Lectures on Advanced Quantum Mechanics

M. Zirnbauer
Institut für Theoretische Physik
Universität zu Köln

WS 10/11, WS 17/18

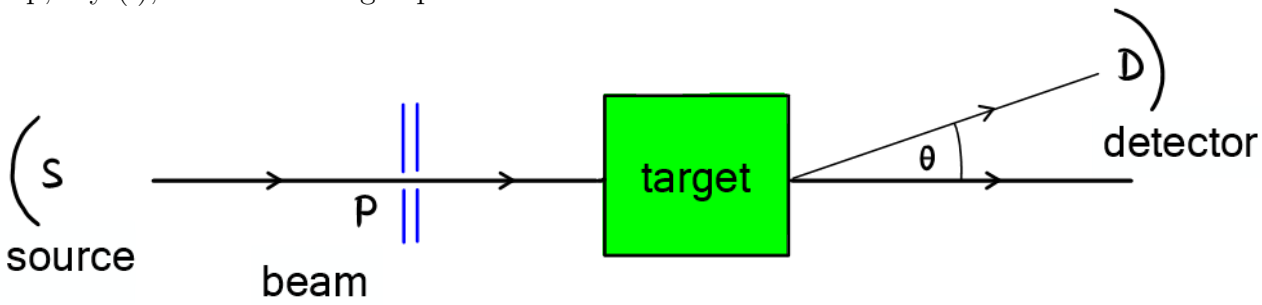
Contents

1 Scattering theory

Literature. A good fraction of the material of this chapter, in particular the introduction and the section on time-dependent scattering, is taken from the book by W.O. Amrein, J.M. Jauch and K.B. Sinha, *Scattering Theory in Quantum Mechanics* [Benjamin, Reading (Mass.), 1977].

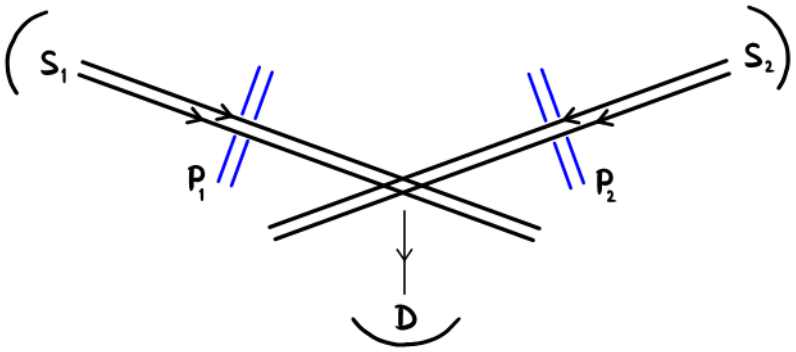
1.1 Preliminaries

The formalism of scattering theory plays an important role in physics, as it allows to predict experimental observations from the fundamental interactions postulated by theory. A typical setup, say (i), for a scattering experiment looks as follows:



Thus there is a source for a beam of particles which, after preparation of the beam kinematics, hit a target and then get observed in a detecting device. Note that thick targets cause multiple scattering events; while these might be hard to analyze in the case of an amorphous target, they lead to characteristic interference patterns and thus structural information in the case of a crystal (for the target).

Next, let us look at the symmetric situation, (ii), of a scattering experiment with two colliding beams:



In this setting one wants to place the detector far from the collision zone for good angular resolution. One does not place the detector in the forward direction of an incoming beam.

Remark. In the case of scattering of massive particles, the situation (i) can be reduced to the situation (ii) by a change of inertial reference frame.

Here comes a list of large accelerators:

- Fermilab Chicago, Tevatron (1983; $p + p$; ca. 2 TeV; top quark discovered; 6 km)
- Stanford Linear Accelerator ('SLAC'; 3 km)
- Brookhaven RHIC (relativistic heavy ion collider; quark-gluon plasma)

- DESY Hamburg (HERA $p + e$; PETRA $e^+ + e^-$)
- CERN Geneva (LEP; LHC = Large Hadron Collider, 17 TeV?)

1.2 Types of scattering process

We briefly introduce some nomenclature:

1. *Elastic scattering* is schematically depicted as

   $$a + b \longrightarrow a + b\,.$$

   In this case all energy is returned to the state of motion.

2. In the case of *inelastic scattering*,

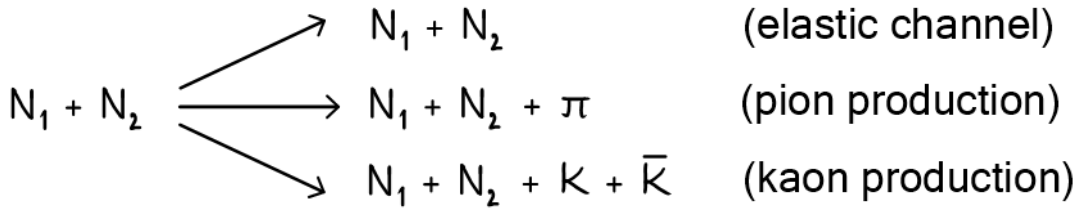
   $$a + b \longrightarrow a' + b'\,,$$

   internal degrees of freedom of the particles get excited (e.g., rotational or vibrational degrees of freedom).

3. In the general category of *rearrangement scattering*, the identity of the scattered particles is altered:

   $$a + b \longrightarrow c + d + e + \ldots$$

   An example is the break-up of the deuteron into its constituents (namely, one proton and one neutron) during a collision with another particle. A second example is nucleon-nucleon scattering, where one has the possibilities

   plus various other "channels" not shown here. (The theoretical analysis of rearrangement scattering requires the formalism of multichannel scattering, which will not be treated in this lecture course.)

4. By *resonance scattering* one means a process that involves the formation and subsequent decay of an unstable (but possibly very long-lived) intermediate state.

5. In scattering theory one may also consider the decay of an unstable particle ($a \longrightarrow b + c + \ldots$). For example, a free neutron decays into a proton, an electron, and an anti-neutrino:

   $$n \longrightarrow p + e + \bar{\nu}_e\,.$$

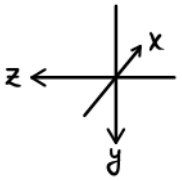
1.3 Observables

Here are some quantities that are observable by scattering experiments:

- the (differential) scattering cross section;
- the life time of an unstable particle; the resonance width;
- branching ratios in multichannel scattering processes.

We now introduce the differential scattering cross section, assuming the situation (i) of Sect. 1.1. Let the (probability) current density I of the incoming particles be homogeneous and directed along the z-axis for a Cartesian coordinate system $\{x, y, z\}$:

$$I = i\, dxdy\,.$$



The physical dimension of the constant i is

$$[i] = \frac{\text{particle number}}{\text{area} \times \text{time}}\,.$$

At large distances $r = \sqrt{x^2 + y^2 + z^2}$ from the scattering region, the current density J of the outgoing (scattered) particles becomes

$$J = j\, \sin\theta\, d\theta\, d\phi\,, \quad j = j(\Omega)\,, \quad \Omega = \frac{\vec{r}}{r} \quad \text{(point on the unit sphere)},$$

where the polar angle θ and the azimuthal angle ϕ relate to the axis of the incoming beam (here, the z-axis). The coefficient j has the physical dimension

$$[j] = \frac{\text{particle number}}{\text{solid angle} \times \text{time}}\,.$$

The differential scattering cross section is then defined as the ratio

$$\frac{d\sigma}{d\Omega} = \frac{j}{i} \qquad \text{(not quite the usual derivative)}.$$

Its physical dimension is

$$\left[\frac{d\sigma}{d\Omega}\right] = \text{area}\,.$$

The total scattering cross section is obtained by integrating over all solid angles ($d\Omega \equiv \sin\theta\, d\theta\, d\phi$):

$$\sigma_{\text{tot}} = \int \frac{d\sigma}{d\Omega}\, d\Omega = \frac{1}{i} \int_{S^2} J.$$

Remark. In these lecture notes we will be concerned with the scattering of quantum mechanical particles. We mention in passing that the notions of differential and total scattering cross section already make sense in the setting of classical mechanics. For example, the total cross section for hard balls (as target particles) of radius R is $\sigma_{\text{tot}} = \pi R^2$ in the classical limit.

We now state a practical recipe by which to calculate the differential scattering cross section for potential scattering of Schrödinger particles with energy $E = \hbar^2 \vec{k}^2/2m$ and wave vector $\vec{k}$.

Prescription. Find the solution of Schrödinger's equation $H\psi = E\psi$ with the asymptotic form

$$\psi(\vec{r}) \overset{|\vec{r}|=r\to\infty}{\longrightarrow} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{r}} + f_E(\Omega)\,\frac{\mathrm{e}^{\mathrm{i}kr}}{r}, \quad k = |\vec{k}|\,,$$

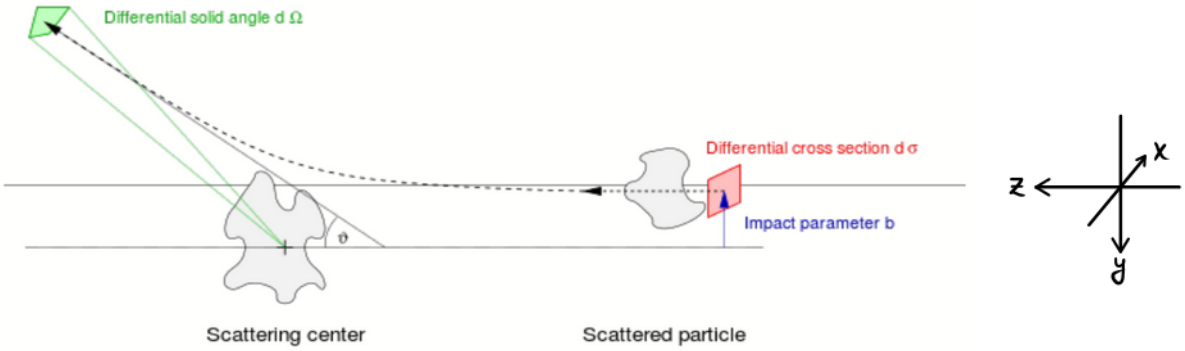
where $f_E(\Omega) = f_E(\vec{r}/r)$ is called the *scattering amplitude.* Then the differential cross section is

$$\frac{d\sigma}{d\Omega} = |f_E(\Omega)|^2\,.$$

Problem. Verify this formula from the definition of $d\sigma/d\Omega$.

1.3.1 Classical scattering

Although our topic here is quantum scattering, let us briefly touch on the differential scattering cross section for classical scattering. In the classical case, one has a scattering map $S:\ \mathbb{R}^2 \to \mathrm{S}^2$, which is determined by the classical dynamics and sends any impact parameter $(x, y) \in \mathbb{R}^2$ of the incoming particles to the corresponding direction of motion $\vec{r}/r \in \mathrm{S}^2$ of the outgoing particles. The situation is illustrated by the following figure from the Wikipedia entry "Cross section (physics)".



Let us now assume that the scattering map S has an inverse, $S^{-1} \equiv T:\ \mathrm{S}^2 \to \mathbb{R}^2$. (In general situations, such as the cases of rainbow scattering and spiral scattering, S^{-1} exists only locally; see, e.g., `users.physics.harvard.edu/~morii/phys151/lectures/Lectures07.pdf`). Now, incoming flux through $A \subset \mathbb{R}^2$ translates to outgoing flux through $S(A)$. Given $T = S^{-1}$, we can invert the flux-to-flux correspondence to obtain $\int_{T(\Sigma)} I = \int_\Sigma J$, valid for any region $\Sigma \subset \mathrm{S}^2$. It follows by the substitution rule (abstractly written as $\int_{T(\Sigma)} I = \int_\Sigma T^*I$) that $I = i\,dxdy \equiv i\,d\sigma$ transforms into J under T. To make this transformation explicit, let ξ, η be any two local coordinates for S^2, and let us write the solid-angle element as $d\Omega = \omega_{\xi\eta}\,d\xi\,d\eta$ and express T by a pair of functions $x = f(\xi, \eta)$ and $y = g(\xi, \eta)$. We then have

$$j\,d\Omega \equiv J \overset{!}{=} T^*I = i\,T^*(d\sigma) = i\left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right| d\xi\,d\eta\,.$$

Thus the differential scattering cross section appears as the Jacobian of the transformation (by the inverse map $T = S^{-1}$) from the area element $d\sigma$ for $\mathbb{R}^2$ to the solid-angle element $d\Omega$ for S^2:

$$\frac{d\sigma}{d\Omega} \equiv \frac{j}{i} = \left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right| \frac{1}{\omega_{\xi\eta}}\,.$$

(This motivates the notation $d\sigma/d\Omega$; indeed, j/i is seen to be a kind of generalized derivative.)

1.4 Lippmann-Schwinger equation

Consider a time-independent system with Hamiltonian $H = H_0 + V$ where H_0 is a Hamiltonian of free motion (or some other Hamiltonian for which the solutions of Schrödinger's equation are already known) and V is a perturbation which is in some sense small. (To be on safe ground, we should require bounded support or, at least, rapid decay at infinity.)

Let ψ_0 be a solution of the free problem with energy E:

$$H_0\psi_0 = E\psi_0\,.$$

For $\varepsilon = 0+$ let $G_0 := (E + \mathrm{i}\varepsilon - H_0)^{-1}$ (operator inverse) the free Green operator or resolvent.

Claim. If ψ satisfies

$$\psi = \psi_0 + G_0 V\psi \qquad \text{(abstract form of Lippmann-Schwinger equation)},$$

then ψ is a solution of Schrödinger's equation $H\psi = E\psi$.

Verification. We start by rewriting the Lippmann-Schwinger equation as

$$\psi_0 = \psi - G_0V\psi = (1 - G_0V)\psi = G_0(G_0^{-1} - V)\psi.$$

Now let

$$G := (E + \mathrm{i}\varepsilon - H)^{-1}$$

be the Green operator (or resolvent) of the full Hamiltonian. Then

$$G^{-1} = G_0^{-1} - V \quad \text{and} \quad \psi_0 = G_0G^{-1}\psi,$$

so that

$$\psi = GG_0^{-1}\psi_0 = G\,(E + \mathrm{i}\varepsilon - H_0)\psi = \mathrm{i}\varepsilon G\psi_0\,.$$

It follows that

$$(E - H)\psi = \mathrm{i}\varepsilon\frac{E - H}{E + \mathrm{i}\varepsilon - H}\,\psi_0 \xrightarrow{\varepsilon\to 0} 0\,, \quad \text{since} \quad \left\|\frac{E - H}{E + \mathrm{i}\varepsilon - H}\right\|_{\mathrm{op}} < 1\,.$$

In other words: $H\psi = E\psi$ as claimed.

Now let

$$H_0 = \frac{p^2}{2m} = -\frac{\hbar^2}{2m}\nabla^2, \quad V = V(\vec{r}) \qquad (\text{"potential scattering"}).$$

Then the operator G_0 has the integral kernel (or "Green's function")

$$\langle \vec{r} \mid G_0 \mid \vec{r}\,'\rangle = -\frac{\mathrm{e}^{\mathrm{i}k|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}\,\frac{2m}{\hbar^2},$$

and the abstract equation for ψ takes the explicit form of an integral equation:

$$\psi(\vec{r}) = \psi_0(\vec{r}) - \frac{2m}{\hbar^2}\int_{\mathbb{R}^3}\frac{\mathrm{e}^{\mathrm{i}k|\vec{r}-\vec{r}\,'|}}{4\pi|\vec{r}-\vec{r}\,'|}\,V(\vec{r}\,')\,\psi(\vec{r}\,')\,d^3r'$$

(Lippmann-Schwinger equation in coordinate representation).

1.4.1 Born approximation

If $\| G_0 V \|_{\text{op}} < 1$, one can expand the solution in a geometric series (called the Born series):

$$\psi = (1 - G_0 V)^{-1} \psi_0 = \psi_0 + G_0 V \psi_0 + G_0 V G_0 V \psi_0 + \dots \, .$$

For $\| G_0 V \|_{\text{op}} \ll 1$ (high energy, or weak potential) one may use the lowest-order (or first) Born approximation,

$$\psi \approx \psi_0 + G_0 V \psi_0 \, .$$

To write the first Born approximation in explicit form, let us take $\psi_0(\vec{r}) = \mathrm{e}^{\mathrm{i}kz}$. Then

$$(G_0 V \psi_0)(\vec{r}) = -\frac{m}{2\pi\hbar^2} \int_{\mathbb{R}^3} \frac{\mathrm{e}^{\mathrm{i}k|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|} \, V(\vec{r}\,') \, \mathrm{e}^{\mathrm{i}kz'} d^3 r' .$$

For V of finite range and $\vec{r}$ far from the scattering center of V we may expand

$$k|\vec{r}-\vec{r}\,'| = kr \left| \frac{\vec{r}}{r} - \frac{\vec{r}\,'}{r} \right| \approx kr \left(1 - \Omega \cdot \frac{\vec{r}\,'}{r} \right) = kr - k\Omega \cdot \vec{r}\,' .$$

We then see that

$$(G_0 V \psi_0)(\vec{r}) \overset{r\to\infty}{\longrightarrow} f_E^{(1)}(\Omega) \, \frac{e^{\mathrm{i}kr}}{r} ,$$

where

$$f_E^{(1)}(\Omega) = -\frac{m}{2\pi\hbar^2} \int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}k\Omega\cdot\vec{r}\,' + \mathrm{i}kz'} V(\vec{r}\,') \, d^3 r' .$$

Rewriting this in a form which is independent of the choice of Cartesian basis, we obtain

$$f_E^{(1)}(\Omega) = -\frac{m}{2\pi\hbar^2} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}(\vec{k}_i - \vec{k}_f)\cdot\vec{r}\,'} V(\vec{r}\,') \, d^3 r' .$$

We thus see that the scattering amplitude in the first Born approximation, $f_E^{(1)}(\Omega)$, is essentially given by the Fourier transform of the potential V.

1.4.2 Examples

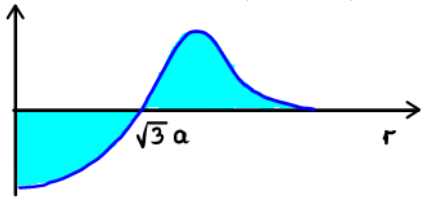
Let us illustrate the first Born approximation at two examples.

1. $V(\vec{r}) = V_0 \lambda^3 \delta(\vec{r})$. Here V_0 has the physical dimension of energy, and λ has the physical dimension of length. δ is the Dirac δ-function (actually, δ-distribution), with support at zero (the origin of the coordinate system). The Fourier transform of the δ-function is simply a constant, so
$$f_E^{(1)}(\Omega) = -\frac{m}{2\pi\hbar^2} V_0 \lambda^3 = -\frac{\lambda}{4\pi} \frac{V_0}{\hbar^2/(2m\lambda^2)} ,$$
independent of the energy E and the outgoing direction Ω. Note that $E_{\text{kin}} = \hbar^2/(2m\lambda^2)$ is a rough estimate of the kinetic energy of a Schrödinger particle with mass m confined to a box of size λ. The ratio V_0/E_{kin} is a dimensionless measure of the strength of the scattering potential. We have written the answer for f_E in a form which makes it transparent that f_E has the physical dimension of length. We notice that the scattering amplitude f_E is negative in the repulsive case ($V_0 > 0$) and positive in the attractive case ($V_0 < 0$).

2. $V(\vec{r}) = V_0\,\lambda^5\nabla^2\delta(\vec{r})$. We first give some explanation of what it means to apply the Laplacian ∇^2 to a δ-function. (As a side remark: this operation is mathematically well-defined if δ is regarded as a Schwartz distribution.) For this purpose we use a Gaussian regularization of the δ-function:

$$\nabla^2\delta(\vec{r}) = \lim_{a\to 0+} \nabla^2 \mathrm{e}^{-\frac{r^2}{2a^2}}/(2\pi a^2)^{3/2} = \lim_{a\to 0+} (2\pi a^2)^{-3/2}\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right)\mathrm{e}^{-\frac{r^2}{2a^2}}$$

$$= \lim_{a\to 0+} (2\pi a^2)^{-3/2}\left(\frac{\partial}{\partial r} + \frac{2}{r}\right)\left(-\frac{r}{2a^2}\right)\mathrm{e}^{-\frac{r^2}{2a^2}} = \lim_{a\to 0+} (2\pi a^2)^{-3/2}\left(\frac{r^2}{a^4} - \frac{3}{a^2}\right)\mathrm{e}^{-\frac{r^2}{2a^2}}.$$

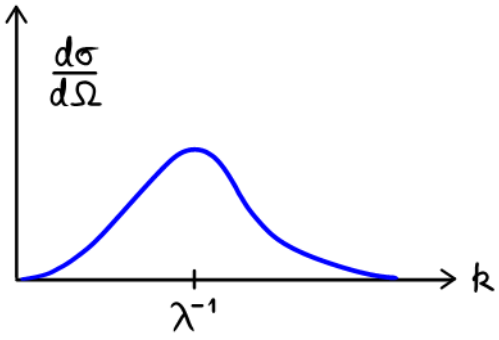


$$\int_{\mathbb{R}^3} \nabla^2\delta(\vec{r})\,d^3r = 0.$$

By using the rule $\nabla \to \mathrm{i}k$ for the Fourier transform, we obtain the following expression for the first Born approximation to the scattering amplitude:

$$f_E^{(1)} = +\frac{m}{2\pi\hbar^2} V_0\,\lambda^5(\vec{k}_i - \vec{k}_f)^2 = \frac{m}{\pi\hbar^2} V_0\,\lambda^5 k^2(1-\cos\theta)\,.$$

If λ is the range of the scattering potential (i.e., we take the Gaussian regularization parameter a to be $\lambda \equiv a$), then from the properties of the Fourier transform we expect the following qualitative picture (for our second example) of the scattering cross section:



1.5 Scattering by a centro-symmetric potential (partial waves)

In this section we consider Hamiltonians of the form

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(r)\,, \quad r = \sqrt{x^2+y^2+z^2},$$

where the potential $V(r)$ is invariant under all rotations fixing a center (which we take to be the origin of our Cartesian coordinate system x, y, z). We assume that $V(r)$ decreases faster than $1/r$ in the limit of $r \to \infty$.

Our goal here is to explain the 'method of partial waves', which is one of the standard methods of scattering theory. To get started, we recall a few facts known from the basic course on quantum mechanics. Using spherical polar coordinates

$$x = r\sin\theta\cos\phi\,, \quad y = r\sin\theta\sin\phi\,, \quad z = r\cos\theta\,,$$

we have the following expression for the Laplacian:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2}\left(\frac{\partial}{\sin\theta\,\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right).$$

If the incoming wave of the scattering wave function ψ is a plane wave $\mathrm{e}^{\mathrm{i}kz}$ traveling in the z direction, then we expect the scattering amplitude to be independent of the azimuthal angle ϕ (by the rotational symmetry of the potential V). In fact, the azimuthal angle ϕ will never appear in the following discussion.

We make an ansatz for the wave function of the form

$$\psi = \sum_{l=0}^{\infty} \psi_l(r)\, P_l(\cos\theta)\,,$$

where P_l is the Legendre polynomial of degree l. We recall that Legendre polynomials are eigenfunctions of the angular part of the Laplacian:

$$-\frac{\partial}{\sin\theta\,\partial\theta}\sin\theta\frac{\partial}{\partial\theta}\,P_l(\cos\theta) = l(l+1)\,P_l(\cos\theta)\,. \tag{1.1}$$

We now write the energy E of the Schrödinger particle in the form $E = \frac{\hbar^2k^2}{2m}$. Our Ansatz for ψ then leads to the following differential equation for the radial functions $\psi_l(r)$:

$$\left(-\frac{1}{r}\frac{d^2}{dr^2}\circ r + \frac{l(l+1)}{r^2} + \frac{2m}{\hbar^2}V(r)\right)\psi_l(r) = k^2\psi_l(r)\,. \tag{1.2}$$

For large values of r this equation and its general solution simplify to

$$-\frac{d^2}{dr^2}\,r\psi_l = k^2 r\psi_l\,, \quad \psi_l(r) = A\,\frac{\mathrm{e}^{\mathrm{i}kr}}{r} + B\,\frac{\mathrm{e}^{-\mathrm{i}kr}}{r}\,.$$

Now it is a basic property (called conservation of probability or 'unitarity' for short) of quantum mechanics that the divergence of the probability current density $\mathbf{j} = \frac{\hbar}{m}\mathfrak{Im}\,\bar\psi\nabla\psi$ must vanish everywhere in space for a solution ψ of the time-independent Schrödinger equation $H\psi = E\psi$. Assuming the large-r behavior $\psi = (A\,\mathrm{e}^{\mathrm{i}kr} + B\,\mathrm{e}^{-\mathrm{i}kr})/r$, a quick computation shows that the total

probability flux through the surface of a ball centered at $r = 0$ is $(4\pi\hbar k/m)(|A|^2 - |B|^2)$. By Gauss' theorem this surface integral must be equal to the integral of $\operatorname{div}\mathbf{j} = 0$ over the ball. Hence, unitarity requires that A and B have the same magnitude $|A| = |B|$.

Thus the large-r asympotics of any solution of the time-independent Schrödinger equation $H\psi = E\psi$ with azimuthal symmetry has to be

$$\psi \overset{r\to\infty}{\longrightarrow} \sum_{l=0}^{\infty} \frac{A_l\, \mathrm{e}^{\mathrm{i}kr} + B_l\, \mathrm{e}^{-\mathrm{i}kr}}{2\mathrm{i}kr}\, (2l+1)\, P_l(\cos\theta)\,, \quad |A_l| = |B_l|\,. \tag{1.3}$$

The factor $(2l+1)/(2\mathrm{i}k)$ has been inserted for later convenience.

Our boundary conditions for the scattering problem dictate that

$$\psi \overset{r\to\infty}{\longrightarrow} \mathrm{e}^{\mathrm{i}kz} + f_E(\theta)\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\,.$$

Lemma. In order for ψ to be of this asymptotic form, the amplitudes B_l in (1.3) must be

$$B_l = (-1)^{l+1}.$$

To prove this lemma, we need an understanding of how $\mathrm{e}^{\mathrm{i}kz}$ expands in partial waves. From the basic quantum theory of angular momentum we recall that the Legendre polynomials have the orthogonality property

$$\int_0^\pi P_l(\cos\theta)\, P_{l'}(\cos\theta)\, \sin\theta\, d\theta = \frac{2\,\delta_{ll'}}{2l+1}\,. \tag{1.4}$$

Moreover, the Legendre polynomials form a complete system of functions on the interval $[-1,+1] \ni \cos\theta$. We may therefore expand $\mathrm{e}^{\mathrm{i}kz} = \mathrm{e}^{\mathrm{i}kr\cos\theta}$ as

$$\mathrm{e}^{\mathrm{i}kr\cos\theta} = \sum_{l=0}^{\infty} \mathrm{i}^l j_l(kr)\, (2l+1)\, P_l(\cos\theta)\,, \tag{1.5}$$

$$j_l(kr) = \frac{\mathrm{i}^{-l}}{2}\int_0^\pi \mathrm{e}^{\mathrm{i}kr\cos\theta} P_l(\cos\theta)\sin\theta\, d\theta\,. \tag{1.6}$$

The factor of i^l has been inserted in order to make the function $j_l(kr)$ coincide with the so-called spherical Bessel functions. The lowest-order spherical Bessel functions are

$$j_0(\xi) = \frac{\sin\xi}{\xi}\,, \quad j_1(\xi) = \frac{\sin\xi}{\xi^2} - \frac{\cos\xi}{\xi}\,.$$

For small values of the argument, the spherical Bessel functions behave as

$$j_l(\xi) \sim \xi^l.$$

This behavior follows from the orthogonality property (1.4) of the Legendre polynomials and the fact that any polynomial in $\cos\theta$ of degree l can be expressed as a linear combination of the Legendre polynomials $P_{l'}(\cos\theta)$ of degree $l' \le l$.

For large values of the argument, the spherical Bessel functions behave as

$$j_l(\xi) \simeq \frac{\sin(\xi - l\pi/2)}{\xi}\,. \tag{1.7}$$

We will motivate this important relation at the end of the present subsection.

Proof of Lemma. Using the expansion of $\mathrm{e}^{\mathrm{i}kr\cos\theta}$ and the large-r asymptotics of the spherical Bessel functions we have

$$\mathrm{e}^{\mathrm{i}kr\cos\theta} \overset{r\to\infty}{\longrightarrow} \sum_{l=0}^{\infty} \mathrm{i}^l \, \frac{\sin(kr - l\pi/2)}{kr} \, (2l+1) \, P_l(\cos\theta) \, .$$

The radially incoming part for angular momentum l is given by

$$\left(\mathrm{i}^l \, \frac{\sin(kr - l\pi/2)}{kr} \right)_{\text{incoming}} = -\mathrm{i}^l \, \frac{\mathrm{e}^{-\mathrm{i}(kr - l\pi/2)}}{2\mathrm{i}kr} = -\,\mathrm{e}^{\mathrm{i}l\pi} \frac{\mathrm{e}^{-\mathrm{i}kr}}{2\mathrm{i}kr} \, .$$

Thus by comparing coefficients we obtain the desired result $B_l = (-1)^{l+1}$. □

We now consider the difference $\psi - \mathrm{e}^{\mathrm{i}kz}$. By construction, this is a sum of radially outgoing waves:

$$\psi - \mathrm{e}^{\mathrm{i}kz} \overset{r\to\infty}{\longrightarrow} \sum_{l=0}^{\infty} (A_l - 1) \frac{\mathrm{e}^{\mathrm{i}kr}}{2\mathrm{i}kr} \, (2l+1) P_l(\cos\theta) \, ,$$

which is of the expected form $f_E(\theta)\mathrm{e}^{\mathrm{i}kr}/r$. We already know that by unitarity we must have $|A_l| = |B_l| = 1$. It is customary to put $A_l = \mathrm{e}^{2\mathrm{i}\delta_l}$ and call δ_l the *phase shift* (in the channel of angular momentum l). We then have the following result for the scattering amplitude:

$$f_E(\theta) = \sum_{l=0}^{\infty} \frac{\mathrm{e}^{2\mathrm{i}\delta_l(E)} - 1}{2\mathrm{i}k(E)} \, (2l+1) P_l(\cos\theta) \, , \tag{1.8}$$

where we have emphasized the energy dependence of the phase shift $\delta_l(E)$ and the wave number $k(E) = \sqrt{2mE}/\hbar$.

1.5.1 Optical theorem

We now use the formula (1.8) for the scattering amplitude $f_E(\theta)$ to compute the total scattering cross section. By the orthogonality property (1.4) of the Legendre polynomials we obtain

$$\sigma_{\text{tot}}(E) = \int_{\mathrm{S}^2} \frac{d\sigma}{d\Omega} \, d\Omega = \int_{\mathrm{S}^2} |f_E(\Omega)|^2 \, d\Omega = \frac{4\pi}{k^2(E)} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l(E) \, . \tag{1.9}$$

On the other hand, since $(\mathrm{e}^{2\mathrm{i}\delta_l} - 1)/(2\mathrm{i}) = \mathrm{e}^{\mathrm{i}\delta_l} \sin\delta_l$ has imaginary part $\sin^2\delta_l$, the imaginary part of the scattering amplitude in the limit of looking in the forward direction is

$$\mathfrak{Im} \, f_E(\theta \to 0) = \frac{1}{k(E)} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l(E) \, .$$

Thus the total cross section and the forward scattering amplitude are related by

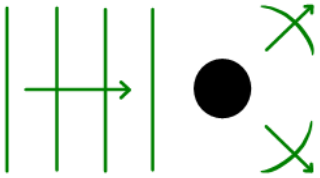
$$\sigma_{\text{tot}}(E) = \frac{4\pi}{k} \, \mathfrak{Im} \, f_E(0) \, . \tag{1.10}$$

This relation is called the optical theorem.

1.5.2 Example: scattering from a hard ball

We now illustrate the method of partial waves at the example of scattering from a hard ball:

$$V(r) = \begin{cases} 0 & r > R\,, \\ +\infty & r < R\,. \end{cases}$$



The goal is to find the scattering phase shifts δ_l. Having calculated these, we get the scattering amplitude and the cross section from the formulas of the previous subsection.

The scattering wave function must vanish identically inside the ball ($r < R$) where the potential is repulsive and infinite. Outside the ball ($r > R$) the motion is that of a free particle. The continuity of the wave function implies Dirichlet boundary conditions at the surface of the ball:

$$\psi\Big|_{r=R} = 0\,.$$

In the exterior of the ball, where the motion is free, we look for solutions $\psi_l(r)P_l(\cos\theta)$ of the Schrödinger equation for a free particle of angular momentum l. We recall that the equation for the radial functions $\psi_l(r)$ reads

$$\left(-\frac{1}{r}\frac{d^2}{dr^2}\circ r + \frac{l(l+1)}{r^2}\right)\psi_l(r) = k^2\psi_l(r)\,. \tag{1.11}$$

Solutions of this equation are the spherical Bessel functions $\psi_l(r) = j_l(kr)$. Indeed, we know that the plane wave $\mathrm{e}^{\mathrm{i}kr\cos\theta}$ is a solution of the free Schrödinger equation, and by expanding $\mathrm{e}^{\mathrm{i}kr\cos\theta} = \sum \mathrm{i}^l j_l(kr)(2l+1)P_l(\cos\theta)$ and using the eigenfunction property (1.1) of the Legendre polynomials, we see that $j_l(kr)$ solves the radial equation (1.11).

Since the radial equation is of second order, a single solution is not enough to express the most general solution. We need a linearly independent second solution. To find it, we recall that $j_l(kr) \sim r^l$ for $r \ll k^{-1}$. Using this, it is easy to see that $\Psi_l(\vec{r}) = j_l(kr)P_l(\cos\theta)$ in the limit of $r \to 0$ contracts to a solution of the Laplace equation $\nabla^2\Psi_l = 0$. Now from the chapter on multipole expansion in electrostatics, we know that there exists a second angular momentum l solution $r^{-l-1}P_l(\cos\theta)$ of Laplace's equation. (Solutions of Laplace's equation are also called harmonic functions.) We therefore expect that there exists a solution, say $n_l(kr)$, of the radial equation (1.11) with the corresponding small-r asymptotics:

$$\left(-\frac{1}{r}\frac{d^2}{dr^2}\,r + \frac{l(l+1)}{r^2}\right)n_l(kr) = k^2 n_l(kr)\,, \quad n_l(\xi) \sim \xi^{-l-1} \quad (\xi \to 0)\,. \tag{1.12}$$

Such a solution $n_l(kr)$ does exist, and it is called the spherical Neumann function of degree l. The spherical Neumann functions of lowest degree are

$$n_0(\xi) = -\frac{\cos\xi}{\xi}\,, \quad n_1(\xi) = -\frac{\cos\xi}{\xi^2} - \frac{\sin\xi}{\xi}\,.$$

We now make an ansatz for the wave function of free motion outside the hard ball:

$$\psi = \sum_{l=0}^{\infty} \mathrm{i}^l\big(a_l\, j_l(kr) + b_l\, n_l(kr)\big)(2l+1)P_l(\cos\theta) \qquad (r > R)\,.$$

The spherical Neumann functions have the following large-r asymptotic behavior:

$$n_l(kr) \overset{r\to\infty}{\longrightarrow} -\frac{\cos(kr - l\pi/2)}{kr}\,.$$

By using also the large-r behavior of the spherical Bessel functions we find the asymptotics

$$\psi \overset{r\to\infty}{\longrightarrow} (2ikr)^{-1} \sum_{l=0}^{\infty} \left((a_l - \mathrm{i}b_l)\,\mathrm{e}^{\mathrm{i}kr} - (-1)^l (a_l + \mathrm{i}b_l)\,\mathrm{e}^{-\mathrm{i}kr}\right)(2l+1)P_l(\cos\theta)\,.$$

Comparison with the general expression (1.3), where $A_l = \mathrm{e}^{2\mathrm{i}\delta_l}$ and $B_l = -(-1)^l$, then yields $a_l - \mathrm{i}b_l = \mathrm{e}^{2\mathrm{i}\delta_l}$, $a_l + \mathrm{i}b_l = 1$, and hence $a_l = \mathrm{e}^{\mathrm{i}\delta_l}\cos\delta_l$, $b_l = -\mathrm{e}^{\mathrm{i}\delta_l}\sin\delta_l$. Thus our ansatz takes the form

$$\psi = \sum_{l=0}^{\infty} \mathrm{i}^l \mathrm{e}^{\mathrm{i}\delta_l}\big(\cos(\delta_l)\,j_l(kr) - \sin(\delta_l)\,n_l(kr)\big)(2l+1)P_l(\cos\theta) \qquad (r > R)\,. \tag{1.13}$$

We now impose the Dirichlet boundary condition $\psi\big|_{r=R} = 0$. This gives the following condition:

$$\tan(\delta_l) = \frac{j_l(kR)}{n_l(kR)}\,,$$

which determines the scattering phase shifts δ_l.

Here we specialize to the long wave length limit $kR \ll 1$. In this limit $\delta_l \sim (kR)^{2l+1}$. Thus scattering is appreciable only in the s-wave channel ($l = 0$). The exact value of the s-wave scattering phase shift is

$$\delta_0 = \arctan\left(\frac{j_0(kR)}{n_0(kR)}\right) = -kR\,.$$

From Eq. (1.9) we then have the total scattering cross section

$$\sigma_{\mathrm{tot}} \approx \frac{4\pi}{k^2}\sin^2\delta_0 \approx 4\pi R^2.$$

Notice that this is larger (by a factor of four) than the geometric cross section πR^2 of classical scattering from a hard ball. (There is no inconsistency in this, as the long wave length limit $kR \ll 1$ is just the opposite of the classical limit $kR \gg 1$.)

1.5.3 Asymptotics of the spherical Bessel functions

Here we fill in a gap which was left in the argument above: the asymptotic behavior (1.7).

For this purpose we use the following integral representation of the Legrendre polynomials:

$$P_l(\cos\theta) = \int_0^{2\pi} \frac{d\phi}{2\pi}\,(\cos\theta - \mathrm{i}\sin\theta\cos\phi)^l\,, \tag{1.14}$$

which can be verified by checking that the integral on the right-hand side satisfies the differential equation (1.1). In combination with (1.6) this gives a representation of the spherical Bessel functions as an integral over the unit sphere S^2 :

$$j_l(\xi) = \frac{\mathrm{i}^{-l}}{4\pi}\int_{\mathrm{S}^2} \mathrm{e}^{\mathrm{i}\xi\cos\theta + l\,\ln(\cos\theta - \mathrm{i}\sin\theta\cos\phi)}\sin\theta\,d\theta\,d\phi\,. \tag{1.15}$$

For large values of ξ, the integrand oscillates strongly everywhere on S^2 with the exception of two points: the 'north pole' ($\theta = 0$) and the 'south pole' ($\theta = \pi$). These are the points where $\cos\theta$ has

vanishing derivative. The contribution to the integral from each of these points can be calculated by using the method of stationary phase (not explained here). Consider first the north pole $\theta = 0$. To compute its contribution to the integral, it is helpful to change coordinates as follows:

$$\cos\theta = \sqrt{1 - n_1^2 - n_2^2}\,, \quad \sin\theta\,\cos\phi = n_1\,, \quad \sin\theta\,\sin\phi = n_2\,, \quad \sin\theta\,d\theta\,d\phi = \frac{dn_1\,dn_2}{\sqrt{1 - n_1^2 - n_2^2}}\,.$$

The stationary-phase contribution to $j_l(\xi)$ from $\theta = 0$ is then found to be

$$\approx \frac{\mathrm{e}^{\mathrm{i}\xi}}{4\pi \mathrm{i}^l} \int_{\mathbb{R}^2} \mathrm{e}^{-\mathrm{i}\xi(n_1^2+n_2^2)/2} dn_1 dn_2 = \frac{\mathrm{e}^{\mathrm{i}(\xi - l\pi/2)}}{2\mathrm{i}\xi}\,.$$

Similarly, the contribution from the south pole $\theta = \pi$ is

$$\approx \frac{\mathrm{e}^{-\mathrm{i}(\xi - l\pi/2)}}{-2\mathrm{i}\xi}\,.$$

By adding these two contributions, we get the claimed behavior (1.7).

1.6 Time-dependent scattering theory

So far, our considerations have been based on the time-*in*dependent Schrödinger equation, which is appropriate if the physical situation is stationary or approximately stationary. A more fundamental approach, which we adopt here, is to start from the investigation of the process of time-dependent scattering of localized wave packets — with the option of passing to the stationary situation by increasing the width of the wave packet and thereby approaching the limit of a plane wave for the incoming state.

As before, we assume that we are given two Hamiltonians: one (namely H_0, the generator of 'free motion') which we understand and another one (H, including interactions) which we are trying to understand. In the following we denote by U_t the time-evolution operator for free motion:

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\, U_t = H_0\, U_t\,, \qquad U_{t=0} = \mathrm{Id}\,,$$

and by V_t the time-evolution operator for the full interacting dynamics:

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\, V_t = H\, V_t\,, \qquad V_{t=0} = \mathrm{Id}\,.$$

By construction, these time-evolution operators are unitary operators on the Hilbert space $\mathcal{H}$ of our physical system with Hermitian scalar product $\langle \cdot, \cdot \rangle$:

$$\forall \psi, \psi' \in \mathcal{H}: \quad \langle U_t\,\psi, U_t\,\psi' \rangle = \langle \psi, \psi' \rangle = \langle V_t\,\psi, V_t\,\psi' \rangle\,.$$

We note that a Hilbert space comes with a norm $\|\,\psi\,\| := \sqrt{\langle \psi, \psi \rangle}$.

In scattering theory one makes the basic assumption that the dynamics generated by the full Hamiltonian H approaches that of the free Hamiltonian H_0 in the limit of the distant past ($t \to -\infty$) and the far future ($t \to +\infty$). This assumption leads to the following idea.

Given a vector $\psi \in \mathcal{H}$ we look for vectors $\psi_\pm \in \mathcal{H}$ with the approximation property

$$\|\,V_t\,\psi - U_t\,\psi_\pm\,\| \stackrel{t\to\pm\infty}{\longrightarrow} 0\,,$$

i.e., evolving ψ by the full time evolution V_t for a very long time $t \to \pm\infty$ leads to the same state as evolving $\psi_\pm$ by the free time evolution U_t. This idea is illustrated in the figure shown below.

It should be noted, however, that it will not always be possible to find such $\psi_\pm$ for every ψ. Indeed, any free motion $U_t\,\psi_\pm$ (if it is a free motion in the usual sense) must escape to infinity. On the other hand, if there exist bound states for H and the vector ψ has nonzero projection on the space of bound states, then $V_t\,\psi$ contains a component which does *not* escape to infinity.

To accommodate this complication, we use the unitarity of V_t (for finite t) to write

$$\| V_t\,\psi - U_t\,\psi_\pm \| = \| \psi - V_t^\dagger U_t\,\psi_\pm \| \,.$$

Here and in what follows we assume that H_0 has no bound states. It is then reasonable to postulate the existence of the following limits:

$$W_\pm := \lim_{t\to\pm\infty} V_t^\dagger U_t \,. \tag{1.16}$$

The operators $W_\pm$ are called Møller operators, or wave operators. They are isometric, i.e., they satisfy $\| W_\pm\psi \| = \| \psi \|$ for all $\psi \in \mathcal{H}$. By the remark above they are not surjective when bound states exist for H. In other words, each of $W_\pm$ has a left inverse but not necessarily a right inverse.

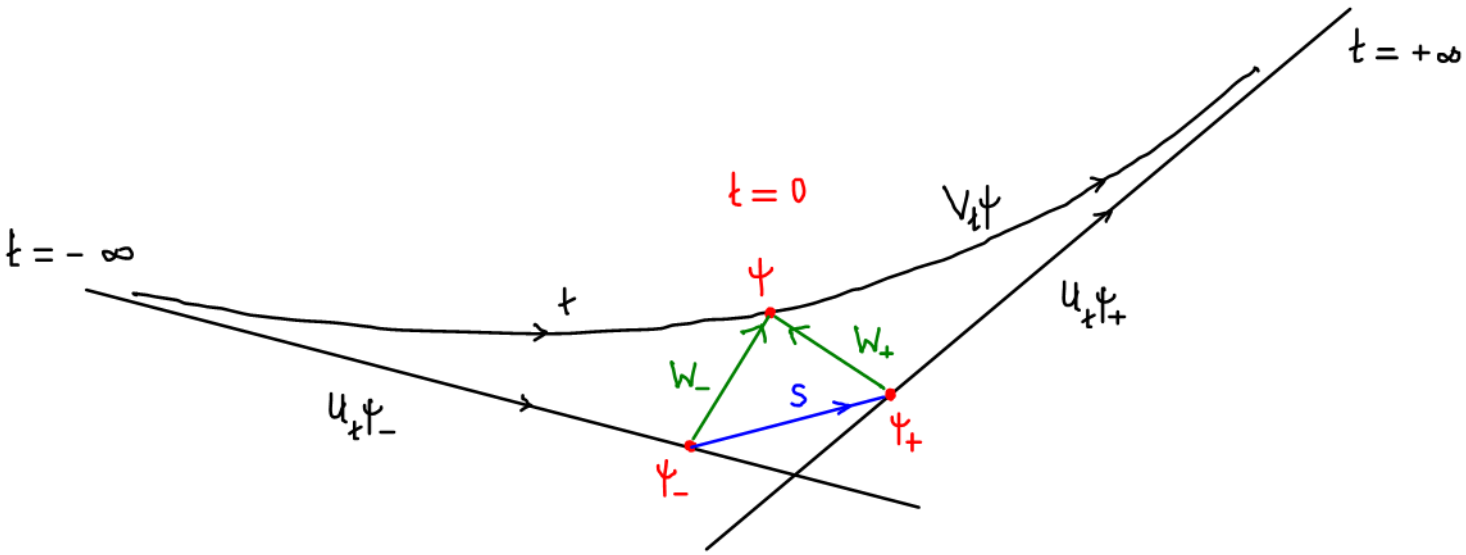
The definitions of U_t, V_t, and $W_\pm$ make sense even if the Hamiltonians H_0 and/or H depend on time. In the case of time-independent H_0 and H we have the explicit formulas

$$U_t = \mathrm{e}^{-\mathrm{i}tH_0/\hbar}\,, \quad V_t = \mathrm{e}^{-\mathrm{i}tH/\hbar}\,, \quad W_\pm = \lim_{t\to\pm\infty} \mathrm{e}^{+\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_0/\hbar}\,.$$

For the next step we require that the range of W_- (i.e., the image of $\mathcal{H}$ under W_-) be contained in the domain of definition of $W_+^\dagger$:

$$\text{range}\,(W_-) \subset \text{domain}\,(W_+^\dagger)\,.$$

This requirement is physically reasonable, and will be discussed in Section 1.6.3 below. If it is met, one may form the composition $W_+^\dagger W_-$. The situation is sketched in the following diagram.



Definition. With the pair $(H_0\,, H)$ we associate an operator

$$S := W_+^\dagger\, W_-\,, \tag{1.17}$$

called the scattering operator.

Remark. By the assumed absence of bound states for H_0, the scattering operator S exists on all of $\mathcal{H}$ and always satisfies $S^\dagger S = \mathrm{Id}_{\mathcal{H}}$. The scattering operator is unitary, i.e. has a right inverse $S^{-1} = S^\dagger$, if domain $(W_+^\dagger)$ = range (W_-). It fails to be unitary if range (W_-) is strictly smaller than domain $(W_+^\dagger)$.

For time-independent $H_0\,, H$ one has the explicit formula

$$S = \lim_{t\to\infty} \mathrm{e}^{+\mathrm{i}tH_0/\hbar}\mathrm{e}^{-2\mathrm{i}tH/\hbar}\mathrm{e}^{+\mathrm{i}tH_0/\hbar}.$$

In the case of time-independent $H_0\,, H$ the Møller operators satisfy the intertwining relations

$$V_t W_\pm = W_\pm U_t\,. \tag{1.18}$$

Heuristically, these relations are motivated by the following computation:

$$W_\pm U_t = \lim_{T\to\pm\infty} V_T^\dagger U_T\; U_t = \lim_{T\to\pm\infty} V_T^\dagger U_{T+t} = \lim_{T\to\pm\infty} V_{T-t}^\dagger U_T = V_t \lim_{T\to\pm\infty} V_T^\dagger U_T = V_t W_\pm\,.$$

As an important corollary of the relations (1.18), the scattering operator commutes with the free time evolution:

$$S\,U_t = U_t\,S\,. \tag{1.19}$$

This is proved as follows:

$$S\,U_t = W_+^\dagger W_-\,U_t = W_+^\dagger V_t\,W_- = (V_{-t}\,W_+)^\dagger W_- = (W_+ U_{-t})^\dagger W_- = U_t\,W_+^\dagger W_- = U_t\,S\,.$$

Now by differentiating the relations (1.19) (which are valid, as we recall, in the case of time-independent $H_0\,, H$), we deduce that the scattering operator commutes with the free Hamiltonian:

$$H_0\,S = S H_0\,.$$

From spectral theory one then knows that H_0 and S can be brought to diagonal form simultaneously. (Note that the eigenstates of H_0 usually fail to be square-integrable. Thus they do not lie inside the Hilbert space $\mathcal{H}$.)

1.6.1 Example: potential scattering $d = 1$

We illustrate the above at the example of potential scattering in one dimension, with

$$H_0 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}, \quad H = H_0 + V(x)\,,$$

and $V(x)$ vanishing (or decaying sufficiently fast) outside a finite region $\mathrm{supp}\,V \subset \mathbb{R}$. The eigenspace of H_0 with energy $E > 0$ is two-dimensional, being spanned by the two functions

$$\phi_\pm(x) := \mathrm{e}^{\pm\mathrm{i}kx}, \quad k = \sqrt{2mE}/\hbar > 0\,.$$

In order for the scattering operator S to commute with H_0, it must leave the two-dimensional H_0-eigenspace (with fixed energy E) invariant. In other words, applying S to either one of the functions $\phi_\pm(x) = \mathrm{e}^{\pm \mathrm{i}kx}$ we must get a linear combination of the same two functions:

$$S\phi_+ = \rho\,\phi_- + \tau\,\phi_+\,, \tag{1.20}$$

$$S\phi_- = \tau'\phi_- + \rho'\phi_+\,. \tag{1.21}$$

The complex numbers $\rho\,,\rho'$ (resp. τ,τ') are called reflection coefficients (resp. transmission coefficients). Due to $S^\dagger S = \mathrm{Id}_{\mathcal{H}}$ they satisfy the unitarity relations

$$|\rho|^2 + |\tau|^2 = 1 = |\rho'|^2 + |\tau'|^2\,, \quad \bar\rho\tau' + \bar\tau\rho' = 0\,. \tag{1.22}$$

In particular, the probabilities $|\rho|^2$ for reflection and $|\tau|^2$ for transmission sum up to unity. We are now left with the question of how to compute the matrix elements ρ, τ, ρ', τ' of the scattering operator. This question is answered in the sequel.

We fix an energy $E = \hbar^2k^2/2m > 0$. For such an energy, we know that the eigenspace of H_0 is two-dimensional, and so is the eigenspace of H. As before, let $\phi_\pm(x) = \mathrm{e}^{\pm \mathrm{i}kx}$ denote the corresponding eigenfunctions of H_0. We now introduce two sets of basis vectors for the two-dimensional eigenspace of H with energy E. First, consider the state vectors

$$\psi^{\mathrm{in}}_\pm := W_-\phi_\pm\,, \tag{1.23}$$

obtained by applying the Møller operator $W_- = \lim_{t\to-\infty} \mathrm{e}^{\mathrm{i}Ht/\hbar}\mathrm{e}^{-\mathrm{i}H_0t/\hbar}$ to the plane waves $\phi_\pm$. These functions $x \mapsto \psi^{\mathrm{in}}_\pm(x)$ are solutions of the Schrödinger equation $H\psi^{\mathrm{in}}_\pm = E\psi^{\mathrm{in}}_\pm$. Indeed, by differentiating the intertwining relation (1.18) at $t = 0$ we obtain

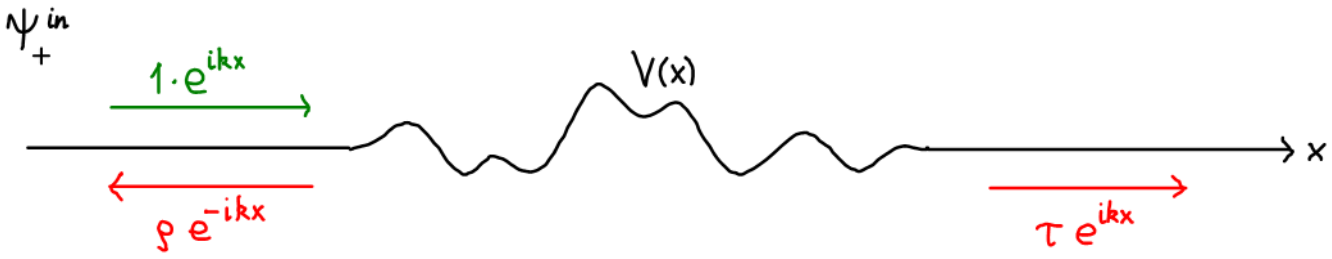
$$HW_\pm = W_\pm H_0\,, \tag{1.24}$$

which implies that if ϕ is an eigenfunction of H_0 with energy E, then $W_\pm\,\phi$ is an eigenfunction of H (with the same energy).

Now we can say more about the solutions $\psi^{\mathrm{in}}_\pm$. Consider, e.g., $\psi^{\mathrm{in}}_+ = W_-\phi_+$. The right factor $\mathrm{e}^{-\mathrm{i}tH_0/\hbar}$ of W_- in the limit of $t \to -\infty$ sends the plane wave $\phi_+(x) = \mathrm{e}^{\mathrm{i}kx}$ (after superposition of a narrow range of k-values to form a localized wave packet) to $x \to -\infty$. The left factor $\mathrm{e}^{\mathrm{i}Ht/\hbar}$ then returns the wave to the scattering region near $x = 0$. This means that ψ^{in}_+ is a stationary scattering state which originates from a wave $\mathrm{e}^{+\mathrm{i}kx}$ moving in the positive x-direction and coming in from $x = -\infty$. In particular, ψ^{in}_+ cannot have a component $\mathrm{e}^{-\mathrm{i}kx}$ at $x \to +\infty$. Thus ψ^{in}_+ must be of the asymptotic form

$$\psi^{\mathrm{in}}_+(x) \overset{\text{asympt}}{\longrightarrow} \begin{cases} A\cdot\mathrm{e}^{+\mathrm{i}kx} + 0\cdot\mathrm{e}^{-\mathrm{i}kx} & x\to+\infty\,, \\ 1\cdot\mathrm{e}^{+\mathrm{i}kx} + D\cdot\mathrm{e}^{-\mathrm{i}kx} & x\to-\infty\,. \end{cases}$$

Here we used the fact that scattering solutions of $H\psi = E\psi$ become superpositions $A\,\mathrm{e}^{\mathrm{i}kx} + B\,\mathrm{e}^{-\mathrm{i}kx}$ of plane waves in the limit of $|x| \to \infty$.



Similarly, $\psi_-^{\rm in}$ is a stationary scattering state which originates from a plane wave $\phi_-(x) = {\rm e}^{-{\rm i}kx}$ moving in the negative x-direction and coming in from $x = +\infty$. There cannot be any incoming component at $x \to -\infty$, so $\psi_-^{\rm in}$ must be of the asymptotic form

$$\psi_-^{\rm in}(x) \overset{\rm asympt}{\longrightarrow} \begin{cases} A' \cdot {\rm e}^{+{\rm i}kx} + 1 \cdot {\rm e}^{-{\rm i}kx} & x \to +\infty\,, \\ 0 \cdot {\rm e}^{+{\rm i}kx} + D' \cdot {\rm e}^{-{\rm i}kx} & x \to -\infty\,. \end{cases}$$

One refers to $\psi_\pm^{\rm in}$ as stationary scattering states satisfying *incoming-wave* boundary conditions.

The second set of states, $\psi_\pm^{\rm out}$, is defined by using the other Møller operator, W_+ :

$$\psi_\pm^{\rm out} := W_+\,\phi_\pm\,.$$

It follows by inversion that

$$W_+^\dagger \psi_\pm^{\rm out} = \phi_\pm\,.$$

By the intertwining relations for W_+, the states $\psi_\pm^{\rm out}$ again solve the time-independent Schrödinger equation $H\psi_\pm^{\rm out} = E\psi_\pm^{\rm out}$. Called scattering states with *outgoing-wave* boundary conditions, they have the asymptotics

$$\psi_+^{\rm out}(x) \overset{\rm asympt}{\longrightarrow} \begin{cases} 1 \cdot {\rm e}^{+{\rm i}kx} + B \cdot {\rm e}^{-{\rm i}kx} & x \to +\infty\,, \\ C \cdot {\rm e}^{+{\rm i}kx} + 0 \cdot {\rm e}^{-{\rm i}kx} & x \to -\infty\,; \end{cases}$$

$$\psi_-^{\rm out}(x) \overset{\rm asympt}{\longrightarrow} \begin{cases} 0 \cdot {\rm e}^{+{\rm i}kx} + B' \cdot {\rm e}^{-{\rm i}kx} & x \to +\infty\,, \\ C' \cdot {\rm e}^{+{\rm i}kx} + 1 \cdot {\rm e}^{-{\rm i}kx} & x \to -\infty\,. \end{cases}$$

By comparing these asymptotic forms with those of the scattering states $\psi_\pm^{\rm in}$ we directly infer that

$$\psi_+^{\rm in} = D\psi_-^{\rm out} + A\psi_+^{\rm out}, \tag{1.25}$$

$$\psi_-^{\rm in} = D'\psi_-^{\rm out} + A'\psi_+^{\rm out}. \tag{1.26}$$

Finally, combining all this information we can make a statement about the scattering operator S. By using the definition $W_-\phi_+ = \psi_+^{\rm in}$ and the relations (1.25) and $W_+^\dagger\psi_\pm^{\rm out} = \phi_\pm$ we compute

$$S\phi_+ = W_+^\dagger W_-\phi_+ = W_+^\dagger\psi_+^{\rm in} = W_+^\dagger(D\psi_-^{\rm out} + A\psi_+^{\rm out}) = D\phi_- + A\phi_+\,,$$

and, similarly,

$$S\phi_- = D'\phi_- + A'\phi_+\,.$$

In view of Eqs. (1.20,1.21) we arrive at the identifications

$$D = \rho\,, \quad A = \tau\,, \quad D' = \tau'\,, \quad A' = \rho'\,.$$

Thus the problem of computing the matrix elements of S is reduced to finding the asymptotics of stationary scattering states with incoming-wave boundary conditions.

