Lecture Notes on

Undergraduate Physics

Kevin Zhou

`kzhou7@gmail.com`

These notes are a review of the undergraduate physics curriculum, with an emphasis on quantum mechanics. Nothing in these notes is original; they have been compiled from a variety of sources. The primary sources were:

- David Tong's Classical Dynamics lecture notes. A friendly set of notes that covers Lagrangian and Hamiltonian mechanics with neat applications, such as the gauge theory of a falling cat.
- Arnold, *Mathematical Methods of Classical Mechanics*. The classic advanced mechanics book. The first half of the book covers Lagrangian mechanics compactly, with nice and tricky problems, while the second half covers Hamiltonian mechanics geometrically.
- David Tong's Electrodynamics lecture notes. Covers electromagnetism at the standard Griffiths level. Especially nice because it does the most complex calculations in index notation, when vector notation becomes clunky or ambiguous.
- David Tong's Statistical Mechanics lecture notes. Has an especially good discussion of phase transitions, which leads in well to a further course on statistical field theory.
- Blundell and Blundell, *Concepts in Thermal Physics*. An ideal first statistical mechanics book filled with applications, touching on information theory, non-equilibrium thermodynamics, the Earth's atmosphere, and much more.
- David Tong's Applications of Quantum Mechanics lecture notes. A conversational set of notes, with a focus on solid state physics. Also contains a nice section on quantum foundations.
- Robert Littlejon's Physics 221 notes. An exceptionally clear set of quantum mechanics notes, with a focus on atomic physics. Every important point and pitfall is discussed carefully, and complex material is developed elegantly, often in a cleaner and more rigorous way than in any of the standard textbooks. These notes form the bulk of these notes.

The most recent version is here; please report any errors found to kzhou7@gmail.com.

Contents

1 Classical Mechanics

1.1 Lagrangian Formalism

We begin by carefully considering generalized coordinates.

- Consider a system with Cartesian coordinates x_A. Hamilton's principle states that solutions of the equations of motion are critical points of the action $S = \int L\, dt$. Then we have the Euler-Lagrange equation
$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x}$$
where we have dropped indices for simplicity. Here, we have $L = L(x, \dot{x})$ and the partial derivative is defined by holding all other arguments of the function constant.

- It is clear that the Euler-Lagrange equations are preserved by any invertible coordinate change, to generalized coordinates $q_a = q_a(x_A)$, because the action is a property of a path and hence is extremized regardless of the coordinates used.

- It is a little less obvious this holds for time-dependent transformations $q_a = q_a(x_A, t)$, so we will prove this explicitly. Again dropping indices,
$$\frac{\partial L}{\partial q} = \frac{\partial L}{\partial x}\frac{\partial x}{\partial q} + \frac{\partial L}{\partial \dot{x}}\frac{\partial \dot{x}}{\partial q}$$
where we have $\dot{q} = \dot{q}(x, \dot{x}, t)$ and hence by invertibility $x = x(q, t)$ and $\dot{x} = \dot{x}(q, \dot{q}, t)$, and
$$\dot{x} = \frac{\partial x}{\partial q}\dot{q} + \frac{\partial x}{\partial t}.$$
This yields the 'cancellation of dots' identity
$$\frac{\partial \dot{x}}{\partial \dot{q}} = \frac{\partial x}{\partial q}.$$

- As for the other side of the Euler-Lagrange equation, note that
$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\frac{\partial \dot{x}}{\partial \dot{q}}\right) = \frac{\partial L}{\partial x}\frac{\partial x}{\partial q} + \frac{\partial L}{\partial \dot{x}}\frac{d}{dt}\frac{\partial x}{\partial q}$$
where, in the first step, we used $\partial x/\partial \dot{q} = 0$ since x is not a function of $\dot{q}$, and in the second step, we used cancellation of dots and the Euler-Lagrange equation.

- To finish the derivation, we note that
$$\frac{d}{dt}\frac{\partial q}{\partial x} = \frac{\partial \dot{q}}{\partial x}$$
which may be shown by direct expansion.

- It is a bit confusing why these partial derivatives are allowed. The point is that we are working on the tangent bundle of some manifold, where the position and velocity are independent. They are only necessarily related when we consider a specific path $x(t)$, and note that all total time derivatives implicitly refer to such a path.

Next we show that if constraints exist, we can work in a reduced set of generalized coordinates.

- A holonomic constraint is a relationship of the form
$$f_\alpha(x_A, t) = 0$$
which must hold on all physical paths. Holonomic constraints are useful because each one can be used to eliminate a generalized coordinate; note that inequalities are not holonomic.

- Velocity-dependent constraints are holonomic if they can be 'integrated'. For example, consider a ball rolling without slipping. In one dimension, this is holonomic, since $v = R\dot{\theta}$. In two dimensions, it's possible to roll the ball in a loop and have it come back in a different orientation. Formally, a velocity constraint is holonomic if there is no nontrivial holonomy.

- To find the equations of motion, we use the Lagrangian
$$L' = L(x^A, \dot{x}^A) + \lambda_\alpha f_\alpha(x^A, t).$$
We think of the λ_α as additional, independent coordinates; then the Euler-Lagrange equation $\partial L'/\partial \lambda_\alpha = 0$ reproduces the constraint. The Euler-Lagrange equations for the x^A now have constraint forces equal to the Lagrange multipliers.

- Now we switch coordinates from x^A, λ_α to $q^a, f_\alpha, \lambda_\alpha$, continuing to use the Lagrangian L'. The Euler-Lagrange equations are simply
$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q^a}} - \frac{\partial L}{\partial q_i} = \frac{\partial}{\partial q^a}\lambda_\alpha f_\alpha = 0, \quad \lambda_\alpha = f_\alpha = 0.$$
Thus, in these generalized coordinates, the constraint forces have disappeared. We may restrict to the coordinates q^a and use the original Lagrangian L. Note that in such an approach, we cannot solve for the values of the constraint forces.

