponding result if the potential is absent. The potential pulls in (for the attractive case, $\delta_0 > 0$) or pushes out (for the repulsive case, $\delta_0 < 0$) the wavefunction by a distance $|\delta_0|/k$.

Graphs of the $l = 0$ wavefunction normalized by a factor of the radial coordinate are shown in Figure 20.7.2 for the attractive case $V_0 < 0$ so that $k' > k$, and the repulsive case $V_0 > 0$ so that $k' < k$, respectively. In the attractive case, $\delta_0$ is positive, and the potential "pulls in" the $l = 0$ wave by a shift $\delta_0/k$ in the position of the nodes. In contrast, for the repulsive case, $\delta_0$ is negative, and the potential can be thought of as "pushing out" the $l = 0$ wave by an amount $|\delta_0|/k$. Things are a little more complicated for the partial waves with $l > 0$, and for potentials that are not piecewise constant. However, the essential qualitative feature remains that outgoing partial waves are pulled in by attractive potentials (because the Schrödinger equation dictates that the wavefunction oscillates faster in the region where the potential energy is smaller) and pushed out by repulsive potentials (because the wavefunction oscillates more slowly, or not at all, in regions with larger potential energy).

The phase shift $\delta_0$ is small in magnitude for $E \gg |V_0|$. In the case of an attractive potential ($V_0 < 0$), the phase shift increases as $E$ decreases until it reaches $\delta_0 = \pi/2$, where the $s$-wave cross-section is as large as it can possibly be, for a given energy,

$$\sigma_{l=0}\Big|_{\text{max}} = \frac{4\pi}{k^2}\sin^2(\pi/2) = \frac{2\pi\hbar^2}{mE}. \tag{20.7.19}$$

In this case, the scattering has saturated the partial wave unitarity bound. Decreasing $E$ even more, eventually one may reach $\delta_0 = \pi$, for which

$$\sigma_{l=0}\Big|_{\text{min}} = \frac{4\pi}{k^2}\sin^2(\pi) = 0. \tag{20.7.20}$$

This is the Ramsauer–Townsend effect; there is no scattering even though $V(r)$ is non-negligible and attractive, because the de Broglie wavelength happens to match the characteristic length scale of the potential. We have seen this type of behavior already in a one-dimensional scattering example, at the end of section 6.6, and in the real world it is seen in a suppression of the cross-section for electron scattering from inert gas atoms Ar, Kr, and Xe.

In the low-energy limit, one can find the scattering length and effective range defined in eq. (20.5.36), by expanding eq. (20.7.18) for small $k$. The results are

$$a = R - \tan(k_0 R)/k_0, \tag{20.7.21}$$

$$r_0 = R - \frac{R^3}{3a^2} - \frac{1}{ak_0^2}, \tag{20.7.22}$$

where we have defined

$$k_0 = \sqrt{-2mV_0}/\hbar. \tag{20.7.23}$$

If $k_0 R$ happens to be slightly larger than $\pi/2$, then the scattering length can be large, $a \gg R$. This corresponds to the existence of an $s$-wave bound state with small binding energy $E_{\text{bound}} \approx -\hbar^2/2ma^2$, and then $\sigma \approx 4\pi a^2/(1 - E/E_{\text{bound}})$ as discussed in the previous section.

## 20.8 Neutron-proton scattering and the deuteron

The low-energy scattering of neutrons and protons provides a famous practical illustration of some of the ideas discussed in sections 20.5–20.7. The masses of the proton and neutron are respectively $m_p = 938.272$ MeV/$c^2$ and $m_n = 939.565$ MeV/$c^2$, with a reduced mass in the two-body problem (see section 4.2),

$$\mu = \frac{m_n m_p}{m_n + m_p} \approx 469.459\,\text{MeV}/c^2. \tag{20.8.1}$$

The potential between nucleons arises from the rather complicated strong nuclear force, but the spectrum of bound states for two nucleons is very simple. There are no $pp$ or $nn$ bound states, and there is only one bound state for the neutron and the proton. This is the deuteron ($d$), a state with spin 1 and binding energy

$$-E^d_{\text{bound}} = 2.2246\,\text{MeV}, \tag{20.8.2}$$

which is determined by measuring the energy of the photon ($\gamma$) emitted when a proton captures a neutron, $n+p \to d+\gamma$. The binding energy 2.2246 MeV is rather small in comparison to other nuclear binding energies per nucleon pair. Since the deuteron is weakly bound, there should be a pole in the neutron-proton scattering cross-section when extrapolated to negative energy.

Let us now understand the angular momentum and parity quantum numbers of the deuteron, as a prerequisite to figuring out its role in neutron-proton scattering. Since the full Hamiltonian must be invariant under rotations, the energy eigenstates are also eigenstates of the square of the total angular momentum operator $\vec{J}$, obtained by combining the spins of both of the constituents $\vec{S}_p$ and $\vec{S}_n$ and their total orbital angular momentum $\vec{L}$ in the center-of-mass frame. The statement that the deuteron has spin 1 means that the operator $J^2$ has eigenvalue $\hbar^2 J(J+1)$, where $J = 1$. The neutron and proton each have spin 1/2, so the possible total constituent spin combinations for the deuteron are $S = 0$ and $S = 1$. The $J = 1$ spin of the deuteron is then obtained by combining this with the total orbital angular momentum quantum number $L$ in the center-of-mass frame, which can therefore only be $L = 0$, 1, or 2. Now, the magnetic dipole moment of the deuteron is known experimentally to be $\mu_d = 0.85744\mu_N$, which is very close to the sum of the dipole moments of the neutron and proton, $\mu_p + \mu_n = 2.79285\mu_N - 1.91304\mu_N = 0.87981\mu_N$. This indicates that the deuteron magnetic moment is very nearly realized when the spins are aligned in the $S = 1$ state, without much contribution from the orbital motion of the charged proton. Therefore, the deuteron must have predominantly $L = 0$. However, there are also two clear experimental indications that it cannot be a pure $L = 0$ eigenstate. First, there is the small but significant discrepancy in the magnetic moment sum noted above. Second, the deuteron is found experimentally to have a non-zero electric quadrupole moment, which would be inconsistent with the perfect spherical symmetry of a pure $L = 0$ eigenstate. The strong interaction Hamiltonian responsible for binding the deuteron is known (from its more fundamental formulation, quantum chromo-dynamics, or QCD) to commute with parity. Therefore, eq. (8.7.8) applies, and it must be possible to assign the deuteron a definite parity eigenvalue $\eta_d = (-1)^L \eta_p \eta_n$. Here, $\eta_p$ and $\eta_n$ are the intrinsic parities of the proton and neutron, which are conventionally taken to be $\eta_p = \eta_n = 1$. To have a definite parity eigenvalue $\eta_d = \pm 1$, the deuteron must be a superposition of states with either all $L$ even, or all $L$ odd. Since we already know that the state is mostly $L = 0$, this implies $\eta_d = +1$ and rules out the possibility of any $L = 1$ component. Therefore, the deuteron must be a linear combination of $L = 0$ and $L = 2$. In order to give $J = 1$, it follows that the combination of constituent spins can only be pure $S = 1$. In spectroscopic notation, the deuteron state is predominantly ${}^{2S+1}L_J = {}^3S_1$, but with a few percent ${}^3D_1$ component.

For our present purposes, the most important feature of the previous paragraph is that the deuteron is a pure $S = 1$ combination. The absence of a bound state in the $S = 0$ combination shows that the strong nuclear force is spin-dependent, which means that we should also keep track of the $S = 0$ and $S = 1$ combinations separately in scattering. The data for low-energy scattering of neutrons and protons in the $S = 0$ and $S = 1$ combinations yield scattering lengths and effective ranges, as defined in eq. (20.5.36), of

$$a^{S=1} = 5.42\,\text{fm}, \qquad r_0^{S=1} = 1.75\,\text{fm}, \tag{20.8.3}$$

$$a^{S=0} = -23.7\,\text{fm}, \qquad r_0^{S=0} = 2.7\,\text{fm}, \tag{20.8.4}$$

in units of 1 fm $= 10^{-15}$ meters. The positive scattering length in the $S = 1$ channel supports the hypothesis of a bound state. Solving eq. (20.6.4) gives a prediction for the bound state of

$$\kappa = 0.231\,\text{fm}^{-1}, \tag{20.8.5}$$

which then yields the estimate

$$E_{\text{bound}} = -\frac{\hbar^2\kappa^2}{2\mu} = -2.22\,\text{MeV}, \tag{20.8.6}$$

in good agreement[†] with the experimental deuteron binding energy quoted in eq. (20.8.2). Now, using the facts that the $E = 0$ cross-section for $s$-wave scattering is $4\pi a^2$, and that it must have a pole at $E = E^d_{\text{bound}}$, we have as in eq. (20.6.8),

$$\sigma^{S=1} = \frac{4\pi\left(a^{S=1}\right)^2}{1 - E/E^d_{\text{bound}}} \approx \frac{3.69\,\text{b}}{1 + E/2.2246\,\text{MeV}}, \tag{20.8.7}$$

where we have used the traditional nuclear physics unit of cross-section

$$1\,\text{b} = 100\,\text{fm}^2 = 10^{-28}\,\text{meters}^2, \tag{20.8.8}$$

called[‡] a **barn**.

In the $S = 0$ channel for neutron-proton scattering, the fact that the scattering length $a^{S=0}$ is negative confirms that there is no bound state. Still, we can write the cross-section in a form similar to eq. (20.8.7), by using eq. (20.5.38),

$$\sigma^{S=0} = \frac{4\pi\left(a^{S=0}\right)^2}{1 + (ka^{S=0})^2(1 - r_0^{S=0}/a^{S=0})} \approx \frac{70.6\,\text{b}}{1 + E/(0.066\,\text{MeV})}. \tag{20.8.9}$$

However, it should be emphasized that there really is no neutron-proton bound state with energy near $-0.066$ MeV. Such a pole in the scattering cross-section at negative real $E$, but

---

[†] Neglecting $r_0^{S=1}$ here would give much worse agreement. However, it must be admitted that the excellence of the agreement as presented here is partly accidental. Since eq. (20.5.36) was truncated beyond order $k^2$, rather than solving the quadratic equation (20.6.4) for $\kappa$, one could just as consistently approximate it by eq. (20.6.7), which would have given instead $\kappa = 0.214\,\text{fm}^{-1}$ and thus $E_{\text{bound}} = -1.91$ MeV.

[‡] The origin of the unit name refers to the fact that, by nuclear physics standards, this is not a small area, jokingly considered as easy to hit as "the broad side of a barn" in idiomatic North American English. Collider experiments at the present high-energy frontier often discuss cross-sections in units of picobarns or femtobarns.

with a negative scattering length $a$ and therefore not associated with an actual bound state, is sometimes called a **virtual bound state**. It can be interpreted as an indication that if the $S = 0$ neutron-proton potential were just slightly more attractive, then the scattering length would be positive and a bound state would exist. Such virtual bound state poles should not be confused with the quasi-bound state resonances with positive energy discussed in section 20.6.

The initial neutron and proton spins are often random and unmeasured, so that the four spin states (1 for $S = 0$ and 3 for $S = 1$) are equally likely. Averaging over them, the prediction for the total neutron-proton cross-section in the $E \rightarrow 0$ limit is

$$\sigma \;=\; \frac{1}{4}\sigma^{S=0} + \frac{3}{4}\sigma^{S=1} \;=\; 20.4\,\text{b}, \tag{20.8.10}$$

in good agreement with experimental observation.

The data eqs. (20.8.3) and (20.8.4) can also be used to make a crude model for the neutron-proton potential in the form of a spherical well as in (20.7.12), by solving eqs. (20.7.21)–(20.7.23) numerically for $R$ and $k_0 = \sqrt{-2\mu V_0}/\hbar$. The results are

$$R^{S=1} \;=\; 2.07\,\text{fm}, \qquad k_0^{S=1} \;=\; 0.91\,\text{fm}^{-1}, \qquad V_0^{S=1} \;=\; -34\,\text{MeV}, \tag{20.8.11}$$

$$R^{S=0} \;=\; 2.59\,\text{fm}, \qquad k_0^{S=0} \;=\; 0.58\,\text{fm}^{-1}, \qquad V_0^{S=0} \;=\; -14\,\text{MeV}. \tag{20.8.12}$$

From the analysis of section 9.4, the condition for such a spherical-well potential to have $n$ bound states with $l = 0$ can be written as [see eq. (9.4.12), and recall that $V_0$ and $a$ there are $-V_0$ and $R$ here]

$$k_0 R/\pi \;>\; n - 1/2. \tag{20.8.13}$$

Putting in the numbers, the $S = 1$ potential model predicts exactly one bound state, and the $S = 0$ potential model barely misses having a bound state, in accord with the observed facts. The deuteron is surprisingly weakly bound, in the sense that the binding energy (2.2246 MeV) is more than an order of magnitude smaller than the depth of the potential (34 MeV).

## 20.9 Scattering of identical particles

So far, we have neglected the possibility that the scattering particles might be identical. To remedy this, first consider the scattering of two identical bosons with no spin (for example, $\alpha$ particles, also known as $^4$He nuclei). The total wavefunction will then be of the form

$$\Psi(\vec{r}_1, \vec{r}_2) \;=\; e^{i\vec{p}\cdot(\vec{r}_1+\vec{r}_2)/2\hbar}\psi(\vec{r}), \tag{20.9.1}$$

where, following eqs. (4.2.4) and (4.2.5), $\vec{p}$ is the total momentum of the two particles, $(\vec{r}_1+\vec{r}_2)/2$ is the center-of-mass position, and

$$\vec{r} = \vec{r}_1 - \vec{r}_2 \tag{20.9.2}$$

is the relative position vector. By choosing the center-of-mass reference frame, we can take $\vec{p} = 0$. Because the particles are spin-less bosons, this wavefunction must be symmetric under $\vec{r}_1 \leftrightarrow \vec{r}_2$, so the relative-position wavefunction must obey

$$\psi(\vec{r}) \;=\; \psi(-\vec{r}), \tag{20.9.3}$$

in other words, it must have even parity. Therefore, instead of a scattering wavefunction proportional to $e^{ikz} + f_k(\theta)\frac{e^{ikr}}{r}$, we must have (dropping the overall normalization in this section)

$$\psi(\vec{r}) \;=\; e^{ikz} + e^{-ikz} + [f_k(\theta) + f_k(\pi - \theta)]\,\frac{e^{ikr}}{r} \tag{20.9.4}$$

for the scattering wavefunction in the center-of-momentum frame. In this way, both incoming particles, and both outgoing particles, are described on an equal footing. The differential cross-section is therefore of the form

$$\frac{d\sigma}{d\Omega} \;=\; |f_k(\theta) + f_k(\pi-\theta)|^2 \;=\; |f_k(\theta)|^2 + |f_k(\pi-\theta)|^2 + 2\mathrm{Re}\left[f_k^*(\theta) f_k(\pi-\theta)\right]. \tag{20.9.5}$$

This incorporates the fact that if one outgoing particle is detected at angle $\theta$, then by momentum conservation the other must be found at angle $\pi - \theta$. It is a necessary feature that the differential cross-section should be equal at those two angles, due to the intrinsic indistinguishability of the particles.