Summary. If the solutions of $H\psi_\pm^{\text{in}} = E\psi_\pm^{\text{in}}$ with incoming-wave boundary conditions are of the asymptotic form

$$\psi_+^{\text{in}} \longrightarrow \begin{cases} \tau \cdot \mathrm{e}^{+\mathrm{i}kx} + 0 \cdot \mathrm{e}^{-\mathrm{i}kx} & x \to +\infty\,, \\ 1 \cdot \mathrm{e}^{+\mathrm{i}kx} + \rho \cdot \mathrm{e}^{-\mathrm{i}kx} & x \to -\infty\,, \end{cases}$$

$$\psi_-^{\text{in}} \longrightarrow \begin{cases} \rho' \cdot \mathrm{e}^{+\mathrm{i}kx} + 1 \cdot \mathrm{e}^{-\mathrm{i}kx} & x \to +\infty\,, \\ 0 \cdot \mathrm{e}^{+\mathrm{i}kx} + \tau' \cdot \mathrm{e}^{-\mathrm{i}kx} & x \to -\infty\,, \end{cases}$$

then we have $S\phi_+ = \rho\,\phi_- + \tau\phi_+$ and $S\phi_- = \tau'\phi_- + \rho'\phi_+$.

1.6.2 Scattering by a centro-symmetric potential in $d = 3$

We return to the example (Section 1.5) of scattering by a centro-symmetric potential $V = V(r)$ in three dimensions. In this case the scattering operator S commutes not just with the free Hamiltonian $H_0 = -\hbar^2\nabla^2/2m$ but also with the square L^2 of the total angular momentum and its projection L_z on the z-axis (or any other axis):

$$L^2 = -\frac{\hbar^2}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} - \frac{\hbar^2}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\,, \qquad L_z = \frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial\phi}\,.$$

If Y_{lm} denotes the spherical harmonic of angular momentum l and magnetic quantum number m, the functions

$$\phi_{k,l,m}(r,\theta,\phi) = j_l(kr)Y_{lm}(\theta,\phi)$$

are joint eigenfunctions of the set of operators H_0, L^2, and L_z :

$$H_0\,\phi_{k,l,m} = \frac{\hbar^2k^2}{2m}\,\phi_{k,l,m}\,, \qquad L^2\phi_{k,l,m} = \hbar^2 l(l+1)\,\phi_{k,l,m}\,, \qquad L_z\,\phi_{k,l,m} = \hbar m\,\phi_{k,l,m}\,.$$

The joint eigenspace $\mathcal{E}_{k,l,m}$ with these eigenvalues is one-dimensional: $\mathcal{E}_{k,l,m} = \mathbb{C}\cdot\phi_{k,l,m}$. Therefore, since S commutes with each of H_0, L^2, and L_z, the function $\phi_{k,l,m}$ is an eigenfunction of S :

$$S\,\phi_{k,l,m} = \mathrm{e}^{2\mathrm{i}\delta_l(k)}\phi_{k,l,m}\,.$$

(In some sense, the present situation is simpler than that of $d = 1$, as the basis of functions $\phi_{k,l,m}$ completely diagonalizes S.)

We now claim that the phases $\delta_l(k)$ are the phase shifts of Section 1.5. To verify this, we recall that a scattering solution $\psi_{k,l,m}(r,\theta,\varphi) = R_{k,l}(r)Y_{lm}(\theta,\varphi)$ of the Schrödinger equation $H\psi_{k,l,m} = E\psi_{k,l,m}$ has the asymptotic behavior

$$R_{k,l}(r) \overset{r\to\infty}{\longrightarrow} (2\mathrm{i}kr)^{-1}\left(\mathrm{e}^{2\mathrm{i}\delta_l(k)}\mathrm{e}^{\mathrm{i}(kr-l\pi/2)} - \mathrm{e}^{-\mathrm{i}(kr-l\pi/2)}\right).$$

[Warning: here we have adjusted our overall phase convention!] This solution is a solution with incoming-wave character in the sense that its radially incoming wave component $\mathrm{e}^{-\mathrm{i}kr}/r$ is exactly the same as the corresponding component of the free solution. We therefore expect that $W_-\phi_{k,l,m} = \psi_{k,l,m}$. Indeed, the right factor of $W_- = \lim_{t\to-\infty}\mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_0/\hbar}$ sends $\phi_{k,l,m}$ (or, rather a localized

wave packet formed by superposition of k-values in a narrow range) to an incoming wave $\mathrm{e}^{-\mathrm{i}kr}/r$ at $r = \infty$ in the distant past, and the left factor then produces the full scattering state with a phase-shifted radially outgoing wave component (but unchanged radially incoming wave component).

We still have to figure out what happens to $\psi_{k,l,m} = W_- \, \phi_{k,l,m}$ when the adjoint Møller operator $W_+^\dagger$ is applied. We know that $W_+^\dagger \psi_{k,l,m} = S\phi_{k,l,m}$ is a unitary number times $\phi_{k,l,m}$, but what is that number? To find it, we look at the radially outgoing wave component $\mathrm{e}^{2\mathrm{i}\delta_l}\mathrm{e}^{\mathrm{i}(kr - l\pi/2)}/(2\mathrm{i}kr)$ of $\psi_{k,l,m}$. The operator $W_+^\dagger$ sends this component to $r = \infty$ by the full time evolution and then sends it back in by the free time evolution. In this journey to infinity and back, no scattering takes place. Therefore, whereas W_- left the radially incoming wave component unchanged, it is the radially outgoing wave component that remains unchanged under $W_+^\dagger$. In this way, by comparing expressions, we see that $S\phi_{k,l,m} = W_+^\dagger \psi_{k,l,m} = \mathrm{e}^{2\mathrm{i}\delta_l(k)}\phi_{k,l,m}$.

1.6.3 On the condition $\mathrm{range}(W_-) \subset \mathrm{domain}(W_+^\dagger)$

Let us make a small remark about the condition in the title of this subsection. For this we recall from basic quantum theory that if U and V are Hilbert spaces with Hermitian scalar products $\langle \cdot, \cdot \rangle_U$ resp. $\langle \cdot, \cdot \rangle_V$, then the adjoint of a linear operator $A : \; U \to V$ is the linear operator $A^\dagger : \; V \to U$, $v \mapsto A^\dagger v$ defined by

$$\langle A^\dagger v, u \rangle_U = \langle v, Au \rangle_V$$

for all $u \in U$. Now if $\mathcal{H}$ is the Hilbert space of our problem with Hamiltonian H, let $\mathcal{H}_{\mathrm{sc}} \subset \mathcal{H}$ denote the subspace of scattering states (i.e., the orthogonal complement of the subspace of bound states). The adjoint of the Møller operator $W_+ : \; \mathcal{H} \to \mathcal{H}_{\mathrm{sc}}$ then is an operator

$$W_+^\dagger : \; \mathcal{H}_{sc} \to \mathcal{H} \, .$$

In other words, $\mathrm{domain}(W_+^\dagger) = \mathrm{range}(W_+) = \mathcal{H}_{\mathrm{sc}}$. The condition in the title can therefore be reformulated as

$$\mathrm{range}(W_-) \subset \mathrm{range}(W_+) \, . \tag{1.27}$$

If the Hamiltonian H is time-reversal invariant (see the next subsection), then one can show the identity

$$\mathrm{range}(W_-) = \mathrm{range}(W_+) \, ,$$

so condition (1.27) is always satisfied in that case. According to a remark made after Definition (1.17), it follows that the scattering operator S for a time-reversal invariant system is always unitary: $S^\dagger S = \mathrm{Id}_{\mathcal{H}}$ and $SS^\dagger = \mathrm{Id}_{\mathcal{H}}$.

1.7 Time reversal and scattering

We have already mentioned the fact that unitary symmetries of the Hamiltonian ($UHU^{-1} = H$ and $UH_0U^{-1} = H_0$) give rise to unitary symmetries of the scattering operator ($USU^{-1} = S$). Here we will tell a related story, describing the consequences of time-reversal symmetry (not a unitary symmetry; see below) for scattering.

We begin with a brief discussion of time reversal in classical mechanics. In a classical phase space with position variables q and momenta p, the operation of inverting the time (called 'time inversion' or 'time reversal' for short) is the anti-canonical transformation $T_{\rm cl}$ defined by

$$T_{\rm cl}: \ (q,p) \mapsto (q,-p)\,.$$

It is anti-canonical because it reverses the sign of the Poisson bracket. Clearly, time reversal is an involution, which is to say that $T_{\rm cl}^2$ is the identity transformation. A classical Hamiltonian system is called time-reversal invariant if the Hamiltonian function satisfies $H = H \circ T_{\rm cl}$, i.e.,

$$H(q,p) = H(q,-p)\,. \tag{1.28}$$

An example of a time-reversal invariant Hamiltonian function is the kinetic energy $H(q,p) = p^2/2m$. For charged particles in a magnetic field B this Hamiltonian function changes to

$$H(q,p) = \frac{(p-eA)^2}{2m}\,,$$

where A is a vector potential for B. We observe that in the presence of a magnetic field, time-reversal symmetry in the sense of (1.28) is broken. (Here we take the viewpoint of regarding the magnetic field B as 'external' or fixed. Time reversal continues to be a symmetry even for $B \neq 0$ if, along with transforming q and p, we also time-reverse $B \mapsto -B$ and $A \mapsto -A$.)

The process of quantization is known to take canonical transformations of the classical phase space into unitary transformations of the quantum Hilbert space, $\mathcal{H}$. Now since time reversal fails to be canonical in the classical theory, we should not expect it to be represented by a unitary operator in the quantum theory.

Rather, time reversal in quantum mechanics will turn out to be an *anti*-unitary operator (the definition is spelled out below). To motivate this fact, consider the time-dependent Schrödinger equation (with position variables x and time variable t):

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\psi(x,t) = -\frac{\hbar^2}{2m}\nabla^2\psi(x,t) + V(x)\psi(x,t)\,.$$

By taking the complex conjugate on both sides and inverting the time argument, we see that if $(x,t) \mapsto \psi(x,t)$ is a solution of this equation, then so is $(x,t) \mapsto \overline{\psi(x,-t)}$. We therefore expect that the operator T of time reversal in the position representation $\mathcal{H} = L^2(\mathbb{R}^3)$ (and for the present case of Schrödinger particles) is simply complex conjugation:

$$(T\psi)(x) = \overline{\psi(x)}\,. \tag{1.29}$$

This is in fact true.

Problem. Deduce from (1.29) that the time-reversal operator on wave functions $\widetilde{\psi}(p)$ in the momentum representation is given by $\widetilde{\psi}(p) \mapsto \overline{\widetilde{\psi}(-p)}$. □

We now infer two properties of the time-reversal operator T which are independent of the representation used. The first property,

$$T(z\psi) = \bar{z}\, T\psi \quad (z \in \mathbb{C}\,,\ \psi \in \mathcal{H})\,, \tag{1.30}$$

is called complex anti-linearity. It says that a complex number z goes past the time-reversal operator T as the complex conjugate, $\bar{z}$. Notice that this property distinguishes T from the usual type of complex-linear operator, say A, which obeys the commutation rule $A\, z = z A$.

A stronger consequence of the formula (1.29) is that T preserves the Hermitian scalar product

$$\langle \psi, \psi' \rangle = \int_{\mathbb{R}^3} \overline{\psi(x)}\, \psi'(x)\, d^3x$$

up to complex conjugation:

$$\forall \psi, \psi' \in \mathcal{H}\ :\quad \langle \psi, \psi' \rangle = \overline{\langle T\psi, T\psi' \rangle} = \langle T\psi', T\psi \rangle\ . \tag{1.31}$$

Definition. An $\mathbb{R}$-linear operator $T : \mathcal{H} \to \mathcal{H}$ with the property (1.31) is called anti-unitary.

Problem. (i) Show that the second property (1.31) actually implies the first property (1.30). (ii) Show that the product of two anti-unitary operators is unitary. □

Next we observe that the operator T defined by (1.29) is an involution: $T^2 = \mathrm{Id}_{\mathcal{H}}$. One may ask whether there is a fundamental reason for T to be an involution. We will shortly see that the answer is: no, there exists another possibility.

If an operator T acts on vectors ψ in Hilbert space by $\psi \mapsto T\psi$, then it acts on quantum observables A by conjugation $A \mapsto TAT^{-1}$. Now by the correspondence principle, the action on observables should have a classical limit ($\hbar \to 0$). Since time reversal in the classical theory is an involution, we infer that the action $A \mapsto TAT^{-1}$ of time reversal on quantum observables must also be an involution. Thus we must have

$$T^2 A\, T^{-2} = A$$

for any A. Assuming that the algebra of observables A acts irreducibly on $\mathcal{H}$, this implies that $T^2 = z\,\mathrm{Id}_{\mathcal{H}}$ where z is some complex number. Since T^2 is unitary, the number z must be unitary.

The possible values of the unitary number $z = \mathrm{e}^{\mathrm{i}\alpha}$ are further constrained by associativity of the operator product $T^2 \cdot T = T \cdot T^2$:

$$z\, T\psi = T^2(T\psi) = T(T^2\psi) = T(z\psi) = \bar{z}\, T\psi\ .$$

It follows that our unitary number $z = \mathrm{e}^{\mathrm{i}\alpha}$ has to be real. This leaves but two possibilities: $z = 1$, and $z = -1$. We have already encountered a situation where $z = 1$. The other case of $z = -1$ also occurs in physics.

Fact. The operator of time reversal on a spinor $\psi = \begin{pmatrix} \psi_\uparrow \\ \psi_\downarrow \end{pmatrix}$ (i.e., the wave function of a particle with spin 1/2) is realized by

$$(T\psi)_\uparrow(x) = \overline{\psi_\downarrow(x)}\,, \quad (T\psi)_\downarrow(x) = -\overline{\psi_\uparrow(x)} \tag{1.32}$$

in the position representation. (This fact will be explained in the chapter on Dirac theory.)

After this brief introduction to time reversal, we turn to the consequences of time-reversal symmetry for scattering.

Definition. A quantum Hamiltonian system is called time-reversal invariant if the Hamiltonian stays fixed under conjugation by the time reversal operator: $H = THT^{-1}$. □

We know that the scattering operator S is obtained by taking a limit of products of time evolution operators. Therefore, we now look at what happens to time evolution operators under conjugation by T. By using the relations $T(AB)T^{-1} = (TAT^{-1})(TBT^{-1})$, $T\mathrm{e}^A\,T^{-1} = \mathrm{e}^{TAT^{-1}}$ and $T\mathrm{i}T^{-1} = -\mathrm{i}$, we get

$$T\mathrm{e}^{-\mathrm{i}tH/\hbar}\,T^{-1} = \mathrm{e}^{+\mathrm{i}t\,(THT^{-1})/\hbar}\,,$$

so for a time-reversal invariant system it follows that

$$T\mathrm{e}^{-\mathrm{i}tH/\hbar}\,T^{-1} = \mathrm{e}^{+\mathrm{i}tH/\hbar} = \left(\mathrm{e}^{-\mathrm{i}tH/\hbar}\right)^{-1}\,.$$

Let now the Hamiltonian H_0 for free motion be time-reversal invariant as well. Then by the same calculation we have $T\mathrm{e}^{-\mathrm{i}tH_0/\hbar}\,T^{-1} = \left(\mathrm{e}^{-\mathrm{i}tH_0/\hbar}\right)^{-1}$ and hence

$$T\mathrm{e}^{\mathrm{i}tH_0/\hbar}\,\mathrm{e}^{-2\mathrm{i}tH/\hbar}\,\mathrm{e}^{\mathrm{i}tH_0/\hbar}\,T^{-1} = \left(\mathrm{e}^{\mathrm{i}tH_0/\hbar}\,\mathrm{e}^{-2\mathrm{i}tH/\hbar}\,\mathrm{e}^{\mathrm{i}tH_0/\hbar}\right)^{-1}\,.$$

By taking the limit $t \to \infty$ we conclude that

$$TST^{-1} = S^{-1}. \tag{1.33}$$

Corollary. Assume that H_0 has no bound states and that H_0 and H are time-reversal invariant. Then the scattering operator is unitary on the full Hilbert space: $S^\dagger S = SS^\dagger = \mathrm{Id}_{\mathcal{H}}$. (This conclusion is not tied to time reversal but holds if the pair $H_0\,, H$ has any anti-unitary symmetry.)

In concrete applications one usually looks at matrix elements of the scattering operator in certain subspaces with fixed quantum numbers. One then wants to understand the consequences of time-reversal invariance at the level of matrix elements. In this endeavor it is possible to get confused. Indeed, you might make the following (incorrect) argument. You might say that since T for spinless particles is just complex conjugation, the result (1.33) implies that the scattering matrix is symmetric: $S^t = \bar{S}^\dagger = \bar{S}^{-1} = T\bar{S}T^{-1} = S$. Copying from one of the standard textbooks, you would write the scattering matrix, say for potential scattering in one dimension, as

$$S = \begin{pmatrix} \tau & \rho' \\ \rho & \tau' \end{pmatrix}\,.$$

[You would probably argue that this, not (1.22), is the 'correct' way of arranging the scattering matrix elements. After all, in the limit of vanishing potential, where we have zero reflection

$\rho = \rho' = 0$ and full transmission, $\tau = \tau' = 1$, the scattering matrix should turn into the identity matrix.] The symmetry $S = S^t$ of the scattering matrix would then seem to imply that $\rho \stackrel{?}{=} \rho'$. This is false. The correct statement is that $\tau = \tau'$ due to time-reversal invariance.

What went wrong with our argument? The answer is that we were not careful enough to translate the result (1.33) for the operator S into a correct statement about the matrix of S.

Problem. Get the argument straightened out to show that $\tau = \tau'$. □

In the sequel we will explain how the notion of 'symmetry' of the scattering operator S can be formulated in an invariant (or basis-free) manner. For this purpose we take a time-out in order to review some basic linear algebra.

1.7.1 Some linear algebra

Let V be a vector space over the number field $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$. We recall that the dual, V^*, of V is the vector space of linear functions $f : V \to \mathbb{K}$. Let now $L : V \to W$ be a $\mathbb{K}$-linear mapping between two $\mathbb{K}$-vector spaces V and W. The canonical transpose of L is the mapping $L^t : W^* \to V^*$ defined by

$$(L^t f)(v) := f(Lv) \, .$$

We call it the 'transpose' because, if L (resp. L^t) is expressed with respect to bases of V and W (resp. the dual bases of V^* and W^*), then the matrix of L^t is the transpose of the matrix of L.

Consider now the special case of $W = V^*$. One then has $W^* = (V^*)^* = V$ (for this, to be precise, we should require the vector space dimension to be finite) and the canonical transpose of $L : V \to V^*$ is another mapping $L^t : V \to V^*$ between the same vector spaces. In this situation we can directly compare L with L^t and give a natural meaning to the word 'symmetric'.

Definition. A linear mapping $L : V \to V^*$ is called *symmetric* if $L = L^t$. It is called *skew* if $L = -L^t$.

Remark. In the case of $V = W$, there is *no* canonical definition of 'symmetric' linear map $L : V \to V$. (The matrix of L with respect to some basis of V may be symmetric, but this property is not preserved by a change of basis in general.) To speak of a symmetric map in this context, one needs an identification of V with V^*, e.g., by a non-degenerate quadratic form on V.

Examples (for $\mathbb{K} = \mathbb{R}$):

1. Velocity in three-dimensional space is a vector $v \in V \equiv \mathbb{R}^3$. Momentum is not a vector (at least not fundamentally so) but rather a form or linear function on vectors: $p \in V^* = (\mathbb{R}^3)^*$. The invariant pairing $p(v) := \sum_i p_i v^i$ between the momentum $p \in V^*$ and the velocity $v \in V$ of a particle has the invariant physical meaning of (twice the) kinetic energy of the particle. The mass m (or mass tensor m in an anisotropic medium) is a symmetric linear mapping

$$m : V \to V^* \, , \quad v \mapsto m(v) = p \, .$$

The symmetric nature of m is expressed by $m(v)(v') = p(v') = p'(v) = m(v')(v)$.

2. A rigid body in motion has an angular velocity $\omega \in \mathfrak{so}_3$ (where $\mathfrak{so}_3 \simeq \mathbb{R}^3$ is the Lie algebra of the rotation group SO_3 fixing some point, e.g., the center of mass of the rigid body). The angular momentum L of the body is an element $L \in \mathfrak{so}_3^*$ of the dual vector space. The pairing $L(\omega)$ computes twice the energy of rotational motion of the body. The tensor I of the moments of inertia of the body is a symmetric linear mapping

$$I : \mathfrak{so}_3 \to \mathfrak{so}_3^* , \quad \omega \mapsto I(\omega) = L .$$

3. A homogeneous electric field E is a form $E \in V^*$ ($V = \mathbb{R}^3$), while a homogeneous electric current density j is a vector $j \in V$ (or can be canonically identified with a vector once a homogeneous charge density has been given). The invariant pairing between j and E has the meaning of power, i.e., the rate of energy transfer between the electric field and the matter current. The d.c. electrical conductivity σ of a metal in the Ohmic regime is a linear mapping

$$\sigma : V^* \to V , \quad E \mapsto \sigma(E) = j .$$

If the metal has time-reversal invariance, the conductivity is symmetric: $\sigma^t = \sigma$. If time-reversal symmetry is broken by a magnetic field, σ acquires a skew component $\sigma_H = -\sigma_H^t$ called the Hall conductivity. Notice that any skew (linear) mapping $L : V^* \to V$ for $\dim V = 3$ (or any other odd dimension) must have a vector e which is a null vector, i.e., $L(e) = 0$. In the case of σ_H this vector e coincides with the axis of the magnetic field. Since $\sigma_H(E)(E') = -\sigma_H(E')(E) = 0$ vanishes for $E = E'$, the Hall part of the conductivity does not contribute to the power.

After this list of examples, we continue our review of some basic linear algebra. Let now V be a Hermitian vector space; in other words, V is a complex vector space ($\mathbb{K} = \mathbb{C}$) equipped with a Hermitian scalar product $\langle \cdot, \cdot \rangle_V$. We then have a canonical anti-linear bijection

$$c_V : V \to V^* , \quad v \mapsto \langle v, \cdot \rangle_V . \tag{1.34}$$

In the language of Dirac, this is called the ket-bra bijection, $|v\rangle \mapsto \langle v|$.

Definition. Let $L : V \to W$ be a complex linear mapping between two Hermitian vector spaces V and W. The Hermitian adjoint $L^\dagger : W \to V$ is defined as the composition

$$L^\dagger = c_V^{-1} \circ L^t \circ c_W : \quad W \xrightarrow{c_W} W^* \xrightarrow{L^t} V^* \xrightarrow{c_V^{-1}} V . \tag{1.35}$$

Problem. Show that L and $L^\dagger$ are related by the equation

$$\langle L^\dagger w, v \rangle_V = \langle w, Lv \rangle_W$$

for all $v \in V$ and $w \in W$. □

1.7.2 T-invariant scattering for spin 0 and 1/2

We are now ready to describe in what sense the scattering operator S of a time-reversal invariant system is symmetric. Let us associate with $S : \mathcal{H} \to \mathcal{H}$ a complex linear operator $\widetilde{S} : \mathcal{H} \to \mathcal{H}^*$ by

$$\widetilde{S} = c_{\mathcal{H}} \circ T \circ S : \quad \mathcal{H} \xrightarrow{S} \mathcal{H} \xrightarrow{T} \mathcal{H} \xrightarrow{c_{\mathcal{H}}} \mathcal{H}^* . \tag{1.36}$$

Fact. The scattering operator S of a time-reversal invariant system of particles with spin zero (resp. spin 1/2) is symmetric (resp. skew) in the sense that $\widetilde{S} = +\widetilde{S}^t$ (resp. $\widetilde{S} = -\widetilde{S}^t$).

Proof. We evaluate $\widetilde{S}$ on a pair of vectors $\psi, \psi' \in \mathcal{H}$:

$$\widetilde{S}(\psi)(\psi') = \langle TS\psi, \psi' \rangle .$$

By using the relation (1.33) for a system with time-reversal invariance we obtain

$$\langle TS\psi, \psi' \rangle = \langle S^{-1}T\psi, \psi' \rangle = \langle T\psi, S\psi' \rangle ,$$

where the second equality results from the unitarity $S^{-1} = S^\dagger$ of the scattering operator. In the next step we use the anti-unitary property (1.31) of T :

$$\langle T\psi, S\psi' \rangle = \overline{\langle T^2\psi, TS\psi' \rangle} = \langle TS\psi', T^2\psi \rangle = \widetilde{S}(\psi')(T^2\psi) .$$

To summarize, writing $T^2 = \epsilon_T \mathrm{Id}_{\mathcal{H}}$ we have

$$\widetilde{S}(\psi)(\psi') = \epsilon_T \widetilde{S}(\psi')(\psi) .$$

Thus $\widetilde{S}$ is symmetric for spinless particles ($\epsilon_T = 1$) and skew for spin-1/2 particles ($\epsilon_T = -1$). □

Example. A current topic of active research are so-called topological insulators with strong spin-orbit scattering and time-reversal invariance in $d = 2$ or $d = 3$ dimensions. The one-dimensional boundary of such an insulator in $d = 2$ may house a propagating spin-1/2 mode with one right-moving and one left-moving component. Skew symmetry constrains the scattering matrix for such a two-component mode to be of the form

$$\widetilde{S} = \begin{pmatrix} 0 & \mathrm{e}^{\mathrm{i}\vartheta} \\ -\mathrm{e}^{\mathrm{i}\vartheta} & 0 \end{pmatrix} .$$

This means that the reflection coefficient vanishes identically and the absolute value of the transmission coefficient $\tau = \mathrm{e}^{\mathrm{i}\vartheta}$ always remains unity as the length of the one-dimensional boundary is increased. (Only the phase ϑ changes). Remarkably, this 'perfectly conducting' property of the boundary channel is stable with respect to the introduction of any kind of (time-reversal invariant) disorder.

2 Relativistic quantum mechanics: Dirac equation

2.1 Motivation

We now turn to a very fundamental theme in relativistic quantum mechanics and quantum field theory: the Dirac equation. Its special importance derives from the fact that all known elementary particles of matter (the so-called leptons and quarks), are described by the Dirac equation or its quantum field-theoretic version. The elementary particle we have in mind here, in particular, is the electron.

We begin our motivation with the relativistic equation relating the positive energy E of a free particle of mass m to its momentum p :

$$E = \sqrt{(mc^2)^2 + (pc)^2}. \tag{2.1}$$

In attempting to reconcile quantum mechanics with special relativity, one looks for a relativistic wave equation with the property that this relation is reproduced. By the quantum-theoretic correspondences $E \leftrightarrow \mathrm{i}\hbar\,\partial/\partial t$ and $p \leftrightarrow \hbar\nabla/\mathrm{i}$, a first proposal for a relativistic wave equation (of the electron, say) might be

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} \stackrel{?}{=} \sqrt{(mc^2)^2 - (\hbar c)^2\nabla^2}\;\psi\,.$$

It is, however, not clear how to make physical sense of the square root of an expression involving the Laplacian ∇^2. Defining the square root by its power series, one would end up with a 'non-local' differential equation (involving derivatives up to arbitrarily high order).

To avoid the problems caused by taking a square root, an alternative approach might be to start from the energy-momentum relation (2.1) in the squared form

$$E^2 = (mc^2)^2 + (pc)^2. \tag{2.2}$$

By using again the correspondences $E \leftrightarrow \mathrm{i}\hbar\,\partial/\partial t$ and $p \leftrightarrow \hbar\nabla/\mathrm{i}$ one gets the wave equation

$$-\hbar^2\frac{\partial^2}{\partial t^2}\,\psi = (mc^2)^2\psi - (\hbar c)^2\nabla^2\psi\,, \tag{2.3}$$

which is known as the Klein-Gordon equation. The continuity equation naturally associated to it is $\dot\rho + \mathrm{div} j = 0$ (leading to a conservation law $\int \rho\, d^3x = \mathrm{const}$) with

$$j = \frac{\hbar}{2\mathrm{i}m}\left(\overline{\psi}\,\nabla\psi - \overline{\nabla\psi}\,\psi\right), \quad \rho = \frac{\mathrm{i}\hbar}{2mc^2}\left(\overline{\psi}\,\partial_t\psi - \overline{\partial_t\psi}\,\psi\right) \qquad (\partial_t \equiv \partial/\partial t)\,. \tag{2.4}$$

For solutions with time dependence $\psi \sim \mathrm{e}^{-\mathrm{i}\omega t}$ and positive frequency ω one has $\rho = \frac{\hbar\omega}{mc^2}|\psi|^2 \geq 0$ but for $\psi \sim \mathrm{e}^{+\mathrm{i}\omega t}$ the same quantity becomes negative. This means that $\rho\, d^3x$ in the present case, unlike the Schrödinger case, cannot be interpreted as a probability density. (By the way: in a quantum-field theoretic setting $\rho\, d^3x$ does have an interpretation as a charge density.)

It is not difficult to understand why the positivity of ρ is lacking for the Klein-Gordon equation: it is because the expression for ρ contains the time derivative ∂_t. This, in turn, is a consequence of the Klein-Gordon equation being of second order in ∂_t. (The Schrödinger equation, in contrast, is of first order in ∂_t.)

2.2 Dirac equation

The lesson taken from the Klein-Gordon equation is that, in trying to construct a relativistic generalization of the Schrödinger equation, one should retain the first-order-in-∂_t nature of the equation. Special relativity then suggests that the new equation should also be of first order in the spatial derivatives $\partial/\partial x_j$. In fact, Dirac (1928) pioneered the idea of looking for a first-order differential operator

$$D = \beta mc^2 + \sum_{j=1}^{3} \alpha_j \, p_j c \,, \qquad p_j = \frac{\hbar}{\mathrm{i}} \frac{\partial}{\partial x_j} \,, \tag{2.5}$$

with algebraic objects $\beta, \alpha_1, \alpha_2, \alpha_3$ that remain to be specified. If these satisfy the relations

$$\beta^2 = 1 \,, \quad \beta\alpha_j + \alpha_j\beta = 0 \,, \quad \alpha_i\alpha_j + \alpha_j\alpha_i = 2\delta_{ij} \qquad (i, j = 1, 2, 3) \,, \tag{2.6}$$

then D squares to

$$D^2 = (mc^2)^2 - (\hbar c)^2 \nabla^2, \qquad \nabla^2 = \sum_{j=1}^{3} \frac{\partial^2}{\partial x_j^2} \,. \tag{2.7}$$

Moreover, if ψ is any solution of the equation

$$\mathrm{i}\hbar \frac{\partial \psi}{\partial t} = D\psi \,, \tag{2.8}$$

then ψ by iteration is also a solution of the equation

$$-\hbar^2 \frac{\partial^2 \psi}{\partial t^2} = D^2 \psi \,. \tag{2.9}$$

By using the formula (2.7) for D^2 we see that the latter is nothing but the Klein-Gordon equation (2.3). Thus for plane wave solutions ψ of (2.8) with frequency $\omega = E/\hbar$ and wave vector $k = p/\hbar$ one gets the desired energy-momentum relation (2.1). (At the same time, one gets expressions for ρ and j of a more desirable form; see below.)

The question now is whether one can realize the algebraic relations (2.6), and if so, how. It is certainly impossible to satisfy these relations while clinging to the Schrödinger viewpoint of complex numbers $\beta, \ldots, \alpha_3$ multiplying a wave function ψ with values in $\mathbb{C}$.

Therefore, following Dirac we now abandon the Schrödinger viewpoint and allow that ψ may take values in a more general vector space, say $\mathbb{C}^n$ with $n \geq 1$. With that generalization, we can take $\beta, \ldots, \alpha_3$ to be $n \times n$ matrices multiplying the n-component vector ψ. It then turns out to be possible to realize the relations (2.6) for $n \geq 4$. Indeed, one possible choice for $n = 4$ is

$$\beta = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \quad \alpha_j = \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix}, \quad j = 1, 2, 3, \tag{2.10}$$

where $\mathbf{1} \equiv \mathbf{1}_2$ is the 2×2 unit matrix and σ_j are the Pauli matrices:

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

Problem. Check that the choice (2.10) satisfies the relations (2.6). □

We will elaborate on the theoretical background behind the relations (2.6) in a later subsection. For now, we record that there exists at least one possible realization for $n = 4$. In our further arguments, we will often refer to this realization for concreteness.

Definition. The Dirac equation for a free particle of mass m reads

$$\mathrm{i}\hbar\,\frac{\partial\psi}{\partial t} = mc^2\beta\psi + \frac{\hbar c}{\mathrm{i}}\sum_{j=1}^{3}\alpha_j\,\frac{\partial\psi}{\partial x_j}\,, \tag{2.11}$$

where $\psi(x,t)$ takes values in $\mathbb{C}^4$ and the 4×4 matrices $\beta,\alpha_1,\alpha_2,\alpha_3$ are subject to (2.6).

2.3 Relativistic formulation

For some purposes it is useful to write the Dirac equation in a form which puts space and time on a similar footing. The standard physics convention is to introduce

$$x^0 := ct\,,\quad x^j := x_j\,,\quad \gamma^0 := \beta\,,\quad \gamma^j := \beta\alpha_j = -\alpha_j\beta\quad (j=1,2,3)\,. \tag{2.12}$$

With these conventions, the Dirac equation (2.11) takes the relativistic form

$$\left(\gamma^0\frac{\partial}{\partial x^0} + \sum\nolimits_{j=1}^{3}\gamma^j\frac{\partial}{\partial x^j} + \mathrm{i}\frac{mc}{\hbar}\right)\psi = 0\,. \tag{2.13}$$

The parameter $mc/\hbar$ has the physical dimension of an inverse length. Its reciprocal $\hbar/(mc)$ is called the (reduced) Compton wave length. For the electron with mass $m \approx 0.5\,\mathrm{MeV}/c^2$ one has

$$\frac{\hbar}{mc} = \frac{\hbar c}{mc^2} \approx \frac{200\,\mathrm{MeV}\cdot\mathrm{fm}}{0.5\,\mathrm{MeV}} \approx 0.4\times 10^{-12}\mathrm{m}\,. \tag{2.14}$$

In the jargon of physics the matrices $\gamma^0,\ldots,\gamma^3$ are called the gamma matrices. We note that for the choice (2.10) they have the expressions

$$\gamma^0 = \begin{pmatrix}\mathbf{1} & 0\\ 0 & -\mathbf{1}\end{pmatrix}\,,\quad \gamma^j = \begin{pmatrix}0 & \sigma_j\\ -\sigma_j & 0\end{pmatrix}\,,\quad j=1,2,3. \tag{2.15}$$

It is customary to use Greek letters for space-time indices; i.e., $\mu = 0,1,2,3$.

Problem. Adopting this notation, show that the algebraic relations (2.6) take the concise form

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbf{1}\,, \tag{2.16}$$

where $\mathbf{1}\equiv\mathbf{1}_4$ is the 4×4 unit matrix, and

$$g^{00} = 1\,,\quad g^{11} = g^{22} = g^{33} = -1\,,\quad g^{\mu\nu} = 0\quad\text{for}\quad \mu\neq\nu\,, \tag{2.17}$$

are the components of the Minkowski metric tensor. □

It is also customary in the present context to use the Einstein summation convention, which says that repeated Greek indices are understood to be summed over.

Summary. The relativistic (or covariant) form of the free-particle Dirac equation (2.11) is

$$\left(\gamma^\mu\frac{\partial}{\partial x^\mu} + \mathrm{i}\frac{mc}{\hbar}\right)\psi = 0\,. \tag{2.18}$$

2.4 Non-relativistic reduction

Dirac's theory is intended to be a relativistic quantum theory of the electron. We already have a *non*-relativistic quantum theory of the electron, namely the Schrödinger equation or, including spin, the Pauli equation. By the principles of theory building, in order for a new theory to be acceptable it must be consistent with the old theory which is already known to be true (within its limits of validity). Therefore the logical step to be taken next is to verify that the Dirac equation reduces to the Schrödinger/Pauli equation in the non-relativistic limit.

For this purpose we write the Dirac equation in the following block-decomposed form:

$$0 = \begin{pmatrix} \mathrm{i}\frac{mc}{\hbar} + \frac{1}{c}\frac{\partial}{\partial t} & \sum \sigma_j \frac{\partial}{\partial x_j} \\ -\sum \sigma_j \frac{\partial}{\partial x_j} & \mathrm{i}\frac{mc}{\hbar} - \frac{1}{c}\frac{\partial}{\partial t} \end{pmatrix} \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix}, \quad \psi_\pm = \begin{pmatrix} \psi_{\pm,\uparrow} \\ \psi_{\pm,\downarrow} \end{pmatrix}. \tag{2.19}$$

At present, the symbols $\uparrow$ and $\downarrow$ are just some fancy notation to label the two components of $\psi_\pm(x,t) \in \mathbb{C}^2$. (Later we will see that they do, in fact, reflect the spin of the electron.)

The non-relativistic limit is $|v| \ll c$, or $|k| \ll mc/\hbar$. By the correspondence $k \leftrightarrow \nabla/\mathrm{i}$ this means that the off-diagonal blocks of the matrix in (2.19) are to be considered as being much smaller than the diagonal blocks. In zeroth-order approximation we neglect the off-diagonal blocks altogether to obtain

$$\left(\mathrm{i}\frac{mc}{\hbar} + \frac{1}{c}\frac{\partial}{\partial t}\right)\psi_+^{(0)} = 0\,, \quad \psi_-^{(0)} = 0\,. \tag{2.20}$$

We are setting $\psi_-^{(0)} = 0$ by *fiat* because we intend to identify ψ_+ with the spinor wave function of the Pauli equation, and there is no room for additional degrees of freedom in the non-relativistic limit. (We will learn later that ψ_- describes the positron, the antiparticle of the electron. In the present context, we envisage a situation with no positrons present. Hence our choice $\psi_-^{(0)} = 0$.) Note that the first equation in (2.20) implies that $\psi_+^{(0)}$ has the time dependence

$$\psi_+^{(0)} \sim \mathrm{e}^{-\mathrm{i}mc^2 t/\hbar}.$$

We now turn to a first-order (or improved) approximation. For this we write the system of equations (2.19) in the form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} = 0\,,$$

or equivalently,

$$A\psi_+ + B\psi_- = 0\,, \quad C\psi_+ + D\psi_- = 0\,.$$

We will see that the operator $D = \mathrm{i}\frac{mc}{\hbar} - \frac{1}{c}\frac{\partial}{\partial t}$ has an inverse when acting on $C\psi_+$. We can therefore solve the second equation for ψ_-, and by inserting the solution $\psi_- = -D^{-1}C\psi_+$ into the first equation we obtain an equation solely for ψ_+ :

$$(A - BD^{-1}C)\,\psi_+ = 0\,,$$

or explicitly,

$$\left(\mathrm{i}\frac{mc}{\hbar} + \frac{1}{c}\frac{\partial}{\partial t}\right)\psi_+ + \left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)\left(\mathrm{i}\frac{mc}{\hbar} - \frac{1}{c}\frac{\partial}{\partial t}\right)^{-1}\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)\psi_+ = 0\,.$$

Now since we know from the zeroth-order approximation that ψ_+ has the leading time dependence $\psi_+^{(0)} \sim \mathrm{e}^{-\mathrm{i}mc^2t/\hbar}$, we have $-\frac{1}{c}\frac{\partial}{\partial t}\psi_+ \approx \mathrm{i}\frac{mc}{\hbar}\psi_+$, and we may replace $-D^{-1}C\psi_+$ by

$$-D^{-1}C\psi_+ = \left(\mathrm{i}\frac{mc}{\hbar} - \frac{1}{c}\frac{\partial}{\partial t}\right)^{-1}\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)\psi_+ \to \frac{\hbar}{2\mathrm{i}mc}\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)\psi_+ .$$

By multiplying the equation for ψ_+ by $\mathrm{i}\hbar c$ and slightly rearranging the terms, we then arrive at the improved approximation $\psi_+ \approx \psi_+^{(1)}$ where $\psi_+^{(1)}$ satisfies

$$\mathrm{i}\hbar\frac{\partial\psi_+^{(1)}}{\partial t} = mc^2\psi_+^{(1)} - \frac{\hbar^2}{2m}\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)^2 \psi_+^{(1)}.$$

In the final step we use the relations $\sigma_j\sigma_l + \sigma_l\sigma_j = 2\delta_{jl}$ for the Pauli matrices to find

$$\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)^2 = \sum \frac{\partial^2}{\partial x_j^2} = \nabla^2.$$

Altogether we have

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\psi_+^{(1)} = mc^2\psi_+^{(1)} - \frac{\hbar^2}{2m}\nabla^2\psi_+^{(1)}. \tag{2.21}$$

This is indeed the free-particle Schrödinger equation with a constant shift of the energy by the rest mass mc^2. Thus the Dirac equation has passed its first test.

Note added. At the same level of approximation, the equation $\psi_- = -D^{-1}C\psi_+$ yields

$$\psi_-^{(1)} = \frac{\hbar}{2\mathrm{i}mc}\left(\sum \sigma_j \frac{\partial}{\partial x_j}\right)\psi_+^{(1)}.$$

Thus $\psi_-^{(1)}$ is smaller than $\psi_+^{(1)}$ by, roughly speaking, a factor of $\hbar|k|/mc$ or $|v|/c$.

2.5 Enter the electromagnetic field

In order to turn the free-particle Dirac equation (2.18) into an equation for charged particles such as the electron, we need to introduce the coupling to the electromagnetic field. [Recall from the course on classical electrodynamics that the electromagnetic field strength tensor, also known as the Faraday tensor, is given by

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}, \tag{2.22}$$

where A_μ are the components of the 4-vector of the electromagnetic gauge field.] The form of this coupling is determined by the principles of gauge invariance and minimal substitution, as follows.

If χ is some space-time dependent function (of physical dimension action/charge), the Faraday tensor is invariant under gauge transformations

$$A_\mu \to A_\mu + \frac{\partial}{\partial x^\mu}\chi. \tag{2.23}$$

The principle of gauge invariance now says the following. If ψ is the wave function of a quantum particle with electric charge e, then the physics of the coupled system (i.e., the particle interacting with the electromagnetic field) must be invariant under the gauge transformation (2.23) of the gauge field in combination with the gauge transformation

$$\psi \to \mathrm{e}^{\mathrm{i}e\chi/\hbar}\psi \tag{2.24}$$

of the matter field ψ. It is easy to see that the expression

$$\left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x^\mu} - eA_\mu\right)\psi$$

is gauge invariant in this sense. The principle of minimal substitution then tells us to enforce gauge invariance by making in the Dirac equation (or any other charged quantum wave equation, for that matter) the substitution

$$\frac{\partial}{\partial x^\mu}\,\psi \to \left(\frac{\partial}{\partial x^\mu} - \frac{\mathrm{i}e}{\hbar}A_\mu\right)\psi\,. \tag{2.25}$$

Doing so, we arrive at the final form of the Dirac equation.

Definition. The Dirac equation for a particle of mass m and charge e in the presence of an electromagnetic field (described by the gauge field A_μ) is

$$\gamma^\mu\left(\frac{\partial}{\partial x^\mu} - \frac{\mathrm{i}e}{\hbar}A_\mu\right)\psi + \mathrm{i}\frac{mc}{\hbar}\psi = 0\,. \tag{2.26}$$

Problem. By following the steps of Section 2.4, show that the full Dirac equation (2.26) in the non-relativistic limit reduces to the Pauli equation (i.e., the Schrödinger equation including the Pauli coupling of the spin of the charged particle to the magnetic field B):

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\,\psi_+^{(1)} = (mc^2 + e\,\Phi)\psi_+^{(1)} - \frac{\hbar^2}{2m}\sum_j\left(\frac{\partial}{\partial x_j} - \frac{\mathrm{i}e}{\hbar}A_j\right)^2\psi_+^{(1)} - \frac{e\hbar}{2m}\sum_j B_j\,\sigma_j\,\psi_+^{(1)}. \tag{2.27}$$

Here $\Phi = -cA_0$ is the electric scalar potential, and A_j are components of the magnetic vector potential $\mathbf{A}$ obeying $\mathrm{rot}\,\mathbf{A} = B$.

Remark. By pushing the process of non-relativistic reduction to higher order, one gets corrections to the Schrödinger equation beyond that of the Pauli term. Of these, we wish to mention one specific term of importance in contemporary physics: the spin-orbit interaction (or spin-orbit coupling, SOC). It has long been understood that SOC is of relevance for nuclear structure physics, where it affects the pattern of nuclear shell closure by lowering the energy of the $j_+ = l + 1/2$ sub-shells relative to those of the $j_- = l - 1/2$ sub-shells (cf. Nobel Prize in Physics 1963 for M. Goeppert-Mayer and H. Jensen). In recent years, SOC has also acquired considerable notoriety in condensed matter physics, as its presence may lead to band inversion, turning band insulators into topological insulators that feature gapless surface states. A quick heuristic for SOC is to note that the Pauli equation attributes to a Dirac particle a spin-magnetic moment of $\vec{\mu} = -e\vec{S}/m$ with $S_l = \hbar\,\sigma_l/2$. Now, special relativity demands that a spin-magnetic moment μ_{magn} in a state of motion with velocity v acquires an electric dipole moment $\vec{\mu}_{\mathrm{el}} = -\vec{\mu}_{\mathrm{magn}} \times \vec{v}/c$. The electric dipole moment due to relativistic motion is acted upon by an electric field E via the dipole-field coupling $-\vec{\mu}_{\mathrm{el}} \cdot \vec{E}$. This coupling can be re-expressed as the triple scalar product of $\vec{\mu}_{\mathrm{magn}}$, $\vec{v}/c$, and $\vec{E}$. SOC is obtained from here by implementing the spin-magnetic moment as a quantum operator; in the case of the electron one has $\vec{\mu}_{\mathrm{magn}} = -e\hbar\,\vec{\sigma}/2m$.

2.6 Continuity equation

At this stage of the theoretical development, one might hope that the probabilistic interpretation of the square $|\psi|^2$ of the Schrödinger wave function could be carried over without any essential change to the Dirac equation (as a single-particle theory). In the present subsection we substantiate this optimistic thought. Later, however, we will see that there are serious problems with this interpretation, and we will indicate what needs to be changed to end up with a satisfactory theory.

The probabilistic interpretation of the Schrödinger wave function $|\psi|^2 =: \rho$ rests on the continuity equation $\dot{\rho} + \operatorname{div} j = 0$ together with positivity, $\rho \geq 0$. Let us now transcribe the Schrödinger derivation of this equation (which students know from basic quantum mechanics) to the Dirac case. For this purpose we start from the Dirac equation (2.26) in the form

$$\left(\frac{\partial}{\partial t} + \frac{\mathrm{i}e}{\hbar}\Phi\right)\psi + c\sum_l \left(\frac{\partial}{\partial x_l} - \frac{\mathrm{i}e}{\hbar}A_l\right)\alpha_l\psi + \mathrm{i}\frac{mc^2}{\hbar}\beta\psi = 0\,. \tag{2.28}$$

ψ has four components, as we recall, and thus takes values in $\mathbb{C}^4$. We now assume that the vector space $\mathbb{C}^4$ is Hermitian, i.e., is equipped with a Hermitian scalar product $\mathbb{C}^4 \times \mathbb{C}^4 \to \mathbb{C}$. Using this structure we define the Hermitian adjoint $\psi^\dagger$ with values in the dual vector space $(\mathbb{C}^4)^*$, and $\psi^\dagger\psi$ with values in $\mathbb{C}$ (actually, $\mathbb{R}$).

Now we observe that the matrices β and α_l in (2.10) are Hermitian: $\beta = \beta^\dagger$ and $\alpha_l = \alpha_l^\dagger$ $(l = 1, 2, 3)$. We promote this observation to an axiom of the theory, i.e., we demand that any permissible choice of β and α_l not only obeys the algebraic relations (2.6) but must also have the property of being Hermitian. By dualizing the equation (2.28) we then obtain the following equation for $\psi^\dagger$:

$$\left(\frac{\partial}{\partial t} - \frac{\mathrm{i}e}{\hbar}\Phi\right)\psi^\dagger + c\sum_l \left(\frac{\partial}{\partial x_l} + \frac{\mathrm{i}e}{\hbar}A_l\right)\psi^\dagger\alpha_l - \mathrm{i}\frac{mc^2}{\hbar}\psi^\dagger\beta = 0\,. \tag{2.29}$$

Next we contract the equation (2.28) for the vector ψ with the dual vector $\psi^\dagger$, and similarly the equation (2.29) for $\psi^\dagger$ with ψ. Afterwards we add the two resulting scalar equations. The terms containing $\mathrm{i} = \sqrt{-1}$ all cancel since their signs are changed by taking the Hermitian adjoint. So we get

$$\frac{\partial}{\partial t}\,\psi^\dagger\psi + c\sum_l \frac{\partial}{\partial x_l}\,\psi^\dagger\alpha_l\psi = 0\,. \tag{2.30}$$

This has the form of a continuity equation $\dot{\rho} + \operatorname{div} j = 0$ if we let

$$\rho := \psi^\dagger\psi\,, \quad j_l := c\,\psi^\dagger\alpha_l\psi\,. \tag{2.31}$$

Summary. We record that if ψ is a solution of the Dirac equation (with or without electromagnetic field), then the scalar $\rho = \psi^\dagger\psi$ and the vector j with components $j_l = c\,\psi^\dagger\alpha_l\psi$ satisfy the continuity equation

$$\dot{\rho} + \operatorname{div} j = 0\,. \tag{2.32}$$

By a standard argument using the divergence theorem (a.k.a. Gauss' theorem) it follows that the total space integral $\int \rho\, d^3x$ is conserved.

Remark. Coming from Schrödinger quantum mechanics, it is natural to think that ρ (actually, $\rho\, d^3x$) is the probability density for a relativistic electron, and j is the vector of the corresponding probability current density. However, it will turn out that this (wishful) thinking is untenable. Dirac's theory in fact will have to be reformulated (in the framework of quantum field theory) so as to give ρ the interpretation of charge density of the electron (actually of the quantum field encompassing the electron as well as the positron).

2.7 Clifford algebra

In this and the following subsection we provide some theoretical background concerning the four-component nature of the wave function ψ of the Dirac equation. A question which was left open in Section 2.3 is this: how can we say a priori that the algebraic relations

$$\gamma^\nu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$$

are realizable for $n\times n$ matrices with $n \geq 4$, and how can such a matrix realization be constructed? To answer this question, we shall take the liberty of going into more detail than is offered in most physics textbooks, as the very same formalism will turn out to be relevant for the procedure of second quantization of many-particle quantum mechanics.

In the sequel we will be concerned with a vector space V over the real number field $\mathbb{K} = \mathbb{R}$ or the complex number field $\mathbb{K} = \mathbb{C}$. We assume that V comes with a non-degenerate symmetric $\mathbb{K}$-bilinear form (also referred to as a 'quadratic form' for short)

$$Q \,:\, V\times V \to \mathbb{K}\,, \quad (v,v') \mapsto Q(v,v') = Q(v',v)\,. \tag{2.33}$$

Examples. We give two examples, the number field being $\mathbb{K} = \mathbb{R}$ in both cases. The first example is the Euclidean vector space $V = \mathbb{R}^3$ equipped with the Euclidean scalar product Q,

$$Q(v,v') = |v|\,|v'|\,\cos\angle(v,v')\,.$$

The second example is the example of relevance for the Dirac equation: the Lorentzian vector space $V = \mathbb{R}^4$ with the Minkowski scalar product Q given by (summation convention!)

$$Q(v,w) = Q\left(v^\mu e_\mu\,, w^\nu e_\nu\right) = g_{\mu\nu}\, v^\mu w^\nu = v^0w^0 - v^1w^1 - v^2w^2 - v^3w^3 \tag{2.34}$$

in any standard basis $\{e_0, e_1, e_2, e_3\}$. □

To define what is meant by the Clifford algebra of a vector space V with quadratic form Q, we need the following basic concept.

Definition. An associative algebra is a vector space, say $\mathcal{A}$, with the additional structure of an associative product $\mathcal{A}\times\mathcal{A}\to\mathcal{A}$, $(a,b)\mapsto ab$, which distributes over addition: $a(b+c) = ab + ac$.

Remark. As usual, associativity of the product means that there is no need to use parentheses in multiple products such as $abc = (ab)c = a(bc)$. The main examples for an associative algebra

are provided by matrices: the $\mathbb{K}$-vector space of matrices of size $n \times n$ (say) with matrix elements taken from $\mathbb{K}$, is an associative algebra with the product being the usual matrix multiplication.

Definition. The Clifford algebra $\mathrm{Cl}(V,Q)$ of the vector space V with quadratic form Q is the associative algebra generated by $V \oplus \mathbb{K}$ with relations

$$vw + wv = 2Q(v,w)\,. \tag{2.35}$$

Remark. The words 'associative algebra generated by $V \oplus \mathbb{K}$' mean that the elements of $\mathrm{Cl}(V,Q)$ are polynomial expressions in the elements of V with coefficients taken from the number field $\mathbb{K}$. The relations (2.35) imply, e.g.,

$$uvw = -uwv + 2Q(v,w)u = -vuw + 2Q(u,v)w \qquad (u,v,w \in V)\,.$$

Example. Let $(V = \mathbb{R}^4, Q)$ be the Lorentzian vector space with Minkowski scalar product Q. Take $\{e_0, e_1, e_2, e_3\}$ to be some standard basis of V, so that $Q(e_\mu\,, e_\nu) = g_{\mu\nu}$ where $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$. Then some examples of elements in $\mathrm{Cl}(V,Q)$ are

$$e_0 e_0 = 1\,, \quad e_1 e_1 = -1\,, \quad e_0 e_1 = -e_1 e_0\,, \quad e_0 e_1 e_0 = -e_1\,.$$

All this may seem abstract and unfamiliar. It can be made more tangible as follows.

Let us think of the vectors $e_0\,, \ldots, e_3$ as basis elements in some four-dimensional subspace $V \simeq \mathbb{R}^4$ of the linear space (or vector space) of 4×4 matrices with quadratic form

$$Q(v,v') = \tfrac{1}{4}\mathrm{Tr}\,(vv')\,. \tag{2.36}$$

More precisely, let

$$e_0 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}\,, \quad e_j = \begin{pmatrix} 0 & -\sigma_j \\ \sigma_j & 0 \end{pmatrix} \quad (j = 1,2,3)\,.$$

One easily verifies the scalar products $Q(e_\mu, e_\nu) = g_{\mu\nu}$. Thus $V \simeq \mathbb{R}^4$ becomes the four-dimensional real vector space of matrices of the form

$$v = \begin{pmatrix} v^0 & 0 & -v^3 & -v^1 + \mathrm{i}v^2 \\ 0 & v^0 & -v^1 - \mathrm{i}v^2 & v^3 \\ v^3 & v^1 - \mathrm{i}v^2 & -v^0 & 0 \\ v^1 + \mathrm{i}v^2 & -v^3 & 0 & -v^0 \end{pmatrix}\,. \tag{2.37}$$

Because we have realized our vectors $v = v^\mu e_\mu$ as matrices, the vector space $V \simeq \mathbb{R}^4$ now carries the additional structure of an associative algebra with matrix multiplication playing the role of the product. If Id denotes the 4×4 unit matrix, one sees that the relations

$$vw + wv = 2Q(v,w)\,\mathrm{Id} \tag{2.38}$$

are satisfied by the matrices (2.37).

Thus, in the present example $\mathrm{Cl}(V,Q)$ can be realized as the algebra of 4×4 matrices that arise by multiplication of generators of the form (2.37). Using mathematical language one calls such a concrete realization by matrices a representation of the abstractly defined algebra. Essentially

the same representation was given in an earlier subsection [with only a minor sign difference due to the relativistic convention of raising and lowering indices]. □

Problem. Show that if $\dim V = n$ and $\{e_1, \ldots, e_n\}$ is a Q-orthogonal basis of V, then the following Clifford algebra elements constitute a basis of $\mathrm{Cl}(V, Q)$ as a vector space:

$$1\,, \quad e_j \quad (j = 1, \ldots, n)\,, \quad e_i e_j \quad (i < j)\,, \quad e_i e_j e_l \quad (i < j < l)\,, \quad \ldots\,, \quad e_0 e_1 \cdots e_n\,. \tag{2.39}$$

Count the number of these elements to deduce that $\dim \mathrm{Cl}(V, Q) = 2^{\dim V}$.