- In problems with symmetry, there will be conserved quantities, which may be formally written as constraints on the positions and velocities. However, it's important to remember that they are not genuine constraints, because they only hold on-shell. Treating a conserved quantity as a constraint and using the procedure above will give incorrect results.

- We may think of the coordinates q^a as contravariant under changes of coordinates. Then the conjugate momenta are covariant, so the quantity $p_i \dot{q}^i$ is invariant. Similarly, the differential form $p_i dq^i$ is invariant.

- We say a Lagrangian is regular if
$$\det \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \neq 0.$$
In this case, the equation of motion can be solved for $\ddot{q}$. We'll only deal with regular Lagrangians, but irregular ones can appear in relativistic particle mechanics.

Example. Purely kinetic Lagrangians. In the case
$$L = \frac{1}{2} g_{ab}(q_c) \dot{q}^a \dot{q}^b$$

the equation of motion is the geodesic equation

$$\ddot{q}^a + \Gamma^a_{bc}\dot{q}^b\dot{q}^c = 0, \quad \Gamma^a_{bc} = \frac{1}{2}g^{ad}\left(\partial_c g_{bd} + \partial_b g_{cd} - \partial_d g_{bc}\right)$$

where we have assumed the metric is invertible, and symmetrized the geodesic coefficients. This works just like the derivation in general relativity, except that in that context, the metric and velocities include a time component, so the solutions have stationary proper time.

Example. A particle in an electromagnetic field. The Lagrangian is

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - e(\phi - \dot{\mathbf{r}} \cdot \mathbf{A}).$$

With a little index manipulation, this reproduces the Lorentz force law, with

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad \mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}.$$

The momentum conjugate to $\mathbf{r}$ is

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{r}}} = m\dot{\mathbf{r}} + e\mathbf{A}$$

and is called the canonical momentum, in contrast to the kinetic momentum $m\dot{\mathbf{r}}$. The canonical momentum is what becomes the gradient operator in quantum mechanics, but it is not gauge invariant; instead the kinetic momentum is. The switch to covariant derivatives in gauge theory is analogous to switching to kinetic momentum.

Example. A single relativistic particle. The Lagrangian should be a Lorentz scalar, and the only one available is the proper time τ. Then we have

$$L = -mc^2\sqrt{1 - \dot{r}^2/c^2} - V(\mathbf{r})$$

for which the momentum is $\gamma m\mathbf{v}$, as expected, and the action is proportional to proper time. However, this is not entirely satisfying because the potential action term $\int V(\mathbf{r}(t))\, dt$ is not manifestly a Lorentz scalar; the potential must have some nontrivial transformation, but it is left unspecified.

For example, in the special case of electromagnetism, the potential is

$$L = -mc^2\sqrt{1 - \dot{r}^2/c^2} - e(\phi - \dot{\mathbf{r}} \cdot \mathbf{A})$$

where A^μ transforms as a four-vector, and one can check directly that the action is a Lorentz scalar. This is also a way to put a particle in a generic potential V relativistically – we augment the V into a four-potential analogous to A^μ. The procedure fails for gravitational forces, for which general relativity is needed.

Alternatively, we can formulate the problem in a Lorentz-invariant way by specifying the configuration as $x^\mu(\lambda)$, where λ is an arbitrary parameter on the worldline. Then we have

$$L = -mc\sqrt{\frac{dx^\mu}{d\lambda}\frac{dx_\mu}{d\lambda}} - V(x^\mu(\lambda)) = -mc^2\frac{d\tau}{d\lambda} - V, \quad S = -\int mc^2\, d\tau - \int V d\lambda.$$

For electromagnetism, the potential is $eA_\mu dx^\mu/d\lambda$, giving the manifestly Lorentz invariant action

$$S = -\int mc^2\, d\tau - e\int A_\mu dx^\mu.$$

Then the action is manifestly a Lorentz scalar. The equation of motion is

$$m\frac{d^2x^\mu}{d\tau^2} = eF^{\mu\nu}\frac{dx_\nu}{d\tau}, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

where $F_{\mu\nu}$ is the field strength tensor.

1.2 Rigid Body Motion

We begin with the kinematics of rigid bodies.

- A rigid body is a collection of masses constrained so that $\|\mathbf{r}_i - \mathbf{r}_j\|$ is constant for all i and j. Then a rigid body has six degrees of freedom, from translations and rotations.
- If we fix a point to be the origin, we have only the rotational degrees of freedom. Define a fixed coordinate system $\{\widetilde{\mathbf{e}}_a\}$ as well as a moving body frame $\{\mathbf{e}_a(t)\}$ which moves with the body. Both sets of axes are orthogonal and thus related by an orthogonal matrix,
$$\mathbf{e}_a(t) = R_{ab}(t)\widetilde{e}_b(t), \quad R_{ab} = \mathbf{e}_a \cdot \widetilde{\mathbf{e}}_b.$$
Since the body frame is specified by $R(t)$, the configuration space C of orientations is $SO(3)$.
- Every point $\mathbf{r}$ in the body can be expanded in the space frame or the body frame as
$$\mathbf{r}(t) = \widetilde{r}_a(t)\widetilde{\mathbf{e}}_a = r_a \mathbf{e}_a(t).$$
Note that the body frame changes over time as
$$\frac{d\mathbf{e}_a}{dt} = \frac{dR_{ab}}{dt}\widetilde{\mathbf{e}}_b = \left(\frac{dR}{dt}R^{-1}\right)_{ab}\mathbf{e}_b$$
This prompts us to define the matrix $\omega = \dot{R}R^{-1}$, so that $\dot{\mathbf{e}}_a = \omega_{ab}\mathbf{e}_b$.
- The matrix ω is antisymmetric, so we take the Hodge dual to get the angular velocity vector
$$\omega_a = \frac{1}{2}\epsilon_{abc}\omega_{bc}, \quad \boldsymbol{\omega} = \omega_a \mathbf{e}_a.$$
Inverting this relation, we have $\omega_a \epsilon_{abc} = \omega_{bc}$. Substituting into the above,
$$\frac{d\mathbf{e}_a}{dt} = -\epsilon_{abc}\omega_b \mathbf{e}_c = \boldsymbol{\omega} \times \mathbf{e}_a$$
where we used $(\mathbf{e}_a)_d = \delta_{ad}$.
- The above is just a special case of the formula
$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$$
which can be derived from simple vector geometry. Using that picture, the physical interpretation of $\boldsymbol{\omega}$ is $\hat{\mathbf{n}}\, d\phi/dt$, where $\hat{\mathbf{n}}$ is the instantaneous axis of rotation and $d\phi/dt$ is the rate of rotation. Generally, both $\hat{\mathbf{n}}$ and $d\phi/dt$ change with time.