As a consequence, there must be constructive interference in the amplitudes for identical spin-less bosons for right-angle scattering,

$$\left.\frac{d\sigma}{d\Omega}\right|_{\theta=\pi/2} \;=\; 4|f_k(\pi/2)|^2. \tag{20.9.6}$$

Also, only even-$l$ partial waves can contribute in the identical boson scattering case, because

$$f_k(\theta) + f_k(\pi-\theta) \;=\; \sum_{l=0}^{\infty}(2l+1)\left[P_l(\cos\theta) + P_l(\cos(\pi-\theta))\right] f_l(k) \tag{20.9.7}$$

$$\;=\; 2\sum_{\text{even } l}(2l+1)P_l(\cos\theta) f_l(k), \tag{20.9.8}$$

where the terms with odd $l$ have canceled because $P_l(\cos(\pi-\theta)) = (-1)^l P_l(\cos\theta)$.

Now consider the case of scattering of electrons, or other identical spin-1/2 fermions. This can be divided into two cases, depending on the total spin quantum number $S$.

First, suppose that $S = 0$. Since that is an antisymmetric spin state, the spatial wavefunction must be symmetric under exchange of the labels 1, 2, so $\psi(\vec{r}) = \psi(-\vec{r})$, and

$$\frac{d\sigma}{d\Omega}^{S=0} \;=\; |f_k(\theta) + f_k(\pi-\theta)|^2, \tag{20.9.9}$$

and so all of the remarks just made for scattering of identical bosons go through as before.

If instead the total spin state is the symmetric $S = 1$ combination, then the spatial wavefunction must be antisymmetric under exchange of the labels $1, 2$, so $\psi(\vec{r}) = -\psi(-\vec{r})$ with odd parity. Therefore, the scattering wavefunction must be proportional to

$$\psi(\vec{r}) \;=\; e^{ikz} - e^{-ikz} + \left[f_k(\theta) - f_k(\pi - \theta)\right] \frac{e^{ikr}}{r}, \tag{20.9.10}$$

which implies

$$\frac{d\sigma}{d\Omega}^{S=1} \;=\; |f_k(\theta)|^2 + |f_k(\pi - \theta)|^2 - 2\mathrm{Re}\left[f_k^*(\theta) f_k(\pi - \theta)\right]. \tag{20.9.11}$$

It follows that there is perfect destructive interference for right-angle scattering,

$$\left.\frac{d\sigma}{d\Omega}^{S=1}\right|_{\theta=\pi/2} \;=\; 0 \tag{20.9.12}$$

for identical fermions in the symmetric spin state. Furthermore, only partial waves with odd $l$ contribute.

In many cases, the fermion spins are random and unmeasured. If so, then four spin states (1 for $S = 0$ and 3 for $S = 1$) are equally likely, so the observed unpolarized cross-section for identical fermions is the weighted average:

$$\begin{aligned} \left.\frac{d\sigma}{d\Omega}\right|_{\text{unpolarized}} &= \frac{1}{4}\frac{d\sigma}{d\Omega}^{S=0} + \frac{3}{4}\frac{d\sigma}{d\Omega}^{S=1} && (20.9.13) \\ &= |f_k(\theta)|^2 + |f_k(\pi - \theta)|^2 - \mathrm{Re}\left[f_k^*(\theta) f_k(\pi - \theta)\right], && (20.9.14) \end{aligned}$$

where the absence of a factor of 2 in front of the interference term is not a typographical error. For right angle scattering, the result is partial destructive interference.

In all of the cases of scattering of identical particles, it is important to note that when integrating the differential cross-section to get the total cross-section, one must include a factor of 1/2 to avoid double counting. This is because a state with one particle at $(\theta, \phi)$ also has an identical particle at $(\pi - \theta, \phi + \pi)$. Therefore, the state specified by the presence of a particle at $(\theta, \phi)$ and the state specified by $(\pi - \theta, \phi + \pi)$ are actually the same state, and should not be counted twice. So, one has

$$\sigma = \frac{1}{2} \int d\Omega \, \frac{d\sigma}{d\Omega} \qquad \text{(identical particles)} \tag{20.9.15}$$

for both the identical boson and identical fermion case. Note that this factor of 1/2 for identical particles comes in at the level of the total cross-section, not the differential cross-section.

# 21 Entanglement and correlations in measurements

## 21.1 The Einstein–Podolsky–Rosen and Bohm problem

Although Einstein played an important role in the early development of quantum mechanics, like many others he was troubled by some of its implications. This was in part due to his discomfort with the reliance of Postulates 4 and 5 on probabilities for the result of a measurement and the subsequent state, rather than completely deterministic time evolution. Even after he grudgingly accepted the possibility of non-determinism, Einstein found the implications of entanglement of quantum states to be difficult to accept, and doubted whether quantum mechanics as then formulated could be regarded as a complete framework.

In 1935, Einstein, Boris Podolsky, and Nathan Rosen [*Phys. Rev.* **47**, 777, (1935)] described a thought experiment that highlighted the perceived weirdness associated with measurements involving entangled states. This problem is often called the EPR paradox, but it should be emphasized at the outset that although it may be troubling to humans accustomed to classical ways of thinking, in the end we will find that there is really nothing paradoxical about it. (Indeed, EPR themselves did not call it a "paradox".) EPR described the problem in terms of measurements of position and momentum, but David Bohm later distilled the basic ideas into a simpler form in terms of measurements of components of spins, and that is the version we will now describe.

Consider a system of two spin-1/2 particles, labeled 1 and 2, in a state with the total spin $\vec{S} = \vec{S}_1 + \vec{S}_2$ equal to 0,

$$|S=0\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\hat{z},-\hat{z}\rangle - |-\hat{z},+\hat{z}\rangle\Big). \tag{21.1.1}$$

Throughout this chapter, we will employ spin operators with a factor of $\hbar/2$ extracted, so

$$\vec{\sigma}_k = \frac{2}{\hbar}\vec{S}_k \tag{21.1.2}$$

for particle $k$, represented by the Pauli matrices. All three components of $\vec{\sigma}_1$ commute with all three components of $\vec{\sigma}_2$, because they operate on distinct one-particle Hilbert spaces. The notation in eq. (21.1.1) is such that, for example, $|+\hat{z},-\hat{z}\rangle \equiv |+\hat{z}\rangle_1 \otimes |-\hat{z}\rangle_2$ is an eigenstate of both $\sigma_{1z}$ and $\sigma_{2z}$, with eigenvalues $+1$ and $-1$, respectively. The state of the system is thus represented by a ket in the tensor product of the Hilbert spaces for the two individual particles. We are purposely avoiding the usual $\uparrow$ and $\downarrow$ notations for the spins, because we will want to consider components of the spins other than the $z$ component. Indeed, since the $S=0$ state is spherically symmetric, there is nothing special about the $\hat{z}$ direction, and up to irrelevant phases we could just as easily write the state in eq. (21.1.1) as

$$|S=0\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\hat{n},-\hat{n}\rangle - |-\hat{n},+\hat{n}\rangle\Big) \tag{21.1.3}$$

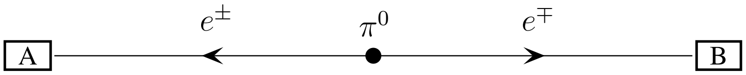

Figure 21.1.1: The setup for Bohm's version of the EPR problem. Decays of neutral pions at rest produce two spin-1/2 particles moving in opposite directions, in an entangled spin state with total spin 0, and with spatial wavefunctions that have no overlap at late times. Two distant observers Alice and Bob can each measure any component they choose of the spin of the particle that arrives at their respective detector A or B.

for any unit vector direction $\hat{n}$. In any case, there is no way to write this total spin-0 ket as a single product of kets in the individual spin-1/2 particle Hilbert spaces; no matter the choice of bases for the individual spins, it is always a non-trivial linear combination of such products. Therefore, it is an example of an entangled state as defined in section 2.9.

The particles also have spatial wavefunctions that we need not write explicitly. The only thing important for us is that they describe localized wavepackets moving away from each other. The two spins, although entangled, are therefore separated by a large distance at late times. This is sometimes called **non-local entanglement**.

One way to realize such a state would be an electron-positron pair from the decay of a neutral pion at rest in a suitable reference frame,[†]

$$\pi^0 \quad \rightarrow \quad e^- e^+. \tag{21.1.4}$$

The pion has spin 0, and since we are in its rest frame, the total angular momentum is also 0. We imagine that this experiment will be conducted many times, with two ideal observers Alice and Bob located on opposite sides of the point where the pions decay, and far away, as shown in Figure 21.1.1. They each have a small detector capable of measuring any desired component of the spin of a particle moving through it. Since the detectors are far apart, where the spatial wavefunctions for the two particles have no overlap, Alice can only measure one of the spins, and Bob can only measure the other. Let us choose the labeling of the particles so that Alice always makes measurements on particle 1 (which on an event-by-event basis might be either the electron or the positron), and Bob always on particle 2. They can each measure any component of that spin, or choose to make no measurement. For simplicity, we assume that the Hamiltonian is just that of free particles, so that the spin states have no Hamiltonian time evolution.

Now suppose that Alice measures $\sigma_{1z}$. The result will be either +1 or −1, with equal probability. If the result was +1, then if Bob measured $\sigma_{2z}$ the result will necessarily be −1.

---

[†] This is a very rare decay, occurring for less than $10^{-7}$ of all neutral pions. Also, it is not easy to produce neutral pions at rest in a laboratory frame in the first place. But this is a thought experiment; we only care that it is possible in principle, not about such practical difficulties.

This follows from the form of the state eq. (21.1.1). We can say that Alice has collapsed the state to $|+\hat{z}, -\hat{z}\rangle$ by her measurement of $\sigma_{1z}$, so that the only possible outcome for Bob is $\sigma_{2z} = -1$. Conversely, if Alice instead measured $\sigma_{1z}$ to be $-1$, then if Bob measures $\sigma_{2z}$ the result must be $+1$. There is a perfect anticorrelation between their measurements.

This perfect anticorrelation in Alice's and Bob's measured results for $\sigma_{1z}$ and $\sigma_{2z}$ is not at all surprising. In fact, it is actually much more general, in the sense that it does not even rely on quantum mechanics being correct. It would inevitably be true in any alternative theory, just as long as angular momentum is conserved. There is a similar perfect anticorrelation between the measured electric charges; if Alice sees that particle 1 is an electron, then she can be sure that Bob's particle 2 will be a positron, due to conservation of charge.

However, spins are much more interesting than electric charges, in the sense that they reveal the nontrivial aspects of measurement correlations in quantum mechanics. This is because spins can point in arbitrary directions, and Alice and Bob could choose to make measurements of different spin components. For example, suppose Alice measures $\sigma_{1x}$ but Bob still measures $\sigma_{2z}$. If Alice obtains the result $\sigma_{1x} = +1$, then from eq. (21.1.3) with $\hat{n} = \hat{x}$, we can say that the total state collapses to $|+\hat{x}, -\hat{x}\rangle$. Then, since $|-\hat{x}\rangle_2 = (|+\hat{z}\rangle_2 - |-\hat{z}\rangle_2)/\sqrt{2}$, Bob will obtain the results $\sigma_{2z} = \pm 1$ randomly, and with equal probability. The same random results will occur for Bob's measurement of $\sigma_{2z}$ if the result of Alice's measurement was instead $\sigma_{1x} = -1$, or if she decided to make no measurement at all.

To be more general, suppose that Alice and Bob measure arbitrary spin components $\hat{a}\cdot\vec{\sigma}_1$ and $\hat{b}\cdot\vec{\sigma}_2$, respectively, for unit vectors $\hat{a}$ and $\hat{b}$. Neither can predict with certainty what the other will measure on an event-by-event basis (unless $\hat{a}$ and $\hat{b}$ are parallel), but there is a statistical correlation between their measurements, which can be expressed as the expectation value

$$C(\hat{a}, \hat{b}) \;=\; \langle \hat{a}\cdot\vec{\sigma}_1 \, \hat{b}\cdot\vec{\sigma}_2 \rangle \tag{21.1.5}$$

in the state $|S = 0\rangle$. This can be straightforwardly evaluated. One labor-saving trick for doing so is to note that when acting on the state $|S = 0\rangle$, one can always replace $\hat{b}\cdot\vec{\sigma}_2$ with $-\hat{b}\cdot\vec{\sigma}_1$, followed by use of the Pauli matrix identity $(\hat{a}\cdot\vec{\sigma})(\hat{b}\cdot\vec{\sigma}) = \hat{a}\cdot\hat{b} + i(\hat{a}\times\hat{b})\cdot\vec{\sigma}$ from eq. (8.2.17). Then, since the expectation value of $\vec{\sigma}_1$ in the $S = 0$ state vanishes, the prediction from quantum mechanics is

$$C(\hat{a}, \hat{b}) \;=\; -\hat{a}\cdot\hat{b}. \tag{21.1.6}$$

This agrees with the special cases that we have already discussed; for $\hat{a} = \hat{b}$, the measurements of Alice and Bob are perfectly anticorrelated, and when $\hat{a}$ and $\hat{b}$ are orthogonal there is no correlation. More generally, $C(\hat{a}, \hat{b})$ is always in the range from $-1$ (perfect anticorrelation) to $+1$ (perfect correlation).