2.8 Spinor representation

While the Clifford algebra $\mathrm{Cl}(V, Q)$ is defined abstractly as an associative algebra with certain binary relations, we saw that $\mathrm{Cl}(V, Q)$ has a concrete realization by 4×4 matrices for the Minkowski example of $V = \mathbb{R}^4$. This realization will be called the spinor representation of $\mathrm{Cl}(\mathbb{R}^4, Q)$ from now on. In Section 2.7 we simply wrote down the spinor representation without explaining where it came from. This gap in our theoretical development is to be closed in the present subsection.

For brevity we will only consider the special case of a real vector space $V \simeq \mathbb{R}^{2n}$ of even dimension with Euclidean scalar product Q. Given this situation, let us fix some orthonormal basis $\{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ of V.

Although everything so far is real ($\mathbb{K} = \mathbb{R}$), we are now going to consider complex linear combinations of the vectors $v \in V$; mathematically speaking we pass to the complexification $V \otimes_{\mathbb{R}} \mathbb{C}$ of V. (The symbol $\otimes_{\mathbb{K}}$ denotes the tensor product of vector spaces over $\mathbb{K}$; see the end of this subsection for a quick exposition.) In the complexification $V \otimes_{\mathbb{R}} \mathbb{C}$ we then form the linear combinations ($\mathrm{i} = \sqrt{-1}$)

$$c_j := (e_j - \mathrm{i} f_j)/2\,, \quad c_j^* := (e_j + \mathrm{i} f_j)/2 \qquad (j = 1, \ldots, n)\,. \tag{2.40}$$

We also introduce the complex vector spaces

$$P = \mathrm{span}_{\mathbb{C}}\{c_1, \ldots, c_n\}\,, \quad P^* = \mathrm{span}_{\mathbb{C}}\{c_1^*, \ldots, c_n^*\}\,, \tag{2.41}$$

and record the decomposition

$$V \otimes_{\mathbb{R}} \mathbb{C} = P \oplus P^*\,. \tag{2.42}$$

Such a decomposition is called a polarization. [The notation indicates that P^* may be regarded as the dual vector space of P via the non-degenerate pairing $P \otimes P^* \to \mathbb{C}$ by $(c, c^*) \mapsto Q(c, c^*)$.]

Notice that the Clifford relations (2.35) imply the following relations for our new generators:

$$c_j c_l + c_l c_j = 0\,, \quad c_j^* c_l^* + c_l^* c_j^* = 0\,, \quad c_j c_l^* + c_l^* c_j = \delta_{jl} \quad (j, l = 1, \ldots, n)\,. \tag{2.43}$$

In physics, these relations are called the canonical anti-commutation relations (CAR). They will play a central role in the formalism of second quantization for many-fermion systems.

To formulate the algorithm for constructing the spinor representation of $\mathrm{Cl}(V, Q)$, we need one more algebraic concept.

Definition. By the exterior algebra $\wedge(U)$ of a $\mathbb{K}$-vector space U one means the associative algebra generated by $U \oplus \mathbb{K}$ with relations

$$\forall u, u' \in U : \quad uu' + u'u = 0 \, . \tag{2.44}$$

The exterior algebra is a direct sum of subspaces of fixed degree:

$$\wedge(U) = \bigoplus_{l=0}^{\dim U} \wedge^l(U) \, , \tag{2.45}$$

where $\wedge^0(U) \equiv \mathbb{K}$, $\wedge^1(U) \equiv U$, $\wedge^2(U)$ consists of all quadratic elements $\sum u_i u_j$, $\wedge^3(U)$ consists of all cubic elements $\sum u_i u_j u_k$, and so on.

Remarks. i) Note that (2.44) implies $u^2 = 0$ for any $u \in U \subset \wedge(U)$. ii) In physics the exterior algebra $\wedge(U)$ is also known as the fermionic Fock space of U. (More precisely, if U is a Hilbert space, then $\wedge(U)$ is another Hilbert space called the fermionic Fock space associated to the single-particle Hilbert space U.) In the Fock space setting the degree has the physical meaning of particle number. iii) The notion of exterior algebra is basic to the calculus of differential forms.

Problem. Show that the exterior algebra $\wedge(U)$ is isomorphic as a vector space to the Clifford algebra $\mathrm{Cl}(U, Q)$. Hint: prove by construction of a basis that $\wedge(U)$ has dimension $2^{\dim U}$. □

We now recall the definition of the complex vector space P^* in (2.41) and make the important observation that the exterior algebra of P^* (or P for that matter) is contained as a subalgebra in $\mathrm{Cl}(V, Q) \otimes \mathbb{C}$. By this observation the exterior algebra $\wedge(P^*)$ can be turned into a representation space for $\mathrm{Cl}(V, Q)$ as follows.

We are looking for an operation of $\mathrm{Cl}(V, Q)$ on $\wedge(P^*)$,

$$\mathrm{Cl}(V, Q) \times \wedge(P^*) \to \wedge(P^*) \, , \quad (a, \xi) \mapsto a \cdot \xi \, , \tag{2.46}$$

which is a representation. In other words, we want a Clifford multiplication rule (2.46) which is compatible with the associative product in the Clifford algebra in the sense that (i) the relation

$$(ab) \cdot \xi = a \cdot (b \cdot \xi) \tag{2.47}$$

holds for all $a, b \in \mathrm{Cl}(V, Q)$, $\xi \in \wedge(P^*)$, and (ii) the Clifford relations (2.35) are preserved, i.e., for all $v, w \in V \subset \mathrm{Cl}(V, Q)$, $\xi \in \wedge(P^*)$, we have

$$v \cdot (w \cdot \xi) + w \cdot (v \cdot \xi) - 2Q(v, w) \cdot \xi = 0 \, . \tag{2.48}$$

Since $\mathrm{Cl}(V, Q)$ is generated by $V \oplus \mathbb{K}$ it is sufficient to specify the Clifford multiplication (2.46) for scalars $k \in \mathbb{C} \subset \mathrm{Cl}(V, Q)$ and vectors $v \in V \subset \mathrm{Cl}(V, Q)$. In the former case the multiplication (2.46) is the natural one and needs no explanation.

In order to describe the multiplication by vectors $v \in V$ (in the following we use the alternative word 'action' for it) we use the polarization $V \otimes_{\mathbb{R}} \mathbb{C} = P \oplus P^*$. First, let $v \in P^*$. The action of such an element $v \in P^*$ on $\xi \in \wedge(P^*)$ is simply exterior multiplication:

$$v \cdot \xi := v\xi \, , \tag{2.49}$$

i.e., the product is computed in the exterior algebra $\wedge(P^*)$. Note that the action of $v \in P^*$ increases the degree: it maps $\wedge^l(P^*)$ into $\wedge^{l+1}(P^*)$.

Second, let $v \in P$. Such elements act by an operation which lowers the degree:

$$P \ni v \,:\, \wedge^l(P^*) \to \wedge^{l-1}(P^*) \,. \tag{2.50}$$

This operation is defined recursively. For the two lowest degrees $l = 0, 1$ one sets

$$l = 0: \quad v \cdot 1 := 0 \,, \tag{2.51}$$

$$l = 1: \quad v \cdot v' := 2Q(v, v') \,. \tag{2.52}$$

The action on higher-degree elements is then defined by the so-called anti-Leibniz rule:

$$v \cdot (\xi\eta) := (v \cdot \xi)\,\eta + (-1)^l \xi\,(v \cdot \eta) \,, \qquad \xi \in \wedge^l(P^*). \tag{2.53}$$

Claim. Equations (2.49)–(2.53) give a representation of $\mathrm{Cl}(V, Q)$ on $\wedge(P^*)$. □

Since all elements of $V \oplus \mathbb{K} \subset \mathrm{Cl}(V, Q)$ act as linear operators and multiplication of linear operators is associative, it is immediately clear that the compatibility condition (2.47) is satisfied. To verify the Claim, what must be shown is that the property (2.48) holds. For this purpose we introduce some more notation. We write the decomposition of $v \in V$ by $V \otimes_{\mathbb{R}} \mathbb{C} = P \oplus P^*$ as $v = v_P + v_{P^*}$. We then denote the degree-increasing action of the P^*-component v_{P^*} by $\varepsilon(v_{P^*})$ and the degree-lowering operation of the P-component v_P as $\iota(v_P)$. [In physics one calls $\varepsilon(v_{P^*})$ the particle creation component and $\iota(v_P)$ the particle annihilation component of v.] Thus

$$v \cdot \xi = \varepsilon(v_{P^*})\xi + \iota(v_P)\xi \,. \tag{2.54}$$

Problem. Show that the anti-commutation relations

$$\begin{aligned} &\varepsilon(x)\varepsilon(x') + \varepsilon(x')\varepsilon(x) = 0 \,, \quad \iota(y)\iota(y') + \iota(y')\iota(y) = 0 \,, \\ &\varepsilon(x)\iota(y) + \iota(y)\varepsilon(x) = 2Q(x, y)\,\mathrm{Id}_{\wedge(P^*)} \end{aligned} \tag{2.55}$$

hold for all $x, x' \in P^*$ and $y, y' \in P$. □

Using the result (2.55) we do the following calculation:

$$\begin{aligned} &v \cdot (w \cdot \xi) + w \cdot (v \cdot \xi) - 2Q(v, w) \cdot \xi \\ &= \big(\varepsilon(v_{P^*}) + \iota(v_P)\big)\big(\varepsilon(w_{P^*}) + \iota(w_P)\big)\xi + \big(\varepsilon(w_{P^*}) + \iota(w_P)\big)\big(\varepsilon(v_{P^*}) + \iota(v_P)\big)\xi - 2Q(v, w)\xi \\ &= \varepsilon(v_{P^*})\iota(w_P)\xi + \iota(w_P)\varepsilon(v_{P^*})\xi + \varepsilon(w_{P^*})\iota(v_P)\xi + \iota(v_P)\varepsilon(w_{P^*})\xi - 2Q(v, w)\xi \\ &= 2Q(v_{P^*}, w_P)\xi + 2Q(v_P, w_{P^*})\xi - 2Q(v, w)\xi = 0 \,. \end{aligned}$$

Thus the condition (2.48) is indeed satisfied and our Claim is true. What we have learned is summarized in the next statement — where we keep the assumption of even dimension of V but drop the (unnecessary) condition that Q be a Euclidean structure.

Definition. Let (V, Q) be an even-dimensional vector space over the reals $\mathbb{R}$ with polarization $V \otimes_{\mathbb{R}} \mathbb{C} = P \oplus P^*$. The spinor representation of the Clifford algebra $\mathrm{Cl}(V, Q)$ is defined to be the representation on the exterior algebra $\wedge(P^*)$ which is given by the action (2.54). □

It is now straightforward to produce a matrix representation of $\mathrm{Cl}(V, Q)$ by fixing a basis of $\wedge(P^*)$ and expanding the action of the Clifford algebra elements w.r.t. this basis.

Example. We illustrate the procedure at the simple example of the Euclidean plane $V = \mathbb{R}^2$ with orthonormal basis $\{e, f\}$. As before, let $c := (e - \mathrm{i}f)/2$ and $c^* := (e + \mathrm{i}f)/2$. The action of c^* and c on the basis $\{1, c^*\}$ of $\wedge(P^*)$ is computed to be

$$c^* \cdot 1 \equiv \varepsilon(c^*)1 = c^* \ , \quad c^* \cdot c^* \equiv \varepsilon(c^*)c^* = (c^*)^2 = 0 \ ,$$

and

$$c \cdot 1 \equiv \iota(c)1 = 0 \ , \quad c \cdot c^* \equiv \iota(c)c^* = 2Q(c, c^*) = \tfrac{1}{2}Q(e - \mathrm{i}f, e + \mathrm{i}f) = 1 \ .$$

Thus if we make the following identifications for the basis vectors $\{1, c^*\}$ of $\wedge(P^*)$:

$$1 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ , \quad c^* \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ ,$$

then the Clifford algebra elements 1, e, f, and ef are represented by the matrices

$$\begin{aligned} \mathrm{matrix}(1) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \\ \mathrm{matrix}(e) = \mathrm{matrix}(c + c^*) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_1 \ , \\ \mathrm{matrix}(f) = \mathrm{matrix}(\mathrm{i}c - \mathrm{i}c^*) &= \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} = \sigma_2 \ , \\ \mathrm{matrix}(ef) &= \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix} = \mathrm{i}\sigma_3 \ . \end{aligned}$$

Problem. By following the same procedure, construct the spinor representation of the Clifford algebra $\mathrm{Cl}(V, Q)$ for the Euclidean vector space $V = \mathbb{R}^4$ with orthonormal basis $\{e_1, e_2, f_1, f_2\}$. □

We finish this subsection with a few final comments:

1. The same construction applied to the Lorentzian vector space $V = \mathbb{R}^4$ with Minkowski scalar product gives the so-called Weyl form of the gamma matrices:

$$\gamma^0 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \ , \quad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix} \qquad (j = 1, 2, 3). \tag{2.56}$$

2. The spinor representation of $\mathrm{Cl}(\mathbb{R}^{2n}, Q)$ is the one and only representation of $\mathrm{Cl}(\mathbb{R}^{2n}, Q)$ which is both non-trivial and irreducible.

3. We recall the dimension formulas $\dim \mathrm{Cl}(V, Q) = 2^{\dim V}$ and $\dim \wedge (P^*) = 2^{\frac{1}{2}\dim V}$. For the case of four dimensions ($V = \mathbb{R}^4$) it so happens that the dimension $2^{\frac{1}{2}\dim V} = 2^2$ of the spinor representation is also four. This coincidence of dimensions $4 = 4$ is an accident.

2.8.1 Tensor product

In the construction of the spinor representation we used the notion of tensor product, which may be unfamiliar to the students of this course. Therefore we add a few words of explanation.

We assume that the notion of direct product (of sets, groups, vector spaces, manifolds, etc.) is understood. Now in the case of two $\mathbb{K}$-vector spaces V and W one has a variant of the direct product $V \times W$ which is called the tensor product (over $\mathbb{K}$) and is denoted by $V \otimes_{\mathbb{K}} W \equiv V \otimes W$. The tensor product is distinguished by the feature that the scalars $k \in \mathbb{K}$ move between factors:

$$kv \otimes w = v \otimes kw \ , \tag{2.57}$$

whereas $(kv, w) \in V \times W$ is not the same as $(v, kw) \in V \times W$.

Example. If V, W are vector spaces over $\mathbb{K}$, then $\mathrm{Hom}(V, W)$ denotes the $\mathbb{K}$-vector space of linear mappings from V to W. Recall from Section 1.7.1 that V^* denotes the vector space dual to V. There is a canonical isomorphism

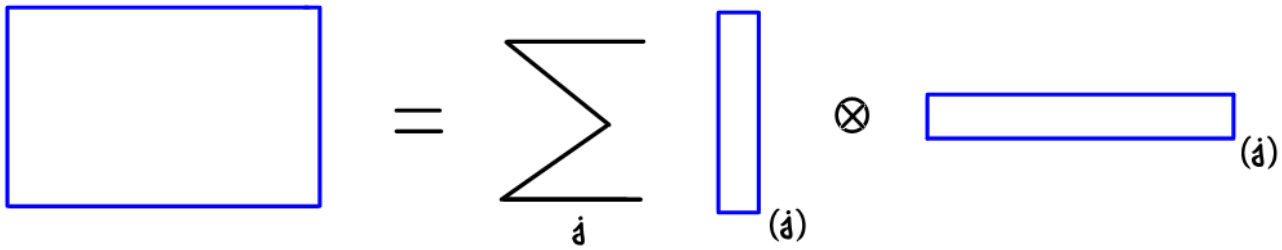
$$I \ : \ W \otimes V^* \to \mathrm{Hom}(V, W) \ ,$$

which is given by

$$I(w \otimes f)(v) := f(v)\, w \ .$$

Note that $I(kw \otimes f)(v) = f(v)\, kw = k\, f(v)\, w = I(w \otimes kf)(v)$, so the tensor product $W \otimes V^*$ (not the direct product $W \times V^*$) in fact is the proper product to appear in this isomorphism.

A corollary of this isomorphism is the fact that every matrix of size $m \times n$ can be decomposed uniquely (up to an ambiguity due to scalar factors moving in the tensor product) as a sum of terms each of which is the tensor product (sometimes called a 'dyadic' product in physics textbooks) of an m-component column vector with an n-component row vector:



2.9 Transforming scalar and vector fields

Our next goal is to describe the precise sense in which the Dirac equation is invariant under Lorentz transformations, i.e., has the same form in all inertial frames. As a preparation, we cover some basic ground concerning transformation behavior under group actions.

So let G be a group, and let M and N be G-spaces, which is to say that M and N are spaces which carry an action of the group G. Thus there exist products $G \times M \to M$, $(g, x) \mapsto g \cdot x$ and $G \times N \to N$, $(g, y) \mapsto g \cdot y$, such that for all $g, h \in G$, $x \in M$, and $y \in N$, we have

$$(gh) \cdot x = g \cdot (h \cdot x) \ , \quad (gh) \cdot y = g \cdot (h \cdot y) \ . \tag{2.58}$$

Let now f be a mapping from M to N. Given the G-actions on M and N, there exists an induced action of G on such mappings $f : M \to N$ by

$$(g \bullet f)(x) := g \cdot f(g^{-1} \cdot x) \, . \tag{2.59}$$

Problem. Show that one has $(gh) \bullet f = g \bullet (h \bullet f)$. □

Now recall from special relativity that the Lorentz group $G = \mathrm{SO}_{1,3}$ acts on space-time vectors $v \in \mathbb{R}^4$ by linear transformations $v \mapsto g \cdot v$ which preserve the Minkowski scalar product. A (complex) scalar field is a function $f : \mathbb{R}^4 \to \mathbb{C}$. The Lorentz group acts on it by

$$(g \bullet f)(v) = f(g^{-1} \cdot v) \, . \tag{2.60}$$

A (space-time) vector field is a mapping $X : \mathbb{R}^4 \to \mathbb{R}^4$. The Lorentz group acts on it by

$$(g \bullet X)(v) = g \cdot X(g^{-1} \cdot v) \, . \tag{2.61}$$

The wave function of the Dirac equation is neither a scalar nor a vector field but a spinor field $\psi : \mathbb{R}^4 \to \wedge(P^*)$, $P = \mathbb{C}^2$. In order to establish the relativistic covariance of the Dirac equation, we will have to understand how the Lorentz group acts on spinors $\xi \in \wedge(\mathbb{C}^2)$.

2.10 Infinitesimal spin transformations

Further on, we will learn that there is a surprising twist to this story: the Lorentz group does not act on spinors ; what does act is a close cousin, the so-called spin group. We again begin with some basic material — this time from Lie theory. The goal here is to introduce the infinitesimal transformations which generate the spin group.

Definition. A Lie algebra over $\mathbb{K}$ is a $\mathbb{K}$-vector space, say $\mathfrak{g}$, equipped with a skew-symmetric bilinear product (called the Lie bracket)

$$\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \, , \quad (X, Y) \mapsto [X, Y] = -[Y, X] \, , \tag{2.62}$$

which satisfies the so-called Jacobi identity,

$$\forall X, Y, Z \in \mathfrak{g} : \quad [X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]] \, . \tag{2.63}$$

Remark. Every associative algebra can be given the structure of a Lie algebra by taking the Lie bracket to be the commutator: $[X, Y] := XY - YX$. The Jacobi identity is automatically satisfied in this case.

Example. The Euclidean vector space $\mathbb{R}^3$ equipped with the standard vector product (or cross product)

$$\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \, , \quad (v, v') \mapsto v \times v' = -v' \times v \, ,$$

is a Lie algebra isomorphic to $\mathfrak{so}_3$, the Lie algebra of the group of proper rotations of $\mathbb{R}^3$. The Jacobi identity in this case reads

$$v \times (v' \times v'') = (v \times v') \times v'' + v' \times (v \times v'') \, .$$

Next we recall that a Lie group G is a group with the structure of a differentiable manifold such that the product $G\times G\to G$ is a differentiable mapping. Every Lie group G comes with a Lie algebra $\mathfrak{g}$, which has the geometric meaning of tangent space at the neutral element of G. In the cases of interest to us, one gets for each $X\in\mathfrak{g}$ a one-parameter subgroup of G by exponentiation, $\mathbb{R}\ni t\mapsto \mathrm{e}^{tX}$. The Lie bracket of $\mathfrak{g}$ then results from the differentiable product of G by

$$[X,Y]:=\frac{\partial^2}{\partial s\partial t}\,\mathrm{e}^{sX}\mathrm{e}^{tY}\mathrm{e}^{-sX}\mathrm{e}^{-tY}\Big|_{s=t=0}\,. \tag{2.64}$$

Definition. If V is a $\mathbb{K}$-vector space with quadratic form Q (here non-degeneracy is important), one defines the orthogonal group $\mathrm{O}(V,Q)$ as the Lie group of $\mathbb{K}$-linear transformations

$$\mathrm{O}(V,Q)=\{g\in\mathrm{End}(V)\mid \forall v,v'\in V:\ Q(gv,gv')=Q(v,v')\} \tag{2.65}$$

which leave Q invariant. The special orthogonal group $\mathrm{SO}(V,Q)\subset\mathrm{O}(V,Q)$ is the connected component containing the neutral element.

Examples. i) For $V=\mathbb{R}^3$ with Euclidean scalar product Q one has $\mathrm{SO}(V,Q)=\mathrm{SO}_3$, the group of proper rotations of the Euclidean vector space $\mathbb{R}^3$. ii) If $V=\mathbb{R}^4$ is equipped with Minkowski scalar product Q, then $\mathrm{SO}(V,Q)$ is the group of Lorentz transformations of V. $\square$.

We now consider the Lie algebra $\mathfrak{so}(V,Q)$ of the Lie group $\mathrm{SO}(V,Q)$. By setting $g_t:=\mathrm{e}^{tX}$ and linearizing the condition $Q(g_tv,g_tv')=Q(v,v')$ at the neutral element $t=0$ we get the following characterization of $\mathfrak{so}(V,Q)$:

$$\mathfrak{so}(V,Q)=\{X\in\mathrm{End}(V)\mid \forall v,v'\in V:\ Q(Xv,v')+Q(v,Xv')=0\}\,. \tag{2.66}$$

Problem. Verify from this definition that $\mathfrak{so}(V,Q)$ is closed under the commutator, i.e., if both X and Y are in $\mathfrak{so}(V,Q)$, then so is $[X,Y]:=XY-YX$. $\square$

The elements of the Lie algebra $\mathfrak{so}(V,Q)$ are skew-symmetric in the following sense. Using that Q is non-degenerate, one has an isomorphism $I:\ V\to V^*$ by $I(v)=Q(v,\cdot)$. One then associates with every element $X\in\mathfrak{so}(V,Q)$ a linear mapping $\tilde{X}:\ V\to V^*$ by $\tilde{X}:=I\circ X$. By the condition (2.66) on X this map is skew: $\tilde{X}^t=-\tilde{X}$.

In the case of the Euclidean vector space $V=\mathbb{R}^3$ with Cartesian basis $\{e_1,e_2,e_3\}$ the matrix of $X\in\mathfrak{so}(V,Q)=\mathfrak{so}_3$ has the form

$$\mathrm{matrix}(X)=\begin{pmatrix}0&X_{12}&X_{13}\\-X_{12}&0&X_{23}\\-X_{13}&-X_{23}&0\end{pmatrix}=\mathrm{matrix}(\tilde{X})\,.$$

In the case of the Lorentzian vector space $V=\mathbb{R}^4$ with Minkowski scalar product Q, the matrices of $X\in\mathfrak{so}(V,Q)$ and $\tilde{X}\in\mathrm{Hom}(V,V^*)$ with respect to a standard basis $\{e_0,e_1,e_2,e_3\}$ look as follows:

$$\mathrm{matrix}(X)=\begin{pmatrix}0&X_{01}&X_{02}&X_{03}\\X_{01}&0&X_{12}&X_{13}\\X_{02}&-X_{12}&0&X_{23}\\X_{03}&-X_{13}&-X_{23}&0\end{pmatrix}\,,\quad \mathrm{matrix}(\tilde{X})=\begin{pmatrix}0&X_{01}&X_{02}&X_{03}\\-X_{01}&0&-X_{12}&-X_{13}\\-X_{02}&X_{12}&0&-X_{23}\\-X_{03}&X_{13}&X_{23}&0\end{pmatrix}\,.$$

We now return to considering the Clifford algebra $\mathrm{Cl}(V,Q)$ and in it the subspace $\mathrm{Cl}_2(V,Q)$ of skew-symmetrized elements of degree two:

$$\mathrm{Cl}_2(V,Q) := \{X \in \mathrm{Cl}(V,Q) \mid X = \sum\nolimits_j (u_j v_j - v_j u_j)\,;\ u_j\,, v_j \in V\}\ . \tag{2.67}$$

Problem. Show that $\mathrm{Cl}_2(V,Q)$ is closed under the commutator. □

Thus $\mathrm{Cl}_2(V,Q)$ has the structure of a Lie algebra. Which Lie algebra?

Fact. $\mathrm{Cl}_2(V,Q)$ is isomorphic as a Lie algebra to $\mathfrak{so}(V,Q)$.

Proof. First of all, notice that $\mathrm{Cl}_2(V,Q)$ and $\mathfrak{so}(V,Q)$ are isomorphic as vector spaces:

$$\dim \mathrm{Cl}_2(V,Q) = \tfrac{1}{2}\dim V\,(\dim V - 1) = \dim \mathfrak{so}(V,Q)\ . \tag{2.68}$$

We will now proceed by explicitly constructing the claimed Lie algebra isomorphism. For this purpose we observe that the commutator of $X = \sum(u_j w_j - w_j u_j) \in \mathrm{Cl}_2(V,Q)$ with $v \in V \subset \mathrm{Cl}(V,Q)$ is another element $[X,v] \in V \subset \mathrm{Cl}(V,Q)$:

$$[X,v] = 2\sum\nolimits_j [u_j w_j\,, v] = 4\sum\nolimits_j \big(Q(w_j\,,v)\,u_j - Q(u_j\,,v)\,w_j\big) \in V\ .$$

We thus have a linear mapping

$$\tau\ :\ \mathrm{Cl}_2(V,Q) \to \mathrm{End}(V)\ , \quad X \mapsto \tau(X) := [X,\,\cdot\,]\ , \tag{2.69}$$

from elements of $\mathrm{Cl}_2(V,Q)$ to linear transformations of V. This mapping is a representation of the Lie algebra $\mathrm{Cl}_2(V,Q)$. Indeed,

$$\tau(X)\tau(Y)v - \tau(Y)\tau(X)v = [X,[Y,v]] - [Y,[X,v]] = [[X,Y],v] = \tau([X,Y])v\ .$$

Now let $\{u,v\} := uv + vu$ denote the anti-commutator and do the following calculation:

$$\begin{aligned} Q(\tau(X)u,v) + Q(u,\tau(X)v) &= Q([X,u],v) + Q(u,[X,v]) \\ &= \tfrac{1}{2}\{[X,u],v\} + \tfrac{1}{2}\{u,[X,v]\} \overset{\text{Jacobi}}{=} \tfrac{1}{2}[X,\{u,v\}] = 0\ . \end{aligned}$$

In the next to last equality we used a kind of generalized Jacobi identity, which is easily checked to be correct. Thus we see that the condition (2.66) for $\tau(X)$ to be in $\mathfrak{so}(V,Q)$ is satisfied.

Now the representation $\tau\ : \mathrm{Cl}_2(V,Q) \to \mathfrak{so}(V,Q)$ is injective (we leave the verification of this statement as a problem for the student). By the equality (2.68) of dimensions it follows that τ is bijective and hence an isomorphism of Lie algebras. □

Summary. The subspace $\mathrm{Cl}_2(V,Q)$ of skew-symmetrized degree-two elements of the Clifford algebra $\mathrm{Cl}(V,Q)$ has the structure of a Lie algebra. We have two representations for it:

1. As a subspace of $\mathrm{Cl}(V,Q)$ the Lie algebra $\mathrm{Cl}_2(V,Q)$ acts on spinors $\xi \in \wedge(P^*)$ by the spinor representation. This representation has dimension $2^{\frac{1}{2}\dim V}$.

2. Via the isomorphism τ the Lie algebra $\mathrm{Cl}_2(V,Q)$ acts on vectors $v \in V$ by the fundamental (or defining) representation of $\mathfrak{so}(V,Q)$. The dimension of this representation is $\dim V$.

2.11 Spin group

By exponentiating a Lie algebra $\mathfrak{g}$ (which sits inside an associative algebra $\mathcal{A}$ or acts on some representation space R by linear operators or matrices, so that products such as X^n for $X \in \mathfrak{g}$ make sense) one gets a Lie group G :

$$\mathfrak{g} \xrightarrow{\exp} G\,, \quad X \mapsto \mathrm{e}^X := 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \ldots + \frac{X^n}{n!} + \ldots\,.$$

(To ensure that the exponential series converges, one poses the requirement that the associative algebra $\mathcal{A}$ or the representation space R be finite-dimensional, so that the Lie algebra elements are represented by matrices of finite size.) In order to verify that exponentiation of a Lie algebra really does yield a group, one argues on the basis of the so-called Baker-Campbell-Hausdorff series:

$$\mathrm{e}^X \mathrm{e}^Y = \mathrm{e}^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]+\frac{1}{12}[[X,Y],Y]+\ldots}, \tag{2.70}$$

which suggests that the product $\mathrm{e}^X \mathrm{e}^Y$ of two exponentiated Lie algebra elements is the exponential of another Lie algebra element $X+Y+\frac{1}{2}[X,Y]+\ldots$. (There indeed exists a recursive construction by which the infinite Baker-Campbell-Hausdorff series can be shown to exist and converge in the finite-dimensional case.)

Example. By exponentiating the Lie algebra $\mathfrak{so}(V,Q)$ in the associative algebra $\mathrm{End}(V)$ we get the Lie group $\mathrm{SO}(V,Q)$.

Definition. The Lie group obtained by exponentiating the Lie algebra $\mathrm{Cl}_2(V,Q)$ in the Clifford algebra is called the spin group $\mathrm{Spin}(V,Q)$. The representation which results from letting $\mathrm{Spin}(V,Q) \subset \mathrm{Cl}(V,Q)$ act on $\wedge(P^*)$ is called the spinor representation of $\mathrm{Spin}(V,Q)$.

Remark. Since $\mathrm{Cl}_2(V,Q)$ is isomorphic as a Lie algebra to $\mathfrak{so}(V,Q)$, one might think that $\mathrm{Spin}(V,Q)$ would be isomorphic as a Lie group to $\mathrm{SO}(V,Q)$. However, as we shall see, this is not the case. □

We now show for the case of even-dimensional $V = \mathbb{R}^{2n}$ and Euclidean Q that the group $\mathrm{Spin}(V,Q)$ contains an element $g = -1$. For this we fix any pair $e, f \in V$ of orthogonal unit vectors and consider the element $X := \frac{1}{2}(ef - fe) \in \mathrm{Cl}(V,Q)$. For any parameter $t \in \mathbb{C}$ we have

$$\mathrm{e}^{t\,(ef-fe)/2} = \mathrm{e}^{tX} = \cosh(tX) + \sinh(tX)\,.$$

We then compute the square of X :

$$X^2 = \tfrac{1}{4}(ef-fe)^2 = \tfrac{1}{4}(efef + fefe - effe - feef) = -1\,,$$

where we used the Clifford relations $e^2 = ee = Q(e,e) = 1$, $f^2 = 1$, and $ef = -fe$. Hence,

$$\mathrm{e}^{tX} = \cosh(tX) + \sinh(tX) = \cos(t) + X\sin(t)\,.$$

Thus, in particular,

$$\mathrm{e}^{\pi X} = -1\,.$$

It follows (for Euclidean $V = \mathbb{R}^{2n}$) that if g is an element of $\mathrm{Spin}(V,Q)$, then so is $-g$.

Next we consider the action of $g = \mathrm{e}^X \in \mathrm{Spin}(V,Q) \subset \mathrm{Cl}(V,Q)$ on $v \in V \subset \mathrm{Cl}(V,Q)$ by conjugation in the Clifford algebra:

$$\mathrm{e}^X v\, \mathrm{e}^{-X} = v + [X,v] + \frac{1}{2!}[X,[X,v]] + \frac{1}{3!}[X,[X,[X,v]]] + \ldots = \sum_{l=0}^{\infty} \frac{\tau(X)^l}{l!}\, v = \mathrm{e}^{\tau(X)} v.$$

Note that since $\tau(X)$ is in $\mathfrak{so}(V,Q)$, the exponential $\mathrm{e}^{\tau(X)}$ is an element of the special orthogonal group $\mathrm{SO}(V,Q)$. Thus we have a mapping

$$\rho \,:\, \mathrm{Spin}(V,Q) \to \mathrm{SO}(V,Q)\,, \quad g \mapsto \rho(g)\,, \quad \rho(g)v := gvg^{-1}\,. \tag{2.71}$$

This mapping is a representation of $\mathrm{Spin}(V,Q)$ on the vector space V. The situation is summarized in the following commutative diagram (commutative here says that $\exp \circ \tau = \rho \circ \exp$):

$$\begin{array}{ccc} \mathrm{Cl}_2(V,Q) & \stackrel{\exp}{\longrightarrow} & \mathrm{Spin}(V,Q) \\ \tau \big\downarrow & & \big\downarrow \rho \\ \mathfrak{so}(V,Q) & \stackrel{\exp}{\longrightarrow} & \mathrm{SO}(V,Q). \end{array}$$

Remark. The representation $\rho : \mathrm{Spin}(V,Q) \to \mathrm{SO}(V,Q)$ is not a bijection, but a so-called $2:1$ covering. In the case of Euclidean $V = \mathbb{R}^{2n}$ this is plausible from $\rho(-g) = \rho(g)$.

Problem. In the special case of Euclidean $V = \mathbb{R}^{2n}$ with orthonormal basis $\{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ show for $X = \frac{1}{4}\sum (e_j f_j - f_j e_j)$ that

$$\rho(\mathrm{e}^{tX})\, e_j = e_j \cos t - f_j \sin t\,, \quad \rho(\mathrm{e}^{tX})\, f_j = e_j \sin t + f_j \cos t\,.$$

2.12 Relativistic covariance of the Dirac equation

We focus now on our case of interest: the Lorentzian vector space $V = \mathbb{R}^4$ with Minkowski scalar product Q. To simplify the notation we write $\mathrm{SO}_{1,3} := \mathrm{SO}(V,Q)$ and $\mathrm{Spin}_{1,3} := \mathrm{Spin}(V,Q)$. We recall that we have met two representations of the spin group: the spinor representation, say

$$\sigma \,:\, \mathrm{Spin}_{1,3} \to \mathrm{GL}\big(\textstyle\bigwedge(P^*)\big) \simeq \mathrm{GL}_4(\mathbb{C})\,, \tag{2.72}$$

by invertible transformations (GL = general linear group) of the exterior algebra $\bigwedge(P^*) \simeq \bigwedge(\mathbb{C}^2) \simeq \mathbb{C}^4$, and the vector representation

$$\rho \,:\, \mathrm{Spin}_{1,3} \to \mathrm{SO}_{1,3}\,. \tag{2.73}$$

If $\{e_0, e_1, e_2, e_3\}$ is a standard basis of $V = \mathbb{R}^4$ and $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are the corresponding Clifford algebra elements in the spinor representation, the two representations are related by

$$\sigma(g)\gamma_\mu \sigma(g)^{-1} = \rho(g)\gamma_\mu = \gamma_\nu\, \rho(g)^\nu{}_\mu\,. \tag{2.74}$$

The gamma matrices γ^μ appearing in the Dirac equation differ from the γ_μ by the position of the index, which is raised or lowered by using the Minkowski metric: $\gamma_\mu = Q(e_\mu, e_\nu)\gamma^\nu =: Q_{\mu\nu}\gamma^\nu$ and $\gamma^\mu = Q^{\mu\nu}\gamma_\nu$ where $Q^{\mu\nu}Q_{\nu\lambda} = \delta^\mu_\lambda$. We therefore have

$$\sigma(g)\gamma^\mu\sigma(g)^{-1} = \rho(g)\gamma^\mu = Q^{\mu\nu}\rho(g)\gamma_\nu = Q^{\mu\nu}Q_{\lambda\tau}\gamma^\tau \rho(g)^\lambda{}_\nu = \rho(g^{-1})^\mu{}_\nu \gamma^\nu. \tag{2.75}$$

We now introduce a notion similar to that of scalar field and vector field.

Definition. A spinor field is a mapping $\psi : \mathbb{R}^4 \to \wedge(P^*) \simeq \mathbb{C}^4$ from space-time (viewed as a vector space after fixing a space-time origin) into the spinor representation space. It transforms under the spin group as

$$(g \cdot \psi)(v) := \sigma(g)\psi(\rho(g)^{-1}v)\,, \quad (g \in \mathrm{Spin}_{1,3}\,,\ v \in \mathbb{R}^4). \quad \square \tag{2.76}$$

The next statement has the implication that the wave function ψ of the Dirac equation is to be viewed as a spinor field.

Proposition. If ψ is a solution of the Dirac equation with gauge potential $A = A_\mu\, dx^\mu$, then $g \cdot \psi$ for $g \in \mathrm{Spin}_{1,3}$ is a solution of the same equation with transformed gauge potential

$$(g \cdot A)_\nu(v) = A_\mu(\rho(g)^{-1}v)\,\rho(g^{-1})^\mu{}_\nu. \tag{2.77}$$

Proof. Let ψ be a solution of

$$\gamma^\mu \left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x^\mu} - eA_\mu\right)\psi + mc\,\psi = 0\,. \tag{2.78}$$

We multiply the equation from the left by $\sigma(g)$ for $g \in \mathrm{Spin}_{1,3}$ and use (2.75) to obtain

$$\rho(g^{-1})^\mu{}_\nu\gamma^\nu \left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x^\mu} - eA_\mu\right)\sigma(g)\psi + mc\,\sigma(g)\psi = 0\,. \tag{2.79}$$

It remains to transform the arguments of ψ and A_μ from v to $\rho(g)^{-1}v$. For this step we use that

$$\mathrm{e}^{-\omega^\mu{}_\nu\, x^\nu \partial_\mu}\,\partial_\lambda\,\mathrm{e}^{\omega^\mu{}_\nu\, x^\nu\partial_\mu} = \partial_\lambda + \omega^\mu{}_\lambda\,\partial_\mu + \ldots = (\mathrm{e}^\omega)^\mu{}_\lambda\,\partial_\mu\,, \qquad \partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \quad \omega^\mu{}_\nu \in \mathbb{R}.$$

This formula shows that a finite rotation $v \mapsto \mathrm{e}^\omega v$ (generated by an infinitesimal rotation ω with matrix elements $\omega^\mu{}_\nu$) is represented on functions by the operator $\mathrm{e}^{-\omega^\mu{}_\nu\, x^\nu\partial_\mu}$. Indeed, applying the latter to a function $v \mapsto f(v)$ we get

$$\big(\mathrm{e}^{-\omega^\mu{}_\nu\, x^\nu\partial_\mu} f\big)(v) = f(\mathrm{e}^{-\omega}\,v) = (\mathrm{e}^\omega \cdot f)(v)\,. \tag{2.80}$$

Now we apply that very operator $\mathrm{e}^{-\omega^\mu{}_\nu\, x^\nu\partial_\mu}$ for $\mathrm{e}^\omega \equiv \rho(g)$ to Eq. (2.79). This gives

$$\rho(g^{-1})^\mu{}_\nu\gamma^\nu \left(\frac{\hbar}{\mathrm{i}}\,\rho(g)^\lambda{}_\mu\,\frac{\partial}{\partial x^\lambda} - eA'_\mu\right)(g\cdot\psi) + mc\,(g\cdot\psi) = 0\,, \tag{2.81}$$

where $A'_\mu(v) = A_\mu(\rho(g)^{-1}v)$. Since $\rho(g)^\lambda{}_\mu\,\rho(g^{-1})^\mu{}_\nu = \rho(gg^{-1})^\lambda{}_\nu = \delta^\lambda_\nu$ we arrive at

$$\gamma^\mu \left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x^\mu} - e\,(g\cdot A)_\mu\right)(g\cdot\psi) + mc\,(g\cdot\psi) = 0\,, \tag{2.82}$$

which is the desired result.

Remark. What we have shown here is that the Dirac equation is covariant under an "active" transformation by $g \in \mathrm{Spin}(V,Q)$. Equivalently, we could have demonstrated its invariance under the corresponding coordinate transformations (also known as "passive" transformations).

2.13 Discrete symmetries of the Dirac equation

2.13.1 Pin group and parity transformation

Both the special orthogonal group $\mathrm{SO}(V,Q)$ and the spin group $\mathrm{Spin}(V,Q)$ are connected. (In fact, Spin has the additional property of being simply connected, which means that every closed path in it is contractible to a point; the latter is not the case for SO.) However the full orthogonal group $\mathrm{O}(V,Q)$ has more than one connected component. In particular, the operation of space reflection, while not contained in $\mathrm{SO}(V,Q)$, does belong to $\mathrm{O}(V,Q)$ for our case of $V=\mathbb{R}^4$ with Minkowski scalar product Q. One may then ask whether there exists a group $\mathrm{Pin}(V,Q)$ which is related to $\mathrm{Spin}(V,Q)$ in the same way that $\mathrm{O}(V,Q)$ is related to $\mathrm{SO}(V,Q)$. The answer, as it turns out, depends on whether the quadratic form Q is definite or indefinite. In the definite (or Euclidean) case, the answer is yes; otherwise it is only partially yes. We now briefly highlight a few salient points in order to motivate the form of the parity operator for the Dirac equation.

The Clifford algebra comes with an automorphism $\alpha:\mathrm{Cl}(V,Q)\to\mathrm{Cl}(V,Q)$ by the $\mathbb{Z}_2$-degree: $\alpha(x)=x$ if x is even, and $\alpha(x)=-x$ if x is odd. The Clifford algebra also comes with an anti-automorphism called the transpose; this is defined by $1=1^t$, $v^t=v$ for $v\in V$, and

$$(xy)^t := y^t x^t\,.$$

Definition. The Clifford group $\Gamma(V,Q)$ is defined as the group of invertible elements $x\in\mathrm{Cl}(V,Q)$ with the property that twisted conjugation

$$\mathrm{Cl}(V,Q)\supset V\ni v\mapsto xv\,\alpha(x)^{-1}$$

stabilizes V, i.e., maps V into itself. The group $\mathrm{Pin}(V,Q)\subset\Gamma(V,Q)$ is the subgroup defined as

$$\mathrm{Pin}(V,Q)=\{x\in\Gamma(V,Q)\mid x^t x=\pm 1\}\,. \tag{2.83}$$

Problem. Check that $\mathrm{Pin}(V,Q)$ contains $\mathrm{Spin}(V,Q)$. □

In the case of the Lorentzian vector space $\mathbb{R}^{1,3}$ (a short-hand notation for $V=\mathbb{R}^4$ with Minkowski scalar product Q) any space-like or time-like unit vector $u=e_\mu$ is in $\mathrm{Pin}_{1,3}\equiv\mathrm{Pin}(V,Q)$. Indeed, $u^t u=e_\mu e_\mu=Q(e_\mu,e_\mu)=\pm1$, and

$$uv\,\alpha(u)^{-1}=uv\frac{-u}{Q(u,u)}=v-2u\frac{Q(u,v)}{Q(u,u)}\in V.$$

We see that the operation $v\mapsto uv\,\alpha(u)^{-1}$ is a reflection at the hyperplane orthogonal to u.

Problem. For $g\in\mathrm{Pin}(V,Q)$ show that $\rho(g):V\to V$ defined by $\rho(g)v=gv\,\alpha(g)^{-1}$ is an orthogonal transformation, $\rho(g)\in\mathrm{O}(V,Q)$. □

Consider now the pin group element $g=e_1e_2e_3$ for a Cartesian basis $\{e_1,e_2,e_3\}$ of a three-dimensional space $\mathbb{R}^3\subset\mathbb{R}^{1,3}$. The corresponding element $\rho(g)\in\mathrm{O}_{1,3}$ is the sequence of three reflections at the hyperplanes orthogonal to e_1, e_2, and e_3. The combined effect of these reflections

is a space reflection, i.e., the operation of inverting each of the three Cartesian coordinates. This operation is also called a 'parity transformation' in physics.

In the context of the Dirac equation one wants to work with conjugation $v \mapsto gvg^{-1}$ rather than twisted conjugation $gv\,\alpha(g)^{-1}$. To correct for the extra minus sign inflicted by twisted conjugation, we multiply $g = e_1e_2e_3$ by a pin group element x with the property that $xv\,x^{-1} = -v$ for all $v \in V$. If e_0 with $Q(e_0, e_j) = 0$ $(j = 1, 2, 3)$ is a time-like unit vector, the product $x = e_0e_1e_2e_3$ is such an element. Thus we arrive at $xg = e_0$. Since e_0 is represented in the spinor representation by γ^0, the next definition is well motivated.

Definition. The operator P of parity transformation acts on Dirac spinors ψ as

$$(P \cdot \psi)(\vec{r}, t) = \gamma^0 \psi(-\vec{r}, t). \tag{2.84}$$

We now look at what happens to the Dirac equation

$$\gamma^\mu \left(\frac{\hbar}{\mathrm{i}} \frac{\partial}{\partial x^\mu} - eA_\mu \right) \psi + mc\,\psi = 0 \tag{2.85}$$

under a parity transformation. By the chain rule of differentiation, reversal $\psi(\vec{r}, t) \mapsto \psi(-\vec{r}, t)$ of the space arguments sends the spatial derivatives $\partial/\partial x_j$ to their negatives. This sign change is compensated by conjugation $\gamma^j \mapsto \gamma^0\gamma^j\gamma^0 = -\gamma^j$ $(j = 1, 2, 3)$. We thus have the following

Fact. The Dirac equation in the absence of electromagnetic fields $(A_\mu = 0)$ is parity-invariant, i.e., if ψ is a solution, then so is $P \cdot \psi$.

Problem. Show that if ψ is a solution of the Dirac equation with gauge potential $A = A_0\,dx^0 + \sum A_j\,dx^j$, then $P \cdot \psi$ is a solution of the Dirac equation with transformed gauge potential

$$(P \cdot A)(\vec{r}, t) = A_0(-\vec{r}, t)\,dx^0 - \sum A_j(-\vec{r}, t)\,dx^j.$$

Warning. The standard procedure of constructing $\mathrm{Pin}(V, Q)$ from the real Clifford group $\Gamma(V, Q)$ has recently been called into question by B. Janssens, arXiv:1709.02742.

2.13.2 Anti-unitary symmetries of the Dirac equation

Going beyond parity, the Dirac equation has some discrete symmetries which do not come (not in a physically satisfactory manner anyway) from the pin group but require complex anti-linear operations. One of these is time reversal, T. In the standard representation (2.15) of the gamma matrices this operation is given by

$$(T \cdot \psi)(\vec{r}, t) = \gamma^1\gamma^3\,\overline{\psi(\vec{r}, -t)}\,. \tag{2.86}$$

Note that $T^2 = -1$ and

$$\gamma^1\gamma^3 = \begin{pmatrix} \mathrm{i}\sigma_2 & 0 \\ 0 & \mathrm{i}\sigma_2 \end{pmatrix}.$$

Remark. By non-relativistic reduction, this formula yields the time-reversal operator (1.32) for spin-1/2 Schrödinger particles. Note also that T normalizes $\mathrm{Spin}_{1,3}$ $(T\,\mathrm{Spin}_{1,3}\,T^{-1} = \mathrm{Spin}_{1,3})$. □

To investigate the behavior of the Dirac equation under time reversal, we take its complex conjugate and observe that γ^0, γ^1 and γ^3 are real matrices while γ^2 is imaginary (in standard repn). We also note that $t \mapsto -t$ sends $\partial_t \mapsto -\partial_t$ and that $\gamma^1\gamma^3\gamma^\mu(\gamma^1\gamma^3)^{-1} = \epsilon\,\gamma^\mu$ where $\epsilon = -1$ for $\mu = 1, 3$ and $\epsilon = 1$ for $\mu = 0, 2$. In this way we find that T transforms a solution ψ of the Dirac equation (2.85) into a solution of the same equation with transformed gauge potential

$$(T \cdot A)(\vec{r}, t) = -\left(-A_0(\vec{r}, -t)\, dx^0 + \sum A_j(\vec{r}, -t)\, dx^j\right).$$

The Dirac equation has another complex anti-linear symmetry, which is called charge conjugation. This operation C (again, in standard representation) is defined by

$$(C \cdot \psi)(\vec{r}, t) = \mathrm{i}\gamma^2\, \overline{\psi(\vec{r}, t)}\,. \tag{2.87}$$

The name derives from the easily verified fact that C transforms a solution ψ of the Dirac equation into a solution $C \cdot \psi$ of the same equation with opposite gauge field $A_\mu \to -A_\mu$ (or conjugated charge $e \to -e$). Notice that $C^2 = 1$, and $CP = -PC$, whereas $TP = PT$.

Warning. It turns out that the process of re-interpreting the Dirac spinor ψ as a quantum field converts the complex anti-linear operator C into a unitary (hence complex linear) symmetry. We will elaborate on this subtle point in the chapter on second quantization. □

Consider now the combination of charge conjugation, parity, and time reversal:

$$(CPT \cdot \psi)(\vec{r}, t) = \gamma_5\, \psi(-\vec{r}, -t)\,, \quad \gamma_5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3. \tag{2.88}$$

This combined operation is more fundamental than its individual factors in that it is defined even in the absence of a splitting $\mathbb{R}^{1,3} = \mathbb{R} \oplus \mathbb{R}^3$ into time and space. Since each factor is a symmetry of the Dirac equation, so is the product CPT. The precise statement is this: if ψ solves the Dirac equation with gauge field $A = A_\mu\, dx^\mu$, then $CPT \cdot \psi$ is a solution with transformed gauge field

$$(CPT \cdot A)_\mu(\vec{r}, t) = -A_\mu(-\vec{r}, -t)\,. \tag{2.89}$$

Remark. A deep result of theoretical physics is the so-called CPT theorem: any quantum field theory which is unitary (i.e., has a Hermitian Hamiltonian for its time-evolution generator), local (i.e., the Lagrangian contains only derivatives up to a finite order), and Lorentz-invariant, must have CPT as a symmetry. In other words: the combined operation of running time backwards, reflecting space, and replacing matter by anti-matter, must always be a symmetry of the fundamental laws of nature. Another remark is that the CPT theorem serves as important input to the proof of the spin-statistics theorem, which says that bosons and fermions have integer and half-integer spin, respectively.

2.13.3 Parity violation by the weak interaction

Although CPT is always a symmetry, it turns out that there do exist fundamental interactions of nature which break one or another of the individual factors. Here we briefly indicate how parity is broken by the weak interaction.

In the limit of zero mass ($m \to 0$), the Dirac equation for the 4-component spinor ψ splits into a pair of equations for two 2-component spinors ψ_L and ψ_R, which are called Weyl or half-spinors. This works as follows. Based on $\gamma_5^2 = 1$ for $\gamma_5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$, one introduces projection operators

$$\Pi_\pm = \tfrac{1}{2}(1 \pm \gamma_5)\,. \tag{2.90}$$

By using $\gamma^\mu \Pi_\pm = \Pi_\mp \gamma^\mu$, one infers from the Dirac equation (2.85) with mass $m = 0$ that

$$0 = \Pi_\pm \gamma^\mu \left(\frac{\hbar}{\mathrm{i}} \frac{\partial}{\partial x^\mu} - eA_\mu \right) \psi = \gamma^\mu \left(\frac{\hbar}{\mathrm{i}} \frac{\partial}{\partial x^\mu} - eA_\mu \right) \Pi_\mp \psi\,. \tag{2.91}$$

Thus both $\psi_L := \Pi_+\psi$ and $\psi_R := \Pi_-\psi$ are solutions of the massless Dirac equation.

To interpret the half-spinors $\psi_R\,, \psi_L$ it is best to choose the so-called Weyl representation for the gamma matrices:

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.92}$$

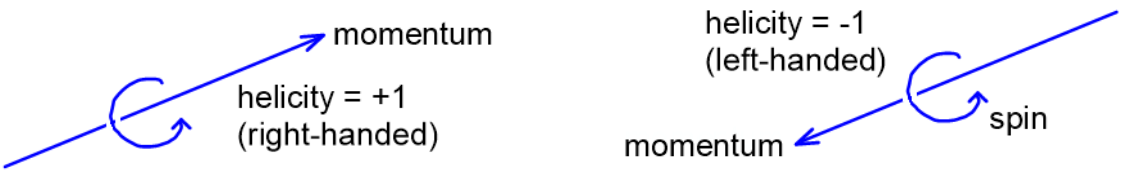
In this representation, (the non-zero part of) ψ_L consists of the upper two components of ψ and ψ_R consists of the lower two components. The decoupled pair of equations for $\psi_{L,R}$ then reads

$$\left(\partial_t + c \sum \sigma_j\, \partial_j\right) \psi_R = 0\,, \quad \left(\partial_t - c \sum \sigma_j\, \partial_j\right) \psi_L = 0 \tag{2.93}$$

in zero electromagnetic field. In order for $\psi_R = \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{x} - |k|ct)}\eta$ to be a solution of the first equation, the constant 2-component spinor η must be an eigenvector

$$\vec{\sigma}\cdot\vec{k}\;\eta = |k|\,\eta\,, \qquad \vec{\sigma}\cdot\vec{k} = \begin{pmatrix} k_3 & k_1 - \mathrm{i}k_2 \\ k_1 + \mathrm{i}k_2 & -k_3 \end{pmatrix},$$

of $\vec{\sigma}\cdot\vec{k}/|k|$ with eigenvalue (also referred to as helicity) $+1$. In such an eigenvector the momentum and spin angular momentum of the particle form a right-handed system (hence the notation ψ_R and the name helicity). Similarly, in order for $\psi_L = \mathrm{e}^{\mathrm{i}(\vec{k}\cdot\vec{x} - |k|ct)}\xi$ to be a solution of the second equation, the spinor ξ must be an eigenvector of $\vec{\sigma}\cdot\vec{k}/|k|$ with eigenvalue -1. In this case the momentum and the spin angular momentum form a left-handed system. Note that $\Pi_+ P = P\Pi_-$, so that $P\psi_L = \psi_R$ and $P\psi_R = \psi_L$, in agreement with the observation that space reflection transforms a right-handed system into a left-handed one and vice versa.



Experiments on β-decay done in the mid-1950's showed a strong right-left asymmetry of the radioactive decay process. This is understood nowadays (from the so-called Standard Model of elementary particle physics) as resulting from a peculiar feature of the electro-weak interaction: left-handed particles carry electro-weak charge, but right-handed particles do not – a clear violation of parity symmetry! More precisely, the left-handed component e_L of the electron constitutes a doublet of half-spinors $\begin{pmatrix} e_L \\ \nu_L \end{pmatrix}$ together with the electron neutrino, and an analogous doublet is

formed by the up and down quarks. These doublets are governed by the second equation of (2.93) with minimal substitution $\partial_\mu \to \partial_\mu - \mathcal{A}_\mu$ where $\mathcal{A}_\mu$, the electro-weak gauge potential, is a 2×2 matrix. The off-diagonal elements of $\mathcal{A}_\mu$ quantize as the $W^\pm$ boson, the diagonal elements as the Z^0 boson. The right-handed component of the electron only couples to the electromagnetic field.

2.14 Dirac electron in a Coulomb potential

A major application of non-relativistic quantum mechanics is the computation of the discrete energy spectrum of the hydrogen atom using the Schrödinger equation. In this section we carry out a similar computation based on the Dirac equation.

The first step is to choose a gauge. We take

$$A = A_\mu \, dx^\mu = A_0 \, dx^0 = -\Phi \, dt \,, \tag{2.94}$$

where $\Phi = Ze/(4\pi\varepsilon_0 r)$ is the Coulomb potential due to a positive nuclear charge Ze placed at the origin of the coordinate system. As usual, $r = \sqrt{x^2 + y^2 + z^2}$ denotes the distance from the origin. Adopting this gauge, the Hamiltonian for the negatively charged Dirac electron is

$$H = \frac{\hbar c}{\mathrm{i}} \sum_l \alpha_l \frac{\partial}{\partial x_l} + \beta mc^2 - \frac{Ze^2}{4\pi\varepsilon_0 \, r} \,. \tag{2.95}$$

We are going to look for bound-state solutions of the stationary Dirac equation

$$H\psi = E\psi. \tag{2.96}$$

Our strategy is the same as in the non-relativistic case: exploiting the spherical symmetry of the Coulomb potential, we will reduce $H\psi = E\psi$ to a radial differential equation. The solution of the latter equation will then be found as a power series.

Comment. By introducing $\lambda = \hbar/mc$ (reduced Compton wavelength) and $\alpha = e^2/(4\pi\varepsilon_0 \, \hbar c)$ (fine structure constant) the scaled Hamiltonian H/mc^2 is expressed in terms of dimensionless coordinates $\xi_l = x_l/r$ as

$$\frac{H}{mc^2} = \beta + \frac{1}{\mathrm{i}} \sum_l \alpha_l \frac{\partial}{\partial \xi_l} - \frac{Z\alpha}{|\xi|} \,.$$

We therefore expect the bound-state energy spectrum (in units of mc^2) to be a function of $Z\alpha$.

2.14.1 Using rotational symmetry

The Coulomb potential is invariant under space rotations fixing the origin (the nucleus). This invariance will allow us to find the solution in the form of a product ansatz (separation of variables).