Example. To get an explicit formula for $R(t)$, note that $\dot{R} = \omega R$. The naive solution is the exponential, but since ω doesn't commute with itself at different times, we must use the path ordered exponential,
$$R(t) = P\,\exp\left(\int_0^t \omega(t')dt'\right).$$
For example, the second-order term here is
$$\int_0^{t''}\left(\int_{t'}^{t}\omega(t'')\,dt''\right)\omega(t')\,dt'$$
where the ω's are ordered from later to earlier. Then when we differentiate with respect to t, it only affects the dt'' integral, which pops out a factor of ω on the left as desired. This exponential operation relates rotations R in $SO(3)$ with infinitesimal rotations ω in $\mathfrak{so}(3)$.

We now turn from kinematics to dynamics.

- Using $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, the kinetic energy is

$$T = \frac{1}{2}\sum m_i \mathbf{v}^2 = \frac{1}{2}\sum m_i \|\boldsymbol{\omega} \times \mathbf{r}_i\|^2 = \frac{1}{2}\sum m_i \left(\omega^2 r_i^2 - (r_i \cdot \boldsymbol{\omega})^2\right).$$

  This implies that

$$T = \frac{1}{2}\omega_a I_{ab} \omega_b$$

  where I_{ab} is the symmetric tensor

$$I_{ab} = \sum_i m_i \left(r_i^2 \delta_{ab} - (\mathbf{r}_i)_a (\mathbf{r}_i)_b\right)$$

  called the inertia tensor. Note that since the components of ω are in the body frame, so are the components of I and $\mathbf{r}_i$ that appear above; hence the I_{ab} are constant.

- Explicitly, for a continuous rigid body with mass density $\rho(\mathbf{r})$, we have

$$I = \int d^3\mathbf{r}\, \rho(\mathbf{r}) \begin{pmatrix} y^2 + z^2 & -xy & -xz \\ -xy & x^2 + z^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{pmatrix}.$$

- Since I is symmetric, we can rotate the body frame to diagonalize it. The eigenvectors are called the principal axes and the eigenvalues I_a are the principal moments of inertia. Since T is nonnegative, I is positive semidefinite, so $I_a \geq 0$.

- Parallel axis theorem: if I_0 is the inertia tensor about the center of mass, the inertia tensor about the point $\mathbf{c}$ is

$$(I_{\mathbf{c}})_{ab} = (I_0)_{ab} + M(c^2 \delta_{ab} - \mathbf{c}_a \mathbf{c}_b).$$

  The proof is similar to the usual parallel axis theorem, with contributions proportional to $\sum m_i \mathbf{r}_i$ vanishing. The extra contribution the inertia tensor we would get if the object's mass was entirely at the center of mass.

- Similarly, the translational and rotational motion of a free spinning body 'factorize'. If the center of mass position is $\mathbf{R}(t)$, then

$$T = \frac{1}{2} M \dot{\mathbf{R}}^2 + \frac{1}{2}\omega_a I_{ab} \omega_b.$$

  This means we can indeed ignore the center of mass motion for dynamics.

- The angular momentum is

$$\mathbf{L} = \sum m_i \mathbf{r}_i \times \mathbf{v}_i = \sum m_i \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i) = \sum m_i (r_i^2 \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \mathbf{r}_i)\mathbf{r}_i).$$

  We thus recognize

$$\mathbf{L} = I\boldsymbol{\omega}, \quad T = \frac{1}{2}\boldsymbol{\omega} \cdot \mathbf{L}.$$

  For general I, the angular momentum and angular velocity are not parallel.

- To find the equation of motion, we use $d\mathbf{L}/dt$ in the center of mass frame, for

$$0 = \frac{dL_a}{dt}\mathbf{e}_a + L_a\frac{d\mathbf{e}_a}{dt} = \frac{dL_a}{dt}\mathbf{e}_a + L_a\boldsymbol{\omega}\times\mathbf{e}_a.$$

  Dotting both sides by $\mathbf{e}_b$ gives $0 = \dot{L}_a + \epsilon_{aij}\omega_I L_j$. In the case of principle axes ($L_1 = I_1\omega_1$),

$$I_1\dot{\omega}_1 + \omega_2\omega_3(I_3 - I_2) = 0$$

  along with cyclic permutations thereof. These are Euler's equations. In the case of a torque, the components of the torque (in the principle axis frame) appear on the right.