The issue that bothered EPR and Bohm is that Alice and Bob might be so far apart from each other that there is no possibility of communication between them when they each choose which spin measurement to make. They could even make their individual decisions about which spin components to measure at random for each individual pion decay after it has occurred and while the electron and positron are already in flight. Despite this, the possible results obtained by Bob would seem to depend on the choices made by Alice, and vice versa. At the last moment before particle 1 arrives, Alice could decide to measure $\sigma_{1z}$. Then, from the results of Alice's measurement alone, she can be immediately certain of the result if Bob measures $\sigma_{2z}$, a measurement occurring far away and quite out of her control. Or, she could decide instead to measure $\sigma_{1x}$, in which case (regardless of the outcome of her own measurement) she would correctly assign a probability of 50% to the outcome of a measurement by Bob of $\sigma_{2z}$. The same follows if she decided to make no measurement at all. This is all despite the fact that no signal carrying the news of her last-moment decision could reach Bob before his measurement occurs. Einstein referred to this sort of thing as "spooky action at a distance", and believed that it pointed to an incompleteness of the quantum theory as a description of reality, but Nature does not care whether we, or even Einstein, find a phenomenon peculiar or spooky. There are two things to check to make sure that there is not a problem with the quantum theory, let alone a true paradox.

The first thing to check is that the predictions of quantum mechanics are really consistent and unambiguous. In the preceding, we have phrased things in terms of Alice's measurement collapsing the state, with implications for what Bob will see. But the situation is symmetrical; we could just as easily phrase any pair of measurements in terms of Bob collapsing the state, with implications for what Alice will see. For the case in which Alice measured $\sigma_{1z} = +1$ and Bob measured $\sigma_{2z} = -1$, we could choose to say that Bob has collapsed the total state to $|+\hat{z}, -\hat{z}\rangle$, which means that the only possible result for Alice's measurement was, indeed, $\sigma_{1z} = +1$. This outcome is fully consistent with the interpretation in which Alice's measurement collapses the state. A third way is to view the collapse symmetrically, by taking Alice's and Bob's measurements to be a single measurement. (This is possible because all components of $\vec{\sigma}_1$ and $\vec{\sigma}_2$ commute with each other.) Again, the state collapses to $|+\hat{z}, -\hat{z}\rangle$. For any given outcome of both measurements, quantum mechanics indeed always gives the same result for the collapse of the state regardless of what order we assign to the measurements.[‡] To see this more generally, suppose that Alice measures $\hat{a} \cdot \vec{\sigma}_1$ to be +1 and Bob measures $\hat{b} \cdot \vec{\sigma}_2$ to be +1. All

---

[‡] This is necessary, because the measurements of Alice and Bob are space-like separated events, so special relativity tells us that there is no invariant way of saying which measurement occurred first. In some inertial reference frames, Alice's measurement came first, but in other reference frames Bob's came first, and in still others they are exactly simultaneous.

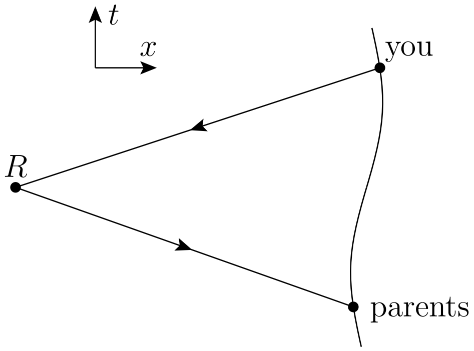

Figure 21.1.2: If instantaneous communication were possible, then you could send a signal to a relay $R$ which would be instantaneous (at constant time) in some other reference frame, but would arrive at an earlier time in your own reference frame. The relay could then send a signal, instantaneous in a different reference frame, which could be used to somehow prevent your parents from meeting, preventing your own birth. Fortunately, instantaneous communication is not a consequence of quantum mechanics.

three ways of describing the state collapse give the same probabilities for this outcome, and the same final state after both measurements. The simple, but essential, reason for this consistency is that the projection operators $P_{\hat{a}\cdot\vec{\sigma}_1} = |+\hat{a}\rangle_1 \langle+\hat{a}|_1$ and $P_{\hat{b}\cdot\vec{\sigma}_2} = |+\hat{b}\rangle_2 \langle+\hat{b}|_2$ commute.

The second thing to check is that quantum mechanics does not provide any instantaneous communication between the distant Alice and Bob, whose measurements are events occurring at a space-like separation in the language of special relativity. In general, an instantaneous (or even faster-than-light) communication between space-like separated points in one inertial reference frame would appear to travel backwards in time in some other inertial frame that is moving with respect to the first. This would imply causality paradoxes, because signals by combinations of actors in different inertial frames can be constructed in such a way that a message could be received on a time-like path before it is sent from a previous point on the same path. For example, you could use instantaneous communications to prevent your parents from meeting, thus preventing your own birth, as illustrated in Figure 21.1.2.

To see that Alice and Bob's experiments do not actually allow instantaneous communication, we need only note that no matter what each of them chooses to measure, the results of their individual measurements will always appear to be completely random, until they meet to compare them or at least send a conventional signal at a speed not exceeding that of light. Even if Alice and Bob agree to seek maximum anticorrelation by measuring $\sigma_{1z}$ and $\sigma_{2z}$ respectively (and even assuming that they do not break their promises!), each will record their data as a random sequence of $+1$ and $-1$, containing no information. *After* communicating by conventional means, they can check that their results do have correlations in agreement with quantum mechanics, but there is no information transmitted in either direction until that happens.

Why, then, did EPR and Bohm view the situation as disturbing? The essential issue is that one must not fall into the trap of thinking of the state of a system as always defined only by local information, a viewpoint sometimes called **local realism**. If you try to think about particle 2 only in terms of non-entangled kets such as $|+\hat{z}\rangle_2$ or $|-\hat{x}\rangle_2$, then it might indeed appear paradoxical that Alice's distant measurement of $S_{1z}$ can affect what Bob measures. However, that way of thinking about the state of particle 2 is explicitly rejected by quantum mechanics, which insists that we take into account the full non-local entanglement in order to get correct and consistent predictions. The real value of the EPRB problem is that it forces us to recognize this essential truth about how quantum mechanics works.

One sometimes sees grand pronouncements, inspired by the EPRB problem, along the lines of "quantum mechanics is non-local". However, one must be careful about what this means, because the word "non-local" has several completely different meanings in physics. For example, it is possible to define a Hamiltonian that has non-local dynamics; that is what quantum field theorists usually mean when they talk about the possibility of non-locality. That is certainly not the case in the preceding discussion, as we took $H = 0$ for the spin degrees of freedom, and even the $H$ describing the time evolution of the spatial wavefunctions was (implicitly) just that of free particle propagation. The entanglement of the state occurred due to perfectly local processes. The rejection of local realism in favor of allowing non-local entanglement of states does not imply that we need to accept a non-local Hamiltonian, for which there is absolutely no experimental evidence.

## 21.2 Hidden variables and Bell's inequality

Having learned in the previous section that standard quantum mechanics rejects local realism in favor of allowing non-local entanglement, it must be recognized that this is a falsifiable hypothesis to be experimentally tested against alternatives. There is a general class of alternatives called **hidden variables** theories, which attempt to incorporate local realism in a way consistent with experiment. In the remaining sections of this chapter, we will consider local hidden variables theories, see how they make predictions that are incompatible with quantum mechanics, and understand why experiments show that they cannot be correct.

Hidden variables are supposed to be quantities that we cannot directly measure or control, but are necessary to completely characterize the state of a system. They could be part of some completely deterministic proposed alternative to quantum mechanics, or could be an extra part of a framework that otherwise looks like conventional quantum mechanics. In the latter version, it is not just a ket $|\psi\rangle$ that describes the state of the system, but rather the state ket together with the hidden variable(s), which we will generically call $\lambda$. They are supposed to be distributed in a way that is either truly random, or just appears to be random because of our lack of knowledge of them. For any given system, the distribution of the hidden variables can be described by a probability density function $\rho(\lambda)$ with the properties

$$\int d\lambda\, \rho(\lambda) = 1, \tag{21.2.1}$$

$$\rho(\lambda) \geq 0, \tag{21.2.2}$$

so that $\rho(\lambda)\, d\lambda$ is the probability that it will be found between $\lambda$ and $\lambda + d\lambda$. The function $\rho(\lambda)$ might depend on the particulars of the system and the type of experiment one is doing.

The presence of $\lambda$ is supposed to be responsible for the illusion of random outcomes of measurements. For some observable $A$, and a state fully characterized by non-hidden properties $\psi$ and hidden variables $\lambda$, the hypothesis is that the outcome of a measurement is not random, but rather determined as some function

$$\text{measured value of } A = f_{A,\psi}(\lambda). \tag{21.2.3}$$

The function $f_{A,\psi}(\lambda)$ can be chosen so that as $\lambda$ runs over all possible values, it returns all allowed values predicted by quantum mechanics for the measurement of $A$ in the state $\psi$. Since we have no way of knowing what $\lambda$ is, the results of individual measurements appear random. Then, the mean value of many measurements of $A$ for apparently identical states $\psi$ will be

$$\overline{A} = \int d\lambda\, \rho(\lambda)\, f_{A,\psi}(\lambda). \tag{21.2.4}$$

This is the counterpart of the expectation value $\langle A \rangle = \langle \psi | A | \psi \rangle$ in standard quantum theory.

For example, let us consider how this would work for a system with two spin-1/2's combined to have total spin $S = 0$ in an entangled state coming from $\pi^0$ decay, as considered in section 21.1. To be general, consider the measurements by Alice and Bob of arbitrary components of the spins of particle 1 and 2, say $\hat{a} \cdot \vec{\sigma}_1$ and $\hat{b} \cdot \vec{\sigma}_2$, where $\hat{a}$ and $\hat{b}$ are unit vectors. Then the outcomes for Alice's and Bob's measurements are respectively determined by some functions

$$\hat{a} \cdot \vec{\sigma}_1 = f_A(\hat{a}, \lambda) = \pm 1, \tag{21.2.5}$$

$$\hat{b} \cdot \vec{\sigma}_2 = f_B(\hat{b}, \lambda) = \pm 1. \tag{21.2.6}$$

To incorporate local realism, the value of $\lambda$ is assumed to be the same for both particles 1 and 2, and is fixed at the moment that the parent pion decays. Thus, in each event, $\lambda$ has an unknown, but specific, value, which already determines what the ensuing measurement of any component of the spins will be, through the functions $f_A$ and $f_B$.

In the case of spins, the functions determining the measurement outcomes must satisfy some constraints. First, a measured value for $\hat{n} \cdot \vec{\sigma}_k$ is the same as the negative of a measured value for $-\hat{n} \cdot \vec{\sigma}_k$, so $f_A(\hat{n}, \lambda) = -f_A(-\hat{n}, \lambda)$ and $f_B(\hat{n}, \lambda) = -f_B(-\hat{n}, \lambda)$, for any $\hat{n}$. Also, to be consistent with angular momentum conservation, it must be true that whenever $\hat{a} = \hat{b} = \hat{n}$, the measurement results are perfectly anticorrelated in the case of the EPRB setup. Thus,

$$f_B(\hat{n}, \lambda) = -f_A(\hat{n}, \lambda) \tag{21.2.7}$$

so that the measurements are consistent with total spin $S = 0$.

We can now write down the hidden-variables prediction for the statistical correlation of Alice's and Bob's measurements. It is given by the product of the possible measurement results, integrated over the hidden variables weighted by their probability density,

$$C(\hat{a}, \hat{b}) \;=\; \int d\lambda\, \rho(\lambda)\, f_A(\hat{a}, \lambda) f_B(\hat{b}, \lambda). \tag{21.2.8}$$

This hidden variables result should be contrasted with the prediction of quantum mechanics that we found in eq. (21.1.6), which was $C(\hat{a}, \hat{b})_{\text{QM}} = -\hat{a} \cdot \hat{b}$.

The hidden variables result for $C(\hat{a}, \hat{b})$ is clearly less specific than the quantum mechanics prediction, because we have not committed to a particular form for the functions $\rho$, $f_A$, and $f_B$. Nevertheless, it is possible to draw some specific conclusions. In the case that $\hat{b} = \hat{a}$, the hidden variables prediction is

$$\begin{aligned} C(\hat{a}, \hat{a}) \;&=\; \int d\lambda\, \rho(\lambda)\, f_A(\hat{a}, \lambda) f_B(\hat{a}, \lambda) \;=\; -\int d\lambda\, \rho(\lambda)\, f_A(\hat{a}, \lambda)^2 \;=\; -\int d\lambda\, \rho(\lambda) \\ &=\; -1, \end{aligned} \tag{21.2.9}$$

where we have used eq. (21.2.7) to get the second equality, then the fact that $f_A(\hat{a}, \lambda) = \pm 1$, and finally eq. (21.2.1). In this special case, the hidden variables prediction does agree with that of quantum mechanics. This had to be true simply because in our hidden variables theory we built in the perfect anticorrelation required by angular momentum conservation, through eq. (21.2.7).

However, for more general $\hat{a}$ and $\hat{b}$, it is not immediately clear whether the hidden variables predictions for $C(\hat{a}, \hat{b})$ can be made to always agree with the predictions of quantum mechanics. In 1964, John S. Bell[†] answered the question by discovering that the correlations predicted by hidden variables theories must satisfy an inequality that is clearly violated by the correlations predicted by quantum mechanics. Surprisingly, this is true for *any* choices of the functions $\rho(\lambda)$, $f_A(\hat{a}, \lambda)$, and $f_A(\hat{b}, \lambda)$ in the hidden variables theory. This provides a way for experiment to decisively settle the question of whether the general hidden variables idea or standard quantum mechanics is true.