To begin, we recall the action of a spin group element $g \in \mathrm{Spin}_{1,3}$ on the spinor field ψ :

$$(g \cdot \psi)(v) = \sigma(g)\psi(\rho(g)^{-1}v).$$

Here we want this transformation law for the special case of $g \in \mathrm{Spin}_3 \subset \mathrm{Spin}_{1,3}$, $\rho \in \mathrm{SO}_3$, and we only need it at the infinitesimal level. Hence we put $g_t = \mathrm{e}^{tX}$ and $\tau(X) = \omega$ and differentiate with respect to the parameter t at $t = 0$.

To compute the contribution from the rotation $v \mapsto \rho(g_t)^{-1}v$ of the argument of ψ, we use the formulas $\rho(g_t)^{-1} = \mathrm{e}^{-t\omega}$ and $f(\mathrm{e}^{-t\omega}\bullet) = \mathrm{e}^{-t\sum\omega_{kl}\,x_l\,\partial_k} f$ [see Eq. (2.80)],

$$\mathrm{e}^{-t\sum\omega_{kl}\,x_l\partial_k} = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}t\sum\omega_j L_j}\,, \quad \omega_{lk} = \sum\nolimits_j \epsilon_{jkl}\,\omega_j\,, \quad L_j = \frac{\hbar}{\mathrm{i}}\sum\nolimits_{kl}\epsilon_{jkl}\,x_k\partial_l = \sum\nolimits_{kl}\epsilon_{jkl}\,x_k p_l\,.$$

Here ϵ_{jkl} is the totally anti-symmetric tensor of $\mathbb{R}^3$, i.e., $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$, and the tensor vanishes if two indices coincide.

There is a second contribution to the derivative at $t = 0$; this comes from multiplication $\psi \mapsto \sigma(g_t)\psi = \sigma(\mathrm{e}^{tX})\psi$ by the spinor representation of g_t. To compute it, we use the formulas

$$X = \tau^{-1}(\omega) = \frac{1}{4}\sum\nolimits_{kl}\omega_{kl}\,\gamma_l\gamma_k = -\frac{\mathrm{i}}{2}\sum\nolimits_j \omega_j\Sigma_j\,, \quad \Sigma_j = \begin{pmatrix}\sigma_j & 0\\ 0 & \sigma_j\end{pmatrix},$$

where γ_j represents the Cartesian basis vector $e_j \in \mathbb{R}^3$ viewed as a generator of the Clifford algebra (in the spinor representation). Putting everything together, we obtain

$$\frac{d}{dt}\,\mathrm{e}^{\tau^{-1}(t\omega)}\cdot\psi\Big|_{t=0} = -\frac{\mathrm{i}}{\hbar}\sum\nolimits_l \omega_l J_l\,\psi, \quad J = L + S, \quad S_l = \frac{\hbar}{2}\Sigma_l\,. \tag{2.97}$$

The interpretation of this result is clear: the total angular momentum J is the sum of orbital angular momentum L and spin angular momentum S.

Problem. Show by direct calculation that the Dirac Hamiltonian (2.95) has vanishing commutator with each of the three components of the total angular momentum $J = L + S$. □

It follows that $[H, J^2] = 0$. We also have $[H, P] = 0$ by the parity symmetry of the Coulomb potential. We thus have a set of four commuting operators $\{H, J^2, J_3\,, P\}$, and we will be looking for joint eigenfunctions of this set. For this purpose, let

$$\varphi^l_{jm} := \big(Y_l \otimes \chi_{1/2}\big)_{jm} = a^l_{jm}\,Y_{l,\,m-1/2}\otimes\begin{pmatrix}1\\0\end{pmatrix} + b^l_{jm}\,Y_{l,\,m+1/2}\otimes\begin{pmatrix}0\\1\end{pmatrix} \tag{2.98}$$

be a 2-component angular momentum eigenfunction:

$$J^2\varphi^l_{jm} = \hbar^2 j(j+1)\varphi^l_{jm}\,, \quad J_3\,\varphi^l_{jm} = \hbar m\,\varphi^l_{jm}\,, \tag{2.99}$$

$$L^2\varphi^l_{jm} = \hbar^2 l(l+1)\varphi^l_{jm}\,, \quad S^2\varphi^l_{jm} = \hbar^2\tfrac{1}{2}(\tfrac{1}{2}+1)\varphi^l_{jm}\,. \tag{2.100}$$

(Note $m \in \mathbb{Z} + \frac{1}{2}$.) The exact details of this construction will be spelled out in the chapter on tensor invariants of the rotation group SO_3; here we just mention/anticipate that the coefficients a^l_{jm} and b^l_{jm} are so-called Clebsch-Gordan coefficients coupling orbital angular momentum (with quantum numbers $l\,m\pm\frac{1}{2}$) and spin angular momentum (with quantum numbers $\frac{1}{2}\ \pm\frac{1}{2}$) to total angular momentum (with quantum numbers $j\,m$):

$$a^l_{jm} = \langle l\,m{-}\tfrac{1}{2}\,,\,\tfrac{1}{2}\,\tfrac{1}{2}\mid j\,m\rangle\,, \quad b^l_{jm} = \langle l\,m{+}\tfrac{1}{2}\,,\,\tfrac{1}{2}\,{-}\tfrac{1}{2}\mid j\,m\rangle\,. \tag{2.101}$$

The main result to be used in the sequel now follows from the identity

$$2\sum\nolimits_i L_i\,S_i = (L+S)^2 - L^2 - S^2 = J^2 - L^2 - S^2.$$

Using it we infer that φ^l_{jm} is an eigenfunction of the operator $\vec{\sigma}\cdot\vec{L} := \sum \sigma_i L_i$:

$$(\vec{\sigma}\cdot\vec{L})\,\varphi^l_{jm} = \hbar\left(j(j+1) - l(l+1) - \tfrac{3}{4}\right)\varphi^l_{jm} = -\hbar\,(1+\kappa)\,\varphi^l_{jm}\,. \tag{2.102}$$

The second equation defines the parameter κ. Note that

$$\kappa = \begin{cases} l & j = l - \frac{1}{2}\,, \\ -l-1 & j = l + \frac{1}{2}\,. \end{cases} \tag{2.103}$$

2.14.2 Reduction by separation of variables

We now proceed with the solution of $H\psi = E\psi$ by making an ansatz for the 4-component spinor ψ as a direct sum of two 2-component spinors ξ and η :

$$\psi = \begin{pmatrix} \xi \\ \eta \end{pmatrix}. \tag{2.104}$$

The stationary Dirac equation with Coulomb potential $V(r) := -Ze^2/(4\pi\varepsilon_0 r)$ then decomposes as a pair of coupled equations for ξ and η :

$$(E - mc^2 - V)\,\xi = c\,\vec{\sigma}\cdot\vec{p}\,\eta\,, \tag{2.105}$$

$$(E + mc^2 - V)\,\eta = c\,\vec{\sigma}\cdot\vec{p}\,\xi\,. \tag{2.106}$$

At this point we recall from (2.98) the angular momentum eigenfunction φ^l_{jm} and separate the variables by making the product ansatz

$$\xi = \mathrm{i}F(r)\varphi^l_{jm}\,, \quad \eta = G(r)\,\frac{\vec{\sigma}\cdot\vec{r}}{r}\,\varphi^l_{jm}\,. \tag{2.107}$$

We pause briefly to subject this ansatz to a consistency check. A short calculation shows that $\vec{\sigma}\cdot\vec{r}$ is invariant under infinitesimal rotations:

$$[J_j\,, \vec{\sigma}\cdot\vec{r}] = \Big[\sum \epsilon_{jkl}\,x_k\,p_l + \tfrac{1}{2}\hbar\sigma_j\,, \vec{\sigma}\cdot\vec{r}\Big] = \frac{\hbar}{\mathrm{i}}\sum \epsilon_{jkl}\,x_k\,\sigma_l + \mathrm{i}\hbar\sum \epsilon_{jkl}\,\sigma_l\,x_k = 0\,.$$

On the other hand, space reflection $\vec{r}\to -\vec{r}$ takes $\vec{\sigma}\cdot\vec{r}$ into its negative. Thus $\vec{\sigma}\cdot\vec{r}$ is what is called a pseudo-scalar in physics. Now a spherical harmonic $Y_{l\,m_l}$ transforms under space reflection as $Y_{l\,m_l}\to(-1)^l\,Y_{l\,m_l}$. Therefore our ansatz has parity $P\psi = (-1)^l\psi$.

To proceed, we use the formula $\sigma_j\sigma_k = \delta_{jk} + \mathrm{i}\sum \epsilon_{jkl}\,\sigma_l$ to deduce the identities

$$(\vec{\sigma}\cdot\vec{r})(\vec{\sigma}\cdot\vec{p}) = \vec{r}\cdot\vec{p} + \mathrm{i}\vec{\sigma}\cdot(\vec{r}\times\vec{p}) = rp_r + \mathrm{i}\vec{\sigma}\cdot\vec{L}\,, \tag{2.108}$$

$$(\vec{\sigma}\cdot\vec{p})(\vec{\sigma}\cdot\vec{r}) = \vec{p}\cdot\vec{r} + \mathrm{i}\vec{\sigma}\cdot(\vec{p}\times\vec{r}) = -3\mathrm{i}\hbar + rp_r - \mathrm{i}\vec{\sigma}\cdot\vec{L}\,. \tag{2.109}$$

With their help we now compute the result of applying the operator $\vec{\sigma}\cdot\vec{p}$ to the two-component spinors ξ and η of (2.107):

$$\vec{\sigma}\cdot\vec{p}\,\xi = \frac{\vec{\sigma}\cdot\vec{r}}{r^2}\left((\vec{\sigma}\cdot\vec{r})\,(\vec{\sigma}\cdot\vec{p})\right)\xi = \hbar\,\frac{\vec{\sigma}\cdot\vec{r}}{r}\left(\frac{\partial}{\partial r} + \frac{1+\kappa}{r}\right)F\varphi^l_{jm}\,, \tag{2.110}$$

$$\vec{\sigma}\cdot\vec{p}\,\eta = \big((\vec{\sigma}\cdot\vec{p})(\vec{\sigma}\cdot\vec{r})\big)\frac{G}{r}\,\varphi^l_{jm} = \mathrm{i}\hbar\left(-\frac{\partial}{\partial r} - \frac{2}{r} + \frac{1+\kappa}{r}\right)G\,\varphi^l_{jm}\,. \tag{2.111}$$

Finally, by making the substitutions $F = r^{-1}f$ and $G = r^{-1}g$ we bring the pair of equations (2.105, 2.106) into the form

$$(E - mc^2 - V)f = \hbar c \left(-\frac{\partial}{\partial r} + \frac{\kappa}{r}\right) g\,, \tag{2.112}$$

$$(E + mc^2 - V)g = \hbar c \left(+\frac{\partial}{\partial r} + \frac{\kappa}{r}\right) f\,. \tag{2.113}$$

We must now solve this pair of ordinary differential equations in the radial variable r.

2.14.3 Solution of radial problem

We first investigate the problem in the limit $r \to \infty$ where the Coulomb energy $V(r) \sim 1/r$ and the rotational energy κ/r become negligible and our pair of equations simplifies to

$$(mc^2 - E)f \approx \hbar c \frac{\partial g}{\partial r}\,, \qquad (mc^2 + E)g \approx \hbar c \frac{\partial f}{\partial r}\,.$$

We easily see that this pair of asymptotic equations is solved by

$$f \approx (mc^2 + E)^{1/2}\, \mathrm{e}^{-\lambda r}\,, \qquad g \approx -(mc^2 - E)^{1/2}\, \mathrm{e}^{-\lambda r}\,,$$

where

$$\lambda = (\hbar c)^{-1}\sqrt{(mc^2)^2 - E^2} \tag{2.114}$$

is a characteristic wave number of the problem. We observe that in the non-relativistic limit $0 < mc^2 - E \ll mc^2$ the expression for λ reduces to

$$\lambda = \hbar^{-1}\sqrt{2m\triangle}\,, \qquad \triangle = mc^2 - E\,,$$

which is the de Broglie (imaginary) wave number for a bound state of binding energy $\triangle$.

These asymptotic considerations motivate the following ansatz:

$$f = \sqrt{1 + \frac{E}{mc^2}}\, \mathrm{e}^{-\lambda r}\,(A + B), \qquad g = \sqrt{1 - \frac{E}{mc^2}}\, \mathrm{e}^{-\lambda r}(A - B). \tag{2.115}$$

with unknown functions $A = A(r)$ and $B = B(r)$ which behave as $A \to 0$ and $B \to$ const for $r \to \infty$. To process this ansatz, it is useful to introduce the following dimensionless quantities:

$$\rho := 2\lambda r\,, \quad k_E := Z\alpha\, \frac{E}{\hbar c \lambda}\,, \quad k_M := Z\alpha\, \frac{mc^2}{\hbar c \lambda}\,, \quad \alpha := \frac{e^2}{4\pi\varepsilon_0 \hbar c}\,, \tag{2.116}$$

where $\alpha \approx 1/137$ is the so-called fine structure constant. (This name is of historical origin. It is nowadays understood that α is not a constant; being the coupling of quantum electrodynamics, α depends logarithmically on the energy cutoff or regularization scale.)

We now insert the ansatz (2.115) into (2.112, 2.113). After some trivial (if somewhat lengthy) algebra we arrive at the following equations for A and B :

$$\frac{\partial A}{\partial \rho} = A - \frac{k_E}{\rho}\, A - \frac{k_M + \kappa}{\rho}\, B, \qquad \frac{\partial B}{\partial \rho} = \frac{k_M - \kappa}{\rho}\, A + \frac{k_E}{\rho}\, B. \tag{2.117}$$

The final step is to find the solution to these equations in the form of power series. To prepare this step, we investigate the behavior in the limit of $\rho \to 0$. Let

$$A \simeq a\rho^\gamma, \quad B \simeq b\rho^\gamma \quad (\rho \to 0).$$

By inserting this small-ρ asymptotics into (2.117) and then sending $\rho \to 0$, we obtain the following system of linear equations for the numbers a and b :

$$(\gamma + k_E)\, a + (\kappa + k_M)\, b = 0\,,$$
$$(\kappa - k_M)\, a + (\gamma - k_E)\, b = 0\,.$$

In order for this system to have a solution other than the trivial one ($a = b = 0$), the determinant of the underlying matrix must vanish:

$$\mathrm{Det} \begin{pmatrix} \gamma + k_E & \kappa + k_M \\ \kappa - k_M & \gamma - k_E \end{pmatrix} = \gamma^2 - k_E^2 - \kappa^2 + k_M^2 = 0\,.$$

From (2.116) we then see that $k_M^2 - k_E^2 = (Z\alpha)^2$. Therefore the value of γ must be

$$\gamma = \gamma(\kappa) = \sqrt{\kappa^2 - Z^2\alpha^2}\,, \tag{2.118}$$

which makes good sense as long as the nuclear charge Z is not too large.

We finally look for the solution in the form of two power series:

$$A = \sum_{n=0}^{\infty} a_n \rho^{\gamma+n}, \qquad B = \sum_{n=0}^{\infty} b_n \rho^{\gamma+n}.$$

By inserting these into (2.117) and comparing coefficients we get

$$(n + \gamma + k_E)\, a_n + (\kappa + k_M)\, b_n = a_{n-1}\,,$$
$$(\kappa - k_M)\, a_n + (n + \gamma - k_E)\, b_n = 0\,.$$

The equation for $n = 0$ (with $a_{-1} = 0$) has a non-trivial solution for a_0 and b_0 by our choice (2.118) for γ. For $n \geq 1$ we eliminate b_n to obtain a recursion relation for a_n alone:

$$a_n = \left((n + \gamma + k_E) - \frac{(\kappa + k_M)(\kappa - k_M)}{n + \gamma - k_E} \right)^{-1} a_{n-1} = \frac{n + \gamma - k_E}{n^2 + 2n\gamma}\, a_{n-1}\,.$$

A sufficient condition for a bound-state solution to emerge is that the power series terminates at some finite order n resulting in square-integrable functions f and g. The condition for termination is that

$$0 = n + \gamma - k_E$$

holds for some positive integer $n \geq 1$. By solving this condition for E, we arrive at the following answer for the (discrete) energy spectrum:

$$E_{n,\kappa} = \frac{mc^2}{\sqrt{1 + \frac{(Z\alpha)^2}{(n+\gamma(\kappa))^2}}}\,. \tag{2.119}$$

Problem. Show that in the non-relativistic limit $c \to \infty$ this formula reproduces the known energy spectrum of hydrogen as computed from the Schrödinger equation. How do $n\,, \kappa$ relate to the quantum numbers of the non-relativistic spectrum? □

3 Second quantization

3.1 Bosons and fermions

We are assuming that the students of this course already have some familiarity with the quantum mechanics of a single particle and thus with quantum theory at the simplest level. As a preparation for the more advanced formalism of 'second quantization' introduced in this chapter, let us summarize by means of a table what are the key elements of the correspondence between classical mechanics and single-particle quantum mechanics.

classical mechanics	quantum mechanics
phase space $\Gamma = T^*\mathbb{R}^3 = \mathbb{R}^3 \times (\mathbb{R}^3)^*$	Hilbert space $V = L^2(\mathbb{R}^3)$
states $x \in \Gamma$	state vectors $\psi \in V$
observables $f : \Gamma \to \mathbb{R}$	self-adjoint operators $\mathrm{Op}(f)$
Poisson bracket: $\{f, g\} = \sum \left(\frac{\partial f}{\partial q_i} \frac{\partial f}{\partial p_i} - \frac{\partial g}{\partial q_i} \frac{\partial f}{\partial p_i} \right)$	commutator: $[\mathrm{Op}(f), \mathrm{Op}(f)] = \mathrm{i}\hbar\, \mathrm{Op}(\{f, g\}) + \mathcal{O}(\hbar^2)$

When there is more than one particle present, new structures and phenomena related to permutation symmetry appear. It turns out that the generalization to n-particle quantum mechanics depends on whether the particles under consideration are distinguishable or indistinguishable. In the former case the Hilbert space $\mathcal{V}$ of the many-particle system is simply the tensor product

$$\mathcal{V} = V^{(1)} \otimes V^{(2)} \otimes \cdots \otimes V^{(n)} \tag{3.1}$$

of the Hilbert spaces $V^{(1)}, \ldots, V^{(n)}$ for the individual particles (labeled by $1, \ldots, n$). Thus the wave function for n distinguishable particles in the Schrödinger representation is a linear combination of products of single-particle wave functions,

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \sum a_{i_1 i_2 \ldots i_n}\, \psi_{i_1}(\mathbf{r}_1) \psi_{i_2}(\mathbf{r}_2) \cdots \psi_{i_n}(\mathbf{r}_n)\,, \tag{3.2}$$

with complex coefficients $a_{i_1 i_2 \ldots i_n}$. Linear operators representing observables such as the energy or the components of momentum, act on these wave functions in the natural way (i.e., one simply tensors up the setting of single-particle quantum mechanics), and there isn't a whole lot of new material for the student to learn.

However, in the case of indistinguishable or identical particles, the many-particle Hilbert space is spanned by wave functions which are either totally symmetric or totally anti-symmetric. Consider first the latter case where we have

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \mathrm{sign}(\pi)\, \Psi(\mathbf{r}_{\pi(1)}, \mathbf{r}_{\pi(2)}, \ldots, \mathbf{r}_{\pi(n)})\,, \tag{3.3}$$

for any permutation $\pi \in \mathrm{S}_n$. [Here $\mathrm{sign}(\pi) = +1$ if the permutation π is even and $\mathrm{sign}(\pi) = -1$ if π is odd.] Particles with this behavior under permutations are called fermions.

Gibbs, in the context of classical statistical mechanics, observed that symmetrization (more precisely: reduction of statistical weight by identification of n-particle configurations that are permutations of one another) is needed in order for the entropy to be additive. The next series of statements gives a quantum-theoretic explanation of this observation.

Fact. The wave functions for (an indefinite number of) fermions with single-particle Hilbert space V, e.g. $V = L^2(\mathbb{R}^3)$, form an exterior algebra, $\wedge(V)$. The wave function for a definite number n of fermions is an element of the degree-n subspace $\wedge^n(V) \subset \wedge(V)$ of the exterior algebra. If Ψ_p and Ψ_q are totally anti-symmetric wave functions describing two subsystems of p resp. q fermions (all of which are identical), then the wave function for the total system of $p+q$ identical fermions is the exterior product $\Psi_p \wedge \Psi_q$ defined by

$$(\Psi_p \wedge \Psi_q)(\mathbf{r}_1, \ldots, \mathbf{r}_{p+q}) := \frac{1}{p!\, q!} \sum_{\pi \in \mathrm{S}_{p+q}} \mathrm{sign}(\pi) \Psi_p(\mathbf{r}_{\pi(1)}, \ldots, \mathbf{r}_{\pi(p)})\, \Psi_q(\mathbf{r}_{\pi(p+1)}, \ldots, \mathbf{r}_{\pi(p+q)})\,. \tag{3.4}$$

The most general n-fermion wave function $\psi \in \wedge^n(V)$ can be expressed as a linear combination of exterior products of single-particle wave functions:

$$\Psi = \sum a_{i_1 i_2 \ldots i_n} \psi_{i_1} \wedge \psi_{i_2} \wedge \cdots \wedge \psi_{i_n}\,, \tag{3.5}$$

with complex coefficients $a_{i_1 i_2 \ldots i_n}$. The vacuum state is $1 \in \wedge^0(V) \equiv \mathbb{C}$. One also writes $1 \equiv |0\rangle$. Wave functions of the special form $\Psi = \psi_{i_1} \wedge \psi_{i_2} \wedge \cdots \wedge \psi_{i_n}$ are called Slater determinants.

Problem. Show that the exterior product is associative: $(\Psi_p \wedge \Psi_q) \wedge \Psi_r = \Psi_p \wedge (\Psi_q \wedge \Psi_r)$. □

We now turn to the alternative situation of totally symmetric wave functions:

$$\Phi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \Phi(\mathbf{r}_{\pi(1)}, \mathbf{r}_{\pi(2)}, \ldots, \mathbf{r}_{\pi(n)}) \tag{3.6}$$

for any permutation $\pi \in \mathrm{S}_n$. Particles with this behavior under permutations are called bosons.

Definition. If V is a $\mathbb{K}$-vector space, one defines the symmetric algebra $\mathrm{S}(V)$ as the associative algebra which is generated by $V \oplus \mathbb{K}$ with relations $vv' = v'v$ for all $v, v' \in V$. Multiplication in the symmetric algebra with these relations is called the symmetric product. When the symmetric product is to be emphasized, we use the expanded notation $v \vee v' = v' \vee v$. The symmetric algebra is graded by the degree:

$$\mathrm{S}(V) = \bigoplus\nolimits_{n \geq 0} \mathrm{S}^n(V)\,, \quad \mathrm{S}^0(V) \equiv \mathbb{K}\,, \quad \mathrm{S}^1(V) \equiv V. \tag{3.7}$$

Problem. Assuming V to be of finite dimension $\dim V = N$, show that $\dim \mathrm{S}^n(V) = \binom{N+n-1}{n}$. [Thus, unlike the exterior algebra $\wedge(V)$, the symmetric algebra $\mathrm{S}(V)$ has infinite dimension even when V is finite-dimensional.]

Fact. The wave functions for (an indefinite number of) bosons with single-particle Hilbert space V form the symmetric algebra $\mathrm{S}(V)$. Any wave function for a definite number n of bosons is an element of the degree-n subspace $\mathrm{S}^n(V) \subset \mathrm{S}(V)$ of the symmetric algebra. If Φ_p and Φ_q are

totally symmetric wave functions describing two subsystems of p resp. q bosons (all of which are identical), then the wave function for the total system of $p+q$ bosons is the symmetric product

$$(\Phi_p \vee \Phi_q)(\mathbf{r}_1, \ldots, \mathbf{r}_{p+q}) = \frac{1}{p!\, q!} \sum_{\pi \in \mathrm{S}_{p+q}} \Phi_p(\mathbf{r}_{\pi(1)}, \ldots, \mathbf{r}_{\pi(p)})\, \Phi_q(\mathbf{r}_{\pi(p+1)}, \ldots, \mathbf{r}_{\pi(p+q)})\,. \tag{3.8}$$

The most general n-boson wave function $\Phi \in \mathrm{S}^n(V)$ can be expressed as a linear combination of symmetric products of single-particle wave functions:

$$\Phi = \sum a_{i_1 i_2 \ldots i_n}\, \phi_{i_1} \vee \phi_{i_2} \vee \cdots \vee \phi_{i_n}\,. \tag{3.9}$$

The boson vacuum state is $|0\rangle \equiv 1 \in \mathrm{S}^0(V) \equiv \mathbb{C}$. □

To do many-particle quantum mechanics, one must understand how to turn physical observables into operators on the many-particle Hilbert spaces $\wedge(V)$ for fermions and $\mathrm{S}(V)$ for bosons. The present chapter will introduce a good algebraic formalism by which to handle the situation, namely 'second quantization'. Second quantization for bosons is particularly suitable as a formalism to quantize the abelian gauge theory of electromagnetism. Another application will be to the Dirac equation: using second quantization for fermions one reinterprets Dirac's original single-electron theory as a quantum field theory which remains stable and physically meaningful even in the presence of coupling to the quantized electromagnetic field.

3.2 Weyl algebra and Clifford algebra

We now introduce a boson-type analog of the Clifford algebra.

Definition. Let U be a symplectic vector space over the number field $\mathbb{K}$, i.e., U carries a non-degenerate skew-symmetric (or alternating) $\mathbb{K}$-bilinear form

$$A : U \times U \to \mathbb{K}\,, \qquad A(u, u') = -A(u', u)\,. \tag{3.10}$$

The Weyl algebra of U is defined to be the associative algebra, $\mathcal{W}(U, A)$, generated by $U \oplus \mathbb{K}$ with relations

$$uu' - u'u = A(u, u')\,. \tag{3.11}$$

Remark. Without changing the content of the theory, we could enhance the similarity with the definition of the Clifford algebra in Section 2.7 by putting $2A$ instead of A on the right-hand side. Nonetheless, in the present context we prefer to use the normalization convention (3.11).

Example 1. Consider the symplectic plane (or phase plane) $U = \mathbb{R}^2$ with basis vectors e, f and symplectic form

$$A(e, e) = A(f, f) = 0\,, \quad A(e, f) = -A(f, e) = -1\,. \tag{3.12}$$

The Weyl algebra $\mathcal{W}(\mathbb{R}^2, A)$ then is the algebra of polynomial expressions in e and f with real coefficients and the relation $fe - ef = 1$ being understood. A concrete realization of this algebra is by a polynomials in a coordinate generator $x \equiv e$ and its derivative $d/dx \equiv f$.

Example 2. We give a second example, which is closer to the setting pursued below. Let q and p be the position and momentum operators of a one-dimensional quantum system. Let $\mathbb{K} = \mathbb{C}$, $U = \mathrm{span}_{\mathbb{C}}\{p, q\}$, and take A to be the Heisenberg commutator:

$$A(q,q) = A(p,p) = 0\,, \qquad A(q,p) = -A(p,q) = \mathrm{i}\hbar\,. \tag{3.13}$$

Then the Weyl algebra $\mathcal{W}(U,A)$ is the algebra of polynomials in the operators q and p. A specific example of such a polynomial is the Hamiltonian $H = p^2/2m + m\omega^2 q^2/2$ of the one-dimensional harmonic oscillator. Introducing the dimensionless operators

$$a := \frac{1}{\sqrt{2}}\left(\frac{q}{\ell} + \mathrm{i}\frac{\ell p}{\hbar}\right), \quad a^\dagger := \frac{1}{\sqrt{2}}\left(\frac{q}{\ell} - \mathrm{i}\frac{\ell p}{\hbar}\right), \tag{3.14}$$

where $\ell = \sqrt{\hbar/(m\omega)}$ is the oscillator length, one can also view the Weyl algebra $\mathcal{W}(U,A)$ as the algebra of polynomials in the operators a and $a^\dagger$ with commutation relations

$$a\,a^\dagger - a^\dagger a = 1\,. \tag{3.15}$$

The oscillator Hamiltonian is known to take the form $H = \hbar\omega(a^\dagger a + \frac{1}{2})$. □

We now start introducing the formalism of second quantization for bosons. For this we realize the notion of Weyl algebra as follows. Let V be the single-boson Hilbert space and let V^* be its dual. (We shall refrain from using the isomorphism $c_V : V \to V^*$, $v \mapsto \langle v, \cdot\rangle_V$ by the Hermitian scalar product of V for the moment.) We take the direct sum $U = V \oplus V^*$ to be equipped with the canonical alternating form $A : U \times U \to \mathbb{C}$,

$$A(v+\varphi, v'+\varphi') = \varphi(v') - \varphi'(v)\,. \tag{3.16}$$

With these definitions, the Weyl algebra $\mathcal{W}(V \oplus V^*, A)$ is the algebraic structure underlying (the formalism of second quantization for) many-boson quantum systems.

Definition. In physics the following notational conventions are standard. One fixes some basis $\{e_1, e_2, \ldots\}$ of the single-boson Hilbert space V with dual basis $\{f_1, f_2, \ldots\}$ of V^*, and one writes $f_i \equiv a_i$ and $e_i \equiv a_i^+$. Using the duality pairing $f_i(e_j) = \delta_{ij}$, one then gets the Weyl algebra relations (3.11) in the form

$$[a_i\,, a_j^+] = \delta_{ij}\,, \quad [a_i\,, a_j] = [a_i^+, a_j^+] = 0 \qquad (i,j = 1,2,\ldots)\,. \tag{3.17}$$

These are called the canonical commutation relations (CCR).

Remark. Later, a_i and a_i^+ will be acting as linear operators in the so-called bosonic Fock space $\mathrm{S}(V)$ with its canonical Hermitian scalar product. In that setting, assuming that $\{e_1, e_2, \ldots\}$ is an orthonormal basis, the operator $a_i^+ \equiv a_i^\dagger$ will turn out to be the Hermitian adjoint of a_i.

Example. By the canonical commutation relations one has

$$a_i^+ a_j\, a_k^+ a_l = a_i^+ a_k^+ a_j\, a_l + \delta_{jk}\, a_i^+ a_l\,. \quad \square \tag{3.18}$$

In view of the close similarity between the Weyl algebra and the Clifford algebra, one may ask whether a development parallel to that of Sections 2.10 and 2.11 exists on the bosonic side. The answer is: yes, to a large extent. Indeed, the analog of $\mathrm{SO}(V,Q)$ is the symplectic Lie group $\mathrm{Sp}(U,A)$, the analog of $\mathfrak{so}(V,Q)$ is the symplectic Lie algebra $\mathfrak{sp}(U,A)$, and the analog of $\mathrm{Cl}_2(V,Q)$ is the subspace

$$\mathcal{W}_2(U,A) := \{X \in \mathcal{W}(U,A) \mid X = \sum\nolimits_i (u_i v_i + v_i u_i)\ ;\ u_i\,, v_i \in U\}$$

of symmetrized degree-2 elements of the Weyl algebra. It is still true that $\mathcal{W}_2(U,A)$ is a Lie algebra which is isomorphic to $\mathfrak{sp}(U,A)$ by letting $X \in \mathcal{W}_2(U,A)$ act on $u \in U \subset \mathcal{W}(U,A)$ by the commutator, and that by exponentiating this commutator action one recovers the symplectic Lie group. There is, however, one difference: it is not possible to exponentiate $\mathcal{W}_2(U,A)$ inside the Weyl algebra to produce an analog of the spin group. The reason is that, by a theorem due to Stone and von Neumann, any non-trivial realization of the Weyl algebra with relations (3.11) must be on an infinite-dimensional representation space where the elements $X \in \mathcal{W}_2(U,A)$ act as unbounded operators. For example, the Hamiltonian H of the harmonic oscillator is an element $H \in \mathcal{W}_2(U,A)$ for our Example 2 above. The exponential e^{-tH} exists for $\mathfrak{Re}\, t > 0$ but does not for $\mathfrak{Re}\, t < 0$. Nonetheless, in a suitable real framework there does exist a bosonic analog of the spin group; it is called the metaplectic group. (For example, the quantum time evolution of the harmonic oscillator is a one-parameter subgroup of the metaplectic group.)

This concludes our introduction of the Weyl algebra and the canonical commutation relations for bosons. We now turn to the fermionic side: the Clifford algebra. Here we will be brief, as we have already given a thorough discussion of the Clifford algebra in Chapter 2.

Definition. The formalism of second quantization for fermions is based on the following. Starting from the single-fermion Hilbert space V, we take the direct sum $U = V \oplus V^*$ and we equip U with the symmetric bilinear form $B : U \times U \to \mathbb{C}$ defined by

$$B(v+\varphi, v'+\varphi') = \varphi(v') + \varphi'(v)\,. \tag{3.19}$$

The Clifford algebra $\mathrm{Cl}(U,B)$ then is the associative algebra generated by $U \oplus \mathbb{C}$ with relations

$$uu' + u'u = B(u,u')\,. \tag{3.20}$$

In physics it is customary to use the following notational conventions. Fixing any basis $\{e_1, e_2, \ldots\}$ of the single-fermion Hilbert space V with dual basis $\{f_1, f_2, \ldots\}$ of V^*, one writes $f_i \equiv c_i$ and $e_i \equiv c_i^+$. By the duality pairing $f_i(e_j) = \delta_{ij}$, one then gets the Clifford algebra relations (3.20) in the form

$$\{c_i\,, c_j^+\} = \delta_{ij}\,, \quad \{c_i\,, c_j\} = \{c_i^+, c_j^+\} = 0 \qquad (i,j = 1,2,\ldots)\,, \tag{3.21}$$

where $\{u,u'\} := uu' + u'u$ denotes the anti-commutator. We recall from Section 2.8 that the relations (3.21) are referred to as the canonical anti-commutation relations (CAR).

3.3 Representation on Fock space

In Section 2.8 we learned that the abstract Clifford algebra $\mathrm{Cl}(V\oplus V^*, B)$ has a concrete realization by linear operators on the spinor representation space $\wedge(V)$. [Actually, we constructed the spinor representation for $V = P^*$. Note that V and V^* are on the same footing in this context, and exactly the same construction would go through for the choice $\wedge(P^*) = \wedge(V^*)$ of representation space.] We will now explain that, in a very similar way, the abstractly defined Weyl algebra $\mathcal{W}(V \oplus V^*, A)$ has a realization by linear operators on the so-called bosonic Fock space $\mathrm{S}(V)$.

Note that $A(v, v') = A(\varphi, \varphi') = 0$ vanishes for any $v, v' \in V$ and $\varphi, \varphi' \in V^*$. One expresses this fact by saying that V and V^* are Lagrangian subspaces of the complex symplectic vector space $U = V \oplus V^*$. It follows from the Weyl algebra relations (3.11) that $vv' = v'v$ and $\varphi\varphi' = \varphi'\varphi$. These are the defining relations of the respective symmetric algebras, namely $\mathrm{S}(V)$ and $\mathrm{S}(V^*)$. We thus see that both $\mathrm{S}(V)$ and $\mathrm{S}(V^*)$ are contained as subalgebras in $\mathcal{W}(V \oplus V^*, A)$.

We can now get a representation of $\mathcal{W}(V \oplus V^*, A)$ by letting it act on $\mathrm{S}(V)$, or $\mathrm{S}(V^*)$, or any other subalgebra generated by a Lagrangian subspace. We focus on $\mathrm{S}(V)$ for concreteness. Thus our goal now is to define an action of $\mathcal{W}(V \oplus V^*, A)$ on $\mathrm{S}(V)$,

$$\mathcal{W}(V \oplus V^*, A) \times \mathrm{S}(V) \to \mathrm{S}(V)\,, \quad (x, \xi) \mapsto x \cdot \xi\,. \tag{3.22}$$

Since $\mathcal{W}(V \oplus V^*, A)$ is generated by $V \oplus V^* \oplus \mathbb{C}$, it suffices to specify the Weyl multiplication (3.22) for vectors $v \in V$ and co-vectors $\varphi \in V^*$. (The multiplication by scalars $k \in \mathbb{C}$ is of course the obvious one given by the structure of a complex vector space.)

Multiplication by vectors $v \in V$ is simply the symmetric product computed inside the symmetric algebra:

$$v \cdot \xi := \mu(v)\,\xi := v \vee \xi \equiv v\,\xi\,. \tag{3.23}$$

The notation $\mu(v)$ is used in order to distinguish the algebra element $v \in V \subset \mathcal{W}(V \oplus V^*, A)$ from its concrete realization as a linear operator $\mu(v)$ on $\mathrm{S}(V)$. Note that this multiplication $\mu(v) : \mathrm{S}^l(V) \to \mathrm{S}^{l+1}(V)$ increases the degree (in physics language: the number of bosons) by one.

Now consider $\varphi \in V^*$. Such elements act by a degree-lowering operation $\delta(\varphi)$:

$$\delta(\varphi) \;:\; \mathrm{S}^l(V) \to \mathrm{S}^{l-1}(V)\,, \quad \xi \mapsto \varphi \cdot \xi \equiv \delta(\varphi)\,\xi\,, \tag{3.24}$$

which is called a derivation of $\mathrm{S}(V)$, as it has the algebraic properties of a derivative. We define it recursively. For the two lowest degrees $l = 0, 1$ we set

$$l = 0: \quad \delta(\varphi)\,1 := 0\,, \tag{3.25}$$

$$l = 1: \quad \delta(\varphi)\,v := \varphi(v)\,. \tag{3.26}$$

The action on higher-degree elements $\xi \vee \eta \in \mathrm{S}(V)$ is then defined by the product rule (or Leibniz rule) known from differential calculus:

$$\delta(\varphi)\,(\xi \vee \eta) := \delta(\varphi)\xi \vee \eta + \xi \vee \delta(\varphi)\eta\,. \tag{3.27}$$

Problem. Show that the operations of symmetric multiplication $\mu(v) : \mathrm{S}^l(V) \to \mathrm{S}^{l+1}(V)$ and derivation $\delta(\varphi) : \mathrm{S}^l(V) \to \mathrm{S}^{l-1}(V)$ satisfy the canonical commutation relations (CCR):

$$\begin{aligned} &\mu(v)\mu(v') - \mu(v')\mu(v) = 0\,, \quad \delta(\varphi)\delta(\varphi') - \delta(\varphi')\delta(\varphi) = 0\,, \\ &\delta(\varphi)\mu(v) - \mu(v)\delta(\varphi) = A(\varphi, v)\,\mathrm{Id}_{\mathrm{S}(V)}\,, \end{aligned} \tag{3.28}$$

for all $v, v' \in V$ and $\varphi, \varphi' \in V^*$. □

Using the canonical commutation relations, an easy calculation

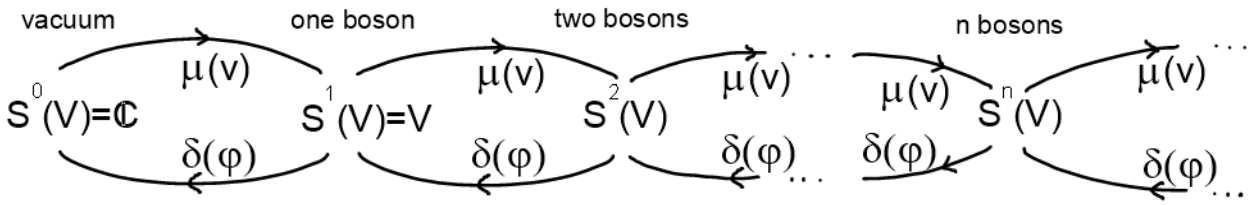
$$\begin{aligned} &(v+\varphi)\cdot((v'+\varphi')\cdot\xi) - (v'+\varphi')\cdot((v+\varphi)\cdot\xi) - A(v+\varphi, v'+\varphi')\cdot\xi \\ &= \big(\mu(v)+\delta(\varphi)\big)\big(\mu(v')+\delta(\varphi')\big)\xi - \big(\mu(v')+\delta(\varphi')\big)\big(\mu(v)+\delta(v)\big)\xi - A(v+\varphi, v'+\varphi')\xi \\ &= \delta(\varphi)\mu(v')\xi - \mu(v')\delta(\varphi)\xi - A(\varphi, v')\xi - \delta(\varphi')\mu(v)\xi + \mu(v)\delta(\varphi')\xi + A(\varphi', v)\xi = 0\,, \end{aligned}$$

shows that the relations (3.11) are satisfied for all $v + \varphi = u$ and $v' + \varphi' = u'$. Thus we have indeed constructed a representation of $\mathcal{W}(V \oplus V^*, A)$.

Definition. For a complex vector space V let the direct sum $V \oplus V^*$ be equipped with the canonical alternating form $A(v+\varphi, v'+\varphi') = \varphi(v') - \varphi'(v)$. The oscillator representation of the Weyl algebra $\mathcal{W}(V \oplus V^*, A)$ is defined by the action

$$(v + \varphi + k)\cdot\xi = \mu(v)\xi + \delta(\varphi)\xi + k\xi \qquad (v \in V,\ \varphi \in V^*,\ k \in \mathbb{C}) \tag{3.29}$$

on the symmetric algebra $\mathrm{S}(V)$. □



Let us now give the translation into the language and notation of physics. We recall that the symmetric algebra $\mathrm{S}(V)$ is referred to as the bosonic Fock space (of the single-particle Hilbert space V) in this context. The neutral element $1 \in \mathrm{S}^0(V) = \mathbb{C}$ is called the vacuum state and is denoted by $1 \equiv |0\rangle$. (To avoid confusion, we emphasize that $|0\rangle$ is not the zero vector!) We also recall that fixing an orthonormal basis $\{e_1, e_2, \ldots\}$ of V together with the dual basis $\{f_1, f_2, \ldots\}$ of V^*, one writes $a_i^+ \equiv e_i$ and $a_i \equiv f_i$ and refers to these as particle creation and particle annihilation operators, respectively. No change of notation is made when these algebraically defined objects act as linear operators on the bosonic Fock space. Thus $a_i \equiv \delta(f_i)$ and $a_i^+ \equiv \mu(e_i)$. The definitions of multiplication $a_i^+ = \mu(e_i) : \mathrm{S}^l(V) \to \mathrm{S}^{l+1}(V)$ and derivation $a_j = \delta(f_j) : \mathrm{S}^l(V) \to \mathrm{S}^{l-1}(V)$ translate to

$$a_j^+ \cdot \big(a_{i_1}^+ a_{i_2}^+ \cdots a_{i_l}^+|0\rangle\big) := a_j^+ a_{i_1}^+ a_{i_2}^+ \cdots a_{i_l}^+|0\rangle, \tag{3.30}$$

$$\begin{aligned} a_j \cdot \big(a_{i_1}^+ a_{i_2}^+ \cdots a_{i_l}^+|0\rangle\big) &:= \delta_{ji_1} a_{i_2}^+ a_{i_3}^+ \cdots a_{i_l}^+|0\rangle \\ &\quad + \delta_{ji_2} a_{i_1}^+ a_{i_3}^+ \cdots a_{i_l}^+|0\rangle + \ldots + \delta_{ji_l} a_{i_1}^+ a_{i_2}^+ \cdots a_{i_{l-1}}^+|0\rangle. \end{aligned} \tag{3.31}$$

The action (3.31) of the particle annihilation operator $a_j = \delta(a_j)$ is also defined by saying that the canonical commutation relations (3.17) hold and

$$a_j|0\rangle = 0\,. \tag{3.32}$$

This concludes, for the moment, our development of the bosonic variant of second quantization, and we turn to a summary and adaptation of the material of Chapter 2 for the fermionic side.

Definition (Review). For a complex vector space V let the direct sum $V \oplus V^*$ be equipped with the canonical symmetric form $B(v+\varphi, v'+\varphi') = \varphi(v') + \varphi'(v)$. The spinor representation of the Clifford algebra $\mathrm{Cl}(V \oplus V^*, B)$ is defined by the action

$$(v + \varphi + k)\cdot\xi = \varepsilon(v)\xi + \iota(\varphi)\xi + k\xi \qquad (v \in V,\ \varphi \in V^*,\ k \in \mathbb{C}) \tag{3.33}$$

on the exterior algebra $\wedge(V)$. In physics one calls $\wedge(V)$ the fermionic Fock space. The vacuum is denoted by $|0\rangle \equiv 1 \in \wedge^0(V)$. Introducing an orthonormal basis as usual, the actions of the particle creation operator $\varepsilon(e_j) \equiv c_j^+$ and particle annihilation operator $\iota(f_j) \equiv c_j$ are given by

$$c_j^+ \cdot \left(c_{i_1}^+ c_{i_2}^+ \cdots c_{i_l}^+|0\rangle\right) := c_j^+ c_{i_1}^+ c_{i_2}^+ \cdots c_{i_l}^+|0\rangle, \tag{3.34}$$

$$\begin{aligned} c_j \cdot \left(c_{i_1}^+ c_{i_2}^+ \cdots c_{i_l}^+|0\rangle\right) &:= \delta_{ji_1} c_{i_2}^+ c_{i_3}^+ \cdots c_{i_l}^+|0\rangle \\ &\quad - \delta_{ji_2} c_{i_1}^+ c_{i_3}^+ \cdots c_{i_l}^+|0\rangle + \ldots + (-1)^{l-1}\,\delta_{ji_l} c_{i_1}^+ c_{i_2}^+ \cdots c_{i_{l-1}}^+|0\rangle. \end{aligned} \tag{3.35}$$

The particle annihilation operators $c_j = \iota(f_j)$ annihilate the vacuum: $c_j|0\rangle = 0\,.$ □

We finish the section with a tabular summary:

particle type	algebra of operators	Fock space
bosons	Weyl algebra $\mathcal{W}(V \oplus V^*, A)$ canonical commutation relations	symmetric algebra: $\mathrm{S}(V) = \mathcal{W}(V \oplus V^*, A)/\mathrm{S}(V^*)$
fermions	Clifford algebra $\mathrm{Cl}(V \oplus V^*, B)$ canonical anti-commutation relations	exterior algebra: $\wedge(V) = \mathrm{Cl}(V \oplus V^*, B)/\wedge(V^*)$

3.4 Fock space scalar product

We are in the process of developing a quantum-theoretic formalism for many bosons and/or fermions. What we have learned so far is how the single-particle Hilbert space V gets promoted to a Fock space $\mathrm{S}(V)$ or $\wedge(V)$ for many particles. In order to turn Fock space into the Hilbert space of many-body quantum theory, we still need to introduce a Hermitian scalar product on it. Given that the single-particle Hilbert space V already carries a Hermitian scalar product $\langle\cdot,\cdot\rangle_V$, there is a natural way of doing this, as follows.

For bosons and fermions alike, the Fock space scalar product is diagonal in the degree or particle number n. For bosons one defines the scalar product $\langle\cdot,\cdot\rangle_{\mathrm{S}^n(V)} : \mathrm{S}^n(V) \times \mathrm{S}^n(V) \to \mathbb{C}$ by

$$\big\langle v_1 \vee v_2 \vee \cdots \vee v_n \,, v'_1 \vee v'_2 \vee \cdots \vee v'_n \big\rangle_{\mathrm{S}^n(V)} = \sum_{\pi \in \mathrm{S}_n} \langle v_1, v'_{\pi(1)} \rangle_V \langle v_2, v'_{\pi(2)} \rangle_V \cdots \langle v_n, v'_{\pi(n)} \rangle_V . \tag{3.36}$$

Notice that in the special case of $v^n = v \vee v \vee \cdots \vee v$ (n factors) one has $\langle v^n, v'^n \rangle_{\mathrm{S}^n(V)} = n!\, \langle v, v' \rangle_V^n$. Using physics notation, this means that

$$\frac{(a_1^+)^{n_1}}{\sqrt{n_1!}} \frac{(a_2^+)^{n_2}}{\sqrt{n_2!}} \cdots \frac{(a_l^+)^{n_l}}{\sqrt{n_l!}} \cdots |0\rangle \tag{3.37}$$

for $n_l \in \mathbb{N} \cup \{0\}$ ($l = 1, 2, \ldots$) is an orthonormal basis of $\mathrm{S}(V)$, provided that the standard convention $a_i^+ = \mu(e_i)$ is in force and $\{e_1, e_2, \ldots\}$ is an orthonormal basis of V.

The Hermitian scalar product $\langle\cdot,\cdot\rangle_{\wedge^n(V)} : \wedge^n(V) \times \wedge^n(V) \to \mathbb{C}$ for fermions is defined by

$$\begin{aligned} \big\langle v_1 \wedge \cdots \wedge v_n \,, v'_1 \wedge \cdots \wedge v'_n \big\rangle_{\wedge^n(V)} &= \sum_{\pi \in \mathrm{S}_n} \mathrm{sign}(\pi)\, \langle v_1, v'_{\pi(1)} \rangle_V \cdots \langle v_n, v'_{\pi(n)} \rangle_V \\ &= \mathrm{Det} \begin{pmatrix} \langle v_1, v'_1 \rangle_V & \cdots & \langle v_1, v'_n \rangle_V \\ \vdots & \ddots & \vdots \\ \langle v_n, v'_1 \rangle_V & \cdots & \langle v_n, v'_n \rangle_V \end{pmatrix} . \end{aligned} \tag{3.38}$$

It can be shown that this has all the properties required of a Hermitian scalar product; in particular, $\langle v_1 \wedge \cdots \wedge v_n \,, v_1 \wedge \cdots \wedge v_n \rangle_{\wedge^n(V)}$ is non-negative and vanishes only if $v_1 \wedge \cdots \wedge v_n$ is the zero vector. The latter happens if and only if the set of vectors $v_1 \ldots, v_n$ is linearly dependent. This feature is a manifestation of the Pauli exclusion principle.

Assuming again the standard physics convention $c_i^+ = \varepsilon(e_i)$ for an orthonormal basis $\{e_1, e_2, \ldots\}$ of V, one finds that the set of Fock space vectors

$$(c_1^+)^{n_1} (c_2^+)^{n_2} \cdots (c_l^+)^{n_l} \cdots |0\rangle \tag{3.39}$$

for $n_l \in \{0, 1\}$ ($l = 1, 2, \ldots$) constitute an orthonormal basis of $\wedge(V)$.

Problem. We recall that for a complex vector space with Hermitian scalar product, one has a canonically defined operation of taking the Hermitian adjoint (denoted by $\dagger$). Show that $(c_j^+)^\dagger = c_j$ (fermions) and $(a_j^+)^\dagger = a_j$ (bosons). $\square$

There exists another perspective on the Hermitian scalar product in Fock space, as follows. We recall that the Hermitian scalar product $\langle \cdot, \cdot \rangle_V$ on single-particle Hilbert space V determines a one-to-one correspondence

$$\gamma_V : V \to V^*, \quad v \mapsto \langle v, \cdot \rangle_V ,$$

called the Dirac ket-bra bijection and written $|v\rangle \mapsto \langle v|$ in Dirac's notation. Consider now the case of fermions for definiteness (for bosons it works the same way). The map c_V induces on the fermionic Fock space $\wedge(V)$ a mapping

$$\gamma_{\wedge(V)} : \wedge(V) \to \wedge(V)^* \cong \wedge(V^*), \quad v_1 \wedge \cdots \wedge v_n \mapsto \gamma_V(v_n) \wedge \cdots \wedge \gamma_V(v_1). \tag{3.40}$$

(Such a mapping is called an anti-homomorphism of exterior algebras, as the order of factors is reversed.) In physics notation (with standard conventions) this reads

$$\gamma_{\wedge(V)} : \quad c_{i_1}^+ c_{i_2}^+ \cdots c_{i_n}^+ |0\rangle \mapsto \langle 0| c_{i_n} \cdots c_{i_2} c_{i_1} .$$

The r.h.s. $\langle 0| c_{i_n} \cdots c_{i_1}$ is a linear function on the fermionic Fock space $\wedge(V)$ and as such an element of $\wedge(V)^*$ in the natural way: to evaluate this linear function on a vector $c_{j_1}^+ \cdots c_{j_n}^+ |0\rangle$ in Fock space we apply the product $c_{i_n} \cdots c_{i_1}$ and then take the vacuum component:

$$\gamma_{\wedge(V)} \left(c_{i_1}^+ c_{i_2}^+ \cdots c_{i_n}^+ |0\rangle \right) \left(c_{j_1}^+ c_{j_2}^+ \cdots c_{j_n}^+ |0\rangle \right) = \langle 0| c_{i_n} \cdots c_{i_2} c_{i_1} c_{j_1}^+ c_{j_2}^+ \cdots c_{j_n}^+ |0\rangle.$$

By using the canonical anti-commutation relations to compute the right-hand side we obtain

$$\gamma_{\wedge(V)} \left(c_{i_1}^+ c_{i_2}^+ \cdots c_{i_n}^+ |0\rangle \right) \left(c_{j_1}^+ c_{j_2}^+ \cdots c_{j_n}^+ |0\rangle \right) = \sum_{\pi \in \mathrm{S}_n} \mathrm{sign}(\pi) \prod_{l=1}^{n} \delta_{i_l, j_{\pi(l)}} ,$$

in agreement with the definition (3.38). Thus we may consider the Fock space Hermitian scalar product (3.38) as equivalent to the natural mapping (3.40).

3.5 Second quantization of one- and two-body operators

Let now $L \in \mathrm{End}(V)$ be a linear operator on the single-particle Hilbert space V. Such an operator is called a one-body operator in the present context. We fix some basis $\{e_1, e_2, \ldots\}$ (not necessarily orthonormal) and denote by $\{f_1, f_2, \ldots\}$ the dual basis of V^*.

Definition. The process of second quantization for fermions or bosons sends the one-body operator $L \in \mathrm{End}(V)$ to the operator $\widehat{L}$ on Fock space $\wedge(V)$ resp. $\mathrm{S}(V)$ which is defined by

$$L \mapsto \widehat{L} = \begin{cases} \sum\limits_i \varepsilon(Le_i)\, \iota(f_i) & \text{(fermions)}, \\ \sum\limits_i \mu(Le_i)\, \delta(f_i) & \text{(bosons)}. \end{cases} \tag{3.41}$$

Fact. The second-quantized operator $\widehat{L}$ extends the one-body operator $L \in \mathrm{End}(V)$ as a derivation of the algebra under consideration, i.e., as a derivation of the exterior algebra $\wedge(V)$ for the case of fermions and of the symmetric algebra $\mathrm{S}(V)$ for the case of bosons.

Proof. Consider the case of fermions and let $v \in V \subset \wedge(V)$ be any single-particle state. Application of $\widehat{L}$ yields

$$\widehat{L}v = \sum\nolimits_i \varepsilon(Le_i)\,\iota(f_i)v = \sum\nolimits_i f_i(v)\,\varepsilon(Le_i)1 = \sum\nolimits_i f_i(v)\,Le_i = Lv.$$

Thus $\widehat{L}$ coincides with L on the single-particle subspace $V = \wedge^1(V)$.

Now consider an n-particle state $\xi = v_1 \wedge v_2 \wedge \cdots \wedge v_n$ which is a product of n single-particle states. By applying the annihilation operator $\iota(f_i)$ you get

$$\iota(f_i)\xi = \iota(f_i)v_1 \wedge v_2 \wedge \cdots v_n - v_1 \wedge \iota(f_i)v_2 \wedge \cdots \wedge v_n + \ldots + (-1)^{n-1} v_1 \wedge v_2 \wedge \cdots \wedge \iota(f_i)v_n\,.$$

Subsequent exterior multiplication by $\varepsilon(Le_i)$ and summation over i gives

$$\widehat{L}\xi = Lv_1 \wedge v_2 \wedge \cdots v_n + v_1 \wedge Lv_2 \wedge \cdots \wedge v_n + \ldots + v_1 \wedge v_2 \wedge \cdots \wedge Lv_n\,.$$

This formula precisely expresses the fact that $\widehat{L}$ is a derivation of the exterior algebra $\wedge(V)$.

For bosons the argument is the same albeit for some sign changes.

Problem. Show that second quantization of one-body operators is a Lie algebra homomorphism, i.e., $\widehat{[L, M]} = [\widehat{L}, \widehat{M}]$. □

Adopting the standard physics conventions, we now choose our basis of V to be an orthonormal basis $\{|i\rangle\}_{i=1,2,\ldots}$ (Dirac notation), and we write $\varepsilon(e_i) \equiv c_i^+$, $\iota(f_i) \equiv c_i$, $\mu(e_i) \equiv a_i^+$, $\delta(f_i) \equiv a_i$. The second quantization rule (3.41) for fermions or bosons then takes the form

$$L \mapsto \widehat{L} = \begin{cases} \sum\limits_{i,j} \langle i|L|j\rangle\, c_i^+ c_j & \text{(fermions)}, \\ \sum\limits_{i,j} \langle i|L|j\rangle\, a_i^+ a_j & \text{(bosons)}. \end{cases} \tag{3.42}$$

Example. Consider Schrödinger particles confined to a cube C with side length L and periodic boundary conditions. The single-particle Hilbert space is $V = L^2(C)$ and an orthonormal basis of V is given by the plane waves

$$\psi_{\mathbf{k}}(\mathbf{r}) = L^{-3/2} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$$

for the discrete set of wave vectors

$$\mathbf{k} = (k_1, k_2, k_3) \in (2\pi\mathbb{Z}/L)^3 =: \Lambda\,.$$

Orthonormality here means that

$$\langle \mathbf{k}|\mathbf{k}'\rangle \equiv \int_C \overline{\psi_{\mathbf{k}}(\mathbf{r})}\,\psi_{\mathbf{k}'}(\mathbf{r})\,d^3r = \delta_{\mathbf{k}\mathbf{k}'}\,.$$

The one-body operator $T := -\hbar^2\nabla^2/2m$ of the Schrödinger kinetic energy has matrix elements

$$\langle \mathbf{k}|T|\mathbf{k}'\rangle = \frac{\hbar^2\mathbf{k}^2}{2m}\,\delta_{\mathbf{k}\mathbf{k}'}\,, \quad \mathbf{k}^2 = k_1^2 + k_2^2 + k_3^2\,,$$

and the second-quantized kinetic energy $\widehat{T}$ is

$$\widehat{T} = \sum_{\mathbf{k}\in\Lambda} \varepsilon_{\mathbf{k}}\, c_{\mathbf{k}}^+ c_{\mathbf{k}}\,, \quad \varepsilon_{\mathbf{k}} = \frac{\hbar^2\mathbf{k}^2}{2m}\,.$$

The operator $c_{\mathbf{k}}^+ c_{\mathbf{k}}$ is called the occupation number operator (for the single-particle state with momentum $\mathbf{p} = \hbar\mathbf{k}$). It has two eigenvalues: 0 (empty state) and 1 (occupied state).