We now analyze the motion of free tops. We consider the time evolution of the vectors $\mathbf{L}$, $\boldsymbol{\omega}$, and $\mathbf{e}_3$. In the body frame, $\mathbf{e}_3$ is constant and points upward; in the space frame, $\mathbf{L}$ is constant, and for convenience we take it to point upward. In general, we know that $\mathbf{L}$ and $2T = \boldsymbol{\omega}\cdot\mathbf{L}$ are constant.

Example. A spherical top. In this trivial case, $\dot{\omega}_a = 0$, so $\boldsymbol{\omega}$ doesn't move in the body frame, nor does $\mathbf{L}$. In the space frame, $\mathbf{L}$ and $\boldsymbol{\omega}$ are again constant, and the axis $\mathbf{e}_3$ rotates about them. As a simple example, the motion of $\mathbf{e}_3$ looks like the motion of a point on the globe as it rotates about its axis.

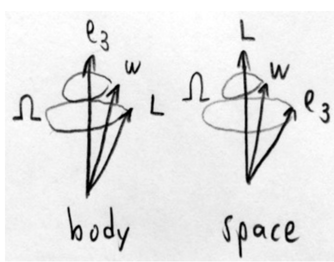
Example. The symmetric top. Suppose $I_1 = I_2 \neq I_3$, e.g. for a top with radial symmetry. Then

$$I_1\dot{\omega}_1 = \omega_2\omega_3(I_1 - I_3), \quad I_2\dot{\omega}_2 = -\omega_1\omega_3(I_1 - I_3), \quad I_3\dot{\omega}_3 = 0.$$

Then ω_3 is constant, while the other two components rotate with frequency

$$\Omega = \omega_3(I_1 - I_3)/I_1.$$

This implies that $|\boldsymbol{\omega}|$ is constant. Moreover, we see that $\mathbf{L}$, $\boldsymbol{\omega}$, and $\mathbf{e}_3$ all lie in the same plane.



In the body frame, both $\boldsymbol{\omega}$ and $\mathbf{L}$ precess about $\mathbf{e}_3$. Similarly, in the space frame, both $\boldsymbol{\omega}$ and $\mathbf{e}_3$ precess about $\mathbf{L}$. To visualize this motion, consider the point $\mathbf{e}_2$ and the case where $\Omega, \omega_1, \omega_2 \ll \omega_3$. Without the precession, $\mathbf{e}_2$ simply rotates about $\mathbf{L}$, tracing out a circle. With the precession, the orbit of $\mathbf{e}_2$ also 'wobbles' slightly with frequency Ω.

Example. The Earth is an oblate ellipsoid with $(I_1 - I_3)/I_1 \approx -1/300$, with $\omega_3 = (1\,\text{day})^{-1}$. Since the oblateness itself is caused by the Earth's rotation, the angular velocity is very nearly aligned with $\mathbf{e}_3$, though not exactly. We thus expect the Earth to wobble with a period of about 300 days; this phenomenon is called the Chandler wobble.

Example. The asymmetric top. If all of the I_i are unequal, the Euler equations are much more difficult to solve. Instead, we can consider the effect of small perturbations. Suppose that

$$\omega_1 = \Omega + \eta_1, \quad \omega_2 = \eta_2, \quad \omega_3 = \eta_3.$$

To first order in η, the Euler equations become

$$I_1 \dot{\eta}_1 = 0, \quad I_2 \dot{\eta}_2 = \Omega \eta_3 (I_3 - I_1), \quad I_3 \dot{\eta}_3 = \Omega \eta_2 (I_1 - I_2).$$

Combining the last two equations, we have

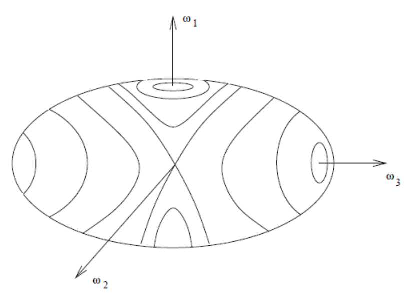
$$I_2 \ddot{\eta}_2 = \frac{\Omega^2}{I_3} (I_3 - I_1)(I_1 - I_2) \eta_2.$$

Therefore, we see that rotation about $\mathbf{e}_1$ is unstable iff I_1 is in between I_2 and I_3. An asymmetric top rotates stably only about the principal axes with largest and smallest moment of inertia.

Note. We can visualize the Euler equations with the Poinsot construction. In the body frame, we have conserved quantities

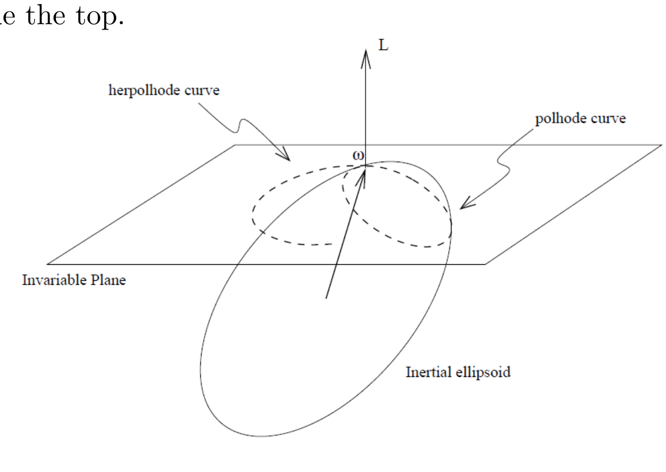
$$2T = I_1 \omega_1^2 + I_2 \omega_2^2 + I_3 \omega_3^2, \quad L^2 = I_1^2 \omega_1^2 + I_2^2 \omega_2^2 + I_3^2 \omega_3^2$$

defining two ellipsoids. The first ellipsoid is called the inertia ellipsoid, and its intersection with the L^2 ellipsoid gives the polhode curve, which contains possible values of $\boldsymbol{\omega}$.