Bell's inequality is remarkably simple to derive. We start with

$$C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c}) \;=\; \int d\lambda\, \rho(\lambda)\, \Big[f_A(\hat{a}, \lambda) f_B(\hat{b}, \lambda) - f_A(\hat{a}, \lambda) f_B(\hat{c}, \lambda)\Big] \tag{21.2.10}$$

$$\;=\; -\int d\lambda\, \rho(\lambda)\, \Big[f_A(\hat{a}, \lambda) f_A(\hat{b}, \lambda) - f_A(\hat{a}, \lambda) f_A(\hat{c}, \lambda)\Big], \tag{21.2.11}$$

---

[†] *Physics*, **1**, 195, (1964), reprinted in J.S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, (1987).

where we have used eq. (21.2.7) to replace $f_B$ in favor of $f_A$. Multiply the second term on the right side by $1 = [f_A(\hat{b}, \lambda)]^2$, and rearrange to get

$$C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c}) \;=\; -\int d\lambda\, \rho(\lambda)\, \Big[1 - f_A(\hat{b}, \lambda) f_A(\hat{c}, \lambda)\Big]\, f_A(\hat{a}, \lambda) f_A(\hat{b}, \lambda). \qquad (21.2.12)$$

Now, take the absolute value of both sides, and use the general fact that the absolute value of any integral is always less than or equal to the integral of the magnitude of the integrand:

$$\left|\int d\lambda\, F(\lambda)\right| \;\leq\; \int d\lambda\, |F(\lambda)|\,. \qquad (21.2.13)$$

Let $F(\lambda)$ be the integrand on the right side of eq. (21.2.12). Then since $f_A(\hat{a}, \lambda) f_A(\hat{b}, \lambda)$ is always equal to $\pm 1$, and both $\rho(\lambda)$ and $1 - f_A(\hat{b}, \lambda) f_A(\hat{c}, \lambda)$ are always non-negative, we have $|F(\lambda)| = \rho(\lambda)\, \Big[1 - f_A(\hat{b}, \lambda) f_A(\hat{c}, \lambda)\Big]$. Therefore we obtain

$$\Big|C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c})\Big| \;\leq\; \int d\lambda\, \rho(\lambda)\, \Big[1 - f_A(\hat{b}, \lambda) f_A(\hat{c}, \lambda)\Big] \qquad (21.2.14)$$

or, using eqs. (21.2.1) and (21.2.7) and (21.2.8) again,

$$\Big|C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c})\Big| \;\leq\; 1 + C(\hat{b}, \hat{c}). \qquad (21.2.15)$$

This is Bell's original inequality governing correlations of spin-1/2 measurements in a total spin $S = 0$ state in hidden variables theories.

To see that Bell's inequality is predicted to be violated by quantum mechanics, consider what happens if we substitute the quantum prediction of the correlation from eq. (21.1.6) into eq. (21.2.15),

$$\Big|\hat{a} \cdot \hat{b} - \hat{a} \cdot \hat{c}\Big| \;\leq\; 1 - \hat{b} \cdot \hat{c}. \qquad (21.2.16)$$

This reduces the question to a purely geometric claim about three arbitrary unit vectors. A single example will be sufficient. Take $\hat{a} = \hat{x}$, $\hat{b} = \hat{y}$, and $\hat{c} = (\hat{x} + \hat{y})/\sqrt{2}$. The left side of eq. (21.2.16) is $1/\sqrt{2}$, while the right side is $1 - 1/\sqrt{2}$, which is less, so the inequality is violated. Thus, quantum mechanics makes a specific prediction that cannot be satisfied in any local hidden variables theory.

## 21.3 Quantum mechanics vs. hidden variables, without inequalities

There are various other versions of Bell's inequality, which show that quantum mechanics and hidden variables theories can be distinguished by measuring correlations. In this section we will discuss another way, which is theoretically even sharper because it does not rely on inequalities or statistics, although in practical terms it is harder to realize experimentally. The result in this example is that quantum mechanics and hidden variables theories make predictions that disagree, not just for correlations on a statistical basis, but for certain measurements on an event-by-event basis with 100% probability.

Consider the following entangled state involving three spin-1/2 particles,[†]

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|{+\hat{z}}, {+\hat{z}}, {+\hat{z}}\rangle - |{-\hat{z}}, {-\hat{z}}, {-\hat{z}}\rangle\Big), \tag{21.3.1}$$

We assume that the three particles, labeled 1, 2, and 3, start from a central point and are spatially separated without disturbing the spins. The particles are observed by Alice, Bob, and Charlie at three remote sites. It is agreed that each of them will independently measure a random choice of either the $x$ or $y$ component of the spin, so that Alice always measures either $\sigma_{1x}$ or $\sigma_{1y}$, while Bob always measures $\sigma_{2x}$ or $\sigma_{2y}$, and Charlie always measures $\sigma_{3x}$ or $\sigma_{3y}$. The experiment is repeated many times, always with the same state $|\psi\rangle$, and then the observers meet to compare their results.

Let us first analyze the situation according to the standard rules of quantum mechanics. The state $|\psi\rangle$ is not an eigenstate of any of the individual spin operators $\sigma_{1x}$, $\sigma_{1y}$, $\sigma_{2x}$, $\sigma_{2y}$, $\sigma_{3x}$, or $\sigma_{3y}$, and each of the individual measurements of Alice, Bob, and Charlie will have a random result ±1 with equal probability. However, you can check that $|\psi\rangle$ has the remarkable property that it is an eigenstate of all eight of the operators of the form

$$\Omega_{abc} = \sigma_{1a}\sigma_{2b}\sigma_{3c}, \tag{21.3.2}$$

where $a, b, c$ are each equal to $x$ or $y$. For the two operators with $a = b = c$, the eigenvalue is $-1$, and for the other six operators the eigenvalue is $+1$.

Before meeting, Alice, Bob, and Charlie have not learned anything except that their own individual measurement results appear completely random. However, consider what they find when they finally get together to compare their results. They decide to first look only at the results for cases in which two of them had chosen to measure the $y$ component of the spin, and the other had measured the $x$ component of the spin. In those cases, we can equivalently say that they had measured one of the observables $\Omega_{xyy}$ or $\Omega_{yxy}$ or $\Omega_{yyx}$, and because

$$\Omega_{xyy}\,|\psi\rangle = \Omega_{yxy}\,|\psi\rangle = \Omega_{yyx}\,|\psi\rangle = |\psi\rangle\,, \tag{21.3.3}$$

they must find that the product of their measured results was definitely $+1$, every time. (Sometimes all three of them will have measured the result to be $+1$, and sometimes two of them will

---

[†] This state and the discussion below follows N. David Mermin *Am. J. Phys.* **58**, 731, (1990), which in turn follows a similar four-spin state proposed by Daniel Greenberger, Michael Horne, and Anton Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe* (1989).

have obtained −1 and the other will have obtained +1.) On the other hand, in the cases where all three had chosen to measure the $x$ component of the spin, we can equivalently say that their choices amounted to measuring $\Omega_{xxx}$. Now, since

$$\Omega_{xxx} \left|\psi\right\rangle = -\left|\psi\right\rangle , \tag{21.3.4}$$

it must be that the product of their three measurements is −1. (Sometimes they will all get the result −1, and sometimes two will get +1 and the other will get −1.)

So far, we have analyzed the experiment assuming the predictions of quantum mechanics. Now let us try to explain these results in terms of a general local hidden variables theory, in which the measurement of $\sigma_{1x}$ is supposed to be determined to be some $f_A(\hat{x}, \lambda) = \pm 1$ and the measurement of $\sigma_{2y}$ is $f_B(\hat{y}, \lambda) = \pm 1$, etc. The hidden variable(s) $\lambda$ are supposed to be different for each repetition of the experiment, giving the illusion of randomness. Then, the observed results for $\Omega_{xyy}, \Omega_{yxy}, \Omega_{yyx}$ would imply that, respectively,

$$\begin{aligned} f_A(\hat{x}, \lambda)\, f_B(\hat{y}, \lambda)\, f_C(\hat{y}, \lambda) &= 1, && (21.3.5) \\ f_A(\hat{y}, \lambda)\, f_B(\hat{x}, \lambda)\, f_C(\hat{y}, \lambda) &= 1, && (21.3.6) \\ f_A(\hat{y}, \lambda)\, f_B(\hat{y}, \lambda)\, f_C(\hat{x}, \lambda) &= 1, && (21.3.7) \end{aligned}$$

for every $\lambda$. Now, multiplying these three equations together, and using $f_A(\hat{y}, \lambda)^2 = f_B(\hat{y}, \lambda)^2 = f_C(\hat{y}, \lambda)^2 = 1$, we discover that

$$f_A(\hat{x}, \lambda)\, f_B(\hat{x}, \lambda)\, f_C(\hat{x}, \lambda) = 1. \tag{21.3.8}$$

This tells us that in those cases where they had all measured the $x$ component of spin, the product of their measurements should have been +1, with 100% certainty, according to the hidden variables theory. This is in direct contradiction to the quantum mechanics prediction.

To recapitulate: if we stipulate that the measurement of $\Omega_{xyy}$, $\Omega_{yxy}$, or $\Omega_{yyx}$ always gives +1, then, according to hidden variables theories, measurement of $\Omega_{xxx}$ must always give +1 as well. However, the prediction of quantum mechanics, for the given state $\left|\psi\right\rangle$, is that $\Omega_{xxx}$ always gives −1. The hidden variables and quantum mechanics hypotheses make contradictory predictions and cannot both be correct.

## 21.4 Aspect's experiments and the demise of local hidden variables

In practice, the most decisive real-world experimental tests, many conducted by Alain Aspect and collaborators, involve polarizations of photons in entangled states, rather than spin-1/2 systems. In a series of increasingly sensitive experiments, the Bell-type inequalities of local hidden variables theories have been put to the test and found to be clearly violated. In this section, we will describe a Bell-type inequality for hidden variables, and the corresponding quantum prediction, as tested by A. Aspect, J. Dalibard, G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982), referred to in the following as the ADR experiment.

Consider observables $A(\hat{a})$ and $B(\hat{b})$ that can only take on the values $\pm 1$, and depend on unit vectors $\hat{a}$ and $\hat{b}$, and are associated with spatially separated detectors. These generalize the spin-1/2 observables $\hat{a} \cdot \vec{\sigma}_1$ and $\hat{b} \cdot \vec{\sigma}_2$ of Alice and Bob in the previous two sections. The experiment can change $\hat{a}$ and $\hat{b}$ independently at any time. According to the hidden variables proposal, the measured values are supposed to be determined by some functions

$$A(\hat{a}) \;=\; f_A(\hat{a}, \lambda) \;=\; \pm 1, \tag{21.4.1}$$

$$B(\hat{b}) \;=\; f_B(\hat{b}, \lambda) \;=\; \pm 1. \tag{21.4.2}$$

The hidden variables $\lambda$ are assumed to be local, implying that they are the same for $A$ and $B$, since they will have originated from a single location and event. Their probability density distribution is some non-negative function $\rho(\lambda)$ just as in eqs. (21.2.1). The correlation between measurement results $C(\hat{a}, \hat{b})$ is again given by the formula eq. (21.2.8). However, since the observables are not necessarily components of angular momenta (and will not be in the ADR experiment), we will not assume or use eq. (21.2.7). This means that the Bell inequality we are about to derive applies more generally than the original one.

Consider the following combination of correlations:

$$\begin{aligned} C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{b}') + C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{b}') \;=\; & \int d\lambda\, \rho(\lambda) \Big[ f_A(\hat{a}, \lambda) f_B(\hat{b}, \lambda) - f_A(\hat{a}, \lambda) f_B(\hat{b}', \lambda) \\ & + f_A(\hat{a}', \lambda) f_B(\hat{b}, \lambda) + f_A(\hat{a}', \lambda) f_B(\hat{b}', \lambda) \Big] \end{aligned} \tag{21.4.3}$$

for arbitrary unit vectors $\hat{a}$ and $\hat{a}'$ associated with observable $A$, and $\hat{b}$ and $\hat{b}'$ associated with observable $B$. Since each of the four terms is bounded by $-1$ and $+1$, in any conceivable theory the left side must be in the range from $-4$ to 4, inclusive. However, in hidden variables theories, a stronger statement holds. This is because a simple brute-force enumeration of all of the possible results for the contents of the square brackets on the right side shows that it can only be 2 or $-2$. (See Table 21.4.3.) Using eq. (21.2.1), this implies that the magnitude of the integral on the right side of eq. (21.4.3) is bounded by 2, and so we have another Bell-type result, the CHSH inequality, named for John Clauser, Michael Horne, Abner Shimony, and Richard Holt,

$$\Big| C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{b}') + C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{b}') \Big| \;\leq\; 2, \tag{21.4.4}$$

which must be obeyed in any local hidden variables theory.

In the ADR experiment, an excited state of a calcium atom with total angular momentum $J = 0$ decays to an intermediate state with $J = 1$ by emitting one photon in an electric dipole

| $f_A(\hat{a}, \lambda)$ | $f_A(\hat{a}', \lambda)$ | $f_B(\hat{b}, \lambda)$ | $f_B(\hat{b}', \lambda)$ | [result] |
|---:|---:|---:|---:|---:|
| 1 | 1 | 1 | 1 | 2 |
| 1 | 1 | 1 | $-1$ | 2 |
| 1 | 1 | $-1$ | 1 | $-2$ |
| 1 | $-1$ | 1 | 1 | $-2$ |
| 1 | 1 | $-1$ | $-1$ | $-2$ |
| 1 | $-1$ | 1 | $-1$ | 2 |
| 1 | $-1$ | $-1$ | 1 | $-2$ |
| 1 | $-1$ | $-1$ | $-1$ | 2 |

Table 21.4.3: Enumeration of possible measurement results pertaining to the ADR experiment, as dictated by eqs. (21.4.1)–(21.4.3), in a theory of local hidden variables. In addition to the eight possibilities shown, there are eight more in which all measurements have the opposite signs. The [result] in the last column is the quantity inside the square brackets on the right side of eq. (21.4.3). It is 2 or $-2$ in all cases.

transition, and then to another $J = 0$ state by emitting a second photon in another electric dipole transition. The photons travel in opposite directions to distant detectors $A$ and $B$, which by a choice of coordinates are located on the positive and negative $z$ axis, respectively. Both photons are emitted with circular polarization, with angular momenta $S_{1z} = \pm\hbar$ and $S_{2z} = \mp\hbar$, in a total spin $S = 0$ state by angular momentum conservation. Thus, if one photon is in a $L$-circular polarization state, then the other also has $L$-circular polarization, because both its momentum direction and its spin are opposite to the first. Similarly, the two photons could both have $R$-circular polarization.