Remark. The homomorphism property $\widehat{[L,M]} = [\widehat{L},\widehat{M}]$ does not single out (3.42) but also holds for other schemes of second quantization. For example, if we take the fermionic (resp. bosonic) Fock space to be $\wedge(V^*)$ (resp. $\mathrm{S}(V^*)$) it holds for

$$L \mapsto \widetilde{L} = \begin{cases} \sum\limits_i \varepsilon(-L^t f_i)\, \iota(e_i) & \text{(fermions)}, \\ \sum\limits_i \mu(-L^t f_i)\, \delta(e_i) & \text{(bosons)}. \end{cases} \tag{3.43}$$

The reason why this works is that $\mathrm{End}(V) \ni L \mapsto -L^t \in \mathrm{End}(V^*)$ is a Lie algebra isomorphism. In Dirac notation with $c_i^+ \equiv \varepsilon(f_i)$, $c_i \equiv \iota(e_i)$, $a_i^+ \equiv \mu(f_i)$, $a_i \equiv \delta(e_i)$ the new scheme reads

$$L \mapsto \widetilde{L} = \begin{cases} -\sum\limits_{i,j} \langle i|L|j\rangle\, c_j^+ c_i & \text{(fermions)}, \\ -\sum\limits_{i,j} \langle i|L|j\rangle\, a_j^+ a_i & \text{(bosons)}. \end{cases} \tag{3.44}$$

The fermionic version of this 'hole quantization' scheme will become relevant in the reinterpretation of the Dirac equation as a quantum field theory. □

We now turn briefly to the scheme of second quantization for two-body operators. By definition, a two-body operator is a linear operator

$$L:\ V \otimes V \to V \otimes V \tag{3.45}$$

defined on the tensor product of two single-particle Hilbert spaces. (While these may be different when the two particles are distinguishable, we continue to focus on the case of identical particles here.) If we work in the Schrödinger representation and $\varphi_i(\mathbf{r})$ $(i = 1, 2, \ldots)$ is a basis of single-particle wave functions, then the two-body matrix elements of L are expressed as

$$\langle ij|L|kl\rangle := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \overline{\varphi_i(\mathbf{r})\,\varphi_j(\mathbf{r}')}\,\big(L(\varphi_k \otimes \varphi_l)\big)(\mathbf{r},\mathbf{r}')\, d^3r\, d^3r'. \tag{3.46}$$

Example. The Coulomb interaction energy U_{C} between two particles with charge e is the two-body operator given by

$$\big(U_{\mathrm{C}}(\varphi_k \otimes \varphi_l)\big)(\mathbf{r},\mathbf{r}') = \frac{e^2}{4\pi\varepsilon_0\,|\mathbf{r}-\mathbf{r}'|}\,\varphi_k(\mathbf{r})\varphi_l(\mathbf{r}'). \tag{3.47}$$

It has two-body matrix elements

$$\langle ij|U_{\mathrm{C}}|kl\rangle = \frac{e^2}{4\pi\varepsilon_0}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\overline{\varphi_i(\mathbf{r})\,\varphi_j(\mathbf{r}')}\,\varphi_k(\mathbf{r})\varphi_l(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d^3r\, d^3r'. \;\square \tag{3.48}$$

Second quantization sends a two-body operator L to an operator $\widehat{L}$ on Fock space. Given a single-particle basis and the corresponding two-body matrix elements $\langle ij|L|kl\rangle$ this works as follows:

$$L \mapsto \widehat{L} = \begin{cases} \frac{1}{2}\sum\limits_{ijkl} \langle ij|L|kl\rangle\, c_i^+ c_j^+ c_l\, c_k & \text{(fermions)}, \\ \frac{1}{2}\sum\limits_{ijkl} \langle ij|L|kl\rangle\, a_i^+ a_j^+ a_l\, a_k & \text{(bosons)}. \end{cases} \tag{3.49}$$

The factors of $\frac{1}{2}$ and reversal of the order of annihilation operators are needed in order for the second-quantized operator $\widehat{L}$ to agree with the defining operator L on the two-particle subspace $\wedge^2(V)$ or $\mathrm{S}^2(V)$. Note that by the commutation relations for fermions/bosons one may (anti)-symmetrize the two-body matrix elements in the expression for $\widehat{L}$:

$$\widehat{L} = \begin{cases} \frac{1}{4}\sum_{ijkl} \big(\langle ij|L|kl\rangle - \langle ij|L|lk\rangle\big)\, c_i^+ c_j^+ c_l\, c_k & \text{(fermions)}, \\ \frac{1}{4}\sum_{ijkl} \big(\langle ij|L|kl\rangle + \langle ij|L|lk\rangle\big)\, a_i^+ a_j^+ a_l\, a_k & \text{(bosons)}. \end{cases} \tag{3.50}$$

The second summand $\langle ij|L|lk\rangle$ is called the exchange matrix element.

3.6 Second quantization of Dirac's theory

3.6.1 Motivation: problems with the Dirac equation

The Dirac equation was originally conceived as a relativistic wave equation for a single charged particle, by the analogy with the Schrödinger equation. In the non-relativistic limit with weak electromagnetic fields, the single-particle interpretation of the Dirac equation works just fine.

However, serious problems of interpretation arise in the presence of strong electromagnetic fields. An example of a problematic situation is the following. Addressing the Dirac equation $H\psi = E\psi$ for stationary states in the case of an electric scalar potential with a step,

$$V = \begin{cases} 0 & z < 0\,, \\ V_0 & z > 0\,, \end{cases}$$

let us look for a plane-wave solution ψ with incoming wave boundary conditions $\psi_{\text{in}} \sim \mathrm{e}^{\mathrm{i}kz}$. Thus we employ an ansatz as a sum

$$\psi = a_+ \mathrm{e}^{\mathrm{i}kz} \begin{pmatrix} E+mc^2 \\ 0 \\ \hbar kc \\ 0 \end{pmatrix} \Theta(-z) + a_- \,\mathrm{e}^{-\mathrm{i}kz} \begin{pmatrix} E+mc^2 \\ 0 \\ -\hbar kc \\ 0 \end{pmatrix} \Theta(-z) + b\,\mathrm{e}^{\mathrm{i}k'z} \begin{pmatrix} E-V_0+mc^2 \\ 0 \\ \hbar k'c \\ 0 \end{pmatrix} \Theta(z)$$

of reflected and transmitted waves. [We use the spinor representation in its standard form with gamma matrices given by (2.10).] The stationary Dirac equation $H\psi = E\psi$ is solved by this ansatz for $z < 0$ if $k = (\hbar c)^{-1}\sqrt{E^2 - (mc^2)^2}$ and for $z > 0$ if

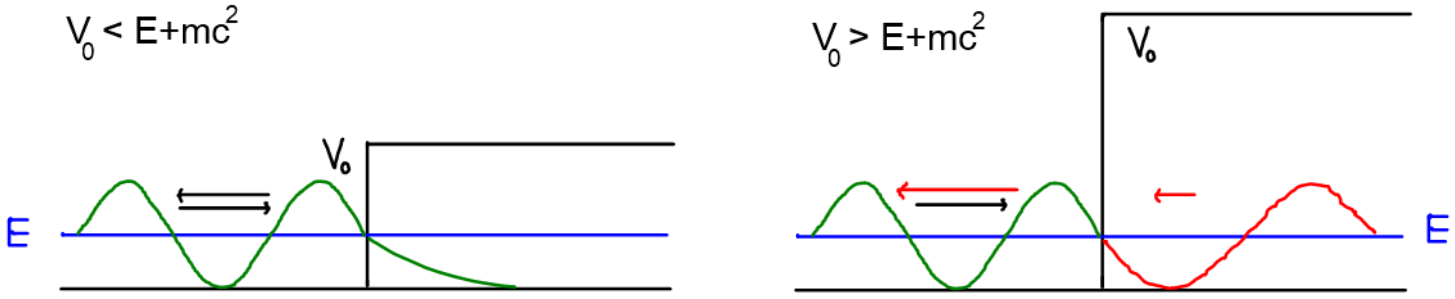
$$k' = \pm(\hbar c)^{-1}\sqrt{(E-V_0)^2 - (mc^2)^2}.$$

The ratios a_-/a_+ and b/a_+ are determined by requiring continuity of the solution ψ at $z = 0$.

The situation of interest now occurs when $V_0 > E + mc^2$. In that case we have a transmitted wave $\sim \mathrm{e}^{\mathrm{i}k'z}$ with real wave number $k' > 0$ and the stationary conserved probability current

$$j = \psi^\dagger \alpha_z \psi = 2(|a_+|^2 - |a_-|^2)(E+mc^2)\hbar kc = 2|b|^2(E - V_0 + mc^2)\hbar k'c < 0$$

is negative. This phenomenon (a.k.a. the Klein paradox) implies, in particular, that the reflection probability $|r|^2 = |a_-/a_+|^2$ exceeds unity! Certainly, no satisfactory interpretation of such a result exists within the confines of unitary single-particle quantum mechanics.



Mathematical analysis of the situation reveals that the paradoxical behavior has its reason in the existence of stationary solutions with negative energy $E - V_0 < -mc^2$.

Furthermore, the Dirac Hamiltonian, as it stands, does not have a ground state: it is bounded neither from below nor from above. When the interaction with the quantized electromagnetic field is switched on, the absence of a lower bound for the Dirac Hamiltonian poses the catastrophic threat that an infinite amount of electromagnetic energy might be released by Dirac particles making a never ending series of transitions to states with lower and lower energy.

These problems force us to abandon the single-particle interpretation of the Dirac equation. In fact, the correct physical interpretation of the result $|a_-/a_+|^2 > 1$ above is in terms of particle production (namely, the creation of electron-positron pairs at the potential step). The paradox is then resolved by reinterpreting the continuity equation for $(\rho, \vec{j}) = (\psi^\dagger\psi, \psi^\dagger\vec{\alpha}\psi)$ – which had originally been intended to be the law of conservation of probability – as the law of conservation of electric charge.

3.6.2 Stable second quantization

As we have tried to indicate, there are serious problems with the original form of Dirac's theory due to the fact that the Dirac Hamiltonian, being unbounded from below, does not have a ground state. Therefore, we now look for a reformulation of Dirac's theory such that the Hamiltonian does have a ground state. In the present section we make preparations for such a reformulation by taking the formalism of second quantization to its most general form.

Let $H \in \mathrm{End}(V)$ be any Hermitian Hamiltonian on a single-particle Hilbert space V; our interest, of course, is in Hamiltonians H which resemble the Dirac Hamiltonian in that they have positive as well as negative spectrum neither of which is bounded. We will describe a procedure of second quantization $H \mapsto \widehat{H}$ which sends H to a non-negative operator $\widehat{H} \geq 0$ on the Fock space $\wedge(P^*)$ for a suitably chosen subspace $P^* \subset V \oplus V^*$. Our procedure consists of two steps.

In the first step we convert $H \in \mathrm{End}(V)$ into an element $H^C \in \mathrm{Cl}(V \oplus V^*, B)$ of the Clifford algebra of the direct sum $V \oplus V^*$ with canonical symmetric form $B(v+\varphi, v'+\varphi') = \varphi(v') + \varphi'(v)$. To describe this step, let $\{e_1, e_2, \ldots\}$ be any basis of V and $\{f_1, f_2, \ldots\}$ be the dual basis of V^*. Expanding $He_j = \sum_i e_i H_{ij}$ with respect to that basis, we define

$$H^C := \sum_{ij} H_{ij}\, e_i\, f_j \tag{3.51}$$

as a quadratic element of the Clifford algebra $\mathrm{Cl}(V \oplus V^*, B)$. The Clifford algebra element H^C is seen to be invariantly defined, i.e., does not depend on the choice of basis. Note that our object H^C is still formal and abtract: we have not yet specified the space which H^C will be acting on.

Our second step is to choose an exterior algebra $\wedge(P^*)$ for $\mathrm{Cl}(V \oplus V^*, B)$ to act on. The choice will be engineered in precisely such a manner that the second-quantized Hamiltonian $\widehat{H}$ representing H^C comes out to be non-negative.

Assuming that H has no zero eigenvalue, let $\Pi_\pm$ be the uniquely defined orthogonal projection operators for the positive and negative eigenspaces $V_+ = \Pi_+ V$ and $V_- = \Pi_-$ of H. In formulas:

$$H = H_+ + H_-\,, \quad H_\pm = \Pi_\pm H \Pi_\pm\,, \quad H_+ > 0 > H_-\,. \tag{3.52}$$

We now recall from Section 2.8 that a spinor representation of the Clifford algebra $\mathrm{Cl}(W, B)$ is constructed by choosing a polarization

$$W = P \oplus P^* \tag{3.53}$$

by Lagrangian subspaces P, P^*. In the present setting, one such polarization $W = V \oplus V^*$ is given a priori. The key point, however, is that we should replace it by another one to obtain a well-behaved second-quantized Hamiltonian $\widehat{H}$. In fact, the good choice of polarization for our purposes is

$$P = V_+^* \oplus V_-\,, \quad P^* = V_+ \oplus V_-^*\,. \tag{3.54}$$

Thus we reverse the roles of vector space and dual space (and hence the definitions of creation and annihilation operator) when switching from the H-positive eigenspace V_+ to the H-negative eigenspace V_-.

We also recall from Section 2.8 that, given a polarization $W = P \oplus P^*$, the Clifford algebra generators $w \in W \subset \mathrm{Cl}(W, B)$ act by the spinor representation on $\xi \in \wedge(P^*)$ as a mixture

$$w \cdot \xi := \iota(w_P)\,\xi + \varepsilon(w_{P^*})\,\xi \tag{3.55}$$

of exterior multiplication $\varepsilon(w_{P^*})$ (or particle creation) for the P^*-component and contraction $\iota(w_P)$ (or particle annihilation) for the P-component of w.

Definition. Let $H = H_+ + H_-$ be the decomposition of a Hermitian Hamiltonian $H \in \mathrm{End}(V)$ into its positive and negative parts, neither of which is bounded. If $V = V_+ \oplus V_-$ is the corresponding decomposition into H-positive and H-negative subspaces, let $\{e_i\}$, $\{\tilde{e}_k\}$ be any two bases of V_+, V_-, and $\{f_j\}$, $\{\tilde{f}_l\}$ the dual bases of V_+^*, V_-^*, respectively. By the positive (or stable) second quantization of H we then mean the operator $\widehat{H}$ on $\wedge(V_+ \oplus V_-^*)$ defined by

$$\widehat{H} := \sum_{ij} (H_+)_{ij}\; \varepsilon(e_i)\, \iota(f_j) - \sum_{kl} (H_-)_{kl}\; \varepsilon(\tilde{f}_l)\, \iota(\tilde{e}_k). \tag{3.56}$$

Thus the elements of $P^* = V_+ \oplus V_-^*$ act as particle creation operators (ε) and those of $P = V_+^* \oplus V_-$ as particle annihilation operators (ι).

Remark. Note that by CAR the expression (3.56) can be rewritten as

$$\widehat{H} = \sum_{ij} (H_+)_{ij}\, \varepsilon(e_i)\, \iota(f_j) + \sum_{kl} (H_-)_{kl} \big(\iota(\tilde{e}_k)\, \varepsilon(\tilde{f}_l) - \delta_{kl} \big). \tag{3.57}$$

Written in this form, stable second quantization clearly implements the standard scheme (3.55) but for the subtraction of an (infinite) constant $\sum_{kl} (H_-)_{kl}\, \delta_{kl} = \operatorname{Tr} H_-$. This subtraction in combination with the adjustment of the order of operators $\iota(\tilde{e}_k)\, \varepsilon(\tilde{f}_l) - \delta_{kl} = -\varepsilon(\tilde{f}_l)\, \iota(\tilde{e}_k)$ is called normal ordering.

Fact. The stable second quantization $\widehat{H}$ defined by (3.56) is positive.

Proof. The operator $\widehat{H}$ is a derivation of the exterior algebra $\wedge(P^*)$. By this token it is positive on $\wedge(P^*)$ if it is positive on the degree-one subspace $P^* \subset \wedge(P^*)$. Now $\widehat{H}$ acts on $V_+ \subset P^*$ as $H_+ > 0$ and on V_-^* as $-H_-^t > 0$. Therefore $\widehat{H}$ is positive on $P^* = V_+ \oplus V_-^*$ as desired.

3.6.3 The idea of hole theory

In the next subsection, the mathematical idea of stable second quantization will be used to fix the problems with the Dirac equation. Before doing so, let us give a simplified discussion.

The main physical idea is to redefine what is meant by the vacuum. To consider a very simple example, let $|0\rangle$ be the (naive) vacuum for a one-dimensional single-particle Hilbert space $V \simeq \mathbb{C}$, and denote by c^+ (c) the particle creation (resp. annihilation) operator as usual. The Hamiltonian in this one-dimensional case is just a real number $H = h \in \mathbb{R}$. If h is positive, the standard scheme $H \mapsto \widehat{H} = h\, c^+ c \geq 0$ of second quantization is fine. On the other hand, for $h < 0$ the scheme must be changed if $\widehat{H}$ is to be non-negative. There exist two equivalent ways of going about this. One is to exchange the roles of $c^+ \leftrightarrow c$ and replace $c^+ c$ by $c c^+ - 1 = -c^+ c$ (that's what we did in Section 3.6.2). Equivalently, we may leave the operators the same and, instead, adjust the definition of the vacuum:

$$|\mathrm{vac}\rangle := c^+ |0\rangle.$$

Indeed, by the Pauli exclusion principle $c^+ c^+ = 0$ we now have $c^+ |\mathrm{vac}\rangle = 0$. Thus c^+ acts on the new vacuum as an annihilation operator. At the same time we have $c|\mathrm{vac}\rangle = |0\rangle \neq 0$, so c now plays the role of a creation operator. And by subtracting a constant from the Hamiltonian

$$\widehat{H} := h c^+ c - h$$

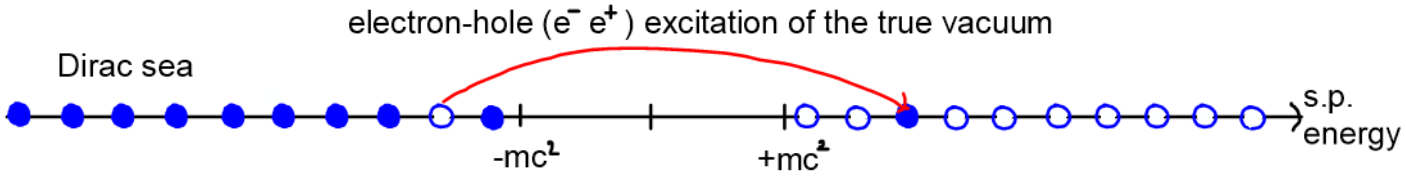
in order to make the energy of the new vacuum vanish (which amounts to a redefinition of the location of zero on the energy axis), we obtain

$$\widehat{H}(c|\mathrm{vac}\rangle) = -h\, (c|\mathrm{vac}\rangle)$$

and thus an excited state $c|\mathrm{vac}\rangle$ with positive energy eigenvalue $-h > 0$.

Going from this example to the setting of the Dirac equation, the idea of Dirac's hole theory is to define the true vacuum to be the state in which each negative-energy stationary solution of the Dirac equation is occupied by one electron (and hence Pauli blocked), while all positive-energy

solutions are unoccupied. The fully occupied negative-energy sector is often referred to as the 'Dirac sea'. Although the true vacuum $|\mathrm{vac}\rangle$ has infinite energy and infinite charge relative to the naive vacuum $|0\rangle$, one still commands the freedom of shifting the observables of energy and charge so that they vanish for $|\mathrm{vac}\rangle$. The elementary excitations of the true vacuum then are particles $c_+^+|\mathrm{vac}\rangle$ in positive-energy states and hole excitations $c_-|\mathrm{vac}\rangle$ of the Dirac sea. The latter have positive energy as well as positive charge (as the negative energy and negative charge of one Dirac-sea electron are missing); they are interpreted as the states of a new elementary particle predicted by this reinterpretation of the Dirac equation: the positron.



3.6.4 Mode expansion of the Dirac field

We now implement the scheme of stable second quantization for the Hamiltonian of the Dirac equation. For pedagogical reasons, we do not emphasize the relativistic covariance of the theory here, but fix an inertial frame and do all constructions at fixed and equal time $t = 0$. Therefore, some of our normalizations will differ from those used in standard text books on relativistic quantum mechanics. Recall now that the first step of our scheme is to convert the Hamiltonian – in the present case, the free Dirac Hamiltonian $H = mc^2\beta + c\sum \alpha_l p_l$ – into a quadratic Clifford algebra element H^C. Using the basis offered by spinors ψ in the position representation, this is achieved (in Dirac notation) by

$$H^C := \int d^3r \int d^3r' \sum_{\tau,\tau'=1}^{4} \psi_\tau^+(\mathbf{r}) \langle \mathbf{r}|H_{\tau\tau'}|\mathbf{r}'\rangle \psi_{\tau'}(\mathbf{r}'), \tag{3.58}$$

where ψ henceforth is called the Dirac field. Its components obey the canonical anti-commutation relations

$$\{\psi_\tau(\mathbf{r}), \psi_{\tau'}(\mathbf{r}')\} = 0\,, \quad \{\psi_\tau^+(\mathbf{r}), \psi_{\tau'}^+(\mathbf{r}')\} = 0\,, \quad \{\psi_\tau(\mathbf{r}), \psi_{\tau'}^+(\mathbf{r}')\} = \delta_{\tau\tau'}\,\delta(\mathbf{r}-\mathbf{r}'). \tag{3.59}$$

The second step is to specify a good Fock space $\wedge(P^*)$ for H^C to act on. For this we need the decomposition of the single-particle Hilbert space $V = L^2(\mathbb{R}^3) \otimes \mathbb{C}^4$ into its H-positive and H-negative subspaces V_+ and V_-. By the translation invariance of H the projection operators

$$(\Pi_\pm \varphi)_\tau(\mathbf{r}) = \int d^3r' \sum_{\tau'} (\Pi_\pm)_{\tau\tau'}(\mathbf{r},\mathbf{r}')\,\varphi_{\tau'}(\mathbf{r}') \tag{3.60}$$

on these subspaces $V_+ = \Pi_+ V$ and $V_- = \Pi_- V$ have Fourier expansions

$$(\Pi_\pm)_{\tau\tau'}(\mathbf{r},\mathbf{r}') = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} (\Pi_\pm)_{\tau\tau'}(\mathbf{k}), \tag{3.61}$$

where $\Pi_\pm(\mathbf{k})$ are the matrices of the projection operators at fixed wave vector $\mathbf{k}$. The latter are found by taking the Dirac Hamiltonian in the momentum representation,

$$H(\mathbf{k}) = \begin{pmatrix} mc^2 & \hbar c\vec{\sigma}\cdot\vec{k} \\ \hbar c\vec{\sigma}\cdot\vec{k} & -mc^2 \end{pmatrix}, \tag{3.62}$$

and diagonalizing it. The 4×4 matrix $H(\mathbf{k})$ squares to $(\hbar\omega(k))^2\,\mathrm{Id}_{\mathbb{C}^4}$ where

$$\hbar\omega(k) = \sqrt{(mc^2)^2 + (\hbar ck)^2}\,, \qquad k = |\mathbf{k}|\,. \tag{3.63}$$

Its eigenvalues are $\pm\hbar\omega(k)$, each with multiplicity two. Let $u_s(\mathbf{k})$ and $v_s(\mathbf{k})$ denote a corresponding set of eigenspinors:

$$H(\mathbf{k})u_s(\mathbf{k}) = +\hbar\omega(k)\,u_s(\mathbf{k})\,, \quad H(\mathbf{k})v_s(\mathbf{k}) = -\hbar\omega(k)\,v_s(\mathbf{k})\,, \tag{3.64}$$

with normalization $u_s^\dagger(\mathbf{k})u_{s'}(\mathbf{k}) = v_s^\dagger(\mathbf{k})v_{s'}(\mathbf{k}) = \delta_{ss'}$ and $u_s^\dagger(\mathbf{k})v_{s'}(\mathbf{k}) = 0$, where $\dagger$ means the adjoint w.r.t. the Hermitian structure of $\mathbb{C}^4$. Because of the two-fold spin degeneracy there remains some freedom in the definition of the quantum number s. One possible choice is to take $u_s(\mathbf{k})$ and $v_s(\mathbf{k})$ to be eigenvectors of the helicity operator $\vec{\Sigma}\cdot\vec{k}/k$ – cf. Eq. (2.97) – with eigenvalue $s = \pm 1$. That is the choice we make here. From the eigenspinors $u_s(\mathbf{k})$ and $v_s(\mathbf{k})$ we then construct the projection matrices

$$\Pi_+(\mathbf{k}) = \sum_{s=\pm1} u_s(\mathbf{k})\otimes u_s^\dagger(\mathbf{k})\,, \qquad \Pi_-(\mathbf{k}) = \sum_{s=\pm1} v_s(\mathbf{k})\otimes v_s^\dagger(\mathbf{k})\,, \tag{3.65}$$

and from these in turn the momentum space representation of the Hamiltonian:

$$H(\mathbf{k}) = \hbar\omega(k)\big(\Pi_+(\mathbf{k}) - \Pi_-(\mathbf{k})\big). \tag{3.66}$$

We are now in a position to specify the structure of fermionic Fock space $\wedge(P^*)$ for the Dirac field ψ in the momentum representation. To implement our general scheme, let $P^* = V_+ \oplus V_-^*$ for $V_\pm = \Pi_\pm V$. We regard the positive eigenspinors $u_s(\mathbf{k})$ as a basis for V_+, with their Hermitian adjoints $u_s^\dagger(\mathbf{k})$ giving the dual basis for V_+^*. We write $c^+_{+,s}(\mathbf{k}) \equiv \varepsilon\big(u_s(\mathbf{k})\big)$ for the creation operators on $\wedge(V_+)$, while $c_{+,s}(\mathbf{k}) \equiv \iota\big(u_s^\dagger(\mathbf{k})\big)$ stands for the corresponding annihilation operators. We say that $c_{+,\,s}(\mathbf{k})$ annihilates an electron with momentum $\hbar\mathbf{k}$ and helicity s. By definition, these creation and annihilation operators satisfy CAR with the non-vanishing anti-commutators being

$$\{c^+_{+,\,s}(\mathbf{k}), c_{+,\,s'}(\mathbf{k}')\} = \delta_{ss'}\int d^3r\,\mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = \delta_{ss'}(2\pi)^3\delta(\mathbf{k}-\mathbf{k}'). \tag{3.67}$$

Turning to the negative-energy sector, we regard the eigenspinors $v_s(\mathbf{k})$ as a basis for V_-, with dual basis $v_s^\dagger(\mathbf{k})$ for V_-^*. We write $c^+_{-,s}(\mathbf{k}) \equiv \varepsilon\big(v^\dagger_{-s}(-\mathbf{k})\big)$ for the creation operators on $\wedge(V_-^*)$; the annihilation operators are $c_{-,s}(\mathbf{k}) \equiv \iota\big(v_{-s}(-\mathbf{k})\big)$. We say that $c^+_{-,\,s}(\mathbf{k})$ creates a positron with momentum $\hbar\mathbf{k}$ and helicity s. The sign reversals $\mathbf{k}\to-\mathbf{k}$ and $s\to-s$ compensate for the overall change of sign (from $L\to -L^t$) in the second-quantization scheme, which affects the energy $L = H$ as well as the momentum $L = \mathbf{p}$ and the helicity $L = \vec{\Sigma}\cdot\vec{p}/p$. The (anti-)commutation relations

for c_- and c_-^+ are no different from those in the positive-energy sector; see Eq. (3.67). The true vacuum is the line $\wedge^0(V_+) \otimes \wedge^0(V_-^*)$ in $\wedge(V_+ \oplus V_-^*)$ which is annihilated by all the c_- and c_+.

By construction [from Eq. (3.56)], the resulting second-quantized Hamiltonian is positive:

$$\widehat{H} = \int \frac{d^3k}{(2\pi)^3}\, \hbar\omega(k) \sum_s \big(c_{+,s}^+(\mathbf{k})\, c_{+,s}(\mathbf{k}) + c_{-,s}^+(\mathbf{k})\, c_{-,s}(\mathbf{k})\big) \geq 0\,. \tag{3.68}$$

The operator for the total momentum carried by the Dirac field is expressed as

$$\widehat{\mathbf{p}} = \int \frac{d^3k}{(2\pi)^3}\, \hbar\mathbf{k} \sum_s \big(c_{+,s}^+(\mathbf{k})\, c_{+,s}(\mathbf{k}) + c_{-,s}^+(\mathbf{k})\, c_{-,s}(\mathbf{k})\big). \tag{3.69}$$

We finally return to the position representation to give an expression for the Dirac field $\psi(\mathbf{r})$. Decomposing $\psi = \Pi_+\psi + \Pi_-\psi$ we obtain

$$(\Pi_+\psi)(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \sum_s u_s(\mathbf{k})\, c_{+,s}(\mathbf{k})\,, \tag{3.70}$$

$$(\Pi_-\psi)(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \sum_s v_{-s}(-\mathbf{k})\, c_{-,s}^+(\mathbf{k})\,, \tag{3.71}$$

which is called the mode expansion of the Dirac field. Altogether we have

$$\psi(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \sum_s \big(u_s(\mathbf{k})\, c_{+,s}(\mathbf{k}) + v_s(\mathbf{k})\, c_{-,-s}^+(-\mathbf{k})\big)\,. \tag{3.72}$$

The corresponding expression for the dual field ψ^+ is

$$\psi^+(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \sum_s \big(c_{+,s}^+(\mathbf{k})\, u_s^\dagger(\mathbf{k}) + c_{-,-s}(-\mathbf{k})\, v_s^\dagger(\mathbf{k})\big)\,. \tag{3.73}$$

Problem 1. Show that the mode expansion (3.72, 3.73) obeys CAR [Eqs. (3.59)]. □

This concludes our discussion of the procedure of second quantization for the free Dirac field. What happens in the presence of an electromagnetic field will be explained in a later section.

Problem 2. From the first-quantized Dirac theory we recall the continuity equation $\dot\rho + \mathrm{div} j = 0$ for $j_l = c\psi^\dagger\alpha_l\psi$ and $\rho = \psi^\dagger\psi \geq 0$, which results in the conservation law $Q = \int d^3r \psi^\dagger\psi = \text{const}$ (independent of time). Show that the second-quantized expression for Q is

$$Q = \int \frac{d^3k}{(2\pi)^3} \sum_s \big(c_{+,s}^+(\mathbf{k})\, c_{+,s}(\mathbf{k}) - c_{-,s}^+(\mathbf{k})\, c_{-,s}(\mathbf{k})\big). \tag{3.74}$$

Concluding comment. Dirac's first intention was to formulate a relativistic quantum theory with positive probability density $\rho = \psi^\dagger\psi \geq 0$ and conserved $Q = \int d^3r\, \psi^\dagger\psi = 1$. This original plan had to be abandoned because it led to serious problems due to the sign-indefinite nature of the (first-quantized) Dirac Hamiltonian H. By going to the second-quantized version (or quantum field-theoretic re-interpretation) of the Dirac Hamiltonian one trades the sign-indefiniteness of H for the sign-indefiniteness of Q. The latter is then re-interpreted as the electric charge, which may be positive or negative.

3.6.5 Charge conjugation symmetry

In Sect. 2.13.2 we learned about charge conjugation C as a complex *anti-linear* symmetry of the Dirac equation. Having re-interpreted the Dirac equation as an equation for the Dirac field ψ, we are now ready for a different perspective on C: by passing to the second-quantized formulation, charge conjugation is turned into a complex *linear* operator and hence a unitary (as opposed to anti-unitary) symmetry of the Dirac Hamiltonian, in the absence of electromagnetic fields.

The principle by which this happens does not depend on the specifics of the Dirac theory, but is universal. We will now sketch that principle for the case of a generic Hamiltonian. The main tool to be introduced for this purpose is that of a *particle-hole transformation.*

Definition. For simplicity, let V be a single-particle Hilbert space of finite dimension N. Then a particle-hole transformation is a complex anti-linear mapping Φ (defined for all $0 \leq n \leq N$)

$$\Phi : \quad \wedge^n(V) \xrightarrow{h} \wedge^n(V)^* \xrightarrow{\omega^{-1}} \wedge^{N-n}(V), \tag{3.75}$$

where the first map is the obvious map $h : \ \Psi \mapsto \langle \Psi, \cdot \rangle_{\wedge(V)}$ given by the Hermitian structure of Fock space, and the second map is the inverse of a $\mathbb{C}$-linear isomorphism $\omega : \ \wedge^{N-n}(V) \to \wedge^n(V)^*$. For the latter one fixes some state of total occupation $\Psi \in \wedge^N(V)$ in order to determine a pairing $\Omega : \ \wedge^{N-n}(V) \times \wedge^n(V) \to \mathbb{C}$ by

$$\Omega(\xi, \xi')\, \Psi = \xi \wedge \xi'. \tag{3.76}$$

For $\xi \in \wedge^{N-n}(V)$ one then sets $\omega(\xi) := \Omega(\xi, \cdot\,) \in \wedge^n(V)^*$. We remark that by particle-hole transforming an operator $L : \ \wedge^n(V) \to \wedge^{n+k}(V)$ that changes the degree by k, one gets an operator $\Phi L\, \Phi^{-1} : \ \wedge^{N-n}(V) \to \wedge^{N-n-k}(V)$ that changes the degree by $-k$.

Problem. Show that for any $|v\rangle \in V$, $\langle u| \in V^*$, and $a\,, b \in \mathbb{C}$ one has

$$\Phi \left(a\, c^+_{|v\rangle} + b\, c_{\langle u|} \right) \Phi^{-1} = \overline{a}\, c_{\langle v|} + \overline{b}\, c^+_{|u\rangle}, \tag{3.77}$$

independent of the choice of $\Psi \in \wedge^N(V)$.

Fact. The second quantization $\widehat{A}$ of any Hermitian and traceless one-body operator A is odd under particle-hole transformation

$$\Phi \widehat{A}\, \Phi^{-1} = -\widehat{A}, \tag{3.78}$$

on the fermionic Fock space $\wedge(V)$. This statement is verified as follows. Let $\widehat{A} = \sum_{ij} \langle i|A|j\rangle\, c_i^+ c_j$ be an arbitrary one-body operator. Its particle-hole transform is computed to be

$$\Phi \left(\sum \langle i|A|j\rangle\, c_i^+ c_j \right) \Phi^{-1} = \sum \overline{\langle i|A|j\rangle}\, c_i c_j^+ = \sum \langle j|A^\dagger|i\rangle (-c_j^+ c_i + \delta_{ij}).$$

Thus if $A^\dagger = A$ and $\mathrm{Tr}_V A = 0$, the claimed property (3.78) follows. $\square$

Let now $A \equiv H = H^\dagger$ be a first-quantized Hamiltonian, and assume that there exists an anti-unitary operator $C : \ V \to V$ with the property

$$CHC^{-1} = -H. \tag{3.79}$$

It follows that if $|v\rangle \in V$ is an eigenvector of H with eigenvalue E, then $C|v\rangle$ is an eigenvector of H with eigenvalue $-E$. Thus C is a spectrum-inverting operation (not a symmetry!) for H. In particular, it follows that H is traceless ($\mathrm{Tr}_V H = 0$).

By second quantization, the spectrum-inverting operation C gives rise to a complex anti-linear operation $\widehat{C}: \wedge^n(V) \to \wedge^n(V)$ for all $0 \leq n \leq N$. The property (3.79) then transcribes to

$$\widehat{C}\widehat{H}\widehat{C}^{-1} = -\widehat{H}, \tag{3.80}$$

and by combining Eq. (3.80) with (3.78) we arrive at

$$(\Phi \circ \widehat{C})\widehat{H}(\Phi \circ \widehat{C})^{-1} = +\widehat{H}. \tag{3.81}$$

Thus the concatenation $\Phi \circ \widehat{C}$ commutes with the second-quantized Hamiltonian $\widehat{H}$; it is therefore a symmetry of $\widehat{H}$. Since both Φ and $\widehat{C}$ are complex anti-linear, the symmetry is complex linear and in fact *unitary*. Note, however, that $\Phi \circ \widehat{C}$ maps $\wedge^n(V)$ to $\wedge^{N-n}$; thus it relates the eigenvalues and eigenvectors of the Hamiltonian for different particle numbers, n and $N-n$.

To reap full benefit from the symmetry $\Phi \circ \widehat{C}$ one assumes that $\ker H = 0$ (i.e., H has no zero eigenvalues). As a consequence, the dimension N must be even. For even better results, one assumes that H has a spectral gap around zero energy. The ground state of $\widehat{H}$ in the distinguished sector of particle number $N/2$ (called "half filling") then is the complex line in $\wedge^{N/2}(V)$ with all negative-energy states occupied (the analog of the filled Dirac sea). This ground state transforms into itself under the application of $\Phi \circ \widehat{C}$. A one-particle excited state with positive energy E in $\wedge^{N/2+1}(V)$ transforms under $\Phi \circ \widehat{C}$ in a one-hole excited state with the same positive energy E in $\wedge^{N/2-1}(V)$. Thus $\Phi \circ \widehat{C}$ is a true unitary symmetry of the system at half filling.

So much for the generic situation. To apply the discussion above to the relativistic case of Dirac fermions, one identifies the spectrum-inverting operation C with the complex anti-linear charge conjugation operator $C: \psi \to \mathrm{i}\gamma^2\bar{\psi}$ of the first-quantized Dirac theory. Indeed, the free Dirac Hamiltonian $H = mc^2\beta + c\sum_l \alpha_l p_l$ (zero electromagnetic field) obeys $CHC^{-1} = -H$, assuming the standard representation for β and α_l. To account for the fact that $\dim V$ is now infinite, the definition (3.75) of particle-hole transformation Φ must be adapted. The main idea here is to set $n = N/2 + q$, $N - n = N/2 - q$ and excise the infinite Dirac sea at half filling, $\wedge^{N/2}(V)$. The particle-hole operator Φ is then defined as a complex anti-linear transformation from the sector with q particles to the sector with q holes (relative to the sea at half filling).

The concatenation $\Phi \circ \widehat{C}$ leaves the true vacuum (i.e., the infinite Dirac sea at half filling) invariant and maps a physical electron state with excitation energy ε to a physical positron state with the same energy ε. By construction, this complex linear operation $\Phi \circ \widehat{C}$ is a unitary symmetry of the free Dirac Hamiltonian. [It becomes the *charge conjugation symmetry* of quantum electrodynamics when understood to act on the electromagnetic field as $(E, B) \mapsto (-E, -B)$.]

3.7 Quantization of the electromagnetic field

Quantization of the electromagnetic field proceeds by the formalism of second quantization for bosons. Let us review that formalism and bring it into a form well suited to our present goals.

3.7.1 Review and consolidation

The general starting point of second quantization for bosons is a real vector space W — the phase space of generalized positions and momenta — with symplectic form $\alpha : W \times W \to \mathbb{R}$. (We here renege on our earlier practice of using the symbol A, as A in the present context stands for the electromagnetic gauge field.) More generally, in geometric quantization one would start from a symplectic manifold W, but since the Maxwell theory of the electromagnetic field is a linear theory, it suffices for our purposes to assume the setting of a vector space. We recall from Section 3.2 that the symplectic vector space (W, α) determines a Weyl algebra $\mathcal{W}(W, \alpha)$.

To quantize the theory and construct a quantum mechanical Hilbert space, it does not suffice just to have a symplectic structure α on W. One also needs a (compatible) complex structure J, i.e., a linear mapping

$$J \in \mathrm{End}(W)\,, \quad J^2 = -\mathrm{Id}_W\,, \tag{3.82}$$

subject to the requirement that J induces a Euclidean scalar product

$$g : W \times W \to \mathbb{R}\,, \quad (w, w') \mapsto g(w, w') = g(w', w)\,, \quad g(w, w) \geq 0\,, \tag{3.83}$$

by the formula

$$\forall w, w' \in W : \quad g(w, w') = \alpha(w, Jw'). \tag{3.84}$$

(In the more general case of a symplectic manifold W, the complex structure J is to induce a Riemannian metric and a compatible Kähler structure.)

Problem. Deduce from (3.84) that $g(Jw, Jw') = g(w, w')$ and $\alpha(Jw, Jw') = \alpha(w, w')$. □

The choice of g isn't a big issue as long as $\dim W < \infty$, but it does become essential in the case of an infinite-dimensional phase space W where different choices of g lead to inequivalent quantizations. In our treatment of the quantized electromagnetic field, the Euclidean structure g will be provided by the electromagnetic energy, which is a positive function.

Example. Consider the symplectic plane

$$W = \mathrm{span}_{\mathbb{R}}\{e_q\,, e_p\} \simeq \mathbb{R}^2$$

with generators e_q and e_p. The symplectic form (dual to the Poisson bracket) is

$$\alpha(e_p\,, e_q) = -\alpha(e_q\,, e_p) = 1\,, \qquad \alpha(e_q\,, e_q) = \alpha(e_p\,, e_p) = 0\,.$$

You should think of the dual basis functions q, p as the variables of position and momentum. In the language of differential forms, one writes $\alpha = dp \wedge dq$.

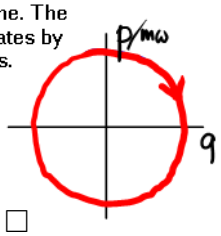
We now equip the vector space W with a Euclidean structure g (i.e., a Euclidean scalar product g) by picking some harmonic oscillator Hamiltonian $H = p^2/2m + m\omega^2 q^2/2$ and adopting the natural interpretation of H as a quadratic form $H : W \to \mathbb{R}$ to set

$$g(e_q\,, e_q) := \frac{2}{\omega}\, H(e_q) = m\omega\,, \qquad g(e_p\,, e_p) := \frac{2}{\omega}\, H(e_p) = (m\omega)^{-1},$$

and $g(e_q\,, e_p) = g(e_p\,, e_q) = 0$. The complex structure $J \in \mathrm{End}(W)$ is then given by

$$Je_q = -m\omega\, e_p\,, \quad Je_p = (m\omega)^{-1} e_q\,,$$

The harmonic oscillator flow is uniform rotation in the phase plane. The operator J rotates by ninety degrees.



or equivalently, in terms of the canonical transpose $J^t \in \mathrm{End}(W^*)$ of J,

$$J^t q = (m\omega)^{-1} p\,, \quad J^t p = -m\omega q\,.$$

Note that J is the operator of time evolution by one quarter of the oscillator period. □

We continue with the general development. The complex structure J determines a polarization

$$W \otimes \mathbb{C} = V \oplus \widetilde{V} \tag{3.85}$$

by complex subspaces V and $\widetilde{V}$ which are acted upon by J as $J|_V = +\mathrm{i}$ and $J|_{\widetilde{V}} = -\mathrm{i}$ (hence the name 'complex structure'). Thus V and $\widetilde{V}$ are the eigenspaces of J corresponding to the eigenvalues $+\mathrm{i}$ and $-\mathrm{i}$ respectively. The complex vector space $\widetilde{V}$ is called the holomorphic part of $W \otimes \mathbb{C}$, while V is the anti-holomorphic part. Note the characterization

$$V = \{w - \mathrm{i}Jw \mid w \in W\}\,, \qquad \widetilde{V} = \{w + \mathrm{i}Jw \mid w \in W\}. \tag{3.86}$$

Thus V is isomorphic to W as a real vector space by $W \xrightarrow{\sim} V$, $w \mapsto w - \mathrm{i}Jw$, and an analogous statement holds for $\widetilde{V}$.

Let the complex linear extension of α to $W \otimes \mathbb{C}$ still be denoted by the same symbol α.

Problem. Show that the pairing

$$\alpha : V \otimes \widetilde{V} \to \mathbb{C}\,, \quad v \otimes \tilde{v} \mapsto \alpha(v, \tilde{v})\,, \tag{3.87}$$

between V and $\widetilde{V}$ is non-degenerate. Show also that the subspaces V and $\widetilde{V}$ are Lagrangian, i.e., $\alpha(v, v') = \alpha(\tilde{v}, \tilde{v}') = 0$ for any two vectors $v, v' \in V$ and $\tilde{v}, \tilde{v}' \in \widetilde{V}$. □

By the pairing between the complex vector spaces V and $\widetilde{V}$ we may identify $\widetilde{V}$ with the dual vector space V^* of V. Planck's constant $\hbar$ has not appeared in the discussion so far, but it now does. The point is that there is some freedom in making the identification $\widetilde{V} \simeq V^*$. Concretely put, one has the freedom of inserting a multiplicative constant $(-\mathrm{i}/\hbar)$ in the isomorphism

$$I : \widetilde{V} \to V^*, \quad \tilde{v} \mapsto -\frac{\mathrm{i}}{\hbar}\, \alpha(\tilde{v}, \cdot). \tag{3.88}$$

By this choice of quantization constant we henceforth identify $\widetilde{V} \ni \tilde{v} \equiv I(\tilde{v}) \in V^*$.

Problem. In the context of our example (the symplectic plane), show that

$$I\,(e_q + \mathrm{i}Je_q) = \frac{\mathrm{i}}{\hbar}\,(p - \mathrm{i}J^t p).$$

The interpretation of this result is that $p = -\mathrm{i}\hbar\, I(e_q) = \frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial q}$. $\square$

Definition/Fact. The holomorphic Lagrangian subspace $\widetilde{V} \subset W \otimes \mathbb{C}$ carries a Hermitian scalar product $h:\ \widetilde{V} \times \widetilde{V} \to \mathbb{C}$ by

$$h(\tilde{v}, \tilde{v}') \equiv h(w + \mathrm{i}Jw, w' + \mathrm{i}Jw') \stackrel{\text{def}}{=} -\frac{\mathrm{i}}{\hbar}\,\alpha(w - \mathrm{i}Jw, w' + \mathrm{i}Jw'). \tag{3.89}$$

Remark. The same can be done on V. Concerning h on $\widetilde{V}$, notice that

$$\mathfrak{Re}\, h(\tilde{v}, \tilde{v}') = \frac{2}{\hbar}\, g(w, w'), \quad \mathfrak{Im}\, h(\tilde{v}, \tilde{v}') = -\frac{2}{\hbar}\,\alpha(w, w'). \tag{3.90}$$

It follows that h has the required properties $\overline{h(\tilde{v}, \tilde{v}')} = h(\tilde{v}', \tilde{v})$ and $h(\tilde{v}, \tilde{v}) \geq 0$. $\square$

The definition (3.89) turns $\widetilde{V} \simeq V^*$ into a complex Hilbert space. The bosonic Fock space then is the symmetric algebra $\mathrm{S}(V^*)$ equipped with the Hermitian scalar product which is induced by that of $V^* \simeq \widetilde{V}$; see Section 3.4 for the details. The Weyl algebra of the symplectic vector space (W, α) acts on $\mathrm{S}(V^*)$ by

$$w \cdot \xi = \delta(w_V)\, \xi + \mu(w_{V^*})\, \xi\,, \qquad w \in W, \quad \xi \in \mathrm{S}(V^*). \tag{3.91}$$

As before in Section 3.4, the symbol $\delta(w_V)$ stands for the operator of derivation by the V-component of w, and $\mu(w_{V^*})$ means multiplication by the V^*-component of w.

Problem. Returning to our previous example, namely the symplectic plane $W = \mathrm{span}_{\mathbb{R}}\{e_q\,, e_p\}$ with mass parameter m and oscillator frequency ω, show that

$$\widetilde{V} \equiv V^* = \mathbb{C} \cdot a^+, \quad a^+ = \frac{1}{\sqrt{2}}\left(\frac{q}{\ell} - \mathrm{i}\frac{\ell p}{\hbar}\right), \quad \ell = \sqrt{\frac{\hbar}{m\omega}}\,.$$

Using the isomorphism $V^* \simeq \widetilde{V}$ and the Hermitian scalar product h prove that a^+ is a unit vector: $h(a^+, a^+) = 1$. Finally, use the Fock space scalar product induced by h to compute the Hermitian adjoint of a^+ and show that

$$(a^+)^\dagger = a = \frac{1}{\sqrt{2}}\left(\frac{q}{\ell} + \mathrm{i}\frac{\ell p}{\hbar}\right). \square$$

To turn our quantization procedure into a tool available for direct use, we must take into account that physical observables are *functions* (not points) on the phase space W. The elementary functions on W are the linear functions of position $q:\ W \to \mathbb{R}$ and momentum $p:\ W \to \mathbb{R}$. For such a function $\varphi \in W^*$ we now ask how it becomes an operator $\widehat{\varphi}$ on the Fock space $\mathrm{S}(V^*)$. For later use, the answer will be written in components with respect to a basis. Thus let $\{e_\lambda\}$ and $\{\tilde{e}_\lambda\}$ be orthonormal bases of V and $\tilde{V}$, with dual bases $\{f_\lambda\}$ and $\{\tilde{f}_\lambda\}$ of V^* and $\tilde{V}^*$, respectively. We arrange for $\tilde{f}_\lambda = I(e_\lambda)$ and $f_\lambda = -I(\tilde{e}_\lambda)$ by the isomorphism (3.88). A general element $\varphi \in W^*$ then expands as

$$\varphi = \sum\nolimits_\lambda \big(\varphi(e_\lambda) f_\lambda + \varphi(\tilde{e}_\lambda)\tilde{f}_\lambda\big). \tag{3.92}$$

Quantization on $\mathrm{S}(V^*)$ sends $f_\lambda \to \mu(f_\lambda) \equiv a_\lambda^+$ and $\tilde{f}_\lambda \to \delta(e_\lambda) \equiv a_\lambda$. Hence,

$$\widehat{\varphi} = \sum\nolimits_\lambda \big(\varphi(e_\lambda)\, a_\lambda^+ + \varphi(\tilde{e}_\lambda)\, a_\lambda\big). \tag{3.93}$$

3.7.2 More preparations

Beginning with a quick reminder of some relevant material from classical electrodynamics, we now apply the general formalism of (second) quantization to the electromagnetic field.

The electromagnetic gauge field is a one-form $\mathcal{A} = -\Phi\,dt + \sum A_j\,dx_j$. From it one obtains the electromagnetic field strength, a two-form $F = B + E \wedge dt$, by taking the exterior derivative $F = d\mathcal{A}$. In components one has

$$E_j = -\frac{\partial \Phi}{\partial x_j} - \frac{\partial A_j}{\partial t}\,, \qquad B_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}\,. \tag{3.94}$$

The components of the excitations $D = \sum_{i<j} D_{ij}\,dx_i \wedge dx_j$ and $H = \sum H_j\,dx_j$ are given by

$$D_{ij} = \varepsilon_0 \sum\nolimits_k \epsilon_{ijk}\,E_k\,, \qquad B_{ij} = \mu_0 \sum\nolimits_k \epsilon_{ijk}\,H_k\,, \tag{3.95}$$

where ε_0 and μ_0 are called the dielectric constant and magnetic permeability of the vacuum, respectively (and ϵ_{ijk} is an orientation-twisted variant of the totally anti-symmetric tensor on $\mathbb{R}^3$). A short-hand form of these so-called constitutive laws is

$$D = \varepsilon_0 \star E\,, \qquad B = \mu_0 \star H\,, \tag{3.96}$$

or $\star D = \varepsilon_0\,E$, $\star B = \mu_0\,H$. The fields D and B are subject to the constraints $dB = 0$ and $dD = \rho$, where $\rho = \tilde{\rho}\,d^3r$ is the three-form of the electric charge density. The dynamical equations of motion are Faraday's law $\dot{B} = -dE$ of induction and the Ampere-Maxwell law $\dot{D} = +dH$.

Notation. Using the traditional vector notation, one expresses the two-form components $D_{ij} = -D_{j\,i}$ as $D_{ij} = \sum \epsilon_{ijk}\,D_k$ and combines the three components into a vector $\mathbf{D} := (D_1, D_2, D_3)$. The same convention is used for the magnetic field strength B. With this convention, the constitutive laws read $\mathbf{D} = \varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu_0\mathbf{H}$. The constraints are $\operatorname{div}\mathbf{B} = 0$ and $\operatorname{div}\mathbf{D} = \tilde{\rho}$, and the equations of motion are $\dot{\mathbf{B}} = -\operatorname{rot}\mathbf{E}$ and $\dot{\mathbf{D}} = \operatorname{rot}\mathbf{H}$. □

We will be using the Hamiltonian formalism (as opposed to the Feynman path integral formalism). The standard textbook approach is to work with the gauge field $\mathcal{A}$ and choose the temporal gauge $\Phi = 0$. In this case the magnetic vector potential $A = \sum A_j\,dx_j$ plays the role of the position variable q of Hamiltonian mechanics, and the electric field strength $E = -\dot{A}$ acquires the meaning of (minus the) velocity $\dot{q}$. The electric excitation D corresponds to the momentum p.

However, the general framework developed in Section 3.7.1 is rather flexible: it allows us to circumvent the introduction of the gauge field (and the ensuing need for gauge fixing) and instead work with the physically observable field strengths and excitations directly. This is what we shall do in the sequel. The equations will be set up in a way that likens the electric excitation D to the position variable and the magnetic field strength B to the momentum variable.

The first step of quantization is to identify the phase space of the electromagnetic field. This will be the space W of all solutions to Maxwell's equations. W is an affine space, as the difference of two solutions in the presence of sources (i.e., electric charges) is a solution of the vacuum Maxwell equations, which form a linear system. In the current section, we will address only the

somewhat simplified problem of quantizing the electromagnetic field in the absence of charged matter, in which case W is a vector space. (Indeed, the sum of two vacuum solutions is another vacuum solution).

3.7.3 Symplectic structure

To quantize the theory following the general scheme of Section 3.7.1, we need a symplectic and a complex structure on the space W of vacuum solutions. We begin with the symplectic structure.

Because every solution is determined by a set of initial data, say $B|_{t=0}$ and $D|_{t=0}$, one can compute the symplectic structure α as an integral over the Euclidean space $\mathbb{R}^3$ at the initial time $t = 0$. For mathematical simplicity, let us assume that the electromagnetic field is confined to a bounded region $U \subset \mathbb{R}^3$ with Dirichlet boundary conditions $D|_{\partial U} = 0 = B|_{\partial U}$ at the surface ∂U. (For a precise treatment, one might also take the domain U to be connected and simply connected.) The following definition of the symplectic structure is modeled after the expression $\alpha = dp \wedge dq$ (or $\alpha = p \wedge q = -q \wedge p$ in the case of a linear system) of Section 3.7.1.

Notation. If B is the two-form of the magnetic field strength (or, for that matter, any closed two-form) then by $d^{-1}B$ we mean any one-form A such that $dA = B$. Equivalently, using the language of vector calculus, if $\mathbf{B}$ is the magnetic field (or, for that matter, any divergenceless vector field) then by $\mathrm{rot}^{-1}\mathbf{B}$ we mean any magnetic vector potential $\mathbf{A}$ such that $\mathrm{rot}\mathbf{A} = \mathbf{B}$.