An inertia ellipsoid with some polhode curves is shown above. Since polhode curves are closed, the motion is periodic in the body frame. This figure also gives an intuitive proof of the intermediate axis theorem: polhodes are small loops near minima and maxima of L^2, but not near the intermediate axis, which corresponds to a saddle point.

Note. The space frame is more complicated, as our nice results for the symmetric top no longer apply. The only constraint we have is that $\mathbf{L} \cdot \boldsymbol{\omega}$ is constant, which means that $\boldsymbol{\omega}$ must lie on a plane perpendicular to $\mathbf{L}$ called the invariable plane. We imagine the inertial ellipsoid as an abstract object embedded inside the top.



Since $\mathbf{L} = \partial T/\partial \boldsymbol{\omega}$, $\mathbf{L}$ is perpendicular to the inertial ellipsoid, which implies that the invariable plane is tangent to the inertial ellipsoid. We can thus imagine this ellipsoid as rolling without slipping on the invariable plane, as shown above. The angular velocity traces a path on this plane called the herpolhode curve, which is not necessarily closed.

1.3 Hamiltonian Formalism

- Hamiltonian mechanics takes place in phase space, and we switch from $(q, \dot{q})$ to (q, p) by Legendre transformation. Specifically, letting F be the generalized force, we have

$$dL = F dq + p d\dot{q}$$

  and so taking $H = p\dot{q} - L$ switches this to

$$dH = \dot{q} dp - F dq.$$

  In the language of thermodynamics, we have $L = L(q, \dot{q})$ and $H = H(q, p)$ naturally. In order to write H in terms of these variables, we must be able to eliminate $\dot{q}$ in favor of p, which is generally only possible if L is convex in $\dot{q}$.

- Plugging in $F = dp/dt$, we arrive at Hamilton's equations,

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i}.$$

  The explicit time dependence just comes along for the ride, giving

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}$$

  where the first equality follows from Hamilton's equations and the chain rule.

- We may also derive Hamilton's equations by minimizing the action

$$S = \int (p_i \dot{q}_i - H)\, dt.$$

  In this context, the variations in p_i and q_i are independent. However, as before, $\delta\dot{q} = (\dot{\delta q})$. Plugging in the variation, we see that δq must vanish at the endpoints to integrate by parts, while δp doesn't have to, so our formulation isn't totally symmetric.

- When L is time-independent with $L = T - V$, and L is a quadratic homogeneous function in $\dot{q}$, we have $p\dot{q} = 2T$, so $H = T + V$. Then the value of the Hamiltonian is the total energy.

Example. The Hamiltonian for a particle in an electromagnetic field is

$$H = \frac{(\mathbf{p} - e\mathbf{A})^2}{2m} - e\phi$$

where $\mathbf{p} = m\dot{\mathbf{r}} + e\mathbf{A}$ is the canonical momentum. We see that the Hamiltonian is numerically unchanged by the addition of a magnetic field (since magnetic fields do no work), but the time evolution is affected, since the canonical momentum is different.

Carrying out the same procedure for our non-covariant relativistic particle Lagrangian gives

$$H = \sqrt{m^2c^2 + c^2(\mathbf{p} - e\mathbf{A})^2} + e\phi.$$

However, doing it for the covariant Lagrangian yields $H = 0$. This is because Hamiltonians are inherently not Lorentz covariant, as they generate time translation in a particular frame; mathematically, it occurs because the covariant Lagrangian is not regular.

Both of the examples above are special cases of the minimal coupling prescription: to incorporate an interaction with the electromagnetic field, we must replace

$$p^\mu \to p^\mu - eA^\mu.$$

In particular, in noncovariant form, this means we should transform

$$E \to E - e\phi, \quad \mathbf{p} \to \mathbf{p} - e\mathbf{A}.$$

In general, minimal coupling is often a good guess, because it is the simplest Lorentz invariant method. In field theory, it translates to adding a term $\int d\mathbf{x}\, J^\mu A_\mu$ where J^μ is the matter 4-current. However, we would need a non-minimal coupling to account for, e.g. the spin of the particle.

Hamiltonian mechanics leads to some nice theoretical results.

- Liouville's theorem: volumes of regions of phase space are constant. To see this, consider the infinitesimal time evolution

  $$q_i \to q_i + \frac{\partial H}{\partial p_i} dt, \quad p_i \to p_i - \frac{\partial H}{\partial q_i} dt.$$

  Then the Jacobian matrix is

  $$J = \begin{pmatrix} I + (\partial^2 H/\partial p_i \partial q_j) dt & (\partial^2 H/\partial p_i \partial p_j) dt \\ -(\partial^2 H/\partial q_i \partial q_j) dt & I - (\partial^2 H/\partial q_i \partial p_j) dt \end{pmatrix}.$$

  Using the identity $\det(I + \epsilon M) = 1 + \epsilon \operatorname{tr} M$, we have $\det J = 1$ by equality of mixed partials.

- In statistical mechanics, we might have a phase space distribution $\rho(q, p, t)$. Then the total/convective derivative $d\rho/dt$ is the rate of density change while comoving with the phase space flow, and $d\rho/dt = 0$ by Liouville's theorem. This implies that

  $$\frac{\partial \rho}{\partial t} = \frac{\partial \rho}{\partial p_i}\frac{\partial H}{\partial q_i} - \frac{\partial \rho}{\partial q_i}\frac{\partial H}{\partial p_i}$$

  which is called Liouville's equation.

- Liouville's theorem holds even if energy isn't conserved, as in the case of an external field. It fails in the presence of dissipation, where there isn't a Hamiltonian description at all.