However, the detectors $A$ and $B$ measure the linear polarizations of the photons, not circular polarizations, as follows. A linear polarizer sends each photon into one photomultiplier if the linear polarization is along a selected direction, and into a different one if the linear polarization is in the orthogonal direction. (The allowed polarizations are always orthogonal to the photon momentum direction.) If detector $A$ measures the linear polarization along the chosen direction $\hat{a}$, then it records $A(\hat{a}) = +1$, and if it instead detects the polarization perpendicular to $\hat{a}$, then it records $A(\hat{a}) = -1$. Similarly, if $B$ sees the linear polarization along the chosen direction $\hat{b}$, then it records $B(\hat{b}) = +1$, and if it detects the polarization perpendicular to $\hat{b}$, then it records $B(\hat{b}) = -1$. The experiment is designed so that detector $A$ can switch between two choices $\hat{a}$ and $\hat{a}'$, and detector $B$ can switch between $\hat{b}$ and $\hat{b}'$. The switching is done randomly and quickly, with times shorter than the light travel time between the two detectors. This avoids any possibility of a bias of the results by some hypothetical mechanism that might relate the choices made by the detectors.

Let us now work out the quantum mechanics prediction for the correlations of measurements of $A(\hat{a})$ and $B(\hat{b})$. As already noted, the photons are produced in a state where the circular polarizations are the same, but could be either both $L$ or both $R$. Since the atomic transitions have equal amplitudes for these two possibilities due to spherical symmetry, we can write

$$|S=0\rangle \;=\; \frac{1}{\sqrt{2}}\big(|L,L\rangle + |R,R\rangle\big). \tag{21.4.5}$$

where the first and second ket labels refer to the photons that are seen by detectors $A$ and $B$, respectively. The single-photon orthobasis states obey $\langle R|R\rangle = \langle L|L\rangle = 1$ and $\langle L|R\rangle = \langle R|L\rangle = 0$. To make contact with the measurements made by the experiment, it is convenient to rewrite the state using a linear polarization orthobasis. Our chosen coordinates imply that the momenta of the photons that reach detectors $A$ and $B$ point in the directions $\hat{k}_A = \hat{z}$ and $\hat{k}_B = -\hat{z}$. Then the transformations from the circular polarization orthobases to the linear polarization orthobases are

$$|L\rangle_A \;=\; \frac{1}{\sqrt{2}}\big(|\hat{x}\rangle_A - i\,|\hat{y}\rangle_A\big), \tag{21.4.6}$$

$$|R\rangle_A \;=\; \frac{1}{\sqrt{2}}\big(|\hat{x}\rangle_A + i\,|\hat{y}\rangle_A\big), \tag{21.4.7}$$

$$|L\rangle_B \;=\; \frac{1}{\sqrt{2}}\big(|\hat{x}\rangle_B + i\,|\hat{y}\rangle_B\big), \tag{21.4.8}$$

$$|R\rangle_B \;=\; \frac{1}{\sqrt{2}}\big(|\hat{x}\rangle_B - i\,|\hat{y}\rangle_B\big), \tag{21.4.9}$$

where $|\hat{x}\rangle$ and $|\hat{y}\rangle$ are states with linear polarizations along the $x$ and $y$ axes respectively, and $\langle\hat{x}|\hat{x}\rangle = \langle\hat{y}|\hat{y}\rangle = 1$ and $\langle\hat{x}|\hat{y}\rangle = \langle\hat{y}|\hat{x}\rangle = 0$ in each case. Using eqs. (21.4.6)–(21.4.9) in eq. (21.4.5),

$$|S=0\rangle \;=\; \frac{1}{\sqrt{2}}\big(|\hat{x},\hat{x}\rangle + |\hat{y},\hat{y}\rangle\big). \tag{21.4.10}$$

Note that this state has even parity, because it is invariant under $(\hat{x},\hat{y}) \rightarrow (-\hat{x},-\hat{y})$, in agreement with the fact that the transition occurred between two even-parity atomic states.

A state $|\hat{n}\rangle$ describing a photon with linear polarization $\hat{n}$ has, up to an arbitrary phase,

$$\langle\hat{n}|\hat{x}\rangle = \hat{n}_x, \qquad \langle\hat{n}|\hat{y}\rangle = \hat{n}_y. \tag{21.4.11}$$

Therefore, if we use $|\hat{a},\hat{b}\rangle$ to denote the state in which the linear polarizations are $\hat{a}$ and $\hat{b}$, then

$$\langle\hat{a},\hat{b}|S=0\rangle \;=\; \frac{1}{\sqrt{2}}\left(\hat{a}_x\hat{b}_x + \hat{a}_y\hat{b}_y\right), \tag{21.4.12}$$

again up to an irrelevant phase. So, the probability for the linear polarizations to be observed as $A(\hat{a}) = 1$, $B(\hat{b}) = 1$ is

$$\frac{1}{2}(\hat{a}\cdot\hat{b})^2 \;=\; \frac{1}{2}\cos^2\theta_{ab}, \tag{21.4.13}$$

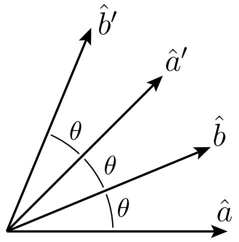

Figure 21.4.1: The configuration for the unit vectors $\hat{a}$, $\hat{a}'$, $\hat{b}$, and $\hat{b}'$ that provides the maximum violation of the Bell-type inequality eq. (21.4.4) in the ADR photon polarization experiment, using the quantum mechanics prediction of eq. (21.4.15). The three angles labeled $\theta$ are all equal to $\pi/8$, so that $\theta_{ab} = \theta_{a'b} = \theta_{a'b'} = \pi/8$, and $\theta_{ab'} = 3\pi/8$.

where $\theta_{ab}$ is the angle between $\hat{a}$ and $\hat{b}$. The same result is obtained for the probability to observe the polarizations both orthogonal to $\hat{a}$ and $\hat{b}$, so that $A(\hat{a}) = -1$, $B(\hat{b}) = -1$. The probabilities to observe $A(\hat{a}) = 1$, $B(\hat{b}) = -1$ and $A(\hat{a}) = -1$, $B(\hat{b}) = 1$ are both

$$\frac{1}{2}|\hat{a} \times \hat{b}|^2 \;=\; \frac{1}{2}\sin^2\theta_{ab}. \tag{21.4.14}$$

Weighting each of the outcomes for $A(\hat{a})B(\hat{b})$ by these respective probabilities, we arrive at

$$\begin{aligned} C(\hat{a},\hat{b}) &= \langle A(\hat{a})B(\hat{b})\rangle \;=\; \frac{1}{2}\cos^2\theta_{ab} + \frac{1}{2}\cos^2\theta_{ab} - \frac{1}{2}\sin^2\theta_{ab} - \frac{1}{2}\sin^2\theta_{ab} \\ &= \cos(2\theta_{ab}). \end{aligned} \tag{21.4.15}$$

This is the quantum mechanics prediction for the correlation.

The local hidden variables Bell inequality, eq. (21.4.4), is incompatible with the quantum mechanics result of eq. (21.4.15) for many choices of $\hat{a}$, $\hat{a}'$, $\hat{b}$, and $\hat{b}'$. (Note that due to the geometry of the experiment and the transverse polarization of photons, these four unit vectors must all lie in the $xy$ plane.) The most extreme violation of the inequality, used in the actual ADR experiment, occurs if one chooses the unit vectors as depicted in Figure 21.4.1, with

$$C(\hat{a},\hat{b}) \;=\; C(\hat{a}',\hat{b}) \;=\; C(\hat{a}',\hat{b}') \;=\; \cos(\pi/4) \;=\; \frac{1}{\sqrt{2}}, \tag{21.4.16}$$

$$C(\hat{a},\hat{b}') \;=\; \cos(3\pi/4) \;=\; -\frac{1}{\sqrt{2}}. \tag{21.4.17}$$

The (idealized) quantum mechanics prediction for that particular geometry is therefore

$$\left|C(\hat{a},\hat{b}) - C(\hat{a},\hat{b}') + C(\hat{a}',\hat{b}) + C(\hat{a}',\hat{b}')\right| \;=\; 2\sqrt{2} \;\approx\; 2.828. \tag{21.4.18}$$

The expected quantum mechanics prediction for the ADR experiment as performed was slightly lower at 2.70 ± 0.05, due to experimental non-ideal realities. This was in excellent agreement with the actual experimental result 2.697 ± 0.015 reported by ADR. This experimental result is also completely incompatible with the upper bound of 2 from eq. (21.4.4) in all theories with local hidden variables.

Further experiments have only sharpened the result that local hidden variables cannot explain observed phenomena, and non-local entanglement of states as predicted by quantum mechanics is both consistent and supported by the experimental evidence. “Local realism” is not real.

Since local hidden variables theories cannot work, one might entertain the possibility of “non-local hidden variables”, which could allow instantaneous interactions between space-like separated points. This is hard to rule out in general because it could predict anything, including causality violation; one might as well explain experiments by invoking sorcery. Furthermore, non-local hidden variables abandon the original motivation of local realism, and in any case are a cure for a disease that does not exist.

# 22 Feynman path integral approach

## 22.1 Propagators

Consider a quantum system with a complete set of commuting observables $Q_a$, where $a$ is an index $1, \ldots, n$. For example, for a single spin-less particle moving in three-dimensional space, we have $n = 3$, and the $Q_a$ with $a = 1, 2, 3$ could be chosen to be the components of the position coordinate vector $\vec{R}$. Alternatively, we could choose the $Q_a$ to be the components of the momentum vector $\vec{P}$. For simplicity, we will often suppress the index $a$, and let the whole set of CSCO eigenvalues be represented simply as the generic symbol $q$. We do this to make room for other subscript labels that will distinguish the values of the $q$'s at various times. Given the state ket $|\psi\rangle$ of the system, we can then write a wavefunction

$$\psi(q, t) \;=\; \langle q|\psi(t)\rangle\,, \tag{22.1.1}$$

where $|q\rangle$ are the CSCO eigenstates in which each of the observables $Q_a$ is known to be equal to the corresponding eigenvalue $q_a$. This is only possible because the observables $Q_a$ are compatible (commuting).

Recall, from section 3.4, that if the state of the system at time $t_0$ was $|\psi(t_0)\rangle$, then the state at time $t$ is

$$|\psi(t)\rangle \;=\; U(t, t_0)\,|\psi(t_0)\rangle\,, \tag{22.1.2}$$

where $U(t, t_0)$ is the unitary time evolution operator. Using completeness of the $Q$ eigenstates at some initial time $t_0$, the wavefunction at any later time $t$ is

$$\psi(q, t) \;=\; \int dq_0\;\langle q|U(t, t_0)|q_0\rangle\,\langle q_0|\psi(t_0)\rangle\,. \tag{22.1.3}$$

It is now useful to define the **propagator** as the function obtained as the relevant matrix element of the unitary time-evolution operator,

$$U(q, t;\, q_0, t_0) \;\equiv\; \langle q|U(t, t_0)|q_0\rangle\,, \tag{22.1.4}$$

so that

$$\psi(q, t) \;=\; \int dq_0\, U(q, t; q_0, t_0)\,\psi(q_0, t_0). \tag{22.1.5}$$

The utility of the propagator function is that, given the wavefunction $\psi(q_0, t_0)$ at an initial time $t_0$, the wavefunction at any other time $t$ can be obtained by doing this integral over $q_0$.

The propagator is the solution to a differential equation, which we can derive by making use of eq. (3.4.2),

$$i\hbar\frac{\partial}{\partial t}U(q, t;\, q_0, t_0) \;=\; i\hbar\,\langle q|\frac{\partial}{\partial t}U(t, t_0)|q_0\rangle\Big] \;=\; \langle q|HU(t, t_0)|q_0\rangle\,. \tag{22.1.6}$$

Consider, for example, the case of a particle moving in three dimensions in a potential $V$, so that the Hamiltonian is $H = P^2/2m + V(\vec{R})$. Interpreting the $q$ as the position $\vec{r}$, eq. (22.1.5) becomes

$$\psi(\vec{r}, t) \;=\; \int d^3\vec{r}_0 \, U(\vec{r}, t;\, \vec{r}_0, t_0)\, \psi(\vec{r}_0, t). \tag{22.1.7}$$

where eq. (22.1.6) tells us that

$$\left[ i\hbar \frac{\partial}{\partial t} + \frac{\hbar^2 \nabla^2}{2m} - V(\vec{r}) \right] U(\vec{r}, t;\, \vec{r}_0, t_0) \;=\; 0. \tag{22.1.8}$$

Since this is a first-order differential equation in $t$, it can in principle be solved given the boundary condition at $t = t_0$, which is $U(\vec{r}, t_0;\, \vec{r}_0, t_0) = \delta^{(3)}(\vec{r} - \vec{r}_0)$. In practice, this might be very difficult, depending on the choice of potential.

Returning to the general case, suppose that the Hamiltonian does not depend explicitly on time, so that $U(t, t_0) = e^{-i(t-t_0)H/\hbar}$, from eq. (3.4.3). If we know all of the $H$ eigenstates $|n\rangle$, with energies $E_n$, then using completeness we get

$$U(q, t;\, q_0, t_0) \;=\; \sum_n e^{-i(t-t_0)E_n/\hbar} \, \langle q|n\rangle \, \langle n|q_0\rangle \;=\; \sum_n e^{-i(t-t_0)E_n/\hbar} \psi_n(q) \psi_n(q_0)^*, \tag{22.1.9}$$

where $\psi_n(q) = \langle q|n\rangle$ are the wavefunctions of the orthonormal Hamiltonian eigenstates with energies $E_n$. Thus, the propagator can be evaluated as a sum over products of energy eigenstate wavefunctions, weighted by phases that vary linearly with time.