Fact. Let D and B be closed (i.e. $dD = dB = 0$). Then the integral

$$\int_U D \wedge d^{-1}B = \int_U \mathbf{D} \cdot \mathrm{rot}^{-1}\mathbf{B}\, d^3r$$

is independent of the choice of one-form $d^{-1}B$ provided that D vanishes on ∂U. Indeed, if A and A' satisfy $B = dA = dA'$ then by the Poincaré lemma there exists a function f such that $A - A' = df$ and

$$\int_U D \wedge A - \int_U D \wedge A' = \int_U D \wedge df = \int_{\partial U} fD - \int_U f dD = 0\,.$$

In vector notation, the same computation for the case of a gauge transformation $\mathbf{A} = \mathbf{A}' + \mathrm{grad} f$ looks as follows:

$$\int_U \mathbf{D} \cdot (\mathbf{A} - \mathbf{A}')\, d^3r = \int_U \mathbf{D} \cdot \mathrm{grad} f\, d^3r = \int_{\partial U} f\, \mathbf{D} \cdot d^2\mathbf{n} - \int_U f\, \mathrm{div}\, \mathbf{D}\, d^3r = 0\,.$$

Remark 1. From magnetostatics one knows that if the equation $\mathbf{B} = \mathrm{rot}\mathbf{A}$ is solved for $\mathbf{A}$ in the Coulomb gauge $\mathrm{div}\mathbf{A} = 0$ (or, in the language of differential forms, $B = dA$ with $d \star A = 0$) then one has (for $U = \mathbb{R}^3$)

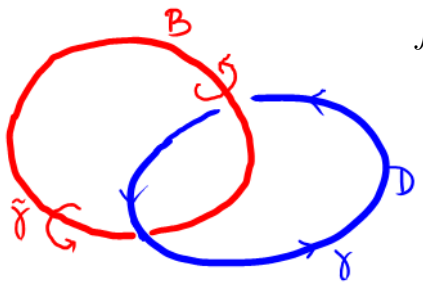
$$\mathbf{A}(\mathbf{r}) = \int_{\mathbb{R}^3} \frac{\mathrm{rot}\mathbf{B}(\mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|}\, d^3r'.$$

Thus with this choice of gauge we obtain the expression

$$\int_{\mathbb{R}^3} \mathbf{D} \cdot \mathrm{rot}^{-1}\mathbf{B}\, d^3r = \int_{\mathbb{R}^3} \frac{\mathbf{D}(\mathbf{r}) \cdot \mathrm{rot}\mathbf{B}(\mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|}\, d^3r' d^3r.$$

Remark 2. Under certain special circumstances the integral $\int D \wedge d^{-1}B$ has a beautiful geometric (actually, topological) interpretation. Suppose that both D and B are localized inside tubular neighborhoods of two curves, say γ and $\widetilde{\gamma}$ respectively. Suppose further that these tubular neighborhoods do not intersect each other. Then it is possible to show that our integral coincides with the Gauss linking number $G(\gamma, \gamma')$ of the two curves, multiplied by the product of electric ($\Phi_D = \int D$) and magnetic ($\Phi_B = \int B$) fluxes circulating in the tubular neighborhoods of the two curves:

$$\int_U D \wedge d^{-1}B = \Phi_D \, \Phi_B \, G(\gamma, \gamma').$$



This figure shows a pair of curves with Gauss linking number $G(\gamma, \widetilde{\gamma}) = 1$.

Fact. The symplectic structure α on the space W of vacuum solutions, as represented by the initial data $D \equiv D|_{t=0}$ and $B \equiv B|_{t=0}$, is expressed by

$$\begin{aligned} \alpha\big((D,B),(D',B')\big) &= \int_U (D' \wedge d^{-1}B - D \wedge d^{-1}B') \\ &= \int_U (\mathbf{D}' \cdot \mathrm{rot}^{-1}\mathbf{B} - \mathbf{D} \cdot \mathrm{rot}^{-1}\mathbf{B}')\, d^3r\,. \end{aligned} \tag{3.97}$$

In particular, α is non-degenerate on W. □

How do we know that α as defined in (3.97) is the proper symplectic form to use? The answer is that α in combination with the Hamiltonian function (the electromagnetic energy) gives the correct Hamiltonian equations of motion, $\dot{B} = -dE$ and $\dot{D} = dH$.

To conclude this subsection, we make a dimension check. Recall that D and B have the physical dimensions

$$[D] = \text{charge}\,, \qquad [B] = [A] = \text{energy/current}\,. \tag{3.98}$$

Equivalently, the components have the physical dimensions

$$[D_{ij}] = \text{charge/area}\,, \quad [B_{ij}] = \text{energy/(current}\times\text{area)}\,, \quad [A_j] = \text{energy/(current}\times\text{length)}\,.$$

It follows that $\int D \wedge d^{-1}B$ has the physical dimension of

$$\text{charge} \times \frac{\text{energy}}{\text{current}} = \text{energy} \times \text{time} = \text{action}\,. \tag{3.99}$$

The process of quantization makes α dimensionless by measuring it in units of Planck's constant.

3.7.4 Complex structure

The complex structure on the space W of vacuum solutions is determined by considering the total electromagnetic energy,

$$\mathcal{H} = \frac{1}{2}\int_U (D \wedge E + B \wedge H) = \frac{1}{2\varepsilon_0}\int_U |\mathbf{D}|^2\, d^3r + \frac{1}{2\mu_0}\int_U |\mathbf{B}|^2\, d^3r \geq 0\,, \tag{3.100}$$

which is the classical Hamiltonian function of the theory. The details are as follows.

The equations of motion $\dot{B} = -dE$ and $\dot{D} = dH$ for the electromagnetic field on the bounded domain U have a fundamental system of periodic solutions, the so-called normal modes. Explicit expressions for these can be given if U has a simple regular shape such as a cube or a ball; otherwise one is only assured of their existence and needs a numerical algorithm (typically involving discretization on a grid) to compute them. To find the normal modes, one looks for solutions of the wave equation

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta\right) f_t = 0\,, \qquad f_t\big|_{\partial U} = 0\,, \tag{3.101}$$

of stationary type. (Δ is the Laplacian, and $c = 1/\sqrt{\varepsilon_0\mu_0}$ is the speed of light.) The ansatz $f_t = f\,\mathrm{e}^{-\mathrm{i}\omega t}$ leads to the Helmholtz equation

$$\left(\omega^2/c^2 + \Delta\right) f = 0\,, \qquad f\big|_{\partial U} = 0\,. \tag{3.102}$$

For the type of domain U which we envisage, solutions of this equation exist only for a discrete set of characteristic frequencies ω_λ.

Problem. Show that if D is a closed electric-type two-form solution of the Helmholtz equation with frequency ω, then $d\star D = \varepsilon_0\, dE$ is a closed magnetic-type two-form solution with the same frequency. In vector notation, the statement is that by taking the curl of a divergenceless vector solution $\mathbf{D}$ one gets a divergenceless (axial) vector solution $\mathrm{rot}\,\mathbf{D}$. □

In the following, let (D^λ, B^λ) denote the electric and magnetic initial data of any two-form solution of the Helmholtz equation with frequency ω_λ. Such a pair (D^λ, B^λ) is called a normal mode. The normal modes for a fixed characteristic frequency ω_λ form a vector space W_λ. If there are no degeneracies in the spectrum of the Laplacian (that's the generic situation for a domain U of arbitrary shape which we henceforth assume) W_λ is two-dimensional. The symplectic form α on W restricts to a non-degenerate symplectic form α_λ on W_λ. In fact, one has

$$\alpha\left((0, d\star D^\lambda), (D^\lambda, 0)\right) = \int_U D^\lambda \wedge \star D^\lambda = \int_U |\mathbf{D}^\lambda|^2\, d^3r \neq 0\,,$$

and similarly, $\alpha\left((d\star B^\lambda, 0), (0, B^\lambda)\right) \neq 0$, unless $D^\lambda = B^\lambda \equiv 0$.

The complex structure J acts diagonally w.r.t. the decomposition $W = \oplus W_\lambda$ by normal modes. To compute its action on any one of the subspaces $W_\lambda \subset W$, one computes the time evolution of the initial data $(D^\lambda, B^\lambda) \subset W_\lambda$ for one quarter of the period $T_\lambda = 2\pi/\omega_\lambda$.

We are now in a position to get specific.

Definition/Fact. The complex structure acts on W_λ as

$$J\big(D^\lambda, B^\lambda\big) = \left(\frac{d\star B^\lambda}{\omega_\lambda\,\mu_0}\,,\; -\frac{d\star D^\lambda}{\omega_\lambda\,\varepsilon_0}\right), \qquad J\big(\mathbf{D}^\lambda, \mathbf{B}^\lambda\big) = \left(\frac{\mathrm{rot}\,\mathbf{B}^\lambda}{\omega_\lambda\,\mu_0}\,,\; -\frac{\mathrm{rot}\,\mathbf{D}^\lambda}{\omega_\lambda\,\varepsilon_0}\right). \tag{3.103}$$

Problem. Verify from the equations of motion $\dot{D} = d\star B/\mu_0$ and $\dot{B} = -d\star D/\varepsilon_0$ that $J(D^\lambda, B^\lambda)$ as specified is indeed the quarter-period time evolution of (D^λ, B^λ). Show also that $J^2 = -\mathrm{Id}$. (Hint: for the latter, the key step is to show that $(d\star)^2$ agrees with the Laplacian $-\Delta$ on closed two-forms. In vector notation this means that $\mathrm{rot} \circ \mathrm{rot} = -\Delta$ on divergenceless vector fields.)

3.7.5 Fock space: multi-photon states

With the symplectic and complex structures in hand, we now take a look at the induced Euclidean scalar product g. By implementing the general prescription of (3.84) we obtain

$$\begin{aligned} g\left((D^\lambda,B^\lambda),(D^\lambda,B^\lambda)\right) &= \alpha\left((D^\lambda,B^\lambda),J(D^\lambda,B^\lambda)\right) \\ &= \frac{1}{\omega_\lambda\,\varepsilon_0}\int_U D^\lambda\wedge\star D^\lambda + \frac{1}{\omega_\lambda\,\mu_0}\int_U B^\lambda\wedge\star B^\lambda = \frac{2}{\omega_\lambda}\,\mathcal{H}(D^\lambda,B^\lambda)\geq 0\,, \end{aligned} \tag{3.104}$$

for any normal mode $(D^\lambda,B^\lambda)\in W_\lambda$. Thus the Euclidean length squared of a normal mode is equal to (twice) the electromagnetic energy of that mode divided by its characteristic frequency.

Problems. (i) Show that the normal mode subspaces W_λ are orthogonal to one another with respect to the Euclidean structure g. (ii) Verify the validity of the invariance properties $g(Jw,Jw')=g(w,w')$ and $\alpha(Jw,Jw')=\alpha(w,w')$ in the present context. □

Next we use the complex structure J to polarize the complexified phase space $W\otimes\mathbb{C}=V\oplus\widetilde{V}$ by two complex Lagrangian subspaces $V=\oplus V_\lambda$ and $\widetilde{V}=\oplus\widetilde{V}_\lambda$. This can be done separately for each normal mode λ. Hence, for every λ let $\widetilde{V}_\lambda$ be the complex one-dimensional space

$$\widetilde{V}_\lambda=\mathbb{C}\cdot\left((D^\lambda,B^\lambda)+\mathrm{i}J(D^\lambda,B^\lambda)\right) \tag{3.105}$$

spanned by any (non-vanishing) normal mode $(D^\lambda,B^\lambda)\subset W_\lambda$. Without loss of generality we may single out the purely electric normal mode in W_λ. A special generator,

$$\tilde{e}_\lambda:=\frac{1}{\sqrt{2}}\left((D_0^\lambda\,,0)+\mathrm{i}J(D_0^\lambda\,,0)\right)\in\widetilde{V}_\lambda\,, \tag{3.106}$$

is then fixed by normalizing to the oscillator ground state energy:

$$\mathcal{H}(D_0^\lambda\,,0)=\tfrac{1}{2}\hbar\omega_\lambda\,. \tag{3.107}$$

By recalling the Hermitian scalar product h which is given by $\mathfrak{Re}\,h=2g/\hbar$ and $\mathfrak{Im}\,h=-2\alpha/\hbar$ on the holomorphic subspace $\widetilde{V}$ (cf. Section 3.7.1), we see that the normalized mode $\tilde{e}_\lambda$ is a unit vector of h. The corresponding generator of the anti-holomorphic part $V_\lambda\subset W_\lambda\otimes\mathbb{C}$ is

$$e_\lambda:=\frac{1}{\sqrt{2}}\left((D_0^\lambda\,,0)-\mathrm{i}J(D_0^\lambda\,,0)\right). \tag{3.108}$$

We now form the orthogonal sums $V:=\oplus_\lambda V_\lambda$ and $\widetilde{V}:=\oplus_\lambda\widetilde{V}_\lambda$ and thus have a polarization

$$W\otimes\mathbb{C}=V\oplus\widetilde{V} \tag{3.109}$$

by Lagrangian holomorphic and anti-holomorphic subspaces. Next, we recall from Sect. 3.7.1 our notation $\{f_\lambda\}$ for the dual basis of $V^*\simeq\widetilde{V}$. The Hilbert space of the quantized electromagnetic field is the symmetric algebra $\mathrm{S}(V^*)$ completed to an L^2-space by the Fock space Hermitian scalar product due to h. The state vectors in the degree-n subspace $\mathrm{S}^n(V^*)$ are called n-photon states. As usual, the multiplication operators $a_\lambda^+=\mu(f_\lambda)$ act on the Fock space $\mathrm{S}(V^*)$ as photon creation operators $a_\lambda^+:\ \mathrm{S}^n(V^*)\to\mathrm{S}^{n+1}(V^*)$, the derivations $a_\lambda=\delta(e_\lambda):\ \mathrm{S}^n(V^*)\to\mathrm{S}^{n-1}(V^*)$ are photon

annihilation operators. The Fock space has a vacuum $1 \in \mathbb{C} = \mathrm{S}^0(V^*)$ which is denoted by $|0\rangle$ in Dirac notation. The commutation relations are the canonical ones (CCR):

$$[a_\lambda\,, a_{\lambda'}^+] = \delta_{\lambda\lambda'}\,, \quad [a_\lambda^+\,, a_{\lambda'}^+] = 0\,, \quad [a_\lambda\,, a_{\lambda'}] = 0\,. \tag{3.110}$$

Based on the general formula (3.93), we can now express the quantized electromagnetic field in the form of a normal mode expansion:

$$\big(\widehat{D}(\mathbf{r}), \widehat{B}(\mathbf{r})\big) = \sum\nolimits_\lambda \big(e_\lambda(\mathbf{r})\, a_\lambda^+ + \tilde{e}_\lambda(\mathbf{r})\, a_\lambda\big)\,. \tag{3.111}$$

By separating the right-hand side into electric and magnetic parts we get

$$\widehat{D}(\mathbf{r}) = \frac{1}{\sqrt{2}} \sum_\lambda D_0^\lambda(\mathbf{r})\,\big(a_\lambda^+ + a_\lambda\big), \qquad \widehat{B}(\mathbf{r}) = \frac{\mathrm{i}}{\sqrt{2}} \sum_\lambda \frac{d \star D_0^\lambda}{\omega_\lambda \varepsilon_0}\,(\mathbf{r})\,\big(a_\lambda^+ - a_\lambda\big). \tag{3.112}$$

Note that in the construction above we could have singled out the purely magnetic normal mode $(0, B_0^\lambda) = J(D_0^\lambda, 0)$ instead of $(D_0^\lambda, 0)$. This would have led to modified photon creation operators b_λ^+ and annihilation operators b_λ by the modified choice of basis vectors in Eqs. (3.106, 3.108). The quantization procedure would have resulted in

$$\widehat{D}(\mathbf{r}) = \frac{-\mathrm{i}}{\sqrt{2}} \sum_\lambda \frac{d \star B_0^\lambda}{\omega_\lambda \mu_0}\,(\mathbf{r})\,\big(b_\lambda^+ - b_\lambda\big), \qquad \widehat{B}(\mathbf{r}) = \frac{1}{\sqrt{2}} \sum_\lambda B_0^\lambda(\mathbf{r})\,\big(b_\lambda^+ + b_\lambda\big). \tag{3.113}$$

This is the same as (3.112) because $B_0^\lambda = -(\omega_\lambda \varepsilon_0)^{-1} d \star D_0^\lambda$ and $\mathrm{i} a_\lambda = b_\lambda$, $\mathrm{i} b_\lambda^+ = a_\lambda^+$. Finally, inserting the expansion (3.112) of the quantum fields $\widehat{D}$, $\widehat{B}$ into the expression (3.100) for the electromagnetic energy, we obtain the quantum Hamiltonian of the electromagnetic field as

$$\widehat{\mathcal{H}} \stackrel{?}{=} \tfrac{1}{2} \sum\nolimits_\lambda \hbar\omega_\lambda \left(a_\lambda^+ a_\lambda + a_\lambda\, a_\lambda^+\right). \tag{3.114}$$

3.7.6 Casimir energy

We are now confronted with the same situation that already existed in the case of the Dirac field: an infinite constant must be subtracted from the Hamiltonian $\widehat{\mathcal{H}}$ in order to make it well-defined on the Fock space $\mathrm{S}(V^*)$. Indeed, applying (3.114) to the Fock vacuum $|0\rangle$, which is annihilated by each of the a_λ, we get an infinite ground state energy

$$\tfrac{1}{2} \sum\nolimits_\lambda \hbar\omega_\lambda = \infty\,. \tag{3.115}$$

The infinite constant to be subtracted (by the procedure of normal ordering) is universal in the sense that it does not depend on the cavity U under consideration: it is the infinite ground state energy of the electromagnetic field for the basic situation that U covers all of space and the boundary conditions are pushed to the edge of the universe. For the case of a finite cavity U, this universal vacuum energy differs by a finite amount E_C from the infinite ground state energy (3.115). Thus the correct formula for the quantum Hamiltonian $\widehat{\mathcal{H}} = \widehat{\mathcal{H}}_U$ is

$$\widehat{\mathcal{H}}_U = \sum\nolimits_\lambda \hbar\omega_\lambda\, a_\lambda^+ a_\lambda + E_\mathrm{C}(U). \tag{3.116}$$

E_C is called the Casimir energy. We mention in passing that over the last decade, advanced techniques have been developed to compute $E_C(U)$ accurately for a cavity U of general shape.

A simple dimensional estimate of E_C can be had for the case of a cavity of high symmetry, say a ball B_L of radius L. For this purpose we note that the characteristic frequencies ω_λ secretly carry a factor of c (speed of light). Therefore E_C must be proportional to $\hbar c$, which has the physical dimension of energy×length. Since the ball radius L is the only length scale in the problem, there exists only one energy that can be formed:

$$E_C(B_L) \sim \frac{\hbar c}{L}\,. \tag{3.117}$$

Thus $E_C(B_L) = K\hbar c/L$, where K is a constant. K turns out to be negative.

The Casimir energy tends to lead to an attractive force between surfaces. This so-called Casimir effect is predicted to be a challenge for future developments of nano-technology, as the attractive Casimir force makes nano-particles clump together, whereas the nano-engineer would like them to be freely movable.

3.7.7 Mode expansion for the torus geometry

Let us now consider a geometry which is simple enough to make it possible for the mode expansion to be given in explicit form: we choose for U a 3-torus T^3, i.e., a cubic domain

$$\mathrm{T}^3 = [0, L_1] \times [0, L_2] \times [0, L_3]$$

with periodic boundary conditions in each direction. Aside from being a mathematical idealization that does not correspond to the real world, this choice has a certain drawback: on T^3 there exist harmonic two-forms, i.e., solutions of the Laplace equation $\Delta D = 0 = \Delta B$. Such solutions require a separate treatment beyond that of Section 3.7.5. For brevity, we will ignore this issue here.

For better overall balance of presentation, we now adopt the vector notation for the electromagnetic field (dropping the bold-face convention for vectors). Our main task then is to construct divergenceless vector field solutions of the Helmholtz equation and thus the unit normal modes e_λ and $\tilde{e}_\lambda$. It is a standard fact that the Helmholtz equation $(\Delta + k^2)f = 0$ for functions f on T^3 has a complete orthonormal system of solutions by plane waves

$$f_k(r) := \frac{\mathrm{e}^{\mathrm{i}k\cdot r}}{\sqrt{\mathrm{vol}}} = (L_1L_2L_3)^{-1/2}\mathrm{e}^{\mathrm{i}(k_1x_1+k_2x_2+k_3x_3)}\,, \qquad k_j \in 2\pi\mathbb{Z}/L_j\,.$$

To obtain the desired solutions for the divergenceless vector fields D and B, we proceed as follows. For every wave vector $k = (k_1, k_2, k_3)$ we choose two so-called polarization unit vectors $\epsilon_s(k)$ $(s = 1, 2)$ with two properties: (i) they form an orthonormal basis of the plane perpendicular to the wave vector k, and (ii) the 3-bein $(\epsilon_1(k), \epsilon_2(k), k)$ is a right-handed system. (Let us mention in passing that such a choice does not exist in a globally smooth manner in the k-continuum. Indeed, on topological grounds any smooth vector field $k \mapsto \epsilon_s(k)$ tangent to the two-sphere must have at least two zeroes. This obstacle is of no concern to us here, as our set of wave vectors is

discrete due to the periodic boundary conditions.) For every (k,s) we then introduce a constant vector field $d_0^{k,s}$ as

$$d_0^{k,s} = \sqrt{\hbar c|k|\,\varepsilon_0}\;\epsilon_s(k). \tag{3.118}$$

Problem. Show that the complex vector field $D_0^{k,s}(r) = d_0^{k,s}\,f_k(r)$ is divergenceless ($\mathrm{div} D_0^{k,s} = 0$) and solves the Helmholtz equation $(\Delta + \omega^2/c^2)\,D_0^{k,s} = 0$ with characteristic frequency $\omega = c|k|$. Verify also that the total energy of the real vector field $\sqrt{2}\,\mathrm{Re} D_0^{k,s}$ viewed as an electric excitation field D is equal to the oscillator zero point energy $\frac{1}{2}\hbar\omega$. □

Using the information from the Problem above and from Section 3.7.5, especially Eq. (3.106), we compute the "holomorphic" unit normal mode $\tilde{e}_{k,s}$ with wave vector k and polarization s as

$$\tilde{e}_{k,s} = \frac{1}{\sqrt{2}}\Big((D_0^{k,s},0) + \mathrm{i}J(D_0^{k,s},0)\Big) = \frac{1}{\sqrt{2}}\Big(D_0^{k,s}, -\mathrm{i}\frac{\mathrm{rot} D_0^{k,s}}{c|k|\,\varepsilon_0}\Big). \tag{3.119}$$

The curl in the magnetic component of $\tilde{e}_{k,s}$ simplifies as $\mathrm{rot} D_0^{k,s} = \mathrm{i}k \times D_0^{k,s}$ (vector product).

We turn to the construction of the "anti-holomorphic" unit vector $e_{k,s} \in V_{k,s}$. While this follows in principle from the defining expression for $\tilde{e}_{k,s}$ by the substitution $+\mathrm{i}J \to -\mathrm{i}J$, we wish to arrange for the dual basis vectors $f_{k,s}$ and $\tilde{f}_{k,s}$ to quantize as mutual Hermitian adjoints: $\mu(f_{k,s}) = a_s^+(k)$ and $\delta(\tilde{f}_{k,s}) = a_s(k)$. Since we have chosen to work with the complex-valued exponential $f_k(r) \propto \mathrm{e}^{\mathrm{i}k\cdot r}$, we must deviate slightly from the blueprint of Section 3.7.5, where we constructed the normal modes over the real number field (in the present context this would correspond to working with the real-valued functions $\mathrm{Re}\,\mathrm{e}^{\mathrm{i}k\cdot r}$ and $\mathrm{Im}\,\mathrm{e}^{\mathrm{i}k\cdot r}$), we start from $\tilde{D}_0^{k,s}(r) = d_0^{k,s}\,\overline{f_k(r)}$, which is still in the normal mode space of frequency $\omega = c|k|$, and obtain

$$e_{k,s} = \frac{1}{\sqrt{2}}\Big((\tilde{D}_0^{k,s},0) - \mathrm{i}J(\tilde{D}_0^{k,s},0)\Big) = \frac{1}{\sqrt{2}}\Big(\tilde{D}_0^{k,s}, \frac{k\times\tilde{D}_0^{k,s}}{c|k|\,\varepsilon_0}\Big). \tag{3.120}$$

From the general prescription given by the formula (3.111) we now have

$$\big(\widehat{D}(r), \widehat{B}(r)\big) = \sum_{k,s}\big(e_{k,s}(r)\,a_s^+(k) + \tilde{e}_{k,s}(r)\,a_s(k)\big)\,. \tag{3.121}$$

Separating this into electric and magnetic parts and introducing the axial vector

$$b_0^{k,s} = \frac{k\times d_0^{k,s}}{c|k|\,\varepsilon_0} \tag{3.122}$$

for short, we arrive at the mode expansion in final form:

$$\widehat{D}(r) = \frac{1}{\sqrt{2\,\mathrm{vol}}}\sum_{k,\,s} d_0^{k,s}\left(\mathrm{e}^{-\mathrm{i}k\cdot r}a_s^+(k) + \mathrm{e}^{\mathrm{i}k\cdot r}a_s(k)\right), \tag{3.123}$$

$$\widehat{B}(r) = \frac{1}{\sqrt{2\,\mathrm{vol}}}\sum_{k,\,s} b_0^{k,s}\left(\mathrm{e}^{-\mathrm{i}k\cdot r}a_s^+(k) + \mathrm{e}^{\mathrm{i}k\cdot r}a_s(k)\right). \tag{3.124}$$

Note that here $[a_s(k), a_{s'}^+(k')] = \delta_{ss'}\delta_{kk'}$, as is appropriate for the present setting with discrete k.

3.8 Matter-field interaction

We have discussed in some detail the quantization of the free Dirac field and the free electromagnetic field. Let us add here a few words about the full interacting theory (called quantum electrodynamics, or QED) of the Dirac field coupled to the electromagnetic field.

For the purpose of writing a formula for the interaction, one introduces a magnetic vector potential $A \in \mathrm{rot}^{-1} B$. (Let us mention in passing that the introduction of A is not forced. One can also work with the physical electric and magnetic fields, by parametrizing the charged matter current by polarization and magnetization fields.) To eliminate the redundant gauge degrees of freedom one fixes the gauge, say by choosing the Coulomb gauge

$$\mathrm{div}\, A = 0\,. \tag{3.125}$$

By direct transcription from (3.123) one then has the mode expansion (still for $U = T^3$)

$$\widehat{A}(r) = \frac{\mathrm{i}}{\sqrt{2\,\mathrm{vol}}} \sum_{k,\,s} \frac{d_0^{k,s}}{c|k|\,\varepsilon_0} \left(\mathrm{e}^{-\mathrm{i}k\cdot r} a_s^+(k) - \mathrm{e}^{\mathrm{i}k\cdot r} a_s(k)\right). \tag{3.126}$$

It is easy to check that $\mathrm{div}\,\widehat{A} = 0$ and $\mathrm{rot}\,\widehat{A} = \widehat{B}$.

In the presence of electric charges it is no longer possible to choose the gauge $\Phi = 0$. Instead, one takes Φ to be the (instantaneous) Coulomb potential determined by Poisson's equation

$$\Delta\Phi(r) = -\tilde{\rho}(r)/\varepsilon_0 = -e\,\psi^\dagger(r)\psi(r)/\varepsilon_0\,. \tag{3.127}$$

The quantized Hamiltonian of the coupled theory then is

$$\widehat{\mathcal{H}} = \int \widehat{\psi}^\dagger \Big(\sum_l \alpha_l \Big(\frac{\hbar}{\mathrm{i}} \frac{\partial}{\partial x_l} - e\widehat{A}_l \Big) + \beta mc^2 \Big) \widehat{\psi}\, d^3r + \frac{1}{2} \int \sum_l \big(\widehat{D}_l \widehat{E}_l + \widehat{B}_l \widehat{H}_l \big)\, d^3r, \tag{3.128}$$

where the quantum field $\widehat{D}$ of the electric excitation is subject to the Gauss-law constraint $d\widehat{D} = \widehat{\rho}$. The matter-field interaction here is in the term $\sum \widehat{A}_l\, \widehat{j}_l$ coupling the magnetic vector potential to the electric current density. By inserting the respective mode expansions, one obtains contributions to the interaction Hamiltonian of the schematic form (omitting spin and polarization)

$$c_+^\dagger(k_1)\, c_-^\dagger(k_2)\, a(k_1 + k_2)\,, \quad a^\dagger(k_1 + k_2)\, c_-(k_2)\, c_+(k_1)\,. \tag{3.129}$$

The first of these creates an electron with momentum k_1 and a positron with momentum k_2 while annihilating a photon with momentum $k_1 + k_2$. The second does the reverse (or adjoint). There are also terms of the schematic form

$$c_+^\dagger(k_1)\, c_+(k_2)\, a(k_1 - k_2)\,, \quad a^\dagger(k_1 - k_2)\, c_+^\dagger(k_2)\, c_+(k_1)\,, \tag{3.130}$$

and similar with $c_+ \to c_-$. These cause scattering of electrons and positrons by the processes of absorption or emission of a photon.

3.9 γ-decay of excited states

We finish the chapter by briefly indicating how one computes the electromagnetic decay rate of unstable states in many-body systems such as molecules, atoms, nuclei, etc. This is a problem of fermions (electrons, nucleons) coupled to bosons (photons). Thus the Hilbert space to work with is a tensor product

$$\mathcal{V} = \wedge(V_F) \otimes \mathrm{S}(V_B) \tag{3.131}$$

of Fock spaces for fermions and bosons, and the Hamiltonian is a sum of three terms:

$$H = H_{\text{matter}} + H_{\text{radiation}} + H_{\text{coupling}} \ . \tag{3.132}$$

We have seen in the previous section how this Hamiltonian ($H \equiv \widehat{\mathcal{H}}$) looks in a situation requiring relativistic treatment. The matter part (the Hamiltonian for the free Dirac field) operates on the fermionic Fock space $\wedge(V_F)$ and is trivial on $\mathrm{S}(V_B)$. (For many purposes it is of course good enough to treat the Dirac field by non-relativistic reduction in the Schrödinger approximation.) The radiation part (the Hamiltonian for the free electromagnetic field) operates on the bosonic Fock space $\mathrm{S}(V_B)$ and is trivial on $\wedge(V_F)$. The coupling Hamiltonian $\sum_l \int j_l \otimes A_l \, d^3r$ has factors j_l operating on $\wedge(V_F)$ and factors A_l operating on $\mathrm{S}(V_B)$.

Consider now some excited state of matter, ψ_i, which decays to another state ψ_f (e.g., the ground state) by the emission of a single photon with wave vector k. The transition or decay rate $\Gamma\big(\psi_f \otimes 1\gamma(k) \leftarrow \psi_i \otimes 0\gamma\big)$ can be computed by using Fermi's golden rule:

$$\Gamma\big(\psi_f \otimes 1\gamma(k) \leftarrow \psi_i \otimes 0\gamma\big) = \frac{2\pi}{\hbar} \big|\big\langle \psi_f \otimes 1\gamma(k) \mid H_c \mid \psi_i \otimes 0\gamma \big\rangle\big|^2 \, \delta(E_f + \hbar\omega_k - E_i) \ , \tag{3.133}$$

where $H_c \equiv H_{\text{coupling}} = \int \sum j_l \, A_l \, d^3r$ is the interaction part of the Hamiltonian. The transition matrix element

$$\big\langle \psi_f \otimes 1\gamma(k) \mid H_c \mid \psi_i \otimes 0\gamma \big\rangle = \int d^3r \, \big\langle \psi_f \mid j_l(r) \mid \psi_i \big\rangle \, \big\langle \text{1-photon}(k) \mid A_l(r) \mid \text{0-photon} \big\rangle \tag{3.134}$$

is essentially the Fourier transform of the transition current density $\big\langle \psi_f \mid j_l(r) \mid \psi_i \big\rangle$ since

$$\big\langle \text{1-photon}(k) \mid A_l(r) \mid \text{0-photon} \big\rangle \sim \mathrm{e}^{-\mathrm{i}k\cdot r}. \tag{3.135}$$

The typical situation is that this Fourier transform can be calculated by multipole expansion. Let us look at the case of a heavy atomic nucleus, for example. The wave length of the emitted photon is

$$\lambda = \frac{2\pi}{|k|} = 2\pi \frac{\hbar c}{\hbar\omega_k} \approx 2\pi \frac{200\,\text{MeV fm}}{E_i - E_f} \ . \tag{3.136}$$

By inserting the typical excitation energy of a low-lying nuclear excited state, one gets a value for λ which is very much bigger than the radius $R \sim 5\,\text{fm}$ of a heavy nucleus such as ^{208}Pb. Therefore, if the transition operator $j_l(r)$ is expanded in multipoles, the leading contributions to the decay rate come from the multipoles of lowest order which are compatible with the angular momenta of the initial and final states of the decay.

The multipole expansion proceeds in terms of so-called tensor operators T_{JM}. Learning more about these is a major motivation for the next chapter.

the fermionic Fock space $\wedge(V_F)$ and is trivial on $\mathrm{S}(V_B)$. (For many purposes it is of course good enough to treat the Dirac field by non-relativistic reduction in the Schrödinger approximation.) The radiation part (the Hamiltonian for the free electromagnetic field) operates on the bosonic Fock space $\mathrm{S}(V_B)$ and is trivial on $\wedge(V_F)$. The coupling Hamiltonian $\int j_l\, A_l\, d^3x$ has factors j_l operating on $\wedge(V_F)$ and factors A_l operating on $\mathrm{S}(V_B)$.

Consider now some excited state of matter, ψ_i, which decays to another state ψ_f (e.g., the ground state) by the emission of a single photon with wave vector k. The transition or decay rate $\Gamma\big(\psi_f \otimes 1\gamma(k) \leftarrow \psi_i \otimes 0\gamma\big)$ can be computed by using Fermi's golden rule:

$$\Gamma\big(\psi_f \otimes 1\gamma(k) \leftarrow \psi_i \otimes 0\gamma\big) = \frac{2\pi}{\hbar}\big|\big\langle \psi_f \otimes 1\gamma(k) \mid H_c \mid \psi_i \otimes 0\gamma\big\rangle\big|^2\, \delta(E_f + \hbar\omega_k - E_i)\ , \tag{3.116}$$

where $H_c \equiv H_{\text{coupling}} = \int \sum j_l\, A_l\, d^3x$ is the interaction part of the Hamiltonian. The transition matrix element

$$\big\langle \psi_f \otimes 1\gamma(k) \mid H_c \mid \psi_i \otimes 0\gamma\big\rangle = \int d^3x\, \big\langle \psi_f \mid j_l(x) \mid \psi_i\big\rangle\, \big\langle \text{1-photon}(k) \mid A_l(x) \mid \text{0-photon}\big\rangle \tag{3.117}$$

is essentially the Fourier transform of the transition current density $\big\langle \psi_f \mid j_l(x) \mid \psi_i\big\rangle$ since

$$\big\langle \text{1-photon}(k) \mid A_l(x) \mid \text{0-photon}\big\rangle \sim \mathrm{e}^{\mathrm{i}k\cdot x}\ . \tag{3.118}$$

The typical situation is that this Fourier transform can be calculated by multipole expansion. Let us look at the case of a heavy atomic nucleus, for example. The wavelength of the emitted photon is

$$\lambda = \frac{2\pi}{|k|} = 2\pi\frac{\hbar c}{\hbar\omega_k} \approx 2\pi\frac{200\,\text{MeV fm}}{E_i - E_f}\ . \tag{3.119}$$

By inserting the typical excitation energy of a low-lying nuclear excited state, one gets a value for λ which is very much bigger than the radius $R \sim 5\,\text{fm}$ of a heavy nucleus such as ^{208}Pb. Therefore, if the transition operator $j_l(x)$ is expanded in multipoles, the leading contributions to the decay rate come from the multipoles of lowest order which are compatible with the angular momenta of the initial and final states of the decay.

The multipole expansion proceeds in terms of so-called tensor operators T_{JM}. Learning more about these is a major motivation for the next chapter.

4 Invariants of the rotation group SO_3

4.1 Motivation

The students of this course already have some familiarity with the quantum theory of angular momentum. Building on this, in the present chapter we will introduce some further material related to quantized angular momentum and the theory of group representations in general.

One of the results we wish to explain is the Wigner-Eckart theorem. Its statement is that the matrix elements of a tensor operator (e.g., the operator for an electromagnetic transition from an excited state to the ground state of a many-body system) between states of definite angular

momenta separates as a product of two factors: a so-called Wigner 3j-symbol (or Clebsch-Gordan coefficient) determined by geometry, and a reduced matrix element containing the information about the intrinsic structure of the many-body states.

4.2 Basic notions of representation theory

We assume the mathematical notions of group and group action to be understood. In the following, $\mathrm{GL}(V)$ denotes the group of invertible $\mathbb{K}$-linear transformations of a $\mathbb{K}$-vector space. (Depending on the situation, V may be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.)

Let G be a group. A (linear) representation of G consists of two pieces of data: (i) a vector space V and (ii) a mapping $\rho : G \to \mathrm{GL}(V)$ with the property

$$\forall g_1 , g_2 \in G : \quad \rho(g_1 g_2) = \rho(g_1)\rho(g_2) . \tag{4.1}$$

In mathematical parlance one says that ρ is a group homomorphism from G into $\mathrm{GL}(V)$. The trivial representation is the representation $\rho(g) \equiv \mathrm{Id}_V$ mapping all group elements to the identity.

Examples.

1. The simplest group is $G = \{e, \pi\}$ with multiplication table $e^2 = e$, $e\pi = \pi e = \pi$, and $\pi^2 = e$. This group has a representation on $V = \mathbb{R}$ by $\rho(e) = +1$ and $\rho(\pi) = -1$.

2. Let G be a finite group, and take V to be the vector space of all scalar-valued functions $f : V \to \mathbb{K}$. The right regular representation, $R \equiv \rho_R$, of G is defined as

$$R : \ G \to \mathrm{GL}(V) , \quad \big(R(g_0)f\big)(g) := f(g g_0) . \tag{4.2}$$

The left regular representation $L \equiv \rho_L : \ G \to \mathrm{GL}(V)$ is

$$\big(L(g_0)f\big)(g) := f(g_0^{-1} g) . \tag{4.3}$$

3. Let $\mathrm{S}^2 \subset \mathbb{R}^3$ be the unit sphere in three dimensions. The rotation group SO_3 acts on S^2 in the fundamental way by

$$\mathrm{SO}_3 \times \mathrm{S}^2 \to \mathrm{S}^2 , \quad (g, v) \mapsto g \cdot v .$$

This action yields a representation of SO_3, say on the Hilbert space $L^2(\mathrm{S}^2)$ of square integrable functions $f : \ \mathrm{S}^2 \to \mathbb{C}$, by

$$\big(\rho(g)f\big)(v) := f(g^{-1} \cdot v) . \quad \square \tag{4.4}$$

A general construction of great importance in representation theory is what is called the induced representation allowing you to build from a known representation of a small group a (new) representation of a bigger group, as follows. Let G be a group and $H \subset G$ a subgroup. H acts on G by right or left multiplication; here we will be concerned with the right action. Assume that you are given a representation (W, π) of the small group H. Then consider the space $\mathrm{Func}(G, W)^H$ of H-invariant functions from G to W, i.e., all functions $f : \ G \to W$ with the property

$$f(g) = \pi(h) f(gh) . \tag{4.5}$$

Since the operators $\pi(h)$ are linear transformations of a vector space, our space of functions $V := \mathrm{Func}(G, W)^H$ is still a vector space. V carries a representation ρ of G by left multiplication:

$$(\rho(g_0)f)(g) = f(g_0^{-1}g) \; . \tag{4.6}$$

This representation is referred to as the induced representation $\mathrm{Ind}_H^G(\pi)$.

Examples.

1. If you take for H the trivial group $H = \{e\}$ then $\mathrm{Ind}_{\{e\}}^G(\mathrm{trivial}) = L$ is the left regular representation of G.

2. Let $G = \mathrm{SO}_3$, and let $H = \mathrm{SO}_2 \subset G$ be the subgroup of rotations about a fixed axis, say the z-axis. Take π to be the trivial H-representation, and let $W = \mathbb{K} = \mathbb{C}$. Then $\mathrm{Func}(\mathrm{SO}_3, \mathbb{C})^{\mathrm{SO}_2}$ can be identified with the vector space of complex-valued functions on the unit sphere $\mathrm{S}^2 \simeq \mathrm{SO}_3/\mathrm{SO}_2$. If we require the functions to be square-integrable, the induced representation $\mathrm{Ind}_H^G(\pi)$ is the representation of Example 3 above. □

Remark. Frobenius reciprocity – a central result in the representation theory of finite groups – is a statement about the induced representation. □

After this brief introduction to the notion of induced representation, let us return to the basics. We wish to explain what is meant by a unitary irreducible representation. Let us begin with the adjective 'unitary'. For this one requires the representation space V of a G-representation $\rho : G \to \mathrm{GL}(V)$ to be Hermitian, i.e., V has to be a complex vector space with Hermitian scalar product $\langle \cdot, \cdot \rangle$. The representation ρ is then called unitary if

$$\langle v, v' \rangle = \langle \rho(g)v, \rho(g)v' \rangle \tag{4.7}$$

holds for all $g \in G$ and $v, v' \in V$.

Example. Let the space of complex-valued functions $f : \mathrm{S}^2 \to \mathbb{C}$ on the unit sphere be equipped with the Hermitian scalar product

$$\langle f_1, f_2 \rangle := \int_{\mathrm{S}^2} \overline{f_2(v)}\, f_1(v)\, d^2v \; , \tag{4.8}$$

where $d^2v = \sin\theta \, d\theta \, d\phi$ is the solid-angle two-form. The SO_3-representation given by the space $L^2(\mathrm{S}^2)$ of square-integrable functions on S^2 is unitary:

$$\langle \rho(g)f_1, \rho(g)f_2 \rangle = \int_{\mathrm{S}^2} \overline{f_2(g^{-1} \cdot v)}\, f_1(g^{-1} \cdot v)\, d^2v = \int_{\mathrm{S}^2} \overline{f_2(v')}\, f_1(v')\, d^2v' = \langle f_1, f_2 \rangle \; .$$

The second equality follows from the substitution $v' := g^{-1} \cdot v$ by the SO_3-invariance of d^2v. □

Next we explain what is meant by saying that a representation is irreducible. For this purpose we need the following preparation of language. Let $\rho : G \to \mathrm{GL}(V)$ be a representation. Then a vector subspace $U \subset V$ is called G-invariant if

$$\forall g \in G \; : \quad \rho(g)U \subset U \; . \tag{4.9}$$

U is called a proper subspace of V if U is neither the zero vector space nor identical with V.

Definition. A representation $\rho : G \to \mathrm{GL}(V)$ is called irreducible if there exist no G-invariant proper vector subspaces of V.

Example. The unitary representation of the previous example is not irreducible as the vector space of, say, the constant functions is an SO_3-invariant proper subspace of $L^2(\mathrm{S}^2)$ of dimension one. However, irreducible SO_3-representations do arise from it by restriction to subspaces of definite total angular momentum. Let

$$V_L := \mathrm{span}_{\mathbb{C}}\{Y_{L,L}\,, Y_{L,L-1}\,, \ldots, Y_{L,-L+1}\,, Y_{L,-L}\}\,, \tag{4.10}$$

where $Y_{L,M}$ denotes the spherical harmonic of total angular momentum L and magnetic quantum number M. One has $\dim V_L = 2L+1$. The SO_3-representation $\rho_L : \mathrm{SO}_3 \to \mathrm{GL}(V_L)$ by

$$(\rho_L(g)Y_{L,M})(v) := Y_{L,M}(g^{-1}\cdot v) = \sum_{M'=-L}^{L} Y_{L,M'}(v)\, \mathcal{D}^L_{M',M}(g)\,, \tag{4.11}$$

is known to be irreducible. The matrix elements $g \mapsto \mathcal{D}^L_{M',M}(g)$ of the representation are referred to as Wigner D-functions in physics. They satisfy the relation

$$\mathcal{D}^L_{M',M''}(g_1 g_2) = \sum_{M=-L}^{L} \mathcal{D}^L_{M',M}(g_1)\, \mathcal{D}^L_{M,M''}(g_2)\,, \tag{4.12}$$

due to the representation property $\rho_L(g_1g_2) = \rho_L(g_1)\rho_L(g_2)$. □

We move on to yet another basic definition. For a group G, let there be two representations $\rho_j : G \to \mathrm{GL}(V_j)$ $(j = 1, 2)$.

Definition. The representations (V_1, ρ_1) and (V_2, ρ_2) are called isomorphic if there exists a linear bijection $\varphi : V_1 \to V_2$ intertwining the representations:

$$\forall g \in G \;:\quad \varphi \circ \rho_1(g) = \rho_2(g) \circ \varphi\,. \quad \square \tag{4.13}$$

Isomorphy of representations is an equivalence relation. The equivalence classes of this equivalence relation are called isomorphism classes. For example, all irreducible representations of SO_3 for a fixed total angular momentum L belong to the same isomorphism class.

We now come to an often cited fundamental result of representation theory.

Schur's lemma. For a group G, let there be two finite-dimensional irreducible representations $\rho_1 : G \to \mathrm{GL}(V_1)$ and $\rho_2 : G \to \mathrm{GL}(V_2)$. If $\varphi \in \mathrm{Hom}(V_1, V_2)$ intertwines these representations, i.e., $\varphi \circ \rho_1(g) = \rho_2(g) \circ \varphi$ for all $g \in G$, then one has a dichotomy: either (i) φ is the zero map, or (ii) φ is invertible.

Proof. Let the linear mapping $\varphi : V_1 \to V_2$ be an intertwiner of the irreducible representations $(V_1\,, \rho_1)$ and $(V_2\,, \rho_2)$. Then $\ker\varphi \subset V_1$ is a G-invariant subspace. Indeed, if $v \in \ker\varphi$ then the intertwining property yields

$$0 = \varphi(v) \quad \Longrightarrow \quad 0 = \rho_2(g)\,\varphi(v) = \varphi(\rho_1(g)\,v)\,,$$

so $\rho_1(g)v \in \ker\varphi$. Now since V_1 is irreducible, the G-invariant subspace $\ker\varphi$ by definition cannot be a proper subspace. Thus there exist only two possibilities: (i) $\ker\varphi \equiv V_1$, in which case φ is the zero map, or (ii) $\ker\varphi \equiv \{0\}$, in which case φ is injective. In the latter case φ is also surjective. Indeed, by similar reasoning as before, $\mathrm{im}\,\varphi \subset V_2$ is a G-invariant subspace of the irreducible G-representation space V_2, so either $\mathrm{im}\,\varphi \equiv V_2$ or $\mathrm{im}\,\varphi = \{0\}$. The latter possibility is ruled out by the known injectivity of φ for the case (ii) under consideration.

Corollary. Let (V,ρ) be a finite-dimensional irreducible G-representation over the complex numbers $\mathbb{C}$. Then every $\mathbb{C}$-linear map $\varphi \in \mathrm{End}(V)$ with the property $\varphi\circ\rho(g) = \rho(g)\circ\varphi$ (for all $g \in G$) is a scalar multiple of the identity.

Proof. Over the complex number field, any linear transformation $\varphi \in \mathrm{End}(V)$ has at least one eigenvalue, say λ. By the definition of what it means for λ to be an eigenvalue, the linear transformation $\varphi - \lambda\cdot\mathrm{Id}_V$ is not invertible. On the other hand, $\varphi - \lambda\cdot\mathrm{Id}_V$ intertwines the irreducible representation (V,ρ) with itself by assumption. Therefore, Schur's lemma implies that $\varphi - \lambda\cdot\mathrm{Id}_V$ must be identically zero. Hence $\varphi = \lambda\cdot\mathrm{Id}_V$ as claimed.

Problem. If $\rho_j : G \to \mathrm{GL}(V_j)$ $(j = 1, 2)$ are two finite-dimensional irreducible representations over $\mathbb{C}$, show that the linear space $\mathrm{Hom}_G(V_1, V_2)$ of intertwiners φ is at most one-dimensional.

4.2.1 Borel-Weil Theorem.

Let us briefly offer a more advanced perspective on induced representations, as follows. If G is a group with subgroup H, the equivalence classes of the right action of H on G are called cosets, and these form a so-called coset space, G/H. Now, given a vector space W carrying an H-representation π, the group H also acts on the direct product $G \times W$ by

$$h\cdot(g,w) := \big(gh^{-1}, \pi(h)w\big)\ . \tag{4.14}$$

The equivalence classes $[g;w] \equiv [gh^{-1};\pi(h)w]$ of this H-action constitute a so-called vector bundle, $G\times_H W$, over the base manifold G/H. The equivalence classes $[g;w]$ for a fixed coset $gH \in G/H$ form a vector space which is isomorphic to W by $[g;w] \mapsto w$; it is called the fiber over gH. A section of the vector bundle is a mapping $s : G/H \to G\times_H W$ with the property that for each $gH \in G/H$ the value $s(gH)$ of the section is a vector $[g;w]$ in the fiber over gH. (Depending on the situation, one may require that s is continuous, or differentiable, or analytic, or holomorphic, etc.). The space of sections s of the vector bundle $G\times_H W$ is denoted by $\Gamma(G\times_H W)$.

Problem. Show that the mapping $\mathrm{Func}(G,W)^H \to \Gamma(G\times_H W)$, $f \mapsto s$, defined by $s(gH) = [g; f(g)]$, is a bijection. □

Thus we have an alternative way of thinking about the induced representation $\mathrm{Ind}_H^G(\pi)$: we may view it as the representation

$$(\rho(g_0)s)(gH) := s(g_0^{-1}gH) \tag{4.15}$$

on sections $s \in \Gamma(G\times_H W)$. This viewpoint has many applications. We mention an especially important one. Let G be a connected and semisimple compact Lie group, and let $T \subset G$ be a

maximal Abelian subgroup (a so-called maximal torus). The coset space G/T can then be shown to be a complex space. (In fact, G/T is a phase space in the sense of Section 3.7.1: it has the attributes of symplectic and complex structure which are required for the geometric quantization procedure described there.) If one is now given any one-dimensional representation $\pi : T \to \mathrm{GL}(W)$ on $W = \mathbb{C}$, then one has a vector bundle $G \times_T W$; it is called a complex line bundle because its fibres are copies of $\mathbb{C}$. The sections of such a line bundle are maps from a complex space into another complex space. One can therefore speak of the subspace of holomorphic sections.

The following theorem holds in the mathematical setting of a connected, semisimple, compact Lie group G as described above.

Borel-Weil Theorem. There is a one-to-one correspondence between (isomorphism classes of) irreducible representations of G and the spaces of holomorphic sections of complex line bundles over G/T. In particular, for every irreducible G-representation $\rho : G \to \mathrm{GL}(V)$ there exists a one-dimensional T-representation $\pi : T \to \mathrm{GL}(W)$ such that ρ can be identified with the induced representation $\mathrm{Ind}_T^G(\pi)$ restricted to the subspace of holomorphic sections of the complex line bundle $G \times_T W$.

Example. Every irreducible representation $(V_L\,, \rho_L)$ of the rotation group $G = \mathrm{SO}_3$ arises in this way. For this purpose one takes $T := \mathrm{SO}_2$, the subgroup of rotations about the z-axis, and has $G/T = \mathrm{SO}_3/\mathrm{SO}_2 = \mathrm{S}^2$, the unit sphere, which is a complex space. (The complex structure J is rotation by $\pi/2 = 90^o$ in the tangent plane). To get the representation $(V_L\,, \rho_L)$ for any total angular momentum $L \in \mathbb{N} \cup \{0\}$, one takes $\pi : \mathrm{SO}_2 \to \mathrm{GL}(W)$ to be the one-dimensional representation on the space of states $W := \mathbb{C} \cdot |L, L\rangle\}$ with maximal projection of the angular momentum:

$$\pi\big(R_z(\phi)\big) = \mathrm{e}^{\mathrm{i}L\phi}\,,$$

where $R_z(\phi)$ is the rotation about the z-axis with rotation angle ϕ. The $(2L+1)$-dimensional irreducible representation $(V_L\,, \rho_L)$ can then be seen as the space of holomorphic sections of the complex line bundle $\mathrm{SO}_3 \times_{\mathrm{SO}_2} \mathbb{C} \cdot |L, L\rangle$. Details of its construction (using ladder operators L^+ and L^-, but omitting the relation to holomorphic sections of a complex line bundle) were discussed in the basic course on quantum mechanics. In particular, the 3-dimensional SO_3-representation corresponding to angular momentum $L = 1$ is the same as the representation of SO_3 on the space of holomorphic vector fields or sections of the tangent bundle

$$T(\mathrm{S}^2) \simeq \mathrm{SO}_3 \times_{\mathrm{SO}_2} \mathbb{C} \cdot |L = 1, M = 1\rangle\,.$$

An example of a holomorphic vector field is the vector field of infinitesimal rotation about the z-axis (or any other axis for that matter).

4.3 Invariant tensors

Let $(V_j\,, \rho_j)$ $(j = 1, \ldots, n)$ be representations of a group G. Then the tensor product

$$\mathcal{V} := V_1 \otimes V_2 \otimes \cdots \otimes V_n \tag{4.16}$$

is a G-representation space in the natural way:

$$g \cdot (v_1 \otimes v_2 \otimes \cdots \otimes v_n) := (\rho_1(g)v_1) \otimes (\rho_2(g)v_2) \otimes \cdots \otimes (\rho_n(g)v_n) \; . \tag{4.17}$$

This definition makes sense because the operators $\rho_j(g)$ are linear transformations.

A tensor of the form $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ is called a pure tensor. The most general tensor $T \in \mathcal{V}$ is a linear combination of pure tensors. In fact, if $\{e_i^{(j)}\}$ is a basis of V_j we can express T as

$$T = \sum T_{i_1, i_2, \ldots, i_n} \, e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \cdots e_{i_n}^{(n)} \; . \tag{4.18}$$

The action $G \times \mathcal{V} \to \mathcal{V}$ on general tensors is defined by linear extension of (4.17).

Definition. A tensor $T \in V_1 \otimes V_2 \otimes \cdots \otimes V_n$ is called G-invariant if $g \cdot T = T$ for all $g \in G$.

Example. Let $V = \mathbb{R}^3$ be the fundamental representation space of $G = \mathrm{SO}_3$. As usual, the SO_3-representation on the dual vector space $V^* = (\mathbb{R}^3)^*$ is defined by $(\rho(g)\varphi)(v) = \varphi(g^{-1}v)$. Consider now the tensor product $V^* \otimes V^*$. The Euclidean scalar product $Q : \; V \otimes V \to \mathbb{R}$ can be regarded as an SO_3-invariant tensor in $V^* \otimes V^*$: if $\{e_1, e_2, e_3\}$ is a Cartesian basis of $V = \mathbb{R}^3$ and $f_i = Q(e_i, \cdot) \in V^*$ are the dual basis forms, then

$$Q = \sum_{i=1}^{3} f_i \otimes f_i \; ,$$

and this satisfies $g \cdot Q = Q$ due to $Q(gv, gv') = Q(v, v')$ for $g \in \mathrm{SO}_3$. Another SO_3-invariant tensor is the triple product $\Omega \in V^* \otimes V^* \otimes V^*$. While the Euclidean scalar product is a symmetric tensor, Ω is totally antisymmetric:

$$\Omega = f_1 \wedge f_2 \wedge f_3 = \sum_{\pi \in \mathrm{S}_3} \mathrm{sign}(\pi) \; f_{\pi(1)} \otimes f_{\pi(2)} \otimes f_{\pi(3)} = \sum \epsilon_{ijk} \, f_i \otimes f_j \otimes f_k \; .$$

Ω is SO_3-invariant because $g \cdot \Omega = \mathrm{Det}(g)^{-1} \, \Omega$ and $\mathrm{Det}(g) = 1$ for $g \in \mathrm{SO}_3$. $\square$

4.3.1 Invariant tensors of degree 2

The situation for degree 2 is rather easy to understand by using Schur's lemma. To that end we recall from Section 2.8.1 the isomorphism

$$\mu : \; V_2 \otimes V_1^* \to \mathrm{Hom}(V_1 \, , V_2) \tag{4.19}$$

which is determined by

$$\mu(v \otimes f)(u) := f(u) \, v \; . \tag{4.20}$$

If V_1 and V_2 are G-representation spaces, the mapping μ takes G-invariant tensors $T \in V_2 \otimes V_1^*$ into G-equivariant linear transformations $\mu(T) \in \mathrm{Hom}_G(V_1 \, , V_2)$. Indeed:

Problem. Show that the invariance property $g \cdot T = T$ for $T \in V_2 \otimes V_1^*$ translates via μ into the intertwining property $\rho_2(g)\mu(T) = \mu(T)\rho_1(g)$. $\square$

In the special case of $V_1 = V_2 = V$ the isomorphism $\mu : \; V \otimes V^* \to \mathrm{Hom}(V,V) = \mathrm{End}(V)$ provides us with a canonical G-invariant tensor $T \in V \otimes V^*$. This is the inverse image of the identity:

$$\mu^{-1}(\mathrm{Id}_V) = \sum_i e_i \otimes f_i \; , \tag{4.21}$$

where $\{e_i\}$ and $\{f_i\}$ are bases of V resp. V^* which are dual to each other by $f_i(e_j) = \delta_{ij}$. This G-invariant tensor $\mu^{-1}(\mathrm{Id}_V)$ is canonical in the sense that it exists for any representation (V,ρ).

Next, consider the general case of $V_1 \neq V_2$ and assume that both representation spaces V_1 and V_2 are G-irreducible. Then there exist only two possibilities for the space $\mathrm{Hom}_G(V_1\,,V_2)$ of intertwiners: if V_1 and V_2 belong to different isomorphism classes, then $\mathrm{Hom}_G(V_1\,,V_2) = 0$; and if V_1 and V_2 are isomorphic, then $\mathrm{Hom}_G(V_1\,,V_2) \simeq \mathbb{C}$ by Schur's lemma (cf. the Problem at the end of Section 4.2). In the former case $V_2 \otimes V_1^*$ has no G-invariant tensors; in the latter case there exists a complex line of G-invariant tensors.

In the following we focus on the important case of $V_2 = V$ and $V_1 = V^*$; i.e., we look for G-invariant tensors in $V^* \otimes V^*$ (or $V \otimes V$; it makes no big difference). This case is relevant, e.g., for the question whether (and if so, how) a rotationally invariant state can be formed from a pair of states carrying angular momentum L in both cases.