- Poincare recurrence: consider a system with bounded phase space. Given an initial point p, every neighborhood D_0 of p contains a point that will return to D_0 in finite time.

  Proof: consider the neighborhoods D_k formed by evolving D_0 with time kT for an arbitrary time T. Since the phase space volume is finite, and the D_k all have the same volume, we must have some overlap between two of them, say D_k and $D_{k'}$. Since Hamiltonian evolution is reversible, we may evolve backwards, yielding an overlap between D_0 and $D_{k-k'}$.

- As a corollary, it can be shown that Hamiltonian evolution is generically either periodic or fills some submanifold of phase space densely.

1.4 Poisson Brackets

The formalism of Poisson brackets is closely analogous to quantum mechanics.

- The Poisson bracket of two functions f and g on phase space is
$$\{f, g\} = \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i}.$$
- Geometrically, it is possible to associate g with a vector field X_g, and $\{f, g\}$ is the rate of change of f along the flow of X_g. For example, for any function $f(p, q, t)$,
$$\frac{df}{dt} = \{f, H\} + \frac{\partial f}{\partial t}$$
where flow along H generates time translation. In particular, if $I(p, q)$ satisfies $\{I, H\} = 0$, then I is conserved. We'll return to exactly what this 'flow' is later.
- The Poisson bracket is antisymmetric, linear, and obeys the product rule
$$\{fg, h\} = f\{g, h\} + \{f, h\}g$$
as expected from the geometric intuition above. It also satisfies the Jacobi identity, so the space of functions with the Poisson bracket is a Lie algebra.
- A related property is the 'chain rule'. If $f = f(h_i)$, then
$$\{f, g\} = \sum \frac{\partial f}{\partial h_i}\{h_i, g\}.$$
This can be seen by applying the regular chain rule and the flow idea above.
- By the Jacobi identity, Lie brackets of conserved quantities are also conserved, so conserved quantities form a Lie subalgebra.

Example. In statistical mechanics, ensembles are time-independent distributions on phase space. Applying Liouville's equation, we require $\{\rho, H\} = 0$. If the conserved quantities of a system are f_i, then ρ may be any function of the f_i, i.e. any member of the subalgebra of conserved quantities.

We typically take the case where only the energy is conserved for simplicity. In this case, the microcanonical ensemble is $\rho \propto \delta(H - E)$ and the canonical ensemble is $\rho \propto e^{-\beta H}$.

Example. The Poisson brackets of position and momentum are always zero, except for
$$\{q_i, p_j\} = \delta_{ij}.$$
The flow generated by momentum is translation along its direction, and vice versa for position.

Example. Angular momentum. Defining $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, we have
$$\{L_i, L_j\} = \epsilon_{ijk}L_k, \quad \{L^2, L_i\} = 0$$
as in quantum mechanics. The first equation may be understood intuitively from the commutation of infinitesimal rotations.

We now consider the changes of coordinates that preserve the form of Hamilton's equations; these are called canonical transformations. Generally, they are more flexible than coordinate transformations in the Lagrangian formalism, since we can mix position and momentum.

- Define $\mathbf{x} = (q_1, \ldots, q_n, p_1, \ldots, p_n)^T$ and define the matrix J as
$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$
Then Hamilton's equations become
$$\dot{\mathbf{x}} = J \frac{\partial H}{\partial \mathbf{x}}.$$
Also note that the canonical Poisson brackets are $\{x_i, x_j\} = J_{ij}$.

- Now consider a transformation $q_i \to Q_i(q, p)$ and $p_i \to P_i(q, p)$, written as $x_i \to y_i(x)$. Then
$$\dot{\mathbf{y}} = (\mathcal{J} J \mathcal{J}^T) \frac{\partial H}{\partial \mathbf{y}}$$
where $\mathcal{J}$ is the Jacobian matrix $\mathcal{J}_{ij} = \partial y_i / \partial x_j$. We say the Jacobian is symplectic if $\mathcal{J} J \mathcal{J}^T$, and in this case, the transformation is canonical.

- The Poisson bracket is invariant under canonical transformations. To see this, note that
$$\{f, g\}_x = (\partial_x f)^T J (\partial_x g)$$
where $(\partial_x f)_i = \partial f / \partial x_i$. By the chain rule, $\partial_x = \mathcal{J}^T \partial_y$, giving the result. Then if we only consider canonical transformations, we don't have to specify which coordinates the Poisson bracket is taken in.

- Conversely, if a transformation preserves the canonical Poisson brackets $\{y_i, y_j\}_x = J_{ij}$, it is canonical. To see this, apply the chain rule for
$$J_{ij} = \{y_i, y_j\}_x = \left(\mathcal{J} J \mathcal{J}^T\right)_{ij}$$
which is exactly the condition for a canonical transformation.

Example. Consider a 'point transformation' $q_i \to Q_i(q)$. We have shown that these leave Lagrange's equations invariant, but in the Hamiltonian formalism, we also must transform the momentum accordingly. Dropping indices and defining $\Theta = \partial Q / \partial q$,
$$\mathcal{J} = \begin{pmatrix} \Theta & 0 \\ \partial P / \partial q & \partial P / \partial p \end{pmatrix}, \quad \mathcal{J} J \mathcal{J}^T = \begin{pmatrix} 0 & \Theta (\partial P / \partial p)^T \\ -\Theta^T \partial P / \partial p & 0 \end{pmatrix}$$
which implies that $P_i = (\Theta^{-1}_{ji}) p_j$, in agreement with the formula $P_i = \partial L / \partial \dot{Q}_i$. Since Θ depends on q, the momentum P is a function of both p and q.

We now consider infinitesimal canonical transformations.