As a simple example, consider the propagator for a free particle of mass $m$, moving in one dimension, with the position $x$ playing the role of $q$ in the preceding. We have already done the work for this case, in section 6.2, and the result of eq. (6.2.9) can be rewritten as

$$U(x, t;\, x_0, t_0) \;=\; \left[ \frac{m}{2\pi i\hbar (t - t_0)} \right]^{1/2} \exp\left[ \frac{im(x - x_0)^2}{2\hbar (t - t_0)} \right]. \tag{22.1.10}$$

Repeating these steps for a spin-less free particle moving in three dimensions gives

$$U(\vec{r}, t;\, \vec{r}_0, t_0) \;=\; \left[ \frac{m}{2\pi i\hbar (t - t_0)} \right]^{3/2} \exp\left[ \frac{im|\vec{r} - \vec{r}_0|^2}{2\hbar (t - t_0)} \right] \tag{22.1.11}$$

for the position-representation propagator.

For a somewhat more involved example, consider the one-dimensional harmonic oscillator, with the familiar Hamiltonian $H = \frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2$. One way of writing the propagator is to use the spectral decomposition as in eq. (22.1.9), which says

$$U(x, t; x_0, t_0) \;=\; \sum_{n=0}^{\infty} \psi_n(x) \psi_n(x_0)^* e^{-i\omega(t-t_0)(n+1/2)}, \tag{22.1.12}$$

where $\psi_n(x)$ are the stationary state wavefunctions of eq. (7.2.28), involving Hermite polynomials. It may therefore come as a surprise that the propagator can be written in a nice form that does not involve Hermite polynomials or an infinite sum,

$$U(x, t;\, x_0, t_0) \;=\; \left[\frac{m\omega}{2\pi i\hbar \sin(\omega(t - t_0))}\right]^{1/2} \exp\left(iS_{\rm cl}[x, t; x_0, t_0]/\hbar\right), \tag{22.1.13}$$

where, for general initial and final positions and times,

$$S_{\rm cl}[x_f, t_f; x_i, t_i] \;=\; \frac{m\omega}{2\sin(\omega(t_f - t_i))}\left\{(x_f^2 + x_i^2)\cos(\omega(t_f - t_i)) - 2x_i x_f\right\}. \tag{22.1.14}$$

Remarkably, this is the action for the classical trajectory that starts at $x_i$ at time $t_i$, and ends at $x_f$ at time $t_f$. Recall that in general the **action** is defined by

$$S = \int_{t_i}^{t_f} dt\, L(x, \dot{x}), \tag{22.1.15}$$

where $L(x, \dot{x})$ is the Lagrangian. In the present case, $L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2 x^2$, and you can check that eq. (22.1.14) follows from

$$x_{\rm cl}(t) \;=\; \frac{x_f \sin(\omega(t - t_i)) - x_i \sin(\omega(t - t_f))}{\sin(\omega(t_f - t_i))}, \tag{22.1.16}$$

which is the classical trajectory that satisfies the equation of motion $\ddot{x} = -\omega^2 x$ and the initial and final boundary conditions.

To derive the form of the quantum harmonic oscillator propagator claimed in eq. (22.1.13) is somewhat non-trivial, but once it has been written down, it is not too hard to verify. This consists of checking that it obeys the first-order (in time) differential equation

$$\left(i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\frac{d^2}{dx^2} - \frac{1}{2}m\omega^2 x^2\right) U(x, t;\, x_0, t_0) \;=\; 0, \tag{22.1.17}$$

as in eq. (22.1.8), and that it satisfies the correct boundary condition

$$U(x, t_0;\, x_0, t_0) \;=\; \delta(x - x_0), \tag{22.1.18}$$

which is checked by taking the $t \to t_0$ limit of eq. (22.1.13), with the help of eq. (2.2.21).

The propagator for the oscillator in eq. (22.1.13) has several features worthy of note. First, it is manifestly periodic in time, with the same period $2\pi/\omega$ as the classical oscillator. Second, one recovers the free particle propagator of eq. (22.1.10) by taking the limit $\omega \to 0$. Finally, it is intriguing that the only dependence of the quantum propagator on the starting and ending positions $x$ and $x_0$ comes through the $e^{iS/\hbar}$ factor for the classical trajectory that connects them. This is a clue for the developments in the next section.

## 22.2 Summing over paths

An alternative way of computing the propagator for a quantum system uses the **Feynman sum over paths** or **sum over histories**, developed by Richard P. Feynman in his 1942 PhD thesis, fulfilling an inspiring but less concrete proposal made earlier by Dirac. Let us write, very schematically for now,

$$U(q_f, t_f; q_i, t_i) \;\sim\; \sum_{\text{paths}} (\text{something}). \tag{22.2.1}$$

In this section, we will denote the initial and final CSCO eigenvalue variables and time by $(q_i, t_i)$ and $(q_f, t_f)$ respectively, in order to distinguish them from intermediate times $t$ and positions $q$ in the following discussion. The "paths" here are trajectories $q(t)$ starting from $q_i$ at time $t_i$ and ending with $q_f$ at time $t_f$. In the case of a spin-less particle in three dimensions, a path could be a trajectory given by a function $\vec{r}(t)$, but it could just as easily be instead a function $\vec{p}(t)$ for the momentum vector of the particle. Once we have chosen the CSCO eigenvalue variables, these paths are arbitrary, in the sense that all paths are included as long as they obey the initial and final boundary conditions.

The task before us now is to make eq. (22.2.1), including its mysterious summand, more precise. To motivate this form, note that

$$U(q_f, t_f;\, q_i, t_i) \;=\; \langle q_f | U(t_f, t_1) U(t_1, t_i) | q_i\rangle \;=\; \int dq_1\; \langle q_f | U(t_f, t_1) | q_1\rangle \langle q_1 | U(t_1, t_i) | q_i\rangle\,, \tag{22.2.2}$$

where we have chosen an arbitrary time $t_1$ between $t_i$ and $t_f$, divided the unitary time evolution operator accordingly using eq. (3.4.24), and then used completeness of the eigenstates $|q_1\rangle$. This can be rewritten as a composition rule for propagators,

$$U(q_f, t_f;\, q_i, t_i) \;=\; \int dq_1\, U(q_f, t_f;\, q_1, t_1)\, U(q_1, t_1;\, q_i, t_i). \tag{22.2.3}$$

In words, to propagate from time $t_i$ to time $t_f$, we can integrate over all of the ways to propagate from $t_i$ to $t_1$, and then from $t_1$ to $t_f$, as indicated in Figure 22.2.1.

We now use the same idea to subdivide the time interval from $t_i$ to $t_f$ into $N$ time steps instead of just two, with intermediate times $t_1, t_2, \ldots, t_{N-1}$, as shown in Figure 22.2.2. At each time step, we sum (actually, integrate) over all possible $q_j$ for $j = 1, 2, \ldots, N-1$, using completeness of the states $|q_j\rangle$ at each step. The result is

$$\begin{aligned} U(q_f, t_f;\, q_i, t_i) \;=\; & \int dq_1 \int dq_2 \cdots \int dq_{N-1}\, U(q_f, t_f;\, q_{N-1}, t_{N-1}) U(q_{N-1}, t_{N-1};\, q_{N-2}, t_{N-2}) \cdots \\ & U(q_2, t_2;\, q_1, t_1) U(q_1, t_1;\, q_i, t_i). \end{aligned} \tag{22.2.4}$$

Although this formula is not very practically useful in its present form, it illustrates the truth of the idea that we can sum over all possible ways to get from $q_i$ at time $t_i$ to $q_f$ at time $t_f$.

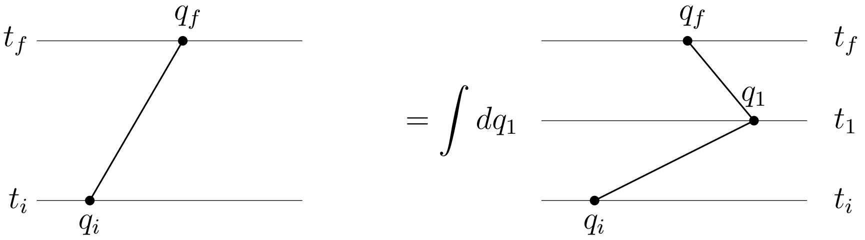

Figure 22.2.1: A diagrammatic representation of the fact that the propagator $U(q_f, t_f; q_i, t_i)$ obeys a composition law eq. (22.2.3); it is equal to the integral over all possible $q_1$ of the product $U(q_f, t_f; q_1, t_1)U(q_1, t_1; q_i, t_i)$, for any choice of $t_1$.

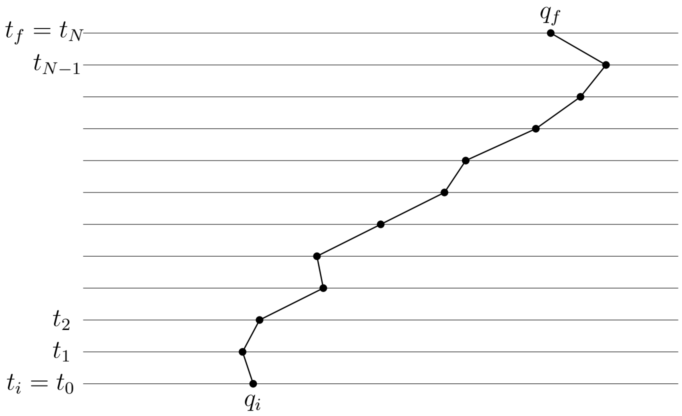

Figure 22.2.2: One possible trajectory, or path, $q(t)$ with initial boundary condition $q_i$ at time $t_i$ and final boundary condition $q_f$ at time $t_f$. Time increases moving up in the figure, and is divided into equal slices with variables $q_1, q_2, \ldots, q_{N-1}$ at intermediate times $t_1, t_2, \ldots, t_{N-1}$. The path integral is obtained by taking the limit of $N \to \infty$ time steps with infinitesimal time intervals $\epsilon = t_j - t_{j-1} = (t_f - t_i)/N$, and integrating over all $q_1, q_2, \ldots, q_{N-1}$.

To realize this idea in a more useful way, Feynman proposed that the propagator can be computed as a sum over all possible space-time paths connecting the initial and final configurations specified by the arguments of the propagator, weighted by a complex phase determined by the classical action for the path. The path is any function $q(t)$ defined for $t_i \leq t \leq t_f$ and constrained to obey the boundary conditions

$$q(t_i) = q_i, \qquad q(t_f) = q_f. \tag{22.2.5}$$

The action for a given path is

$$S[q(t)] \;=\; \int_{t_i}^{t_f} dt\, L(q,\, \dot{q},\, t), \tag{22.2.6}$$

where $L(q,\, \dot{q},\, t)$ is the classical Lagrangian for the system, often the kinetic energy minus the potential energy. For example, for a spin-less particle moving in three dimensions, the position vector $\vec{r}(t)$ plays the role of $q(t)$, and the Lagrangian is

$$L(\vec{r},\, \dot{\vec{r}},\, t) \;=\; \frac{1}{2}m\dot{\vec{r}}^{\,2} - V(\vec{r}). \tag{22.2.7}$$

The action is an example of a **functional**, an object that takes as its input a function [here the path $q(t)$] and returns a number. Feynman therefore proposed that eq. (22.2.1) should be

$$U(q_f, t_f;\, q_i, t_i) \propto \sum_{\text{paths } q(t)} e^{iS[q(t)]/\hbar}, \tag{22.2.8}$$

following a suggestion by Dirac.

Because the number of paths is continuously infinite, it makes more sense to think of integrating over all paths rather than summing them. This suggests the arguably better names **Feynman path integral** or **functional integral**, and the notational change

$$\sum_{\text{paths } q(t)} e^{iS[q(t)]/\hbar} \;\to\; \int d[q(t)]\, e^{iS[q(t)]/\hbar}. \tag{22.2.9}$$

Just as ordinary definite integration takes a function and returns a number, a functional integral takes a functional (in this case, the action) and returns a number. The constant of proportionality implied in eq. (22.2.8) can then be absorbed into the normalization for the functional differential, denoted by $d[q(t)]$. The following can now be taken as an alternative to the Schrödinger equation Postulate 6 of section 3.1:

**Postulate 6′: Time evolution as a sum over paths.** For a system with eigenvalues $\vec{q}$ for a complete set of commuting observables, the propagator is

$$U(\vec{q}_f, t_f;\, \vec{q}_i, t_i) \;=\; \int d[\vec{q}(t)]\, \exp\left(iS[\vec{q}(t)]/\hbar\right). \tag{22.2.10}$$

where $S[\vec{q}(t)] = \int_{t_i}^{t_f} dt\, L(\vec{q}, \dot{\vec{q}})$ is the action functional, and the functional integral is over all paths satisfying $\vec{q}(t_i) = \vec{q}_i$ and $\vec{q}(t) = \vec{q}_f$. Equivalently, the wavefunction satisfies

$$\psi(\vec{q}_f, t_f) \;=\; \int d[\vec{q}(t)]\, \exp\left(iS[\vec{q}(t)]/\hbar\right) \psi(\vec{q}_i, t_i), \tag{22.2.11}$$

but now the functional integral is over all paths satisfying $\vec{q}(t_f) = \vec{q}_f$, but with no constraint on $\vec{q}(t_i)$, since those initial values are being integrated over, as in eq. (22.1.5).

Note that while the Schrödinger version of Postulate 6 requires us to specify the Hamiltonian, the Feynman version Postulate 6′ instead requires that the Lagrangian is the object to be specified to determine the dynamics of the system.

Although path integration has proved to be a very useful tool, especially in quantum field theory, it is not as straightforward to define as the Schrödinger differential equation. One way to define the functional integral, inspired by the composition rule for the propagator as in eq. (22.2.4) and Figure 22.2.2, is to discretize the path into steps, integrate over the values of $q$ on each step, and then take the limit in which the number of steps is infinitely large.