We will use the fact that by the isomorphism $\mu : \; V^* \otimes V^* \to \mathrm{Hom}(V,V^*)$ the G-invariant tensors $T \in V^* \otimes V^*$ are in one-to-one correspondence with intertwiners $\mu(T) \equiv \varphi \in \mathrm{Hom}_G(V,V^*)$. Concerning the latter recall that $\mathrm{Hom}_G(V,V^*) \simeq \mathbb{C}$ if V and V^* are irreducible and belong to the same isomorphism class. Let $\mathrm{Sym}_G(V,V^*)$ and $\mathrm{Alt}_G(V,V^*)$ denote the G-invariant linear transformations from V to V^* which are symmetric resp. skew (or alternating).

Lemma. If an irreducible G-representation V is isomorphic to its dual V^*, then one has either (i) $\mathrm{Hom}_G(V,V^*) = \mathrm{Sym}_G(V,V^*) \simeq \mathbb{C}$ or (ii) $\mathrm{Hom}_G(V,V^*) = \mathrm{Alt}_G(V,V^*) \simeq \mathbb{C}$.

Proof. We recall from Section 1.7.1 that for a linear transformation $\varphi : \; V \to V^*$ there exists a canonical adjoint $\varphi^t : \; V \to V^*$ (the transpose). If $0 \neq \varphi \in \mathrm{Hom}_G(V,V^*)$ is an intertwiner, let $\varphi_s := \frac{1}{2}(\varphi + \varphi^t)$ and $\varphi_a := \frac{1}{2}(\varphi - \varphi^t)$ be the symmetric and alternating parts of φ. Both φ_s and φ_a intertwine the G-representation on V with the G-representation on V^*. Since V is irreducible, we know that the space of such intertwiners is one-dimensional. Now $\varphi_s = +\varphi_s^t$ and $\varphi_a = -\varphi_a^t$ cannot both be non-zero and lie on the same complex line. Therefore we must have either (i) $\varphi_a = 0$ or (ii) $\varphi_s = 0$.

The next statement is an immediate deduction from the isomorphism $V^* \otimes V^* \simeq \mathrm{Hom}(V,V^*)$ of G-representation spaces and the fact that a G-invariant tensor $T \in V^* \otimes V^*$ is the same as a G-invariant bilinear form $Q : \; V \otimes V \to \mathbb{C}$.

Corollary. If a complex vector space V is an irreducible representation space for a group G, then there exists a trichotomy of possibilities for the space, say $\mathcal{Q}$, of G-invariant bilinear forms $Q : \; V \otimes V \to \mathbb{C}$: (i) $\mathcal{Q} = 0$ is trivial, or (ii) $\mathcal{Q}$ is generated by a symmetric bilinear form, or (iii) $\mathcal{Q}$ is generated by an alternating bilinear form.

4.3.2 Example: SO_3-invariant tensors of degree 2

To illustrate the general statements of Section 4.3.1, we now specialize to the case of $G = SO_3$. We have already met the irreducible SO_3-representations on the vector spaces

$$V_L = \mathrm{span}_{\mathbb{C}} \{Y_{L,M}\}_{M=-L,\dots,L} \ ,$$

which are spanned by the spherical harmonics $Y_{L,M}$ of total angular momentum L.

Fact. The tensor product $V_L \otimes V_L$ of irreducible representations of angular momentum L contains an SO_3-invariant symmetric tensor,

$$T = \sum_{M=-L}^{+L} (-1)^M \, Y_{L,M} \otimes Y_{L,-M} \ . \tag{4.22}$$

All other SO_3-invariant tensors in $V_L \otimes V_L$ are scalar multiples of it.

Remark. The property of SO_3-invariance here means that $(g \cdot T)(v_1\,, v_2) = T(g^{-1}v_1\,, g^{-1}v_2) = T(v_1\,, v_2)$ for $T(v_1\,, v_2) = \sum (-1)^M \, Y_{L,M}(v_1)\, Y_{L,-M}(v_2)$. Viewed as (the angular part of) the wave function for a pair of particles with spherical positions v_1 and v_2, the tensor (or wave function) T carries total angular momentum zero. □

We will spend the rest of this subsection explaining where the invariant tensor (4.22) comes from. Let us begin the discussion with the observation that if (V, ρ) is a unitary representation of a group G, then (by the very definition of unitarity) the Hermitian scalar product is G-invariant:

$$\langle \rho(g)v_1\,, \rho(g)v_2 \rangle = \langle v_1\,, v_2 \rangle \ .$$

Alas, the Hermitian scalar product alone cannot deliver an invariant tensor in the sense of the Definition given at the beginning of Section 4.3, as it involves an operation of complex conjugation in the left argument. To repair this feature and produce a complex bilinear form

$$Q: \ V \otimes V \to \mathbb{C}\,, \quad v_1 \otimes v_2 \mapsto \langle \tau v_1, v_2 \rangle \ , \tag{4.23}$$

we need an anti-unitary operator $\tau : \ V \to V$ (akin to the time-reversal operator). We know from Section 1.7.2 that Q is symmetric if $\tau^2 = \mathrm{Id}_V$ and alternating if $\tau^2 = -\mathrm{Id}_V$. If the anti-unitary operator τ commutes with the G-action, i.e., $\tau\rho(g) = \rho(g)\tau$ for all $g \in G$, then the complex bilinear form Q is G-invariant.

Given a G-invariant complex bilinear form Q, we obtain a G-invariant tensor $T \in V^* \otimes V^*$ by starting from the canonical invariant $\mu^{-1}(\mathrm{Id}_V) = \sum e_i \otimes f_i$ and defining

$$T = \sum \widetilde{f}_i \otimes f_i \ , \tag{4.24}$$

where the $\widetilde{f}_i$ are determined by the equation

$$\widetilde{f}_i = Q(e_i\,, \cdot) = \langle \tau e_i\,, \cdot \rangle \ . \tag{4.25}$$

We claim that when the formula (4.24) for the invariant tensor $T \in V^* \otimes V^*$ is specialized to the case of $G = SO_3$ and $V^* = V_L$, one arrives at the invariant tensor $T \in V_L \otimes V_L$ of (4.22).

What is still missing for this argument is the origin and nature of the SO_3-invariant anti-unitary operator $\tau : V_L \to V_L$ needed for the construction of T. In the remainder of the subsection, we shall introduce τ. The point here will be that, although the SO_3-representation space V_L spanned by the degree-L spherical harmonics is complex, this representation can actually be constructed entirely over the real numbers (without ever using $\mathbb{C}$). We now indicate how this goes.

The one-dimensional representation space $V_{0,\mathbb{R}}$ for $L = 0$ is simply given by the constant functions on the unit sphere $\mathcal{S}^2$. The three-dimensional representation space $V_{1,\mathbb{R}} \simeq \mathbb{R}^3$ for $L = 1$ is spanned by the three Cartesian coordinate functions x_1 , x_2 , x_3 restricted to $\mathcal{S}^2$:

$$x_1\big|_{\mathcal{S}^2} = \sin\theta\,\cos\phi\ , \quad x_2\big|_{\mathcal{S}^2} = \sin\theta\,\sin\phi\ , \quad x_3\big|_{\mathcal{S}^2} = \cos\theta\ .$$

To construct the five-dimensional irreducible representation $V_{2,\mathbb{R}}$ for $L = 2$, one starts from the six-dimensional space $\mathrm{S}^2(V_{1,\mathbb{R}})$ (degree-2 part of the symmetric algebra) of functions $x_i\,x_j$ for $i \le j$. This space contains a generator $x_1^2 + x_2^2 + x_3^2$ which becomes trivial upon restriction to $\mathcal{S}^2$. One therefore removes it by passing to a quotient of vector spaces:

$$V_{2,\mathbb{R}} := \mathrm{span}_{\mathbb{R}}\,\{x_i\,x_j\}_{1\le i\le j\le 3}\Big/\mathbb{R}\cdot(x_1^2+x_2^2+x_3^2)\ .$$

(To simplify the notation, restriction to $\mathcal{S}^2$ will henceforth be understood.)

Problem. The degree-n subspace $\mathrm{S}^n(\mathbb{R}^N)$ of the symmetric algebra of $\mathbb{R}^N$ is the space of states of n bosons distributed over N single-boson states. Show that $\dim \mathrm{S}^n(\mathbb{R}^N) = \binom{N+n-1}{n}$. □

For angular momentum $L = 3$, one starts from the space $\mathrm{S}^3(V_{1,\mathbb{R}})$ of cubic monomials and quotients out the three-dimensional subspace of functions $x_l\,(x_1^2 + x_2^2 + x_3^2)$ for $l = 1, 2, 3$:

$$V_{3,\mathbb{R}} := \mathrm{span}_{\mathbb{R}}\,\{x_i\,x_j\,x_k\}_{1\le i\le j\le k\le 3}\Big/\mathrm{span}_{\mathbb{R}}\,\big\{x_l\,(x_1^2+x_2^2+x_3^2)\big\}_{l=1,2,3}\ .$$

By now it should be clear how to continue this construction. For a general value $L \ge 2$ of the angular momentum, one takes the space $\mathrm{S}^L(V_{1,\mathbb{R}})$ of degree-L monomials and quotients out the subspace $\mathrm{S}^{L-2}(V_{1,\mathbb{R}})\cdot(x_1^2+x_2^2+x_3^2) \subset \mathrm{S}^L(V_{1,\mathbb{R}})$:

$$V_{L,\mathbb{R}} := \mathrm{S}^L(V_{1,\mathbb{R}})\Big/\mathrm{S}^{L-2}(V_{1,\mathbb{R}})\cdot(x_1^2+x_2^2+x_3^2)\ . \tag{4.26}$$

This quotient space has dimension

$$\dim V_{L,\mathbb{R}} = \dim \mathrm{S}^L(V_{1,\mathbb{R}}) - \dim \mathrm{S}^{L-2}(V_{1,\mathbb{R}}) = \binom{L+2}{L} - \binom{L}{L-2} = 2L+1\ . \tag{4.27}$$

Why is $V_{L,\mathbb{R}}$ a representation space for SO_3 ? The answer is provided by the following:

Problem. Let V be a G-representation space. Show that if $U \subset V$ is a G-invariant subspace, then the quotient space V/U is a representation space for G. □

It is a true fact stated here without proof (a standard reference is Hermann Weyl's famous treatise *The Classical Groups*) that the SO_3-representation space $V_{L,\mathbb{R}}$ is irreducible. This concludes our quick overview of the theory of real representations of SO_3 .

Now the complex representation spaces V_L are obtained by complexification:

$$V_L = V_{L,\mathbb{R}} \oplus \mathrm{i}V_{L,\mathbb{R}} \ , \tag{4.28}$$

and the action $\mathrm{SO}_3 \times V_L \to V_L$ is defined by $\mathbb{C}$-linear extension of the action $\mathrm{SO}_3 \times V_{L,\mathbb{R}} \to V_{L,\mathbb{R}}$:

$$\rho_L(g) \cdot (u + \mathrm{i}v) := \rho_L(g)u + \mathrm{i}\,\rho_L(g)v \qquad (g \in \mathrm{SO}_3 \ ; \ u, v \in V_{L,\mathbb{R}}) \ . \tag{4.29}$$

For example, for $L = 1$ one has $Y_{1,0} \propto x_3$, $Y_{1,1} \propto x_1 + \mathrm{i}x_2$, $Y_{1,-1} \propto -x_1 + \mathrm{i}x_2$.

We are now in a position to say what is the complex anti-linear operator τ needed for the construction of the SO_3-invariant tensor (4.22): it is just the well-defined operation of complex conjugation with respect to the real structure $V_{L,\mathbb{R}} \subset V_L$:

$$\tau(u + \mathrm{i}v) := u - \mathrm{i}v \ . \tag{4.30}$$

Defined in this way, the operator τ commutes with the G-action on V_L because the latter is already defined on the real subspace $V_{L,\mathbb{R}} \subset V_L$. Here ends our effort to explain the mathematical background behind (4.22).

4.3.3 Example: SU_2-invariant tensors of degree 2

We have seen that all SO_3-invariant tensors of degree 2 are symmetric. For further illustration of the Corollary at the end of Section 4.3.1, we will now describe a related situation in degree 2 where one also finds skew-symmetric tensors.

Our example will be built on the compact Lie group SU_2 of special unitary 2×2 matrices,

$$\mathrm{SU}_2 = \{g \in \mathrm{GL}(\mathbb{C}^2) \mid g^{-1} = g^\dagger \ ; \ \mathrm{Det}(g) = 1\} \ . \tag{4.31}$$

An equivalent description of SU_2 is as follows:

$$\mathrm{SU}_2 = \left\{ \begin{pmatrix} \alpha & \beta \\ -\bar\beta & \bar\alpha \end{pmatrix} \Big| \, \alpha, \beta \in \mathbb{C} \ ; \ |\alpha|^2 + |\beta|^2 = 1 \right\} \ . \tag{4.32}$$

This description shows that SU_2 is isomorphic as a manifold to the three-sphere $\mathcal{S}^3 \subset \mathbb{R}^4 \simeq \mathbb{C}^2$ of solutions of the equation

$$(\mathfrak{Re}\,\alpha)^2 + (\mathfrak{Im}\,\alpha)^2 + (\mathfrak{Re}\,\beta)^2 + (\mathfrak{Im}\,\beta)^2 = 1 \ .$$

Let $\mathrm{U}_1 \subset \mathrm{SU}_2$ be the one-parameter subgroup defined by

$$\mathrm{U}_1 = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \bar\alpha \end{pmatrix} \Big| \, \alpha \in \mathbb{C} \ ; \ |\alpha|^2 = 1 \right\} . \tag{4.33}$$

The coset space $G/T = \mathrm{SU}_2/\mathrm{U}_1$ then is a complex manifold isomorphic to the two-sphere $\mathcal{S}^2$. Consider the representation π_N : $\mathrm{U}_1 \to \mathrm{GL}(\mathbb{C})$ by

$$\pi_N \begin{pmatrix} \alpha & 0 \\ 0 & \bar\alpha \end{pmatrix} = \alpha^N \qquad (N \in \mathbb{N} \cup \{0\}) \ . \tag{4.34}$$

Then according to Section 4.2 there exists an irreducible SU_2-representation on the space of holomorphic sections of the complex line bundle $\mathrm{SU}_2 \times_{\mathrm{U}_1} \mathbb{C}$ over $\mathcal{S}^2$, where U_1 acts on $\mathbb{C}$ by π_N. By the Borel-Weil theorem, all irreducible representations of SU_2 arise in this way as $N = 0, 1, 2, \ldots$ runs through the non-negative integers.

We give a short cut to the outcome of the Borel-Weil construction, as follows. Let SU_2 act on a fundamental pair of coordinate functions $z_1\,, z_2$ for $\mathbb{C}^2$ by

$$g \cdot \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \equiv \begin{pmatrix} \alpha & \beta \\ -\bar\beta & \bar\alpha \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} := \begin{pmatrix} \alpha z_1 + \beta z_2 \\ -\bar\beta z_1 + \bar\alpha z_2 \end{pmatrix} . \tag{4.35}$$

This naturally induces an action of SU_2 on homogeneous polynomials

$$P(z_1\,, z_2) = \sum_{n=0}^{N} a_n\, z_1^{N-n} z_2^n = z_1^N P(1, z_2/z_1)$$

of degree N. The SU_2-action on the ratio $z := z_2/z_1$ is

$$g \cdot z = \frac{\bar\alpha z - \bar\beta}{\beta z + \alpha}\,, \qquad g^{-1} \cdot z = \frac{\bar\beta + \alpha z}{\bar\alpha - \beta z}\,.$$

By setting $P(z_1\,, z_2) = z_1^N \varphi(z_2/z_1)$ one then obtains an SU_2-action ρ_N on holomorphic polynomials $\varphi(z) = \sum_{n=0}^N a_n z^n$ as

$$\big(\rho_N(g)\varphi\big)(z) := (\bar\alpha - \beta z)^N \varphi\left(\frac{\bar\beta + \alpha z}{\bar\alpha - \beta z}\right), \tag{4.36}$$

which turns out to be none other than the induced SU_2-representation on holomorphic sections φ of the complex line bundle $\mathrm{SU}_2 \times_{\mathrm{U}_1} \mathbb{C}$ given by the U_1-representation π_N.

Let us look at the infinitesimal or Lie algebra form of this representation. The representation of the Lie algebra $\mathfrak{su}_2 = \mathrm{Lie}(\mathrm{SU}_2)$ is defined by linearization at unity,

$$(\rho_{N*}(X)\varphi)(z) := \frac{d}{dt}\left(\rho_N(\mathrm{e}^{tX})\varphi\right)(z)\Big|_{t=0}\,. \tag{4.37}$$

For the generators of $\mathfrak{su}_2 = \mathrm{Lie}(\mathrm{SU}_2)$ we take $\mathrm{i}\sigma_1\,,\ \mathrm{i}\sigma_2\,,\ \mathrm{i}\sigma_3$ where σ_j are the standard Pauli matrices. Actually, it is more convenient to work in the complexification $\mathfrak{sl}_2 = \mathfrak{su}_2 \oplus \mathrm{i}\,\mathfrak{su}_2$, as this allows us to present the answer in terms of the $\mathfrak{sl}_2$-generator σ_3 and the ladder operators $\sigma_\pm = \frac{1}{2}(\sigma_1 \pm \mathrm{i}\sigma_2)$. A short computation gives

$$\rho_{N*}(\sigma_3) = 2z\frac{d}{dz} - N, \quad \rho_{N*}(\sigma_+) = -z^2\frac{d}{dz} + Nz\,, \quad \rho_{N*}(\sigma_-) = \frac{d}{dz}\,. \tag{4.38}$$

We observe that for $N = 0$ there exists only one function $\varphi(z) = \mathrm{const}$, which spans the trivial representation. For $N = 1$ we have a pair of basis functions

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \mathrel{\hat{=}} \{z \mapsto \varphi_\uparrow(z) = z\}\,, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrel{\hat{=}} \{z \mapsto \varphi_\downarrow(z) = 1\}\,.$$

These furnish the fundamental representation of SU_2, where

$$\rho_{1*}(\sigma_3) = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,, \quad \rho_{1*}(\sigma_+) = \sigma_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\,, \quad \rho_{1*}(\sigma_-) = \sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\,.$$

For a general value of $N \geq 0$ our representation space is spanned by the functions $\varphi_{N,n}(z) = z^n$ for $n = 0, 1, \ldots, N$. We denote it by $V_N \equiv V_N^{\mathrm{SU}_2}$. Notice that

$$\rho_{N*}(\sigma_3)\varphi_{N,n} = (2n - N)\varphi_{N,n} \ , \quad \rho_{N*}(\sigma_+)\varphi_{N,n} = (N - n)\varphi_{N,n+1} \ , \quad \rho_{N*}(\sigma_-)\varphi_{N,n} = n\,\varphi_{N,n-1} \ .$$

In particular, $\rho_{N*}(\sigma_+)\varphi_{N,N} = 0$ and $\rho_{N*}(\sigma_-)\varphi_{N,0} = 0$; these are called the states of highest resp. lowest weight of the representation V_N.

Problem. Show that the representation space V_N has no SU_2-invariant proper subspaces (the implication being that the representation on V_N is irreducible). □

We now turn to the subject proper of this subsection: the construction of an SU_2-invariant tensor in the tensor product $V_N \otimes V_N$. From the general theory of Section 4.3.1 we know that there exists at most one such tensor (up to multiplication by scalars). Let us find one.

By linearization of the group action (4.17) on tensors, any Lie algebra element $X \in \mathfrak{su}_2$ acts on $V_N \otimes V_N$ as

$$\rho^{(2)}_{N*}(X) = \rho_{N*}(X) \otimes \mathrm{Id}_{V_N} + \mathrm{Id}_{V_N} \otimes \rho_{N*}(X) \ . \tag{4.39}$$

Now the elements of the tensor product $V_N \otimes V_N$ can be realized as polynomials in two complex variables z and z'. Adopting this realization we get from (4.38) the expressions

$$\begin{aligned} \rho^{(2)}_{N*}(\sigma_3) &= 2z\frac{d}{dz} + 2z'\frac{d}{dz'} - 2N \ , \\ \rho^{(2)}_{N*}(\sigma_+) &= -z^2\frac{d}{dz} - z'^2\frac{d}{dz'} + N(z + z') \ , \quad \rho^{(2)}_{N*}(\sigma_-) = \frac{d}{dz} + \frac{d}{dz'} \ . \end{aligned} \tag{4.40}$$

We then look for an SU_2-invariant tensor $T \in V_N \otimes V_N$ by the ansatz

$$T = \sum_{m,\,n=0}^{N} a_{N;\,m,\,n}\, z^n z'^m \tag{4.41}$$

as a polynomial in the two variables z and z'.

Problem. Show that the problem posed by the conditions of infinitesimal invariance,

$$\rho^{(2)}_{N*}(\sigma_3)T = \rho^{(2)}_{N*}(\sigma_+)T = \rho^{(2)}_{N*}(\sigma_-)T = 0 \ ,$$

has the solution

$$a_{N;\,m,\,n} = \delta_{m,\,N-n}\,(-1)^n \binom{N}{n} \ . \quad \square$$

It follows that $\rho^{(2)}_{N*}(X)T = 0$ for every $X \in \mathfrak{su}_2$. Since the Lie group $\mathrm{SU}_2 = \exp(\mathfrak{su}_2)$ is the exponential of its Lie algebra, we conclude that $g \cdot T = T$ for all $g \in \mathrm{SU}_2$.

A neat way of writing the SU_2-invariant tensor T is this:

$$T = (z - z')^N \ . \tag{4.42}$$

From it we immediately see that our invariant tensor T is symmetric for even N and skew for odd N. This concludes our illustration of the Corollary (end of Section 4.3.1) at the example of SU_2.

We finish the subsection by relating the result (4.42) to that of Section 4.3.2. Although the Lie groups SU_2 and SO_3 have identical Lie algebras $\mathfrak{su}_2 \simeq \mathfrak{so}_3$, they are not the same group. It turns out that the relation between them is qualitatively the same as the relation between Spin_{2n} and SO_{2n} (cf. Section 2.11). As a matter of fact, $\mathrm{SU}_2 = \mathrm{Spin}_3$. (In other words, $\mathrm{SU}_2 = \mathrm{Spin}_3$ can be obtained by exponentiating the Lie algebra $\mathfrak{su}_2 = \mathfrak{spin}_3 \simeq \mathrm{Cl}_2(\mathbb{R}^3)$ of skew-symmetrized degree-2 elements in the Clifford algebra of the Euclidean vector space $\mathbb{R}^3$.)

This means that there exists a 2 : 1 covering map $\mathrm{SU}_2 \to \mathrm{SO}_3$. It is constructed just like the 2 : 1 covering map $\mathrm{Spin}_{2n} \to \mathrm{SO}_{2n}$. Let us recall this construction. We conjugate the Pauli matrices σ_j (taking the role of the gamma matrices γ^μ) by $g \in \mathrm{SU}_2$. The result $g\sigma_j\, g^{-1}$ of conjugation is a traceless Hermitian 2×2 matrix. It is therefore expressible in terms of the Pauli matrices:

$$g\sigma_j\, g^{-1} = \sum_i \sigma_i\, R_{ij}(g)\ . \tag{4.43}$$

The real expansion coefficients can be arranged as a real 3×3 matrix $R(g) := \{R_{ij}(g)\}_{i,j=1,2,3}$. By construction, the correspondence $g \mapsto R(g)$ is a group homomorphism: $R(g_1 g_2) = R(g_1)R(g_2)$.

Problem. Show that the linear transformation $R(g)$ with matrix elements $R_{ij}(g)$ has the properties $R(g)^t = R(g)^{-1}$ and $\mathrm{Det}\, R(g) = 1$ of a rotation $r \equiv R(g) \in \mathrm{SO}_3$. □

From the definition (4.43) one sees that $R(-g) = R(+g)$. This is good evidence for the true fact (not proved here) that $R : \mathrm{SU}_2 \to \mathrm{SO}_3$ is a 2 : 1 covering of Lie groups. By this covering map, every representation $r \mapsto \rho_L(r)$ of SO_3 corresponds to a representation $g \mapsto \rho_L(R(g))$ of SU_2. One actually has $V_{2L}^{\mathrm{SU}_2} \simeq V_L^{\mathrm{SO}_3}$, expressing the fact that the angular momentum L translates to the quantum number $N = 2L$ in the present notation.

The converse is not true: the SU_2-representations for odd N do not correspond to representations of SO_3. Morally speaking, they carry half-integer angular momentum (or spin) $S = N/2 \in \mathbb{Z} + \frac{1}{2}$. In physics texts it is sometimes said that they are 'double-valued' representations of SO_3.

4.3.4 Invariant tensors of degree 3

After this rather exhaustive discussion of the situation for degree 2, we turn to degree 3. Here our main analytical tool will be the so-called Haar measure. In the following statement, the word 'compact Lie group' is to be interpreted in its widest sense (which includes finite groups).

Fact. For every compact Lie group G there exists a measure dg (called Haar measure) with the properties of invariance under right and left translation by any group element $g_0 \in G$:

$$\int_G f(g)\, dg = \int_G f(gg_0)\, dg = \int_G f(g_0 g)\, dg\ . \tag{4.44}$$

Here f is any integrable function on G.

Remark. The space of Haar measures for a compact Lie group G is one-dimensional [the invariance property (4.44) is stable under multiplication by scalars]. We fix the normalization by demanding the total mass to be unity: $\int_G dg = 1$.

Examples.

1. In the case of a finite group G of order $\mathrm{ord}(G)$, the integral with Haar measure is a sum

$$\int_G f(g)\,dg := \frac{1}{\mathrm{ord}(G)} \sum_{g\in G} f(g)\ . \tag{4.45}$$

The invariance (4.44) under right and left translations here follows directly from the fact that the mappings $g \mapsto gg_0$ and $g \mapsto g_0 g$ are one-to-one.

2. Let $G = \mathrm{U}_1 = \{z \in \mathbb{C} \mid \bar{z}z = 1\}$ be the group of unitary numbers in $\mathbb{C}$. Writing $z = \mathrm{e}^{\mathrm{i}\varphi}$, the unit-mass Haar measure for U_1 is given by $d\varphi/2\pi$:

$$\int_{\mathrm{U}_1} f(g)\,dg = \frac{1}{2\pi}\int_0^{2\pi} f(\mathrm{e}^{\mathrm{i}\varphi})\,d\varphi = \frac{1}{2\pi}\int_0^{2\pi} f(\mathrm{e}^{\mathrm{i}\varphi+\mathrm{i}\varphi_0})\,d\varphi\ . \tag{4.46}$$

3. Let the rotation group $G = \mathrm{SO}_3$ be parameterized by Euler angles:

$$g = R_3(\phi)R_1(\theta)R_3(\psi)\ , \tag{4.47}$$

where $R_j(\alpha)$ means an angle-α rotation around the j-axis. The Haar measure for SO_3 in these coordinates is expressed by

$$\int_{\mathrm{SO}_3} f(g)\,dg = \frac{1}{8\pi^2}\int_0^{2\pi}\!\!\int_0^{\pi}\!\!\int_0^{2\pi} f\big(R_3(\phi)R_1(\theta)R_3(\psi)\big)\,d\psi\,\sin\theta\,d\theta\,d\phi\ . \tag{4.48}$$

The proof of invariance requires some knowledge of the theory of Lie groups. □

Let now G be a compact Lie group with Haar measure dg and consider the tensor product $V_1 \otimes V_2 \otimes V_3$ of three (finite-dimensional) representations $\rho_l : G \to \mathrm{GL}(V_l)$ $(l = 1, 2, 3)$. In terms of bases for the factors V_l, the most general tensor $T \in V_1 \otimes V_2 \otimes V_3$ is expressed as

$$T = \sum\nolimits_{i,j,k} T_{ijk}\ e_i^{(1)} \otimes e_j^{(2)} \otimes e_k^{(3)}\ .$$

We can produce from T a G-invariant tensor T_{av} by taking the Haar average of its G-translates:

$$T_{\mathrm{av}} := \int_G (g\cdot T)\,dg = \sum_{i,j,k} T_{ijk}\int_G \big(\rho_1(g)e_i^{(1)}\big)\otimes\big(\rho_2(g)e_j^{(2)}\big)\otimes\big(\rho_3(g)e_k^{(3)}\big)\,dg\ . \tag{4.49}$$

Problem. Check the invariance $g\cdot T_{\mathrm{av}} = T_{\mathrm{av}}$ for all $g \in G$. □

In the following, if V is a vector space carrying a representation $\rho : G \to \mathrm{GL}(V)$, then we denote by $V^G \subset V$ the subspace of G-invariants in V. Thus if V_1, V_2, V_3 are G-representation spaces, then $(V_1\otimes V_2\otimes V_3)^G$ means the subspace of G-invariants in the tensor product $V_1\otimes V_2\otimes V_3$.

Now with a tensor $T \in V_1 \otimes V_2 \otimes V_3$ we associate a mapping $\mu(T) : V_3^* \to V_1 \otimes V_2$ in the canonical way. This correspondence $T \leftrightarrow \mu(T)$ is one-to-one. Moreover, if $T = g\cdot T$ is G-invariant, then the homomorphism $\mu(T)$ is G-equivariant, i.e.,

$$\forall g \in G\ :\quad \mu(T)(\varphi) = g\cdot\mu(T)(\varphi\circ g^{-1})\ . \tag{4.50}$$

Thus μ restricts to an isomorphism between the vector space $(V_1\otimes V_2\otimes V_3)^G$ of G-invariant tensors and the vector space $\mathrm{Hom}_G(V_3^*, V_1\otimes V_2)$ of G-equivariant homomorphisms.

4.4 The case of SU_2 : Wigner $3j$-symbol, Clebsch-Gordan coefficient

We now specialize to the case of $G = \mathrm{SU}_2 = \mathrm{Spin}_3$ and recall from Section 4.3.3 that the space

$$V_N = \mathrm{span}_{\mathbb{C}}\{1, z\,, z^2, \ldots, z^N\} \tag{4.51}$$

of degree-N polynomials in a complex variable z is an irreducible representation space for SU_2. The generators $\sigma_3\,,\ \sigma_+\,,\ \sigma_-$ of the complex Lie algebra $\mathfrak{sl}_2 = \mathfrak{su}_2 \oplus \mathrm{i}\,\mathfrak{su}_2$ are represented on V_N by

$$\rho_{N*}(\sigma_3) = 2z\frac{d}{dz} - N, \quad \rho_{N*}(\sigma_+) = -z^2\frac{d}{dz} + Nz\ , \quad \rho_{N*}(\sigma_-) = \frac{d}{dz}\ . \tag{4.52}$$

V_N is called the SU_2-representation for spin $J = N/2$. Its dimension is $\dim V_N = N + 1 = 2J + 1$. The eigenvalues of the generator $\rho_{N*}(\sigma_3/2)$ are $J, J-1, \ldots, -J+1, -J$.

In this section we investigate the space of invariant tensors

$$(V_{N_1} \otimes V_{N_2} \otimes V_{N_3})^{\mathrm{SU}_2} \simeq \mathrm{Hom}_{\mathrm{SU}_2}(V_{N_3}^*\,, V_{N_1} \otimes V_{N_2})\ . \tag{4.53}$$

The following can be regarded as a statement about the dimension of $\mathrm{Hom}_{\mathrm{SU}_2}(V_{N_3}^*\,, V_{N_1} \otimes V_{N_2})$.

Fact. The tensor product $V_{N_1} \otimes V_{N_2}$ of two irreducible SU_2-representation spaces decomposes as

$$V_{N_1} \otimes V_{N_2} \simeq V_{N_1+N_2} \oplus V_{N_1+N_2-2} \oplus \ldots \oplus V_{|N_1-N_2|}\ . \quad \square \tag{4.54}$$

To verify this, one starts from the general result (not proved here) that a tensor product of irreducible representations for a compact Lie group G is (isomorphic to) a direct sum of finitely many irreducible G-representations. In the present case, the irreducible summands occurring in $V_{N_1} \otimes V_{N_2}$ can be identified by the following counting procedure. Let

$$\varphi_{n_1,\,n_2} = z_1^{n_1} z_2^{n_2}\ , \quad 0 \le n_1 \le N_1\ , \quad 0 \le n_2 \le N_2\ ,$$

be a basis of homogeneous polynomials for $V_{N_1} \otimes V_{N_2}$, and let $V^k \subset V_{N_1} \otimes V_{N_2}$ be the eigenspace of eigenvalue k for the (shifted) polynomial-degree operator

$$(\rho_{N_1*} + \rho_{N_2*})(\sigma_3) = 2z_1\frac{\partial}{\partial z_1} + 2z_2\frac{\partial}{\partial z_2} - N_1 - N_2\ .$$

The next table lists the dimension of V^k as a function of k :

k	$\dim V^k$
$-N_1 - N_2$	1
$-N_1 - N_2 + 2$	2
$\vdots$	$\vdots$
$-N_1 - N_2 + 2\min\{N_1\,, N_2\}$	$\min\{N_1\,, N_2\} + 1$
$\vdots$	$\vdots$
$-N_1 - N_2 + 2\max\{N_1\,, N_2\}$	$\min\{N_1\,, N_2\} + 1$
$\vdots$	$\vdots$
$N_1 + N_2 - 2$	2
$N_1 + N_2$	1

We see that $\dim V^k$ increases linearly with k until it reaches a plateau extending from $\min\{N_1\,, N_2\}$ to $\max\{N_1\,, N_2\}$; after that it decreases linearly in such a way that $\dim V^k = \dim V^{-k}$.

By inspection of these dimensions it follows that $V_{N_1} \otimes V_{N_2}$ contains one multiplet of states with spin $J_{\max} = \frac{1}{2}(N_1 + N_2)$, one with spin $J = \frac{1}{2}(N_1 + N_2) - 1$, and so on, until we reach the plateau for the spin value $J_{\min} = \frac{1}{2}|N_1 - N_2|$. This establishes the decomposition (4.54).

Corollary. A statement equivalent to the decomposition (4.54) is

$$\mathrm{Hom}_{\mathrm{SU}_2}(V_{N_3}\,, V_{N_1} \otimes V_{N_2}) \simeq \begin{cases} \mathbb{C} & \text{if } |N_1 - N_2| \leq N_3 \leq N_1 + N_2\ ,\ N_1 + N_2 + N_3 \in 2\mathbb{N}\ , \\ 0 & \text{else}\ . \end{cases} \tag{4.55}$$

So far there was never any need to specify a Hermitian scalar product on V_N; now is a good place to fill this gap. Let $\varphi_n = z^n$ and

$$\langle \varphi_n\,, \varphi_{n'} \rangle_{V_N} := \delta_{n,\,n'} \binom{N}{n}^{-1}\ . \tag{4.56}$$

Problem. Show that with the choice (4.56) of Hermitian scalar product one has

$$\rho_{N*}(\sigma_-)^\dagger = \rho_{N*}(\sigma_+)\ , \tag{4.57}$$

which means that the SU_2-representation ρ_N on V_N is unitary. □

In the sequel we address the problem of coupling two spins $J_1 = N_1/2$ and $J_2 = N_2/2$ to total spin J. This problem is the same as that of constructing the following object.

Definition. A unitary SU_2-equivariant isomorphism

$$\phi_{N_1,\,N_2} : \bigoplus_{n=0}^{\min\{N_1,\,N_2\}} V_{N_1+N_2-2n} \to V_{N_1} \otimes V_{N_2} \tag{4.58}$$

is called a Clebsch-Gordan coefficient (for SU_2). Here the tensor product $V_{N_1} \otimes V_{N_2}$ is equipped with the natural Hermitian structure

$$\langle v_1 \otimes v_2\,, \tilde{v}_1 \otimes \tilde{v}_2 \rangle_{V_{N_1} \otimes V_{N_2}} = \langle v_1\,, \tilde{v}_1 \rangle_{V_{N_1}} \langle v_2\,, \tilde{v}_2 \rangle_{V_{N_2}} \tag{4.59}$$

induced by the Hermitian scalar products on V_{N_1} and V_{N_2}. □

The problem of computing explicit expressions for all Clebsch-Gordan coefficients of SU_2 was solved by the theoretical physicist Racah (Jerusalem, 1948). It seems to be difficult if not impossible to locate a transparent account of Racah's computation in the physics literature. Indeed, typical statements found in textbooks are "... and after some heavy analysis along these lines, Racah succeeded in reducing the expression for the Clebsch-Gordan coefficient to the following [result]". As will be shown below, verification of Racah's formula is quite straightforward, provided that the problem is approached from a good mathematical perspective.

To construct the Clebsch-Gordan coefficients of SU_2 explicitly, let $(N_1\,, N_2\,, N_3) \in (\mathbb{N} \cup \{0\})^3$ be any triple of non-negative integers subject to the constraints

$$|N_1 - N_2| \leq N_3 \leq N_1 + N_2\ , \qquad N_1 + N_2 + N_3 \in 2\mathbb{N}\ . \tag{4.60}$$

We then use the isomorphism (4.53) along with $V_{N_3} \simeq V_{N_3}^*$ to look for an invariant tensor

$$T \in (V_{N_1} \otimes V_{N_2} \otimes V_{N_3})^{\mathrm{SU}_2} \tag{4.61}$$

by a general ansatz of the form

$$T = \sum_{n_1=0}^{N_1} \sum_{n_2=0}^{N_2} \sum_{n_3=0}^{N_3} T_{n_1,n_2,n_3}\, z_1^{n_1} z_2^{n_2} z_3^{n_3} \ . \tag{4.62}$$

As usual, invariance means $g \cdot T = T$ for all $g \in \mathrm{SU}_2$. By linearizing this equation at $g = \mathrm{Id}$ and passing to the complexified Lie algebra $\mathfrak{sl}_2 \simeq \mathbb{C}^3$, we obtain three linearly independent conditions:

$$\sigma_3 \cdot T = \sum_{j=1}^{3} \left(2 z_j \frac{\partial}{\partial z_j} - N_j \right) T = 0 \ , \qquad \sigma_- \cdot T = \sum_{j=1}^{3} \frac{\partial T}{\partial z_j} = 0 \ ,$$
$$\sigma_+ \cdot T = \sum_{j=1}^{3} \left(-z_j^2 \frac{\partial}{\partial z_j} + N_j\, z_j \right) T = 0 \ . \tag{4.63}$$

The first one is easy to implement: it just says that the coefficient T_{n_1,n_2,n_3} vanishes unless

$$n_1 + n_2 + n_3 = \tfrac{1}{2}(N_1 + N_2 + N_3) \ . \tag{4.64}$$

The second condition amounts to

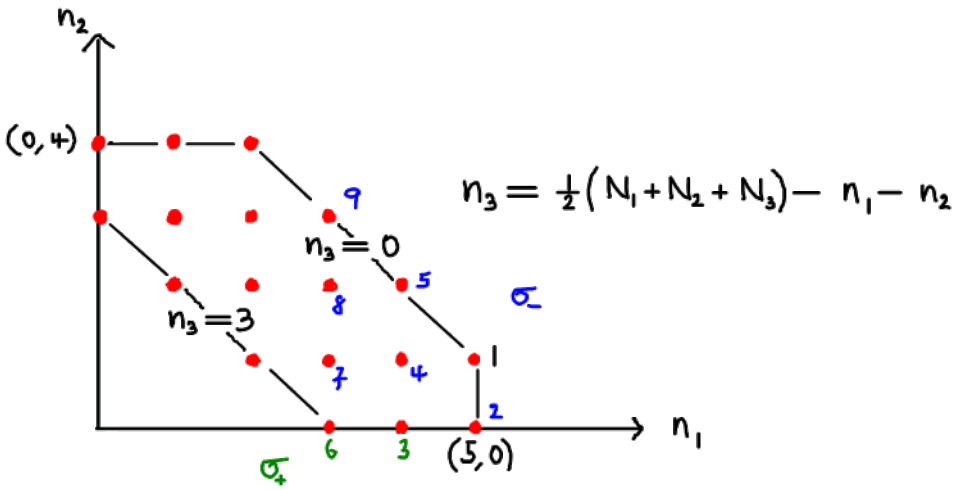
$$k_1\, T_{k_1,k_2-1,k_3-1} + k_2\, T_{k_1-1,k_2,k_3-1} + k_3\, T_{k_1-1,k_2-1,k_3} = 0 \tag{4.65}$$

for all triples $(k_1\,, k_2\,, k_3) \in \mathbb{Z}^3$ subject to $k_1 + k_2 + k_3 = \frac{1}{2}(N_1 + N_2 + N_3) + 2$. Finally, the last condition implies that

$$(N_1 - l_1)\, T_{l_1,l_2+1,l_3+1} + (N_2 - l_2)\, T_{l_1+1,l_2,l_3+1} + (N_3 - l_3)\, T_{l_1+1,l_2+1,l_3} = 0 \tag{4.66}$$

for all triples $(l_1\,, l_2\,, l_3) \in \mathbb{Z}^3$ subject to $l_1 + l_2 + l_3 = \frac{1}{2}(N_1 + N_2 + N_3) - 2$.

Eqs. (4.65) and (4.66) constitute a system of linear equations for the coefficients T_{n_1,n_2,n_3} with indices constrained by (4.64). We know from Corollary (4.55) that the solution space of this system is one-dimensional. Since the equations (4.65) and (4.66) connect only three of the unknowns T_{n_1,n_2,n_3} at a time, one can construct the solution by a recursive procedure. The following diagram shows how one might proceed for the example of $N_1 = 5$, $N_2 = 4$, $N_3 = 3$.



Inspired by Racah's explicit formula, we can actually express the solution in closed form:

Proposition. Let $J_l \equiv N_l/2$ for $l = 1, 2, 3$ and

$$T = \sum_{s_1,\, s_2,\, s_3} \frac{(-1)^{s_1+s_2+s_3}\; z_1^{J_1-s_2+s_3}\, z_2^{J_2-s_3+s_1}\, z_3^{J_3-s_1+s_2}}{(s_1-J_1)!\,(s_2-J_2)!\,(s_3-J_3)!\,(J_1+J_2-s_3)!\,(J_3+J_1-s_2)!\,(J_2+J_3-s_1)!} \tag{4.67}$$

where the sum runs over all triples $s_1\,, s_2\,, s_3$ such that the argument of each of the factorials is a non-negative integer. This tensor is invariant: $T = g \cdot T$ for all $g \in \mathrm{SU}_2\,$.

Proof. It is clear that the polynomial (4.67) is homogeneous of total degree $J_1 + J_2 + J_3$ and thus satisfies the constraint (4.64). Let us now show that it also satisfies the condition (4.65) or, equivalently, the second equation in (4.63). Introducing

$$A_J(x_1, x_2, x_3\,; y_1, y_2, y_3) := \frac{(-1)^{x_1+x_2+x_3}}{(x_1-J_1)!\,(x_2-J_2)!\,(x_3-J_3)!\,(J_1+J_2-y_3)!\,(J_3+J_1-y_2)!\,(J_2+J_3-y_1)!}$$

we have

$$T = \sum_{s_1,\, s_2,\, s_3} A_J(s_1, s_2, s_3\,; s_1, s_2, s_3)\; z_1^{J_1-s_2+s_3}\; z_2^{J_2-s_3+s_1}\; z_3^{J_3-s_1+s_2}\;,$$

and application of the lowering operator σ_- gives

$$\sigma_- \cdot T = \sum_{s_1,\, s_2,\, s_3} \left(\frac{J_1-s_2+s_3}{z_1} + \frac{J_2-s_3+s_1}{z_2} + \frac{J_3-s_1+s_2}{z_3} \right) \times A_J(s_1, s_2, s_3\,; s_1, s_2, s_3)\; z_1^{J_1-s_2+s_3}\; z_2^{J_2-s_3+s_1}\; z_3^{J_3-s_1+s_2}\;. \tag{4.68}$$

Now, by using the relation

$$J_1 - s_2 + s_3 = (J_3 + J_1 - s_2) + (s_3 - J_3)\;,$$

we get the following identity:

$$(J_1 - s_2 + s_3) A_J(s_1, s_2, s_3\,; s_1, s_2, s_3) = A_J(s_1, s_2, s_3\,; s_1, s_2+1, s_3) - A_J(s_1, s_2, s_3-1; s_1, s_2, s_3)\;.$$

Two more identities of the same kind result from cyclic permutations of the index set $\{1, 2, 3\}$. Next, in (4.68) we make a shift of summation index $s_3 \to s_3 - 1$ for the second term (z_2^{-1}) and $s_2 \to s_2 + 1$ for the third term (z_3^{-1}), in order to pull out common powers of z_1, z_2, z_3 and combine the various contributions. In this way we recast the expression (4.68) as

$$\begin{aligned} \sigma_- \cdot T = \sum \Big(& A_J(s_1, s_2, s_3\,; s_1, s_2+1, s_3) - A_J(s_1, s_2, s_3-1; s_1, s_2, s_3) \\ & + A_J(s_1, s_2, s_3-1; s_1, s_2, s_3) - A_J(s_1-1, s_2, s_3-1; s_1, s_2, s_3-1) \\ & + A_J(s_1, s_2+1, s_3\,; s_1+1, s_2+1, s_3) - A_J(s_1, s_2, s_3\,; s_1, s_2+1, s_3) \Big) \\ & \times z_1^{J_1-s_2+s_3-1}\; z_2^{J_2-s_3+s_1}\; z_3^{J_3-s_1+s_2}\;. \end{aligned}$$

We see that the second term cancels the third one, and the first term cancels the last one. The fourth term cancels the fifth one on making a common shift $s_j \to s_j + 1$ ($j = 1, 2, 3$) which leaves the powers of z_1, z_2, z_3 unchanged. This already completes the proof that $\sigma_- \cdot T = 0\,$.

Problem. By an adaptation of the calculation above, show that $\sigma_+ \cdot T = 0$. □

Recall that by the choice (4.56) of Hermitian scalar product for V_N, the basis vectors $\varphi_{N,n} = z^n$ of V_N have squared length

$$\langle \varphi_{N,n}\,,\varphi_{N,n}\rangle = \binom{N}{n}^{-1} = \frac{n!\,(N-n)!}{N!}\,.$$

The standard practice in physics is to work with orthonormal bases. Therefore, we now introduce the unit vectors

$$e_{J,M} := \binom{2J}{J+M}^{1/2} \varphi_{2J,J+M} = \frac{\sqrt{(2J)!}\;\varphi_{2J,J+M}}{\sqrt{(J+M)!\,(J-M)!}} \in V_{2J} \tag{4.69}$$

for $M = -J, -J+1, \ldots, J-1, J$.

Definition. Let $T_{n_1,n_2,n_3} \equiv T^{N_1,N_2,N_3}_{n_1,n_2,n_3}$ be the coefficients which are determined by comparing the ansatz (4.62) with the solution (4.67). The Wigner 3j-symbol $\begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix}$ is defined as

$$\begin{aligned}\begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix} &:= T^{2J_1,2J_2,2J_3}_{J_1+M_1,J_2+M_2,J_3+M_3} \sqrt{\triangle(J_1\,J_2\,J_3)} \\ &\times \sqrt{(J_1+M_1)!\,(J_1-M_1)!\,(J_2+M_2)!\,(J_2-M_2)!\,(J_3+M_3)!\,(J_3-M_3)!}\end{aligned} \tag{4.70}$$

where

$$\triangle(J_1\,J_2\,J_3) := \frac{(J_1+J_2-J_3)!\,(J_2+J_3-J_1)!\,(J_3+J_1-J_2)!}{(J_1+J_2+J_3+1)!} \tag{4.71}$$

is called a triangle coefficient.

Remark. Note that by (4.64) the Wigner 3j-symbol vanishes unless $M_1 + M_2 + M_3 = 0$. Note also that the Wigner 3j-symbol is just a scalar multiple of the coefficients of our tensor T given in (4.67). The following conclusion is therefore immediate.

Corollary. The tensor

$$\sum_{M_1,M_2,M_3} \begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix} e_{J_1,M_1} \otimes e_{J_2,M_2} \otimes e_{J_3,M_3} \tag{4.72}$$

is SU_2-invariant. □

By the isomorphism (4.53) in conjunction with the identification $V^*_{N_3} \simeq V_{N_3}$ due to the invariant pairing $V_N \otimes V_N \to \mathbb{C}$ (cf. Section 4.3.3), we finally arrive at what is called the Clebsch-Gordan coefficient for SU_2. It is given by the formula

$$\langle J_1 J_2\,; M_1 M_2 \mid J_1 J_2\,; J_3 M_3\rangle := (-1)^{J_1-J_2+M_3}\sqrt{2J_3+1}\begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & -M_3 \end{pmatrix}\,. \tag{4.73}$$

The sign change $M_3 \to -M_3$ reflects the isomorphism $V^*_{N_3} \simeq V_{N_3}$ determined by the mapping $f_{J_3,M_3} \to (-1)^{J_3-M_3} e_{J_3,-M_3}$ (dual basis $f_{J,M}$). The remaining sign factor $(-1)^{J_1-J_2-J_3}$ is due to what is known as the phase convention of Condon and Shortley. The normalization factor $\sqrt{2J_3+1}$ is motivated by the properties (4.75) and (4.76) below. A simplified notation is

$$\langle J_1 J_2\,; M_1 M_2 \mid J_1 J_2\,; J_3 M_3\rangle \equiv \langle J_1 J_2 M_1 M_2 \mid J_3 M_3\rangle\;. \tag{4.74}$$

The Dirac-style notation for the Clebsch-Gordan coefficient indicates that it effects a change of orthonormal basis $|J_1 J_2\,;J_3 M_3\rangle \to |J_1 M_1\rangle \otimes |J_2 M_2\rangle$.

Problem. Show that the Clebsch-Gordan coefficient obeys the following relations of orthonormality and completeness:

$$\sum_{M_1 M_2} \langle J_1 J_2 M_1 M_2 \mid J_3 M_3\rangle \, \langle J_1 J_2 M_1 M_2 \mid J_3' M_3'\rangle = \delta_{J_3\,,\,J_3'}\, \delta_{M_3\,,\,M_3'}\,, \tag{4.75}$$

$$\sum_{J_3 M_3} \langle J_1 J_2 M_1 M_2 \mid J_3 M_3\rangle \, \langle J_1 J_2 M_1' M_2' \mid J_3 M_3\rangle = \delta_{M_1\,,\,M_1'}\, \delta_{M_2\,,\,M_2'}\,. \quad \square \tag{4.76}$$

From an informed mathematical perspective, the Clebsch-Gordan coefficient provides us with an SU_2-equivariant homomorphism $\phi_{J_1 J_2} \in \mathrm{Hom}_{\mathrm{SU}_2}(V_{2J_3}\,, V_{2J_1} \otimes V_{2J_2})$ by

$$\phi_{J_1 J_2}(e_{J_3,\,M_3}) = \sum_{M_1 + M_2 = M_3} \langle J_1 J_2 M_1 M_2 \mid J_3 M_3\rangle \;\; e_{J_1,\,M_1} \otimes e_{J_2,\,M_2}\;; \tag{4.77}$$

thus it describes explicitly how the representation V_{2J_3} occurs in the tensor product $V_{2J_1} \otimes V_{2J_2}$.

4.5 Integrals of products of Wigner D-functions

For a compact Lie group G let $\rho:\; G \to \mathrm{GL}(V)$ be an irreducible representation. Fix a basis $\{e_l\}$ of V and define the matrix elements $D_{kl}(g)$ of the representation by

$$\rho(g)\, e_l = \sum_k e_k\, D_{kl}(g)$$

as usual. Let dg be Haar measure of G with total mass $\int_G dg = 1$. In the following we will use that fact that Haar measure is invariant under inversion, i.e. $\int_G f(g)\, dg = \int_G f(g^{-1})\, dg$.

Fact. The matrix elements of an irreducible representation satisfy the orthogonality relation

$$\int_G D_{kl}(g) D_{l'k'}(g^{-1})\, dg = \frac{\delta_{kk'}\delta_{ll'}}{\dim V}\,. \tag{4.78}$$

Proof. Applying $g \in G$ to any tensor $v \otimes \varphi \in V \otimes V^*$ we have

$$g \cdot (v \otimes \varphi) = \rho(g)v \otimes \varphi \circ \rho(g)^{-1}.$$

The Haar average of all translates $g \cdot (v \otimes \varphi)$ is a G-invariant tensor in $V \otimes V^*$. Because V is irreducible, there exists only one such tensor; this is the canonical invariant $\mu^{-1}(\mathrm{Id}_V) = \sum e_i \otimes f_i$; see Eq. (4.21) of Section 4.3.1. It follows that there exists a number $c(v,\varphi) \equiv c$ such that

$$\int_G \rho(g)v \otimes \varphi \circ \rho(g)^{-1} dg = c \sum_i e_i \otimes f_i\,.$$

Now we specialize this relation to $v = e_l$ and $\varphi = f_{l'}$ and then pass to components on each side using $f_k(\rho(g)e_l) = D_{kl}(g)$ and $f_{l'}(\rho(g)^{-1}e_{k'}) = D_{l'k'}(g^{-1})$. The result is

$$\int_G D_{kl}(g) D_{l'k'}(g^{-1})\, dg = c(e_l, f_{l'})\, \delta_{kk'}\,.$$

Next we use the invariance of Haar measure dg under inversion $g \mapsto g^{-1}$ to exchange the roles of the index pairs kk' and ll' and infer that $c(e_l, f_{l'}) = c_0 \delta_{ll'}$.

The constant c_0 is determined by setting $k = k'$, $l = l'$, and summing over l :

$$1 = \int_G dg = \int_G \sum_l D_{kl}(g) D_{lk}(g^{-1})\, dg = c_0 \sum_l \delta_{ll} = c_0 \dim V,$$

which gives the desired result $c_0 = (\dim V)^{-1}$.

Remark. If $D^{(1)}_{kl}(g)$ and $D^{(2)}_{l'k'}(g^{-1})$ are the matrix coefficients for two irreducible representations $V_1 \neq V_2$ in different isomorphism classes, then the integral $\int_G D^{(1)}_{kl}(g)\, D^{(2)}_{l'k'}(g^{-1})\, dg$ vanishes for all k, l, k', l'. Indeed, any non-zero value of the integral would imply a G-invariant tensor in $V_1 \otimes V_2^*$, but we know from Section 4.3.1 that no such tensor exists. □

From now on we restrict the discussion to the case of $G = \mathrm{SU}_2$, where we use the notation $\mathcal{D}^J_{MN}(g) \equiv f_{J,M}(\rho_J(g)\, e_{J,N})$ for the matrix elements (also known in physics as the Wigner D-functions) of the irreducible representation of spin J. For future reference we record that

$$\int_{\mathrm{SU}_2} \mathcal{D}^J_{MN}(g)\, \mathcal{D}^{J'}_{N'M'}(g^{-1})\, dg = \delta_{JJ'} \frac{\delta_{MM'}\, \delta_{NN'}}{2J+1} \,. \tag{4.79}$$

In the remainder of this section, we are going to derive a formula for the Haar integral of a product of three Wigner D-functions. To that end, we consider the Haar average of all SU_2-translates of some tensor $X \in V_{2J_1} \otimes V_{2J_2} \otimes V_{2J_3}$ in the tensor product of three irreducible representations with spins J_1 , J_2 , J_3 . Since the Haar-averaged tensor $\int (g \cdot X)\, dg$ is SU_2-invariant and there exists only one such tensor, namely that of Corollary (4.72), we have

$$\int_{\mathrm{SU}_2} (g \cdot X)\, dg = c_3(X) \sum_{M_1,\, M_2,\, M_3} \begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix} e_{J_1,\, M_1} \otimes e_{J_2,\, M_2} \otimes e_{J_3,\, M_3} \,,$$

with some unknown number $c_3(X)$. Now let $X = e_{J_1,\, N_1} \otimes e_{J_2,\, N_2} \otimes e_{J_3,\, N_3}$. By passing to components with respect to the chosen basis we obtain

$$\int_{\mathrm{SU}_2} \mathcal{D}^{J_1}_{M_1N_1}(g)\, \mathcal{D}^{J_2}_{M_2N_2}(g)\, \mathcal{D}^{J_3}_{M_3N_3}(g)\, dg = c_3(e_{J_1,\, N_1} \otimes e_{J_2,\, N_2} \otimes e_{J_3,\, N_3}) \begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix} \,. \tag{4.80}$$

Our next step is to exchange the roles of vectors and co-vectors. A quick (if dirty) way of doing this is to take the complex conjugate of both sides of the equation and exploit the unitarity of the representation:

$$\overline{\mathcal{D}^J_{MN}(g)} = \mathcal{D}_{NM}(g^{-1}) \,.$$

Haar measure dg does not change under the substitution of integration variable $g \to g^{-1}$ (this is an immediate consequence of Haar measure being unique). Since the Wigner $3j$-symbol is real it follows that

$$\int_{\mathrm{SU}_2} \mathcal{D}^{J_1}_{M_1N_1}(g)\, \mathcal{D}^{J_2}_{M_2N_2}(g)\, \mathcal{D}^{J_3}_{M_3N_3}(g)\, dg = \overline{c_3(e_{J_1,\, M_1} \otimes e_{J_2,\, M_2} \otimes e_{J_3,\, M_3})} \begin{pmatrix} J_1 & J_2 & J_3 \\ N_1 & N_2 & N_3 \end{pmatrix} \,.$$

By comparing this with (4.80) we are led to the following conclusion.