- Consider a canonical transformation $Q_i = q_i + \alpha F_i(q,p)$ and $P_i = p_i + \alpha E_i(q,p)$ where α is small. Expanding the symplectic condition to first order yields
$$\frac{\partial F_i}{\partial q_j} = -\frac{\partial E_j}{\partial p_i}, \quad \frac{\partial F_i}{\partial p_j} = \frac{\partial F_j}{\partial p_i}, \quad \frac{\partial E_i}{\partial q_j} = \frac{\partial E_j}{\partial q_i}.$$
There are all automatically satisfied if
$$F_i = \frac{\partial G}{\partial p_i}, \quad E_i = -\frac{\partial G}{\partial q_i}$$
for some $G(q,p)$, and we say G generates the transformation.

- More generally, consider a one-parameter family of canonical transformations parametrized by α. Then by the above,
$$\frac{dq_i}{d\alpha} = \frac{\partial G}{\partial p_i}, \quad \frac{dp_i}{d\alpha} = -\frac{\partial G}{\partial q_i}, \quad \frac{df}{d\alpha} = \{f, G\}.$$
Interpreting the transformation actively, this looks just like evolution under a Hamiltonian, with G in place of H and α in place of t. The infinitesimal canonical transformation generated by $G(p,q,\alpha)$ is Lie flow under its vector field.

- We now revisit Noether's theorem. We say G is a symmetry of H if the flow generated by G does not change H, i.e. $\{H, G\} = 0$. But this is just the condition for G to be conserved: since the Poisson bracket is antisymmetric, flow under H doesn't change G either!

Example. Using $G = H$ simply generates time translation, $y(t) = x(t - t_0)$. Less trivially, $G = p_k$ generates $q_i \to q_i + \alpha\delta_{ik}$, so momentum generates translations.

1.5 Action–Angle Variables

The additional flexibility of canonical transformations allows us to use even more convenient variables than the generalized coordinates of Lagrangian mechanics. Often, the 'action-angle' variables are a natural choice, and drastically simplify the problem.

Example. The simple harmonic oscillator. The Hamiltonian is
$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2$$
and we switch from (q,p) to (θ, I), where
$$q = \sqrt{\frac{2I}{m\omega}}\sin\theta, \quad p = \sqrt{2Im\omega}\cos\theta.$$
To confirm this is a canonical transformation, we check that Poisson brackets are preserved; the simplest way to do this is to work backwards, noting that
$$\{q,p\}_{(\theta,I)} = 2\{\sqrt{I}\sin\theta, \sqrt{I}\cos\theta\}_{(\theta,I)} = 1$$
as desired. In these new coordinates, the Hamiltonian is simply
$$H = \omega I, \quad \dot\theta = \omega, \ \dot I = 0.$$
We have 'straightened out' the phase space flow into straight lines on a cylinder. This is the simplest example of action angle variables.

- In general, for n degrees of freedom, we would like to find variables (θ_i, I_i) so that the Hamiltonian is only a function of the I_i. Then the I_i are conserved, and $\dot{\theta}_i = \omega_i$, where the ω_i depend on I but are time independent. When the system is bounded, we scale θ_i to lie in $[0, 2\pi)$. The resulting variables are called action-angle variables and the system is integrable.
- Liouville's theorem: if there are n mutually Poisson commuting constants of motion I_i, then the system is integrable.
- Integrable systems are rare and special; chaotic systems are not integrable. The question of whether a system is integrable has to do with global structure, since one can always straighten out the phase space flow lines locally.
- The motion of an integrable system lies on a surface of constant I_i. These surfaces are topologically tori $\mathbb{T}^n$, called invariant tori.

Example. Action-angle variables for a general one-dimensional system. Let

$$H = \frac{p^2}{2m} + V(x).$$

The value of H is the total energy E, so the action variable I must satisfy

$$\dot{\theta} = \omega = dE/dI$$

where the period of the motion is $2\pi/\omega$. Now, by conservation of energy

$$dt = \sqrt{\frac{m}{2}} \frac{dq}{\sqrt{E - V(q)}}.$$

Integrating over a single orbit, we have

$$\frac{2\pi}{\omega} = \sqrt{\frac{m}{2}} \oint \frac{dq}{\sqrt{E - V(q)}} = \oint \sqrt{2m} \frac{d}{dE} \sqrt{E - V(q)}\, dq = \frac{d}{dE} \oint \sqrt{2m(E - V(q))}\, dq = \frac{d}{dE} \oint p\, dq.$$

Note that by pulling the d/dE out of the integral, we neglected the change in phase space area due to the change in the endpoints of the path, because this contribution is second order in dE.

Therefore, we have the nice results

$$I = \frac{1}{2\pi} \oint p\, dq, \quad T = \frac{d}{dE} \oint p\, dq.$$

We can thus calculate T without finding a closed-form expression for θ, which can be convenient. For completeness, we can also determine θ, by

$$\theta = \omega t = \frac{dE}{dI} \frac{d}{dE} \int p\, dq = \frac{d}{dI} \int p\, dq.$$

Here the value of θ determines the upper bound on the integral, and the derivative acts on the integrand.

We now turn to adiabatic invariants.

- Consider a situation where the Hamiltonian depends on a parameter $\lambda(t)$ that changes slowly. Then energy is not conserved; taking $H(q(t), p(t), \lambda(t)) = E(t)$ and differentiating, we have

$$\dot{E} = \frac{\partial H}{\partial \lambda}\dot{\lambda}.$$

  However, certain 'adiabatic invariants' are approximately conserved.

- We claim that in the case

$$H = \frac{p^2}{2m} + V(q; \lambda(t))$$

  the adiabatic invariant is simply the action variable I. Since I is always evaluated on an orbit of the Hamiltonian at a fixed time, it is only a function of E and λ, so

$$\dot{I} = \left.\frac{\partial I}{\partial E}\right|_\lambda \dot{E} + \left.\frac{\partial I}{\partial \lambda}\right|_E \dot{\lambda}.$$

  These two contributions are due to the nonconservation of energy, and from the change in the shape of the orbits at fixed energy, respectively.