As a specific example, consider the case of a spin-less particle moving in a potential $V(\vec{r})$ in three dimensions, with the coordinates $\vec{r}$ playing the role of $\vec{q}$. To compute the path integral, we partition the time interval $(t_i, t_f)$ into $N$ equal steps of duration

$$\epsilon = (t_f - t_i)/N, \tag{22.2.12}$$

so that

$$t_j \;=\; t_i + j\epsilon, \qquad (j = 0, 1, 2, \ldots, N), \tag{22.2.13}$$

with $t_i = t_0$ and $t_f = t_N$. We then define the discretized version of a path from $\vec{r}_i$ to $\vec{r}_f$ by positions $\vec{r}_j$ for each $t_j$ for $j = 1, \ldots, N-1$, with $\vec{r}_0 = \vec{r}_i$ and $\vec{r}_N = \vec{r}_f$. The velocity of the particle in the interval between times $t_{j-1}$ and $t_j$ is $\vec{v} = (\vec{r}_j - \vec{r}_{j-1})/\epsilon$, so the kinetic energy is $\frac{1}{2}mv^2 = \frac{1}{2}m(|\vec{r}_j - \vec{r}_{j-1}|/\epsilon)^2$. Therefore, we can write the discretized version of the action for the path as

$$S[\text{path}] \;=\; S(\vec{r}_i, \vec{r}_1, \vec{r}_2, \ldots, \vec{r}_{N-1}, \vec{r}_f) \;=\; \epsilon \sum_{j=1}^{N} \left[ \frac{m|\vec{r}_j - \vec{r}_{j-1}|^2}{2\epsilon^2} - V(\vec{r}_j) \right]. \tag{22.2.14}$$

In the limit $\epsilon \to 0$ with $N = (t_f - t_i)/\epsilon \to \infty$, this will give the action for the continuous path. We now write eq. (22.2.10) as

$$\psi(\vec{r}_f, t_f) \;=\; \lim_{N\to\infty} \int \frac{d^3\vec{r}_0}{a^3} \int \frac{d^3\vec{r}_1}{a^3} \int \frac{d^3\vec{r}_2}{a^3} \cdots \int \frac{d^3\vec{r}_{N-1}}{a^3} \exp\bigl[iS[\text{path}]/\hbar\bigr]\, \psi(\vec{r}_0, t_0), \tag{22.2.15}$$

where $a$ is a normalization factor, with units of [length], to be found soon. So, removing the integration over $\vec{r}_0$ and the initial wavefunction $\psi(\vec{r}_0, t_0)$,

$$U(\vec{r}_f, t_f;\, \vec{r}_i, t_i) \;=\; \lim_{N\to\infty} \frac{1}{a^3} \int \frac{d^3\vec{r}_1}{a^3} \int \frac{d^3\vec{r}_2}{a^3} \cdots \int \frac{d^3\vec{r}_{N-1}}{a^3} \exp\bigl[iS[\text{path}]/\hbar\bigr], \tag{22.2.16}$$

in which it is important that there is one more factor of $1/a^3$ than the remaining number of integrations, in accord with the fact that the propagator has dimension $1/[\text{length}]^3$. Equation (22.2.16) is one possible meaning of the notation

$$U(\vec{r}_f, t_f;\, \vec{r}_i, t_i) \;=\; \int d[\vec{r}(t)]\, \exp\left(iS[\vec{r}(t)]/\hbar\right) \tag{22.2.17}$$

for the path integral.

Let us now show that the preceding proposal is indeed equivalent to the Schrödinger equation. It is sufficient to consider only time evolution over a single infinitesimal time interval, since longer time intervals can be obtained by using the composition law eq. (22.2.3). For simplicity, we write $t_i = t_0 = t$, and after one infinitesimal time step, $t_f = t + \epsilon$. Then, using the $N = 1$ version of eq. (22.2.15), the wavefunction at the final time in terms of the wavefunction at the initial time is

$$\psi(\vec{r},\, t+\epsilon) \;=\; \int \frac{d^3\vec{r}_0}{a^3} e^{iS(\vec{r},\,\vec{r}_0)/\hbar}\psi(\vec{r}_0, t). \tag{22.2.18}$$

Expanding the left side in small $\epsilon$ gives

$$\text{LS of (22.2.18)} \;=\; \psi(\vec{r},\, t+\epsilon) \;=\; \psi(\vec{r}, t) + \epsilon\frac{\partial}{\partial t}\psi(\vec{r}, t) + \cdots . \tag{22.2.19}$$

The right side of eq. (22.2.18) is

$$\text{RS of (22.2.18)} \;=\; \int \frac{d^3\vec{r}_0}{a^3} \exp\left(\frac{i}{\hbar}\left[\frac{m|\vec{r}-\vec{r}_0|^2}{2\epsilon} - \epsilon V(\vec{r})\right]\right)\psi(\vec{r}_0, t), \tag{22.2.20}$$

or, expanding the exponential part involving $V(\vec{r})$ to order $\epsilon$, and then shifting the integration variable according to $\vec{r}_0 \to \vec{r}_0 + \vec{r}$,

$$\text{RS of (22.2.18)} = \left[1 - i\frac{\epsilon}{\hbar}V(\vec{r})\right]\int \frac{d^3\vec{r}_0}{a^3}\, e^{imr_0^2/2\hbar\epsilon}\psi(\vec{r}+\vec{r}_0, t). \tag{22.2.21}$$

Now, in the limit of small $\epsilon$, there will be almost complete cancellation from the rapidly varying phase inside the integral. The only region that contributes as $\epsilon \to 0$ is where $r_0^2 \approx 0$, so that the phase factor $e^{imr_0^2/2\hbar\epsilon}$ is close to 1. Therefore, we can use the expansion for small $r_0$,

$$\psi(\vec{r}+\vec{r}_0, t) \;=\; \psi(\vec{r}, t) + \vec{r}_0\cdot\vec{\nabla}\psi(\vec{r}, t) + \frac{1}{2}(\vec{r}_0\cdot\vec{\nabla})^2\psi(\vec{r}, t) + \cdots . \tag{22.2.22}$$

Inside the integration over $\vec{r}_0$, this can be simplified because the part odd under $(x_0, y_0, z_0) \to -(x_0, y_0, z_0)$ will cancel. Thus, the linear term does not contribute at all, and using rectangular coordinates and the spherical symmetry of the integrand you can check that inside the integral the last term can be replaced according to $(\vec{r}_0\cdot\vec{\nabla})^2 \to \frac{1}{3}r_0^2\nabla^2$. So, we have

$$\int d^3\vec{r}_0\, e^{imr_0^2/2\hbar\epsilon}\psi(\vec{r}+\vec{r}_0, t) \;=\; \psi(\vec{r}, t)\int d^3\vec{r}_0\, e^{imr_0^2/2\hbar\epsilon} + \frac{1}{6}\nabla^2\psi(\vec{r}, t)\int d^3\vec{r}_0\, r_0^2 e^{imr_0^2/2\hbar\epsilon} + \cdots , \tag{22.2.23}$$

in which the neglected terms have higher powers of $r_0^2$, and so can be checked to contribute only at higher order in $\epsilon$. Using

$$\int d^3\vec{r}_0\, e^{imr_0^2/2\hbar\epsilon} \;=\; 4\pi\int_0^\infty dr_0\, r_0^2\, e^{imr_0^2/2\hbar\epsilon} \;=\; \left(\frac{2\pi i\hbar\epsilon}{m}\right)^{3/2}, \tag{22.2.24}$$

$$\int d^3\vec{r}_0\, r_0^2\, e^{imr_0^2/2\hbar\epsilon} \;=\; 4\pi\int_0^\infty dr_0\, r_0^4\, e^{imr_0^2/2\hbar\epsilon} \;=\; i\frac{3\hbar\epsilon}{m}\left(\frac{2\pi i\hbar\epsilon}{m}\right)^{3/2}, \tag{22.2.25}$$

and plugging into eq. (22.2.21) gives

$$\text{RS of (22.2.18)} = \frac{1}{a^3}\left(\frac{2\pi i\hbar\epsilon}{m}\right)^{3/2}\left[1 + i\frac{\epsilon}{\hbar}\left(\frac{\hbar^2\nabla^2}{2m} - V(\vec{r})\right) + \mathcal{O}(\epsilon^2)\right]\psi(\vec{r},t). \qquad (22.2.26)$$

Now compare eqs. (22.2.19) and (22.2.26). Matching at leading order in $\epsilon$ just informs us that the normalization factor is

$$a = \left(\frac{2\pi i\hbar\epsilon}{m}\right)^{1/2}. \qquad (22.2.27)$$

Then, matching the terms at first order in $\epsilon$ gives

$$i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t) = \left[-\frac{\hbar^2\nabla^2}{2m} + V(\vec{r})\right]\psi(\vec{r},t), \qquad (22.2.28)$$

which is the Schrödinger equation in the position wavefunction representation. This completes the demonstration that the Feynman sum over paths formulation is equivalent to the Schrödinger equation time evolution.

## 22.3 Evaluation of the path integral for the harmonic oscillator

In the previous section we demonstrated the equivalence of the path integral formulation and the Schrödinger equation, but this is not the same as actually computing a useful result directly in terms of the path integral. In this section we will show how to arrive at the propagator for the one-dimensional harmonic oscillator in the path integral approach.

Consider the propagator as a path integral, starting at time $t = 0$ and ending at time $t = T$,

$$U(x_f, T; x_i, 0) = \int d[x(t)]\, e^{iS[x(t)]/\hbar}. \qquad (22.3.1)$$

In the case of the harmonic oscillator, we have a great advantage, that there is a unique classical trajectory, given by eq. (22.1.16) with $t_f = T$ and $t_i = 0$. It will help us greatly to do a change of variables

$$x(t) = x_{\rm cl}(t) + y(t), \qquad (22.3.2)$$

where the new coordinate $y(t)$ can be interpreted as the quantum fluctuation. It satisfies the boundary conditions

$$y(0) = y(T) = 0. \qquad (22.3.3)$$

Now we can evaluate the action functional, with the result

$$S[x(t)] = S[x_{\rm cl}(t)] + \frac{m}{2}\int_0^T dt\,\left(\dot{y}^2 - \omega^2 y^2\right), \qquad (22.3.4)$$

where the cross-terms involving both $x_{\rm cl}$ and $y$ have conveniently canceled. To see this cancelation, note that those cross-terms are

$$m\int_0^T dt\,\left(\dot{x}_{\rm cl}\dot{y} - \omega^2 x_{\rm cl} y\right) \;=\; -m\int_0^T dt\,\left(\ddot{x}_{\rm cl} + \omega^2 x_{\rm cl}\right) y, \tag{22.3.5}$$

where we have integrated by parts, making use of the vanishing of the boundary terms from eq. (22.3.3). The last expression vanishes due to the equation of motion for $x_{\rm cl}$.

Now, since $S[x_{\rm cl}(t)]$ is a constant with respect to the new functional integration variable $y(t)$, we can pull it out of the path integral, to get

$$U(x_f, T; x_i, 0) \;=\; I\, e^{iS[x_{\rm cl}(t)]/\hbar} \tag{22.3.6}$$

where it remains to evaluate the function integral over $y(t)$,

$$I \;\equiv\; \int d[y(t)]\, \exp\left[\frac{im}{2\hbar}\int_0^T dt\,\left(\dot{y}^2 - \omega^2 y^2\right)\right]. \tag{22.3.7}$$

We have already made good progress, since we have successfully obtained the correct $e^{iS[x_{\rm cl}(t)]/\hbar}$ factor in eq. (22.1.13). Indeed, from its definition, the remaining factor $I$ clearly has no dependence on $x_f$ or $x_i$. Our remaining task is to show that, in accord with eq. (22.1.13),

$$I \;=\; \left(\frac{m\omega}{2\pi i\hbar\sin(\omega T)}\right)^{1/2}. \tag{22.3.8}$$

We will do the path integral to get $I$ in two different ways.

First, we use the discretization of the same type used in eq. (22.2.16),

$$I \;=\; \lim_{N\to\infty}\frac{1}{a}\left(\prod_{j=1}^{N-1}\int_{-\infty}^{\infty}\frac{dy_j}{a}\right)\exp\left(\frac{im}{2\hbar\epsilon}\sum_{k=1}^{N}\left[(y_k - y_{k-1})^2 - \epsilon^2\omega^2 y_k^2\right]\right), \tag{22.3.9}$$

where $\epsilon = T/N$, and we have already obtained the normalization $a$ in eq. (22.2.27). Then, letting $z_j = y_j/a$ and assembling $z_1, z_2, \ldots, z_{N-1}$ into an $(N-1)$-dimensional vector $\vec{z}$, we have

$$I \;=\; \lim_{N\to\infty}\frac{1}{a}\int d^N\vec{z}\,\exp\left(-\pi\vec{z}^T B_N\vec{z}\right) \tag{22.3.10}$$

where $B_N$ is the $(N-1)\times(N-1)$ matrix

$$B_N = \begin{pmatrix} 2-\delta & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2-\delta & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2-\delta & -1 & \cdots & 0 \\ 0 & 0 & -1 & 2-\delta & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & 0 & 0 & 0 & -1 & 2-\delta \end{pmatrix}, \tag{22.3.11}$$

which only has non-zero entries on the main diagonal and immediately above and below it, and for convenience we have defined

$$\delta \equiv \epsilon^2\omega^2 = \omega^2 T^2/N^2. \tag{22.3.12}$$

Now, since $B$ is a real symmetric matrix, we can do a rotation in the $(N-1)$-dimensional space to new coordinates $\vec{u} = O\vec{z}$, where $O$ is an orthogonal matrix, in such a way that the rotated matrix $\tilde{B}_N = OB_N O^T$ is diagonal. Since a real symmetric matrix is just a special kind of Hermitian matrix, and an orthogonal matrix is just a special kind of unitary matrix, Theorems 2.6.7 and 2.6.9 apply to tell us that the diagonal entries of $\tilde{B}_N$ are the same as the eigenvalues $\lambda_j$ of $B_N$, so

$$I = \lim_{N\to\infty} \frac{1}{a} \prod_{j=1}^{N-1} \int_{-\infty}^{\infty} du_j\, e^{-\pi\lambda_j u_j^2}. \tag{22.3.13}$$

Now each of the $u_j$ integrals can be done separately, using $\int_{-\infty}^{\infty} du\, e^{-\pi\lambda u^2} = \lambda^{-1/2}$, giving

$$I = \lim_{N\to\infty} \frac{1}{a} \prod_{j=1}^{N-1} (\lambda_j)^{-1/2} = \lim_{N\to\infty} \frac{1}{a} \left(\det B_N\right)^{-1/2} \tag{22.3.14}$$

where we have used the fact from matrix algebra that the determinant of $B_N$ is equal to the product of its eigenvalues.