Fact. For the Haar integral of a product of three Wigner D-functions one has the formula

$$\int_{\mathrm{SU}_2} \mathcal{D}^{J_1}_{M_1N_1}(g)\,\mathcal{D}^{J_2}_{M_2N_2}(g)\,\mathcal{D}^{J_3}_{M_3N_3}(g)\,dg = \begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix} \begin{pmatrix} J_1 & J_2 & J_3 \\ N_1 & N_2 & N_3 \end{pmatrix} . \tag{4.81}$$

Proof. Our previous considerations imply (4.81) up to an unknown constant of proportionality. It remains to show that this constant is unity. For that purpose we set $M_l = N_l$ $(l = 1, 2, 3)$ and sum over magnetic quantum numbers. On the right-hand side we get

$$\sum_{M_1M_2M_3} \begin{pmatrix} J_1 & J_2 & J_3 \\ M_1 & M_2 & M_3 \end{pmatrix}^2 = \sum_{M_3} \sum_{M_1M_2} (2J_3+1)^{-1} \langle J_1J_2M_1M_2 \mid J_3 - M_3\rangle^2 = \sum_{M_3} (2J_3+1)^{-1} = 1$$

from (4.73) and the orthonormality property (4.76) for the Clebsch-Gordan coefficient. To do the integral on the left-hand side, we use the relation

$$\chi_{J_1}(g)\chi_{J_2}(g) = \sum_{J=|J_1-J_2|}^{J_1+J_2} \chi_J(g)\,, \quad \chi_J(g) := \sum_{M=-J}^{J} \mathcal{D}^J_{MM}(g)\,, \tag{4.82}$$

which follows rather easily from

$$\chi_J(\mathrm{e}^{\mathrm{i}\theta\sigma_3}) = \sum_{M=-J}^{J} \mathrm{e}^{\mathrm{i}M\theta} = \frac{\sin\big((2J+1)\theta/2\big)}{\sin(\theta/2)}\,. \tag{4.83}$$

Thus our integral after summation over magnetic quantum numbers becomes

$$\int_{\mathrm{SU}_2} \chi_{J_1}(g)\chi_{J_2}(g)\chi_{J_3}(g)\,dg = \sum_{J=|J_1-J_2|}^{J_3} \int_{\mathrm{SU}_2} \chi_J(g)\chi_{J_3}(g)\,dg = \int_{\mathrm{SU}_2} \chi_{J_3}(g)^2\,dg = 1\,,$$

where in the last two steps we used the orthogonality (4.79).

Fact. A variant of the integral above is

$$\int_{\mathrm{SU}_2} \mathcal{D}^{J_1}_{M_1N_1}(g)\,\mathcal{D}^{J_2}_{M_2N_2}(g)\,\overline{\mathcal{D}^{J_3}_{M_3N_3}(g)}\,dg = \frac{\langle J_1J_2M_1M_2 \mid J_3M_3\rangle\,\langle J_1J_2N_1N_2 \mid J_3N_3\rangle}{2J_3+1}\,. \tag{4.84}$$

Proof. Using completeness (4.76) of the Clebsch-Gordan coefficients we invert the relation (4.77):

$$e_{J_1,\,N_1} \otimes e_{J_2,\,N_2} = \sum_{JN} \langle J_1J_2N_1N_2 \mid JN\rangle\ \phi_{J_1J_2}(e_{J,\,N})\,.$$

We then apply any transformation $g \in \mathrm{SU}_2$ to obtain

$$\sum_{M_1M_2} e_{J_1,\,M_1} \otimes e_{J_2,\,M_2}\,\mathcal{D}^{J_1}_{M_1N_1}(g)\,\mathcal{D}^{J_2}_{M_2N_2}(g) = \sum_{JNM} \langle J_1J_2N_1N_2 \mid JN\rangle\ \phi_{J_1J_2}(e_{J,\,M})\,\mathcal{D}^J_{MN}(g)\,.$$

On the right-hand side we insert the expression (4.77) for $\phi_{J_1J_2}(e_{J,\,M})$ and pass to components:

$$\mathcal{D}^{J_1}_{M_1N_1}(g)\,\mathcal{D}^{J_2}_{M_2N_2}(g) = \sum_{JMN} \langle J_1J_2N_1N_2 \mid JN\rangle\,\langle J_1J_2M_1M_2 \mid JM\rangle\ \mathcal{D}^J_{MN}(g)\,.$$

Finally, we integrate both sides against $\overline{\mathcal{D}^{J_3}_{M_3N_3}(g)} = \mathcal{D}^{J_3}_{N_3M_3}(g^{-1})$ with Haar measure dg. The desired result (4.84) then follows by using the orthogonality relation (4.79) on the right-hand side.

4.6 Tensor operators, Wigner-Eckart Theorem

Recall the action of $g \in \mathrm{SU}_2 = \mathrm{Spin}_3$ on the basis vectors $e_{J,M} \in V_{2J}$ of the representation for spin J:

$$g \cdot e_{J,N} = \sum_M e_{J,M}\, \mathcal{D}^J_{MN}(g) \ .$$

The following definition is motivated by the quantum mechanical fact that if quantum states transform as $\psi \mapsto g\,\psi$ then quantum operators transform by conjugation $\mathrm{Op} \mapsto g\,\mathrm{Op}\,g^{-1}$.

Definition. By an irreducible tensor operator of rank J one means a set of operators $\{T_{JM}\}_{M=-J,\dots,J}$ transforming under rotations $g \in \mathrm{SU}_2 = \mathrm{Spin}_3$ as

$$g\,T_{JN}\,g^{-1} = \sum_M T_{JM}\,\mathcal{D}^J_{MN}(g) \ . \tag{4.85}$$

Example. The operation of multiplying (the angular part of the wave function) by a spherical harmonic Y_{LM} is an irreducible tensor operator of rank $J = L$. Important examples of tensor operators are furnished by the problem of expanding, say, the charge density operator $\hat\rho$ by multipoles with respect to a distinguished point (typically the center of mass). The three (spherical) components of the dipole part of $\hat\rho$ form a tensor operator of rank $J = 1$, the five components of the quadrupole part of $\hat\rho$ form a tensor operator of rank $J = 2$, and so on. □

If $J_x\,, J_y\,, J_z$ are the operators of total angular momentum, and $J_\pm = \frac{1}{2}(J_x \pm \mathrm{i}J_y)$, then the infinitesimal version of (4.85) reads

$$[J_z\,, T_{JM}] = \hbar M\, T_{JM} \ , \quad [J_\pm\,, T_{JM}] = \hbar\sqrt{(J \mp M)(J \pm M + 1)}\; T_{J,M\pm1} \ . \tag{4.86}$$

Wigner-Eckart theorem. Let T_{JM} be (the components of) an irreducible tensor operator. Its matrix elements between two quantum mechanical states of definite spin and spin projection are products

$$\langle J_2 M_2 \mid T_{JM} \mid J_1 M_1\rangle = \langle J_1 J M_1 M \mid J_2 M_2\rangle\ \langle J_1 \parallel T_J \parallel J_2\rangle \tag{4.87}$$

of a Clebsch-Gordan coefficient and a so-called reduced matrix element.

Remark. The dependence of the matrix element on the quantum numbers $M_1\,, M, M_2$ is given entirely by the Clebsch-Gordan coefficient. The main task (possibly difficult) in practical applications is to compute the reduced matrix element $\langle J_1 \parallel T_J \parallel J_2\rangle$.

Proof. We begin by writing

$$\langle J_2 M_2 \mid T_{JM} \mid J_1 M_1\rangle = \left\langle J_2 M_2 \mid g^{-1}\left(g\,T_{JM}\,g^{-1}\right) g \mid J_1 M_1\right\rangle \ .$$

Then we expand $g\,|J_1 M_1\rangle = \sum |J_1 N_1\rangle\ \mathcal{D}^{J_1}_{N_1 M_1}(g)$ and

$$\langle J_2 M_2|\, g^{-1} = \langle g\cdot(J_2 M_2)| = \sum_{N_2} \langle J_2 N_2|\ \overline{\mathcal{D}^{J_2}_{N_2 M_2}(g)} \ ,$$

using the unitarity of the SU_2-action. We thus obtain

$$\langle J_2 M_2 \mid T_{JM} \mid J_1 M_1\rangle = \sum_{N_1 N_2 N} \langle J_2 N_2 \mid T_{JN} \mid J_1 N_1\rangle\, \mathcal{D}^{J_1}_{N_1 M_1}(g)\, \mathcal{D}^J_{NM}(g)\, \overline{\mathcal{D}^{J_2}_{N_2 M_2}(g)} \ .$$

This result holds true for all $g \in \mathrm{SU}_2$ and remains true if we average both sides with SU_2-Haar measure. Hence, by using formula (4.84) we have

$$\langle J_2 M_2 \mid T_{JM} \mid J_1 M_1 \rangle = \langle J_1 J M_1 M \mid J_2 M_2 \rangle \sum_{N_1 N_2 N} \langle J_2 N_2 \mid T_{JN} \mid J_1 N_1 \rangle \, \frac{\langle J_1 J N_1 N \mid J_2 N_2 \rangle}{2J_2 + 1} \, .$$

This gives the formula (4.87) of the Wigner-Eckart theorem along with an expression for the reduced matrix element:

$$\langle J_1 \parallel T_J \parallel J_2 \rangle = (2J_2 + 1)^{-1} \sum_{N_1 N_2 N} \langle J_2 N_2 \mid T_{JN} \mid J_1 N_1 \rangle \, \langle J_1 J N_1 N \mid J_2 N_2 \rangle \, . \tag{4.88}$$

5 Dirac quantization condition

The original plan of this lecture course was to finish with a chapter on the foundations of quantum mechanics (superposition principle: Schrödinger cat states, EPR paradox, violation of Bell's inequality, decoherence, many-worlds interpretation). Since there is not enough time left for a substantial discussion of the subject, we change course and conclude with a short chapter on another foundational theme of quantum mechanics: the quantization of electric charge.

Experimentally one observes that electric charge always occurs as an integer multiple $q_e = ne$ ($n \in \mathbb{Z}$) of a fundamental charge quantum e. Why nature would have it this way is not known; the phenomenon of charge quantization is in fact one of the open questions of theoretical physics.

The situation improves, however, if there are magnetic charges. Allowing for the existence of magnetic monopoles, Dirac (1931) argued that quantum mechanics is consistent if and only if the product of any pair q_e, q_m of electric and magnetic charges is an integer multiple of $2\pi\hbar$:

$$\boxed{q_e \, q_m \in 2\pi\hbar \, \mathbb{Z} \, .} \tag{5.1}$$

This condition, known as the Dirac quantization condition, can be read in two directions. Given a smallest magnetic charge μ, it quantizes the electric charge according to $q_e \in (2\pi\hbar/\mu)\,\mathbb{Z}$. Conversely, given an electric charge quantum e, magnetic charge is quantized by $q_m \in (2\pi\hbar/e)\,\mathbb{Z}$.

The goal of the present chapter is to offer some physical heuristics and mathematical background for the Dirac quantization condition (5.1). Please be advised that this story, albeit a corner stone of theoretical physics, is speculative in the context of physics as an experimental science, as magnetic monopoles have never been observed in the laboratory to this day. [However, quoting string theorist J. Polchinski (arXiv:hep-th/0304042): "the existence of magnetic monopoles seems like one of the safest bets that one can make about physics not yet seen".]

To begin, here is how the equations of electrodynamics (in traditional vector notation) would have to look for a non-zero magnetic charge density ρ_m and magnetic current density $\mathbf{j}_m$:

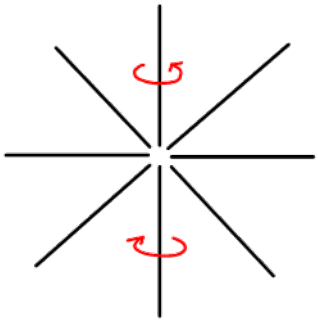
$$\mathrm{rot}\, \mathbf{H} = \dot{\mathbf{D}} + \mathbf{j}_e \, , \quad \mathrm{div}\, \mathbf{D} = \rho_e \, , \quad \mathbf{D} = \varepsilon_0 \, \mathbf{E} \, , \tag{5.2}$$

$$-\mathrm{rot}\, \mathbf{E} = \dot{\mathbf{B}} + \mathbf{j}_m \, , \quad \mathrm{div}\, \mathbf{B} = \rho_m , \quad \mathbf{B} = \mu_0 \, \mathbf{H} \, . \tag{5.3}$$

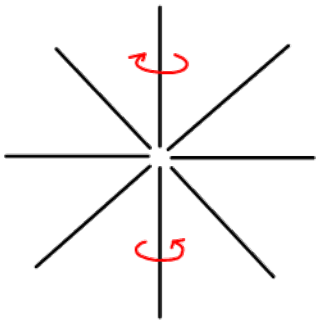
The Lorentz force on a particle of velocity $\mathbf{v}$, electric charge q_e, and magnetic charge q_m, is

$$\mathbf{K} = q_e \left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) + q_m \left(\mathbf{H} - \mathbf{v} \times \mathbf{D}\right) . \tag{5.4}$$

As a particular consequence of the system of equations (5.2) and (5.3), magnetic charge obeys a continuity equation $\dot{\rho}_m + \mathrm{div}\, \mathbf{j}_m = 0$ and hence is conserved. There is no difference between electric and magnetic charges in that respect. However, there does exist a difference of transformation behavior: in order for electrodynamics to be parity-invariant, ρ_m must transform as a pseudo-scalar and $\mathbf{j}_m$ as an axial vector, whereas ρ_e is a scalar and $\mathbf{j}_e$ is a (polar) vector. In other words, while electric charges come with a sign (they are positive or negative), magnetic charges carry a handedness: a magnetic monopole is neither positive nor negative, but is right-handed or left-handed.



magnetic field of a
right-handed monopole



magnetic field of a
left-handed monopole

A quick motivation for the Dirac quantization condition (5.1) is the following. Electric and magnetic charges are sources of electric and magnetic flux, respectively. Therefore, in the simultaneous presence of electric and magnetic monopoles there exist crossed electric and magnetic fields (even in the static limit). These carry momentum (as given by the Poynting vector) as well as angular momentum. For two monopole charges with values q_e and q_m the total angular momentum is proportional to the product $q_e\, q_m$. The condition (5.1) then reflects the fact that angular momentum in quantum mechanics takes quantized values only.

It should be stressed that this line of reasoning was *not* the route by which Dirac arrived at (5.1); his argument will be sketched in the next section.

5.1 Dirac's argument

To make Dirac's point it will be enough to consider a charged particle (which may be relativistic or not, it doesn't matter for our present purposes) moving in a static magnetic field in three-dimensional space. Hence, for simplicity of the argument, we neglect any time dependence of the electromagnetic field and focus on its magnetic part (ignoring the electric field).

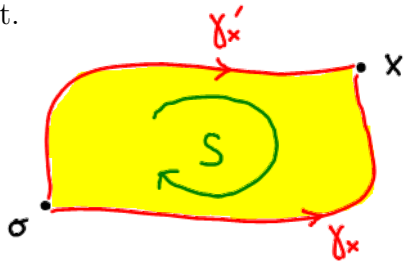
Dirac's argument is based on a generalization of the usual way of thinking about the wave functions of quantum mechanics. The idea is to eliminate the magnetic field from the formalism by relaxing the condition that the wave function has to be single-valued. Needless to say, one cannot eliminate the magnetic field $\mathbf{B}$ by a gauge transformation. It is, however, possible to remove $\mathbf{B}$ by letting the wave function be path-dependent; Dirac spoke of 'non-integrable phases

for wave functions'. It goes like this.

Fixing some reference point $o \in E_3$ (the 'origin' of three-dimensional Euclidean space) one works in a representation in which the wave function at any point $x \in E_3$ depends on not just x but also on a path γ_x connecting o with x. Thus one replaces the usual wave function $x \mapsto \psi(x)$ by a functional on paths, $\gamma_x \to \Psi[\gamma_x]$. This generalization comes with a condition (in order to remove a redundancy which would otherwise ensue): if γ_x and γ'_x are any two paths connecting o with the same end point x, then the values of the wave function for these paths are required to be related by a known phase factor:

$$\Psi[\gamma'_x] = \mathrm{e}^{2\pi \mathrm{i}\, \Phi(S)}\, \Psi[\gamma_x]\,, \qquad \partial S = \gamma'_x - \gamma_x\,, \tag{5.5}$$

where $\Phi(S) = (e/h) \int_S \mathbf{B} \cdot d^2\mathbf{n}$ is the magnetic flux (measured in units of the flux quantum h/e) through any oriented surface S with boundary $\partial S = \gamma'_x - \gamma_x$. The condition (5.5) ensures that the wave function $\Psi[\gamma_x]$ for a single reference path γ_x already determines the wave function for all other paths γ'_x leading to the same end point x. Notice in particular that the absolute value $|\Psi[\gamma_x]| =: a(x)$ is path-independent.



If the magnetic field is bounded, the functional $\gamma_x \mapsto \Psi[\gamma_x]$ is continuous. In other words, small changes of the point x and/or the path γ_x map to small changes in the wave function. The same is true for the phase

$$\arg \Psi[\gamma_x] = (2\mathrm{i})^{-1} \log \left(\Psi[\gamma_x] \,/\, \overline{\Psi[\gamma_x]} \right),$$

as long as the wave function does not vanish.

Zeroes of the wave function need a separate discussion, as follows. Because we are in three dimensions and two conditions ($\mathfrak{Re}\,\Psi = 0$ and $\mathfrak{Im}\,\Psi = 0$) have to be satisfied in order for a complex-valued wave function to vanish, wave function zeroes generically organize into lines of zeroes. Such lines are called nodal lines of the wave function.

The change of phase of the wave function along a small loop encircling a nodal line need not be small; all one can say in general is that the change has to be close to an integer multiple of 2π. For example, if a wave function vanishes on the z-axis ($x = y = 0$) as $(x \pm \mathrm{i}y)^n$, then the change of phase for a path looping once around the z-axis is $\pm 2\pi n$. For present use let us note that the winding of the phase around a nodal line endows that line with a sense of circulation.

The winding of the phase around a nodal line determines a sense of circulation.

With these preparations made, we can state the central point of the discussion: although the setting of quantum mechanics as we know it makes for nodal lines without any beginning or end, there exists no fundamental principle forcing that property, and Dirac suggested the possibility for nodal lines to have points of birth or termination. He called such points 'nodal singularities'.

Dirac's bold step of admitting nodal singularities does not come for free but leads to a constraint, as we shall now explain. When positing the relation

$$\arg \Psi[\gamma_x'] - \arg \Psi[\gamma_x] = 2\pi\Phi(S) \tag{5.6}$$

between phases, we must take into account the contributions from the phase singularities due to nodal lines. Thus we need to augment the (dimensionless) magnetic flux through the surface S by the winding numbers n_i of the nodal lines γ_i that intersect S :

$$\Phi(S) = \frac{e}{h}\int_S \mathbf{B}\cdot d^2\mathbf{n} + \sum \pm n_i \ . \tag{5.7}$$

Here the winding numbers n_i (which are positive by convention) count with a sign factor which is determined by comparing the orientation of the surface S with the sense of circulation of the nodal line γ_i . If these agree, n_i counts with the plus sign; otherwise the minus sign applies.

The said constraint now arises from the indeterminacy of S: there exist many oriented surfaces S with boundary $\partial S = \gamma_x' - \gamma_x$ and one has no way a priori of knowing which to choose. Therefore, in order for Dirac's extended framework (with nodal singularities) to make sense, one has to require that $\Phi(S) = \Phi(S')$ for any pair of oriented surfaces S, S' with the same boundary $\partial S = \partial S'$. This is equivalent to demanding that $\Phi(S)$ in (5.7) vanishes for every closed surface $S = \partial U$. Thus, for every $S = \partial U$ we require that the dimensionless magnetic flux through ∂U is exactly canceled by the winding number flux due to nodal singularities in the enclosed region U.

By further analyzing the situation, Dirac concluded that his extended theoretical framework was consistent if and only if all the wave functions shared the same set of nodal singularities. (The nodal lines at large will of course be different for two different wave functions in general.) Dirac took this as evidence that the nodal singularities are objects of physical reality. By the vanishing of $\Phi(\partial U)$ in (5.7) they are sources or sinks of magnetic flux, so they are to be interpreted as magnetic monopoles. Since the integral $\int_{\partial U} \mathbf{B}\cdot d^2\mathbf{n}$ computes the magnetic flux through ∂U and thus the magnetic charge contained in U, the constraint $\Phi(\partial U) = 0$ amounts to the Dirac quantization condition (5.1). As a historical note: Dirac himself was so confident of his generalized formalism of quantum mechanics in the presence of magnetic monopoles that he wrote "one would be surprised if Nature had made no use of it" [P.A.M. Dirac, *Quantized singularities of the electromagnetic field*, Proc. Roc. Soc. A **133** (1931) 60].

5.2 The treatment of Wu & Yang

Having come to terms with the Dirac quantization condition, we wish to learn how to do actual calculations when magnetic monopoles are present. While Dirac's general framework of path-dependent wave functionals $\gamma_x \mapsto \Psi[\gamma_x]$ is a good tool for conceptual reasoning, it is probably not so convenient for doing concrete calculations. For that, we would like to return to the familiar framework using wave functions $x \mapsto \psi(x)$.

However, a difficulty is now in store for us. To discuss it in very concrete terms, let us consider the problem of a non-relativistic particle of mass m and charge e moving in the magnetic field

$$B = \frac{\mu}{4\pi}\,\omega\,, \qquad \omega = \frac{x\,dy\wedge dz + y\,dz\wedge dx + z\,dx\wedge dy}{(x^2+y^2+z^2)^{3/2}}\,, \tag{5.8}$$

of a static monopole of magnetic charge $\mu = n\,h/e$ (with $n \in \mathbb{Z}$) placed at the origin of a system of Cartesian coordinates $x\,, y\,, z\,$. In vector notation one has the expression

$$\mathbf{B} = \frac{\mu}{4\pi}\,\frac{e_r}{r^2}$$

(which is not fully accurate as e_r is a polar vector whereas $\mathbf{B}$ has to transform as an axial vector). To treat this problem, we need a Hamiltonian H. The textbook quantum mechanics of charged particles would tell us to fix a vector potential $\mathbf{A} \in \mathrm{rot}^{-1}\mathbf{B}$ and write

$$H = \frac{(\mathbf{p} - e\mathbf{A})^2}{2m}\,. \tag{5.9}$$

Unfortunately, in the present situation there simply does not exist any vector potential $\mathbf{A}$ such that $\mathrm{rot}\,\mathbf{A} = \mathbf{B}\,$. Indeed, by integrating $\mathbf{B}$ over a surface $S = \partial U$ enclosing the magnetic monopole, we run into a contradiction:

$$0 \neq \mu = \int_S \mathbf{B}\cdot d^2\mathbf{n} \overset{\text{Gauss}}{=} \int_U \mathrm{div}\,\mathbf{B}\,d^3x = 0\,,$$

since $\mathrm{div}\,\mathbf{B}$ vanishes by $\mathrm{div}\circ\mathrm{rot} = 0$ whenever $\mathbf{B} = \mathrm{rot}\,\mathbf{A}\,$. Thus there isn't any $\mathbf{A}$, and it would hence appear that there exists no Hamiltonian (5.9).

To overcome this difficulty, Dirac invented a trick which is nowadays referred to as the 'Dirac string'. We do not describe it here but give a more modern treatment due to Wu and Yang (1968). The idea is to exclude the singular point $o \in E_3$ and work with two overlapping domains $U^\pm$ (and corresponding coordinate charts) for the punctured space $E_3 \setminus \{o\} = U^+ \cup U^-$. To be concrete, we take U^+ (U^-) to be $E_3 \setminus \{o\}$ with the negative (resp. positive) z-axis removed. On $U^+ \cap U^-$ we then consider a pair of one-forms

$$A^{(\pm)} = \pm\frac{n\hbar}{2e}\,(1 \mp \cos\theta)\,d\phi = \pm\frac{n\hbar}{2e}\left(1 \mp \frac{z}{\sqrt{x^2+y^2+z^2}}\right)\frac{x\,dy - y\,dx}{x^2+y^2}\,. \tag{5.10}$$

Problem. Compute the exterior derivatives $dA^{(\pm)}$ to show that both $A^{(+)}$ and $A^{(-)}$ are magnetic vector potentials for the magnetic field strength $B = dA^{(+)} = dA^{(-)}$ in (5.8). □

The magnetic vector potential $A^{(+)}$ is singular on the negative z-axis ($z \leq 0$) but extends as a smooth form to the positive z-axis ($z > 0$); thus $A^{(+)}$ is good on U^+. For $A^{(-)}$ it is the other

way around (singular for $z \geq 0$ but good on U^-). On the overlap $U^+ \cap U^-$ of their domains of definition, the two forms are related by a gauge transformation:

$$A^{(+)} - A^{(-)} = \frac{n\hbar}{e}\, d\phi = \frac{n\hbar}{e}\, \frac{x\, dy - y\, dx}{x^2 + y^2} \,. \tag{5.11}$$

We now recall that by the principle of gauge invariance in quantum mechanics, a gauge transformation $A \to A + df$ of the magnetic vector potential implies a corresponding gauge transformation $\psi \to \mathrm{e}^{\mathrm{i}(e/\hbar)\, f}\, \psi$ of the wave function. Accordingly, in the present case of (5.11) we obtain

$$\psi^{(+)} = \mathrm{e}^{\mathrm{i}n\phi}\, \psi^{(-)} = \left(\frac{x + \mathrm{i}y}{x - \mathrm{i}y} \right)^{n/2} \psi^{(-)} \,, \tag{5.12}$$

which ensures that $(d - \mathrm{i}eA^{(+)}/\hbar)\psi^{(+)} = \mathrm{e}^{\mathrm{i}n\phi}\, (d - \mathrm{i}eA^{(-)}/\hbar)\psi^{(-)}$ on $U^+ \cap U^-$. Mathematically speaking, the relation (5.12) means that the wave functions $\psi^{(\pm)}$ are local expressions of a globally defined object, ψ, which is called a section of (some) complex line bundle over $E_3 \setminus \{o\}$. Similarly, $A^{(\pm)}$ are the local coordinate expressions for a connection, A, on the same line bundle.

In order to compute the quantum mechanics of our charged particle in the magnetic monopole field (5.8), the prescription of Wu and Yang is to solve the Schrödinger equation on U^+ for $\psi^{(+)}$ using $A^{(+)}$, and on U^- for $\psi^{(-)}$ using $A^{(-)}$. In executing that procedure the relations (5.11) and (5.12) are in force. Some of the computational details will be presented for a somewhat simplified setting in the next subsection. Here we wish to conclude with the remark that the above can be regarded as another derivation of the Dirac quantization condition; indeed, the number n is the magnetic monopole charge (in units of the flux quantum h/e), and the relation (5.12) makes sense (as a well-defined transition between coordinate charts) if and only if n is an integer.

5.3 Special case: monopole charge $n = 2$

In the sequel, we further develop the modern mathematical treatment of the Dirac monopole problem. For pedagogical reasons, we first take a detailed look at the special case of monopole charge $\mu = 2h/e$ – that's twice the magnetic charge quantum (or flux quantum) h/e. This case has the attractive feature that several key objects of the mathematical description can be grasped rather easily by our Euclidean three-dimensional intuition.

To keep the discussion as simple as possible and concentrate on the new things to be learned, we will assume that the motion of our charged particle is constrained to a sphere $M \simeq \mathrm{S}^2$ (or surface of a ball) of radius R centered around the origin o. (If that seems too contrived, I trust you can handle the original problem on $E_3 \setminus \{o\}$ by separation of variables.)

5.3.1 Tangent bundle

We are now going to show that the wave functions for the special case of $\mu = 2h/e$ (with the motion constrained to $M \simeq \mathrm{S}^2$) can be interpreted as vector fields on M. You might object that as a student of quantum theory you were taught to think of the wave function of a charged particle as a function taking values in the complex numbers, $\mathbb{C}$. If so, the following observation should

be helpful: at every point $x \in M$ of a sphere, the vectors $v \in T_xM$ tangent to the sphere span a two-dimensional vector space isomorphic to $\mathbb{R}^2$, which in turn is isomorphic to $\mathbb{C}$.

The argument linking the case of $\mu = 2h/e$ to vector fields goes as follows. Using stereographic projection we introduce two sets of complex coordinate functions $\zeta^{(\pm)} : \mathrm{S}^2 \to \mathbb{C}$ by

$$\xi \equiv \zeta^{(+)} = \tan(\theta/2)\, \mathrm{e}^{\mathrm{i}\phi} , \qquad \eta \equiv \zeta^{(-)} = -\cot(\theta/2)\, \mathrm{e}^{-\mathrm{i}\phi} . \tag{5.13}$$

The coordinate function $\zeta^{(+)} = \xi$ (resp. $\zeta^{(-)} = \eta$) is defined everywhere on S^2 with the exception of the south pole $\theta = \pi$ (resp. north pole $\theta = 0$). Note the relation

$$\xi = -\frac{1}{\eta} . \tag{5.14}$$

One easily checks that our magnetic vector potentials $A^{(\pm)}$ (restricted to S^2) have the expressions

$$A^{(+)} = \frac{\hbar}{e}\,(1 - \cos\theta)\, d\phi = \frac{\hbar}{\mathrm{i}e}\, \frac{\bar\xi d\xi - \xi d\bar\xi}{1 + |\xi|^2} , \quad A^{(-)} = -\frac{\hbar}{e}\,(1 + \cos\theta)\, d\phi = \frac{\hbar}{\mathrm{i}e}\, \frac{\bar\eta\, d\eta - \eta\, d\bar\eta}{1 + |\eta|^2} .$$

The metric tensor pulls back to

$$d\theta^2 + \sin^2\theta\; d\phi^2 = \frac{4\, d\bar\xi\, d\xi}{(1 + |\xi|^2)^2} = \frac{4\, d\bar\eta\, d\eta}{(1 + |\eta|^2)^2} ,$$

which motivates us to introduce the following unit (or normalized) basis vector fields:

$$e_\xi := \frac{1 + |\xi|^2}{2}\, \frac{\partial}{\partial \xi} , \qquad e_{\bar\xi} := \overline{e_\xi} , \qquad e_\eta := \frac{1 + |\eta|^2}{2}\, \frac{\partial}{\partial \eta} , \qquad e_{\bar\eta} := \overline{e_\eta} . \tag{5.15}$$

A short computation using (5.14) then gives

$$e_\xi = \mathrm{e}^{-2\mathrm{i}\phi}\, e_\eta . \tag{5.16}$$

If we now interpret the complex-valued wave functions $\psi^{(\pm)}$ as the components (with respect to the corresponding basis vector fields) of an object v defined invariantly by

$$v = \psi^{(+)} e_\xi + \bar\psi^{(+)} e_{\bar\xi} = \psi^{(-)} e_\eta + \bar\psi^{(-)} e_{\bar\eta} , \tag{5.17}$$

then the relation (5.16) translates to the transition rule (5.12) for $\psi^{(\pm)}$ (in the case of $n = 2$). Thus we learn that the wave functions $\psi^{(\pm)}$ of the Dirac monopole problem for $\mu = 2\,h/e$ are in fact local expressions for (tangent) vector fields v.

Definition. By the tangent bundle TM one means the space of tangent vectors of the manifold M – in our case a two-sphere $M \simeq \mathrm{S}^2$. Locally, i.e., in a small enough neighborhood $U \subset M$ of any base point $x \in M$, the tangent bundle TM is a direct product $TM|_U \simeq U \times \mathbb{R}^2$. The tangent space T_xM is also called the fiber of the tangent bundle at x. It is sometimes denoted by $T_xM \equiv \pi^{-1}(x)$ to indicate that T_xM can be viewed as the inverse image of the projection map $\pi : TM \to M$ defined by $\pi(T_xM) = x$ for all $x \in M$. A section $v \in \Gamma(M, TM)$ of the tangent bundle is a mapping which assigns to every point $x \in M$ a vector $v(x) \in T_xM$. (In other words, $\pi \circ v = \mathrm{Id}$.) A section $v \in \Gamma(M, TM)$ is also called a vector field on M.

Remark. The tangent bundle $T\mathrm{S}^2$ of the two-sphere does not factor globally as a product $\mathrm{S}^2 \times \mathbb{R}^2$. (In contrast, one does have $T\mathrm{S}^3 \simeq \mathrm{S}^3 \times \mathbb{R}^3$.) This fact is expressed by saying that $T\mathrm{S}^2$ is a non-trivial vector bundle. Non-triviality is related to the fact that every smooth vector field $v \in \Gamma(M, TM)$ for $M = \mathrm{S}^2$ has at least two zeroes. (In Dirac's way of thinking, these correspond to the two nodal lines emanating from a nodal singularity of charge $n = 2$.) □

5.3.2 Complex structure

Accepting the change of mathematical description from complex scalar-valued wave functions ψ to real tangent vector fields v, our next question is this: what mathematical object should take the role of $\mathrm{i} = \sqrt{-1} \in \mathbb{C}$ on the left-hand side $\mathrm{i}\hbar\,\partial\psi/\partial t$ of the Schrödinger equation?

Problem. For the complex stereographic coordinate $\zeta^{(+)} = \xi = \tan(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi}$ show that

$$\frac{\partial}{\partial\theta} = \frac{1+|\xi|^2}{2|\xi|}\left(\xi\frac{\partial}{\partial\xi} + \bar{\xi}\frac{\partial}{\partial\bar{\xi}}\right), \qquad \frac{1}{\sin\theta}\frac{\partial}{\partial\phi} = \frac{1+|\xi|^2}{2|\xi|}\left(\mathrm{i}\,\xi\frac{\partial}{\partial\xi} - \mathrm{i}\,\bar{\xi}\frac{\partial}{\partial\bar{\xi}}\right) . \tag{5.18}$$

How do these relations look for $\zeta^{(-)} = \eta = -\cot(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi}$? □

To answer our question, we introduce the basis vector fields

$$e_\theta := \frac{\partial}{\partial\theta}\,, \quad e_\phi := \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\,, \tag{5.19}$$

which are related to those of (5.15) by

$$e_\theta = \frac{\xi}{|\xi|}\,e_\xi + \frac{\bar{\xi}}{|\xi|}\,e_{\bar{\xi}}\,, \quad e_\phi = \mathrm{i}\,\frac{\xi}{|\xi|}\,e_\xi - \mathrm{i}\,\frac{\bar{\xi}}{|\xi|}\,e_{\bar{\xi}}\,. \tag{5.20}$$

(We here focus on the upper hemisphere with basis $e_\xi\,, e_{\bar{\xi}}$ and wave function $\psi^{(+)} \equiv \psi$. In the lower hemisphere the situation is no different.) We now change basis,

$$v = \psi\,e_\xi + \bar{\psi}\,e_{\bar{\xi}} = v_\theta\,e_\theta + v_\phi\,e_\phi\,. \tag{5.21}$$

Using (5.20) we then see that multiplication $(\psi, \bar{\psi}) \mapsto (\mathrm{i}\psi, -\mathrm{i}\bar{\psi})$ by $\mathrm{i} = \sqrt{-1}$ in the complex wave function picture translates to a $\pi/2$ rotation

$$(v_\theta\,, v_\phi) \mapsto (-v_\phi\,, v_\theta) \tag{5.22}$$

in the vector field picture.

Definition. An almost-complex manifold M is a (real) manifold equipped with a smooth tensor field $J \in \Gamma(M, \mathrm{End}\,TM)$, $x \mapsto J_x$, such that its square is $J_x^2 = -\mathrm{Id}_{T_xM}$ for all $x \in M$. A tensor field J with this property is called a complex structure of M.

Example. The two-sphere $M = \mathrm{S}^2$ is an example of an almost-complex manifold. In this case, the complex structure J_x for any point $x \in M$ is rotation by $\pi/2$ in the tangent plane T_xM. □

Returning to our problem of a charged particle moving on S^2 in the field of a charge $n = 2$ magnetic monopole, the Schrödinger equation for the wave function ψ (now: vector field v) takes the form

$$\hbar J_x \dot{v}(x,t) = (Hv)(x,t)\,. \tag{5.23}$$

It remains to transcribe the Hamiltonian H to the vector field picture. For this purpose, we need one further mathematical operation.

5.3.3 Covariant derivative

In this subsection we are going to describe a rule of differentiating vector fields. For brevity, we shall focus on the situation of interest, namely on vector fields on the two-sphere $M = \mathrm{S}^2$ with SO_3-invariant geometry. We refer to differential geometry textbooks for the more general procedure of covariantly differentiating the sections of any vector bundle.

As a warm up we recall the process of differentiating functions on a manifold M. Suppose that we want to calculate the derivative $(df)_x(u)$ of a differentiable function $f : M \to \mathbb{C}$ at some point $x \in M$ in the direction of the tangent vector $u \in T_xM$. To do so, we choose a differentiable curve $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = x$, $\dot\gamma(0) = u$, and compute the said derivative as

$$(df)_x(v) = \frac{d}{dt} f(\gamma(t))\Big|_{t=0} := \lim_{t\to 0} \frac{f(\gamma(t)) - f(\gamma(0))}{t} \,.$$

Now if we naively try to apply the same definition

$$\frac{d}{dt} v(\gamma(t))\Big|_{t=0} \stackrel{??}{=} \lim_{t\to 0} \frac{v(\gamma(t)) - v(\gamma(0))}{t} \,,$$

to a vector field $v \in \Gamma(M, TM)$, we face the problem that the expression $v(\gamma(t)) - v(\gamma(0))$ makes no immediate sense. Indeed, we have $v(\gamma(t)) \in T_{\gamma(t)}M$ and there is no meaning to the difference (nor the sum) of two vectors in different vector spaces $T_{\gamma(t)}M \neq T_{\gamma(0)}M$ (for $t \neq 0$). The point here is that although all tangent spaces $T_xM \simeq \mathbb{R}^2$ are in principle the same, there exists no canonical identification of T_xM with $T_{x'}M$ for $x \neq x'$.

In order to give a meaningful definition of the difference between $v(\gamma(t))$ and $v(\gamma(0))$, we first have to fix some vector space isomorphism $T_{\gamma(t)}M \simeq T_{\gamma(0)}M$. Such an isomorphism is determined by what is called a connection (on a vector bundle). In the case at hand (i.e., magnetic monopole charge $n = 2$, tangent bundle TM), the good choice of isomorphism turns out to be given by parallel transport via the so-called Levi-Civita connection, as follows.

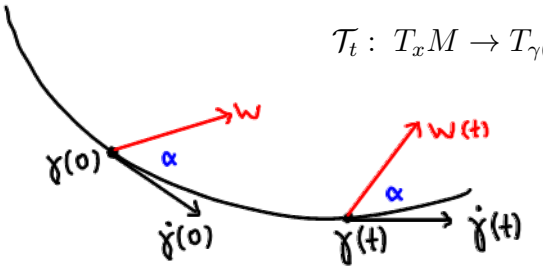
Let $w \in T_xM = T_{\gamma(0)}M$. We wish to introduce a natural scheme of parallel transporting w along the curve γ from $T_{\gamma(0)}M$ to $T_{\gamma(t)}M$. For this we need the following information. Let $M = \mathrm{S}^2 \subset \mathbb{R}^3$ be equipped with the geometry induced by restriction of the standard Euclidean structure of $\mathbb{R}^3$. (Mathematically speaking, this is the geometry given by the Fubini-Study metric of S^2.) Thus for each point $p \in M$ we have a Euclidean scalar product $g_p : T_pM \times T_pM \to \mathbb{R}$ enabling us to measure the lengths of tangent vectors in T_pM and the angles between them. The Fubini-Study metric is SO_3-invariant. In the present context this reflects the fact that the magnetic field of a static monopole is invariant under SO_3 rotations fixing the monopole.

The idea now is to translate $w \in T_xM$ along the curve $\gamma(t)$ in such a way that the length of w and its angle with the tangent vector $\dot\gamma(t)$ of the curve γ stays constant. In formulas, one defines the one-parameter family of parallel translates $w(t) \in T_{\gamma(t)}M$ by the conditions $w(t = 0) = w$ and

$$g_{\gamma(t)}(w(t), w(t)) = g_x(w, w) \,, \quad \frac{g_{\gamma(t)}(w(t), \dot\gamma(t))}{\sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}} = \frac{g_x(w, u)}{\sqrt{g_x(u, u)}} \,, \quad t \in (-\epsilon, \epsilon) \,, \tag{5.24}$$

where $x = \gamma(0)$, $u = \dot\gamma(0)$ as before. This process of parallel translation applies to any $w \in T_xM$. Doing it for all $w \in T_xM$ at once, we have a one-parameter family of isomorphisms

$$\mathcal{T}_t : \; T_xM \to T_{\gamma(t)}M \; , \quad w \mapsto w(t) \; . \tag{5.25}$$



Parallel transport of a vector w along a curve γ. The length of the vector and its angle α with the tangent vector with the curve stays constant.

Definition. The Levi-Civitá connection ∇ is a rule which assigns to any vector field Y and a differentiable vector field v a third vector field $\nabla_Y v$, called the covariant derivative of v in the direction of Y. For any fixed point $x \in M$, the vector $(\nabla_Y v)(x) \in T_xM$ is defined by choosing a differentiable curve $\gamma : \; (-\epsilon, \epsilon) \to M$ with $\gamma(0) = x$ and $\dot\gamma(0) = Y(x)$ and taking the limit

$$(\nabla_Y v)(x) := \lim_{t \to 0} \frac{\mathcal{T}_t^{-1} v(\gamma(t)) - v(x)}{t} \; , \tag{5.26}$$

where $\mathcal{T}_t$ denotes the isomorphism by parallel translation along $\gamma(t)$.

Remark. Unlike the Lie derivative $\mathcal{L}_Y v$, which involves derivatives of Y, the covariant derivative is local, i.e., one has $\nabla_{fY} v = f \nabla_Y v$ for any function f. $\square$

By leaving the vector field argument Y in ∇_Y unspecified, one gets a differential operator ∇ which differentiates vector fields $v \in \Gamma(M, TM)$ to produce sections $\nabla v \in \Gamma(M, T^*M \otimes TM)$ of the tensor product of the tangent bundle TM with the cotangent bundle T^*M. This so-called connection ∇ is compatible with the exterior derivative d in the sense that

$$\nabla(fv) = df \otimes v + f \nabla v \tag{5.27}$$

for any differentiable function f on M.

It is beyond the scope of this lecture course to develop the full calculus of covariant differentiation, so we now take a short cut.

Problem. Make a drawing of the vector fields e_θ, e_ϕ of (5.19). Based on this drawing and the geometric picture of parallel translation, argue that one has the following results for covariant derivatives:

$$\nabla_{e_\theta} e_\theta = \nabla_{e_\theta} e_\phi = 0 \; , \quad \nabla_{e_\phi} e_\theta = \cot\theta \; e_\phi \; , \quad \nabla_{e_\phi} e_\phi = -\cot\theta \; e_\theta \; . \tag{5.28}$$

From these deduce the formulas

$$\nabla e_\theta = \cos\theta \, d\phi \otimes e_\phi \; , \quad \nabla e_\phi = -\cos\theta \, d\phi \otimes e_\theta \; , \tag{5.29}$$

for the Levi-Civitá connection $\nabla : \; \Gamma(M, TM) \to \Gamma(M, T^*M \otimes TM)$. $\square$

By using the product rule (5.27) one gets from (5.29) the following local coordinate expression (w.r.t. the basis e_θ, e_ϕ) for the Levi-Civitá connection:

$$\nabla v = \nabla(v_\theta \, e_\theta + v_\phi \, e_\phi) = (dv_\theta - v_\phi \, \cos\theta \, d\phi) \otimes e_\theta + (dv_\phi + v_\theta \, \cos\theta \, d\phi) \otimes e_\phi \; . \tag{5.30}$$

Now we recall that $e_\phi = Je_\theta$ and $-e_\theta = Je_\phi$. By switching from the vector field v to the complex wave function $\psi := v_\theta + \mathrm{i}v_\phi$ and using the correspondence $J \leftrightarrow \mathrm{i}$, we then obtain the result

$$\nabla\psi = (d + \mathrm{i}\cos\theta\, d\phi)\,\psi\ . \tag{5.31}$$

Notice that the wave function ψ in this equation differs from that of (5.21) by a gauge transformation $\psi = \mathrm{e}^{-\mathrm{i}\phi}\,\psi^{(+)}$, which corresponds to a change of gauge

$$A = A^{(+)} + \frac{\hbar}{e}\,d\phi = -\frac{\hbar}{e}\,\cos\theta\, d\phi \tag{5.32}$$

for the magnetic vector potential A. In the present choice of gauge, A has singularities at $\theta = 0$ and $\theta = \pi$ (reflecting the coordinate singularities due to the choice of basis e_θ, e_ϕ).

We see from (5.31) that the connection ∇ is proportional to the mechanical momentum $p - eA$. More precisely, $\nabla = d - \mathrm{i}eA/\hbar$. Therefore, in view of the expression $H = (p - eA)^2/2m$ for the Hamiltonian, the following statement should now be plausible.

Fact. In the vector field picture the Schrödinger equation reads

$$\hbar\, J\dot{v} = \frac{\hbar^2}{2mR^2}\nabla^\dagger\nabla v\ . \tag{5.33}$$

Here R is the radius of the two-sphere and the symbol $\nabla^\dagger$ denotes the Hermitian adjoint of ∇.

Remark. $\nabla^\dagger$ is a first-order differential operator taking sections of $T^*M \otimes TM$ into sections of TM. (Each of the vector bundles TM and $T^*M \otimes TM$ carries a natural Hermitian structure, by which one defines $\nabla^\dagger$ from ∇ in the usual way; see Section 1.7.1.) The second-order differential operator $\nabla^\dagger\nabla$ on vector fields is called the Bochner Laplacian.

Problem. For $v = e_\theta$ or $v = e_\phi$ verify the expression

$$\nabla^\dagger\big((f\,d\theta + g\sin\theta\, d\phi) \otimes v\big) = -\left(\frac{\partial f}{\partial\theta} + \frac{1}{\sin\theta}\,\frac{\partial g}{\partial\phi} + \cot\theta\,(f + gJ)\right) v\ .$$

By using this expression along with (5.30), show that each of the three-parameter family ($a \in \mathbb{R}^3$) of vector fields

$$v_0^a = (a_x\cos\theta\cos\phi + a_y\cos\theta\sin\phi - a_z\sin\theta)\ e_\theta + (-a_x\sin\phi + a_y\cos\phi)\ e_\phi \tag{5.34}$$

is an eigenvector of $\nabla^\dagger\nabla$ with eigenvalue one.

Remark. Each v_0^a arises by projecting a constant vector field $a_x e_x + a_y e_y + a_z e_z$ from $\mathbb{R}^3$ to S^2. These vector fields span the three-dimensional space of ground states of $\nabla^\dagger\nabla$. Their energy, the so-called lowest Landau level, is

$$E_0 = \frac{\hbar^2}{2mR^2} = \tfrac{1}{2}\hbar\omega\ , \quad \omega = \frac{e|B|}{m}\ , \quad |B| = \frac{2h/e}{4\pi R^2}\ ,$$

where ω is the cyclotron frequency for the magnetic field strength $|B|$.

5.4 Other Cases ($\mu \neq 2h/e$)

Having discussed the case of monopole charge $\mu = 2h/e$ in detail, we now wish to briefly address the more general situation. For that purpose, we introduce another piece of information.

Problem. Let $\Omega = \sin\theta\, d\theta \wedge d\phi$ be the solid-angle (or Fubini-Study) two-form on S^2. Using the formulas for the Levi-Civitá connection ∇ given in the previous section, show that

$$\nabla_X \nabla_Y v - \nabla_Y \nabla_X v - \nabla_{[X,Y]} v = -\Omega(X,Y)\, Jv\ , \tag{5.35}$$

where $[X,Y] = XY - YX$ is the commutator (or Lie bracket) of the two vector fields X, Y viewed as first-order differential operators $X:\ f \mapsto (df)(X)$ on functions.

Remark. Notice that there is no derivative acting on the vector field v on the right-hand side of the formula (5.35). Thus, although

$$F^\nabla(X,Y) := \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]} \tag{5.36}$$

looks very much like a differential operator, it is actually just a tensor field; more precisely, a section of the vector bundle $\mathrm{End}(TM)$. One calls $F^\nabla \equiv R$ the Riemann curvature tensor of the tangent bundle TM with Levi-Civitá connection ∇. More generally, for any vector bundle with a connection ∇, one defines the curvature tensor by the same formula (5.36). □

In all of the discussion up to now, we have been assuming the situation of a magnetic monopole in the static limit, where the magnetic field is centro-symmetric. If our charge $\mu = 2h/e$ monopole is in motion or if divergenceless magnetic fields are superimposed on the magnetic monopole field, the description of the wave function as a vector field continues to apply. The only change is that the Levi-Civitá connection ∇ is replaced by an adapted connection (still denoted by ∇) such that the curvature tensor of the adapted ∇ reflects the variable magnetic field:

$$F^\nabla(X,Y) = -\frac{e}{\hbar} B(X,Y)\, J\ . \tag{5.37}$$

Here B is the two-form of the magnetic field strength with total flux $\int_{S^2} B = 2h/e$.

On the other hand, if the quantized value $\mu = 2h/e$ of the magnetic monopole charge is replaced by a different quantized value $\mu = nh/e$, then the vector bundle itself (not just the connection) undergoes a change. We saw a glimpse of the vector bundles for general $n \in \mathbb{Z}$ in Section 5.2, via the procedure of gluing together two hemispheres by means of the transition function (5.12). We now give a global differential-geometric description of these vector bundles.

To prepare the generalization, we introduce another way of thinking about the tangent bundle of $M = S^2$. For this we fix some point of S^2, say the north pole, o. The rotation group SO_3 acts transitively on S^2, which is to say that if $x \in S^2$ is any other point, we can find some rotation $R \in SO_3$ such that $R \cdot o = x$. Of course R is unique only up right multiplication by elements of the SO_2 subgroup fixing the north pole. Indeed, if $h \cdot o = o$ and $R \cdot o = x$ then $(Rh) \cdot o = x$. Thus $M = S^2$ is identified as a quotient (or coset space) $S^2 \simeq SO_3/SO_2$.

Next, we are going to use this setting to develop a different description of tangent vectors and vector fields. At first, we fix some pair $x \in \mathrm{S}^2$ and $R \in \mathrm{SO}_3$ such that $R \cdot o = x$. The differential $R' : T_oM \to T_xM$ then is a bijective linear mapping from the tangent plane of the north pole to the tangent plane of x. By using its inverse, we can map any tangent vector $v \in T_xM$ to a tangent vector $R'^{-1}v \in T_oM$. Thus we may describe any pair (x, v) with $v \in T_xM$ by some pair $(R\,;u)$ with $u = R'^{-1}v \in T_oM$. Because R is not uniquely defined, this description is not unique. Indeed, if h is an SO_2 rotation fixing o, then the pair $(Rh^{-1}; h'u)$ is as good a description of (x, v) as is $(R\,;u)$. In order to eliminate this redundancy, we consider

$$(Rh^{-1}; h'u) \sim (R\,;u) \tag{5.38}$$

as equivalent $(\sim)$ and denote the equivalence class of all such pairs by $[R\,;u]$. Note that these equivalence classes can be added and multiplied by scalars:

$$[R\,;u_1] + [R\,;u_2] := [R\,;u_1 + u_2]\ , \quad a[R\,;u] := [R\,;au] \qquad (u, u_1, u_2 \in T_oM\ ,\ a \in \mathbb{R})\ .$$

In fact they form a vector space isomorphic to $T_oM \simeq \mathbb{R}^2$.

By performing the construction above for the set of variable points $x \in M$, we arrive at what is called an associated vector bundle over $M = \mathrm{S}^2$. It is denoted by $\mathrm{SO}_3 \times_{\mathrm{SO}_2} \mathbb{R}^2$ (where $\mathbb{R}^2 \equiv T_oM$). The notation reflects the process of construction: we started from the direct product $\mathrm{SO}_3 \times \mathbb{R}^2$ and passed to the quotient by the equivalence relation $\sim$ which is given by the simultaneous SO_2-action on SO_3 and $\mathbb{R}^2$. By construction, our associated vector bundle is the same as the tangent bundle TM : we have the bijective correspondence

$$TM \to \mathrm{SO}_3 \times_{\mathrm{SO}_2} \mathbb{R}^2\ , \quad (x, v) \mapsto [R\,; R'^{-1}v] \qquad (R \cdot o = x)\ . \tag{5.39}$$

What we have described here is a special case of the following general setting.

Definition 1. Let M be a manifold and G be a group. A G-principal bundle P over M is a fiber bundle $\pi : P \to M$ with typical fiber $\pi^{-1}(x) \simeq G$. In any local chart $U \subset M$ the bundle P factors as a direct product $P|_U \simeq U \times G$. The group G acts on P on the right, i.e., $p \cdot (g_1 g_2) = (p \cdot g_1) \cdot g_2$.

Definition 2. Let V be a vector space carrying a G-representation $\rho : G \to \mathrm{GL}(V)$. If P is a G-principal bundle over M, the associated vector bundle $P \times_G V$ over $M \simeq P/G$ is formed by dividing the direct product $P \times V$ by the equivalence relation

$$(p, v) \sim (p \cdot g^{-1}, \rho(g)v) \quad (g \in G)\ . \tag{5.40}$$

Thus the vectors of the vector bundle $P \times_G V$ are equivalence classes $[p\,;v] = [p \cdot g^{-1}; \rho(g)v]$. □

From this general scheme, we get the tangent bundle of $M = \mathrm{S}^2$ by setting $P = \mathrm{SO}_3$, $G = \mathrm{SO}_2$, $V = \mathbb{R}^2$, and taking for ρ the fundamental representation of SO_2 on $\mathbb{R}^2$. Any other vector bundle over S^2 is obtained simply by changing the SO_2-representation.

Fact. Wave functions of the $M = \mathrm{S}^2$ Dirac monopole problem with magnetic charge $\mu = nh/e$ are sections of the vector bundle $\mathrm{SO}_3 \times_{\mathrm{SO}_2} \mathbb{R}^2$ which is associated to SO_3 by the SO_2-representation

$$\rho_n \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos(n\phi/2) & -\sin(n\phi/2) \\ \sin(n\phi/2) & \cos(n\phi/2) \end{pmatrix}\ . \tag{5.41}$$

Remark. This makes immediate sense for $n \in 2\mathbb{Z}$. For the case of odd $n \in 2\mathbb{Z}+1$ it makes sense by the observation that $S^2 = SO_3/SO_2 = Spin_3/Spin_2 = SU_2/U_1$. Taking the square root $(\phi \to \phi/2)$ is well-defined in $Spin_2$, which is a $2:1$ covering of SO_2.

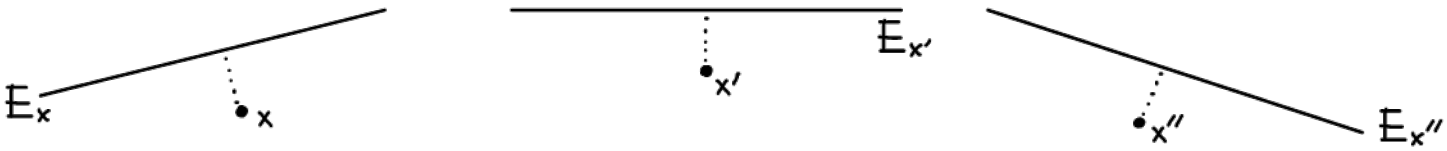
5.5 Lesson learned

Let us finish with a summary of the main message of this chapter. Basic textbooks tell us to think of the wave function of Schrödinger quantum mechanics as a function taking values in $\mathbb{C}$. However, the challenge of Dirac's magnetic monopole problem has revealed that this textbook viewpoint is not always the best one to take. What we have learned is that each point x of position space should be associated with its own individual copy, say $E_x \simeq \mathbb{C}$, of the complex number field. The Schrödinger wave function then is a function $\psi : x \mapsto E_x$.

A priori we have no way of comparing the complex lines E_x for different points x, but the absence or presence of a magnetic field does determine isomorphisms $E_{\gamma(t)} \simeq E_{\gamma(t')}$ along curves γ and thus a covariant derivative ∇. In order to express this differential operator locally as $\nabla = d - ieA/\hbar$ one fixes a real axis $\mathbb{R}$ and hence a frame $E_x \xrightarrow{\sim} \mathbb{R} \oplus i\mathbb{R}$ for each x. The freedom of making an x-dependent change of frame corresponds to the freedom of making a gauge transformation $\psi(x) \mapsto e^{i\theta(x)}\psi(x)$ of the wave function and $A \mapsto A + e\, d\theta/\hbar$ of the vector potential.

The set E of all complex lines E_x over position space $U \subset \mathbb{R}^3$ is what is called a complex line bundle. As long as magnetic monopoles are absent (that's the case in nature as we know it), the complex line bundle E is trivial. It is then possible to fix some universal fiber $E_x \equiv \mathbb{C}$ and decompose $E = U \otimes \mathbb{C}$. In that case the line bundle formalism is just a nice alternative to the usual description of the wave function as a $\mathbb{C}$-valued function.

However, in the presence of magnetic monopoles the line bundle E becomes non-trivial; it is then no longer possible to factor it as $U \otimes \mathbb{C}$. Any attempt to force such a factorization by declaring the wave function to be $\mathbb{C}$-valued leads to singular wave functions and singular magnetic vector potentials. It should be stressed that these singularities are not fundamental but are artifacts due to an improper description. Indeed, the wave function (properly understood as a section of the complex line bundle E) remains well defined and so does ∇. In this sense the complex line bundle picture is more fundamental than the usual one taught in basic quantum mechanics.



Although Dirac did not have the proper mathematical apparatus at his disposal, he recognized the possibility for quantum mechanics to remain consistent in the presence of (quantized) magnetic charges, and he established computational control of that challenging situation. In the present chapter we have met several ways of arriving at the quantization condition named after him. Its final interpretation is that of an integrality condition which lets the complex line bundle and hence the wave function exist as globally defined objects free of singularities.