Let us now find the determinant of the $(N-1)\times(N-1)$ matrix $B_N$, temporarily taking $\delta$ to be general even though in our case it depends on $N$ through eq. (22.3.12). We have

$$B_2 = 2-\delta, \tag{22.3.15}$$

$$B_3 = (2-\delta)^2 - 1, \tag{22.3.16}$$

and for larger $N$ we can evaluate $B_N$ in terms of its minors for the two non-zero entries in the first row. Conveniently, these involve determinants of versions of the same matrix with dimension smaller by 1 and 2. This gives the recurrence relation

$$B_N = (2-\delta)B_{N-1} - B_{N-2}, \qquad (N = 4, 5, 6, \ldots). \tag{22.3.17}$$

It is possible to solve this recurrence relation systematically, but we will take the shortcut of writing down the answer and checking that it works. To do so, we define numbers $\beta_{1,2}$ by

$$\beta_1 + \beta_2 = 2-\delta, \qquad \beta_1\beta_2 = 1. \tag{22.3.18}$$

Then we claim that

$$B_N = \frac{\beta_1^N - \beta_2^N}{\beta_1 - \beta_2}. \tag{22.3.19}$$

This is easily checked to work for $N = 2$ and 3, and then for the recurrence relation eq. (22.3.17), by direct substitution.

We are now ready to take the large $N$ limit. First, solving eq. (22.3.18), we find that

$$\beta_{1,2} \;=\; 1 - \frac{\delta}{2} \pm i\sqrt{\delta - \delta^2/4} \;=\; 1 \pm i\omega T/N + \ldots, \tag{22.3.20}$$

where the ellipses involves terms of higher order in $1/N$. Therefore, for large $N$,

$$\det B_N \;=\; \frac{(1 + i\omega T/N)^N - (1 - i\omega T/N)^N}{2i\omega T/N} \;=\; \frac{e^{i\omega T} - e^{-i\omega T}}{2i\omega T/N} \;=\; \frac{N\sin(\omega T)}{\omega T}, \tag{22.3.21}$$

where the definition of the exponential function, $e^x = \lim_{N\to\infty}(1 + x/N)^N$, was used to get the second equality. Using this in eq. (22.3.14) along with $a = (2\pi i\hbar T/mN)^{1/2}$, we find

$$I = \lim_{N\to\infty}\left(\frac{mN}{2\pi i\hbar T}\right)^{1/2}\left(\frac{\omega T}{N\sin(\omega T)}\right)^{1/2} = \left(\frac{m\omega}{2\pi i\hbar\sin(\omega T)}\right)^{1/2}, \tag{22.3.22}$$

which agrees with the result found before.

Let us now consider a second way to calculate $I$, using a different realization of the path integral. Instead of enumerating the possible paths by discretizing in time, we write the quantum fluctuation as a general Fourier series expansion of terms that satisfy the boundary conditions $y(0) = y(T) = 0$,

$$y(t) \;=\; \sum_{n=1}^{\infty} y_n \sin(n\pi t/T). \tag{22.3.23}$$

We then interpret the path integral in eq. (22.3.7) as an integration over all of the Fourier coefficients $y_n$,

$$I \;=\; \frac{1}{c}\left(\prod_{n=1}^{\infty}\int_{-\infty}^{\infty}\frac{dy_n}{b_n}\right)\exp\left[\frac{im}{2\hbar}\int_0^T dt\;\left(\dot{y}^2 - \omega^2 y^2\right)\right] \tag{22.3.24}$$

where we have introduced normalization factors $c$ and $b_n$ with dimensions of [length]. Now the action integral over $t$ in the exponent can be easily computed, yielding

$$I \;=\; \frac{1}{c}\left(\prod_{n=1}^{\infty}\int_{-\infty}^{\infty}\frac{dy_n}{b_n}\right)\exp\left[-\sum_{n=1}^{\infty}\frac{imT}{4\hbar}\left(\omega^2 - n^2\pi^2/T^2\right)y_n^2\right]. \tag{22.3.25}$$

This nicely separates into one-dimensional integrals over the $y_n$, as

$$I \;=\; \frac{1}{c}\prod_{n=1}^{\infty}\left(\int_{-\infty}^{\infty}\frac{dy_n}{b_n}\exp\left[-\frac{imT}{4\hbar}\left(\omega^2 - n^2\pi^2/T^2\right)y_n^2\right]\right). \tag{22.3.26}$$

Doing the individual Gaussian integrals,

$$I \;=\; \frac{1}{c}\prod_{n=1}^{\infty}\frac{1}{b_n}\left(\frac{4\pi i\hbar}{mT(n^2\pi^2/T^2 - \omega^2)}\right)^{1/2}. \tag{22.3.27}$$

To make further progress, we make use of a famous formula due to Euler,

$$\prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2\pi^2}\right) = \frac{\sin(x)}{x}, \tag{22.3.28}$$

with $x = \omega T$ in our present case, to obtain

$$I = \frac{1}{c}\left(\frac{\omega T}{\sin(\omega T)}\right)^{1/2} \prod_{n=1}^{\infty} \frac{1}{b_n}\left(\frac{4i\hbar T}{n^2\pi m}\right)^{1/2} \tag{22.3.29}$$

Comparing with eq. (22.3.22), we see that we have correctly obtained the $\omega$ dependence, and that the whole formula for $I$ is correct provided that the normalization factors are

$$b_n = \left(\frac{4i\hbar T}{n^2\pi m}\right)^{1/2}, \qquad c = \left(\frac{2\pi i\hbar T}{m}\right)^{1/2}. \tag{22.3.30}$$

The advantage of this Fourier expansion method for the harmonic oscillator path integral is that the separation into individual integrals was automatic. The disadvantage is that the $\omega$-independent normalization factors $1/c$ and $1/b_n$ associated with the path integral had to be obtained by comparison to the known result for $I$. In the discretized version of the path integral, the normalization factor $1/a$ was known already from the case of infinitesimal $T$ in section 22.2.

The case of a free particle can be obtained from the $\omega \to 0$ limit. Beyond that, the path integral method is usually not very efficient compared to other methods in non-relativistic quantum mechanics. Note that in the preceding, we were able to make use of a very special property of the free particle and the harmonic oscillator, namely that the action is quadratic in the configuration variable(s) and its time derivatives, leading to Gaussian integrations in either eq. (22.3.13) or eq. (22.3.26), which we were therefore able to do analytically. This applies only for the very limited number of situations in which the Lagrangian is quadratic, so in other cases approximation methods will be necessary. However, the path integral approach is still extremely valuable, for in addition to the beautiful conceptual understanding it brings, it turns out to provide a practically useful way of systematically organizing such approximation methods. In the case of relativistic quantum field theories, this includes both Feynman rules for perturbative calculations, and other methods for non-perturbative ones, topics beyond the scope of this book.

## 22.4 Classical limit of the sum over paths

Consider a path $q(t)$, and suppose we make a small change $\delta q(t)$ to it,

$$q(t) \to q(t) + \delta q(t), \tag{22.4.1}$$

as illustrated in Figure 22.4.1. This will result in a change in the action, which we write as

$$S[q(t)] \to S[q(t)] + \delta S. \tag{22.4.2}$$

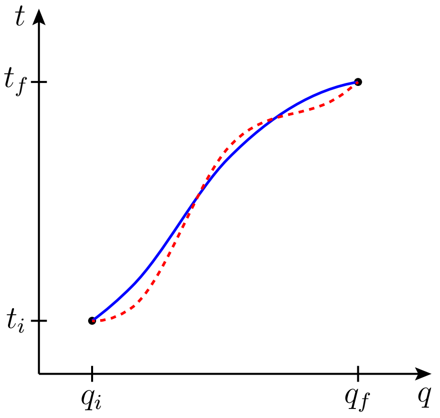

Figure 22.4.1: A path $q(t)$ (solid line), and a nearby path $q(t) + \delta q(t)$ (dashed line), both subject to the boundary condition $\delta q(t_i) = \delta q(t_f) = 0$. The path is a stationary path if the change in the action $\delta S$ vanishes for every infinitesimal change in path $\delta q(t)$.

Because we must maintain the boundary conditions at $t = t_i$ and $t = t_f$ when computing the propagator, we must require

$$\delta q(t_i) \;=\; \delta q(t_f) \;=\; 0. \tag{22.4.3}$$

In general, if $\delta S$ is non-zero, then the integrand of the path integral will acquire an additional phase $e^{i\delta S/\hbar}$ compared to the result from the original path. If we take $\hbar \to 0$, which we associate with the classical limit, this additional phase varies rapidly as $\delta S$ changes, leading to very efficient cancellation between the contributions for nearby paths.

Therefore, in the classical limit, the propagator is dominated by **stationary paths**, which are defined to be those for which $\delta S = 0$ for every $\delta q(t)$ that satisfies the boundary conditions eq. (22.4.3). The point is that those are the only paths that do not suffer efficient cancellations from the neighboring paths, with cancellation becoming perfectly efficient in the $\hbar \to 0$ limit. We can thus write the condition for a stationary path $q(t)$ as the vanishing of the **functional derivative**,

$$\frac{\delta S[q(t)]}{\delta q(t)} \;\equiv\; \lim_{\delta q(t)\to 0} \frac{S[q(t) + \delta q(t)] - S[q(t)]}{\delta q(t)} \;=\; 0. \tag{22.4.4}$$

In quantum mechanics, the path integral samples all paths, but in the classical limit $\hbar \to 0$, only stationary paths contribute non-negligibly.

Suppose we are given a Lagrangian $L(q, \dot{q}, t)$ for a system, so that

$$S[q(t)] \;=\; \int_{t_i}^{t_f} dt\, L(q, \dot{q}, t). \tag{22.4.5}$$

The change in the action due to the change in path $\delta q(t)$ can be written as

$$\delta S = \int_{t_i}^{t_f} dt \left[\delta q(t) \frac{\partial L}{\partial q} + \delta \dot{q}(t) \frac{\partial L}{\partial \dot{q}}\right], \tag{22.4.6}$$

or, using $\delta \dot{q}(t) = \frac{d}{dt} \delta q(t)$,

$$\delta S = \int_{t_i}^{t_f} dt \left[\delta q(t) \frac{\partial L}{\partial q} + \frac{d}{dt} \delta q(t) \frac{\partial L}{\partial \dot{q}}\right]. \tag{22.4.7}$$

Integrating the second term by parts, we have

$$\delta S = \int_{t_i}^{t_f} dt\, \delta q(t) \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right] + \left(\delta q(t) \frac{\partial L}{\partial \dot{q}}\right)\Bigg|_{t_i}^{t_f}. \tag{22.4.8}$$

The last surface term is 0 because of the boundary conditions eq. (22.4.3). In order to have $\delta S = 0$ for all small variations $\delta q(t)$, we conclude that the integrand in eq. (22.4.8) must vanish,

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0. \tag{22.4.9}$$

This is the Lagrangian equation of motion familiar from classical mechanics. More generally, if there are multiple variables $q_a$ with $a = 1, \ldots, N$, with a Lagrangian $L(q_a, \dot{q}_a, t)$, then one can show by the same argument that the classical limit implies the equations of motion

$$\frac{\partial L}{\partial q_a} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_a}\right) = 0 \tag{22.4.10}$$

for each $a$. Thus we have derived classical mechanics from the $\hbar \to 0$ limit of Feynman's path integral formulation of quantum mechanics.

In the case of a single spin-less particle moving in three dimensions, we could of course take our CSCO eigenvalues to be $(q_1, q_2, q_3) \to \vec{r}$. However, in more general situations, we often need to describe processes that change the number of particles. Such processes include decays of one particle into two or more particles, like the decay of a Higgs boson into a quark-antiquark pair, or two gluons, or two photons. They also include the joining of two particles to make particles of a different type, for example the annihilation of an electron and positron pair to make a $Z$ boson, which then turns into a quark-antiquark pair. In such cases, the way we have been treating quantum mechanics is rather clumsy, because even the number of coordinates needed to describe the positions or momenta of the particles changes depending on the state; in other words, they are not a good CSCO.

Furthermore, special relativity teaches us that time and space are on the same footing, while we have been treating spatial coordinates as operators, but using the time coordinate as a parameter, not an operator. This is particularly embarrassing when we realize that the symmetries of special relativity mix space and time coordinates when we change our inertial reference frame.

The simultaneous resolution of these issues is that instead of associating the position coordinates $\vec{r}$ with operators, they are taken to be labels, just like the time $t$. In other words, there are different operators

$$\Phi(\vec{r}, t) \tag{22.4.11}$$

for each label $(\vec{r}, t)$. The operators $\Phi(\vec{r}, t)$, called **quantum fields**, are associated with the different types of particles in the theory, and can also carry indices associated with their spin and other quantum numbers. The field operators can act on a vacuum state $|0\rangle$ (which describes empty space with no particles, and is the lowest energy state) to give other states that have any number of particles and antiparticles. In the path integral formulation, the classical action functional looks schematically like

$$S[\phi(\vec{r}, t)] \;=\; \int_{t_i}^{t_f} dt \int d^3\vec{r}\; \mathcal{L}(\phi, \dot{\phi}, \vec{\nabla}\phi), \tag{22.4.12}$$

where $\mathcal{L}$ is called the classical Lagrangian density, and $\phi$ is the classical field corresponding to the quantum field $\Phi$.

The resulting type of theory, which also obeys the other postulates of section 3.1, is called a **quantum field theory**. Relativistic quantum field theories have the important property that energies are bounded from below, which is not the case in relativistic one-particle quantum mechanics. In quantum field theory, the Schrödinger equation formulation is valid, but often less useful than the equivalent path integral formulation, which provides elegant methods for both conceptual understanding and practical calculations, using both perturbative and non-perturbative methods. But that's another story…

# Index