- When λ is constant, $E = E(I)$ as before, so

$$\left.\frac{\partial I}{\partial E}\right|_\lambda = \frac{1}{\omega(\lambda)} = \frac{T(\lambda)}{2\pi}.$$

  As for the second term, we have

$$\left.\frac{\partial I}{\partial \lambda}\right|_E = \frac{1}{2\pi}\oint \left.\frac{\partial p}{\partial \lambda}\right|_E dq = \frac{1}{2\pi}\oint \left.\frac{\partial p}{\partial \lambda}\right|_E \left.\frac{\partial H}{\partial p}\right|_{\lambda,q} dt'$$

  where we applied Hamilton's equations, and neglected a higher-order term from the change in the endpoints.

- To simplify the integrand, take $H(q, p(q, \lambda, E), \lambda) = E$ and differentiate with respect to λ at fixed E. Then

$$\left.\frac{\partial H}{\partial q}\right|_{\lambda,p}\left.\frac{\partial q}{\partial \lambda}\right|_E + \left.\frac{\partial H}{\partial p}\right|_{\lambda,q}\left.\frac{\partial p}{\partial \lambda}\right|_E + \left.\frac{\partial H}{\partial \lambda}\right|_{q,p,E} = 0.$$

  By construction, the first term is zero. Then we conclude that

$$\left.\frac{\partial I}{\partial \lambda}\right|_E = -\frac{1}{2\pi}\oint \left.\frac{\partial H}{\partial \lambda}\right|_E dt'.$$

  Finally, combining this with our first result, we conclude

$$\dot{I} = \left(T(\lambda)\left.\frac{\partial H}{\partial \lambda}\right|_E - \int \left.\frac{\partial H}{\partial \lambda}\right|_E dt'\right)\frac{\dot{\lambda}}{2\pi}.$$

  Taking the time average of $\dot{I}$ and noting that the change in λ is slow compared to the period of the motion, the two quantities above cancel, so $\langle \dot{I}\rangle = 0$ and I is an adiabatic invariant.

Example. The simple harmonic oscillator has $I = E/\omega$. Then if ω is changed slowly, the ratio E/ω remains constant. The above example also manifests in quantum mechanics; for example, for quanta in a harmonic oscillator, we have $E = n\hbar\omega$. If the ω of the oscillator is changed slowly, the energy can only remain quantized if E/ω remains constant, as it does in classical mechanics.

Example. The adiabatic theorem can also be proved heuristically with Liouville's theorem. We consider an ensemble of systems with fixed E but equally spaced phase θ, forming a ring in phase space. Under any time variation of λ, the ring remains a ring, and the area inside it is conserved.

Now suppose λ is varied extremely slowly. Then every system on the ring should be affected in the same way, so the final ring remains a curve of constant energy E'. Then the area inside curves of constant energy is conserved, proving the theorem.

Example. A particle in a magnetic field. Consider a particle orbiting in a magnetic field $\mathbf{B} = B(x,y)\hat{\mathbf{z}}$. If the field is slowly varying over space, and the particle orbit drifts slowly, then the adiabatic theorem holds. Integrating over a cycle,

$$I = \frac{1}{2\pi}\oint \mathbf{p}\cdot d\mathbf{q} \propto \int m\mathbf{v}\cdot d\mathbf{q} - e\int \mathbf{A}\cdot d\mathbf{q} = \frac{2\pi}{\omega}mv^2 - e\Phi_B.$$

Using $v = R\omega$ and $\omega = eB/m$, we find that these two terms are proportional to each other, so the magnetic flux is conserved! Alternatively, since $\Phi_B = AB$ and $B \propto \omega$, the magnetic moment of the particle is conserved; this is what plasma physicists called the 'first adiabatic invariant'. This is the principle behind magnetic mirrors, and it implies that charged particles in a magnetic field can be heated by increasing the field.

1.6 The Hamilton-Jacobi Equation

We begin by defining Hamilton's principal function.

- Given initial conditions (q_i, t_i) and final conditions (q_f, t_f), there can generally be multiple classical paths between them. Often, paths are discrete, so we may label them with a branch index b. However, note that for the harmonic oscillator we need a continuous branch index.
- For each branch index, we define Hamilton's principal function as
$$S_b(q_i, t_i; q_f, t_f) = A[q_b(t)] = \int_{t_i}^{t_f} dt\, L(q_b(t), \dot{q}_b(t), t)$$
where A stands for the usual action. We suppress the branch index below, so the four arguments of S alone specify the entire path.
- Consider an infinitesimal change in q_f. Then the new path is equal to the old path plus a variation δq with $\delta q(t_f) = \delta q_f$. Integrating by parts gives an endpoint contribution $p_f \delta q_f$, so
$$\frac{\partial S}{\partial q_f} = p_f.$$
- Next, suppose we simply extend the existing path by running it for an additional time dt_f. Then we can compute the change in S in two ways,
$$dS = L_f dt_f = \frac{\partial S}{\partial t_f}dt_f + \frac{\partial S}{\partial q_f}dq_f$$
where $dq_f = \dot{q}_f dt_f$. Therefore,
$$\frac{\partial S}{\partial t_f} = -H_f.$$
By similar reasoning, we have
$$\frac{\partial S}{\partial q_i} = -p_i, \quad \frac{\partial S}{\partial t_i} = H_i.$$

- The results above give $p_{i,f}$ in terms of $q_{i,f}$ and $t_{i,f}$. We can then invert the expression for p_i to write $q_f = q_f(p_i, q_i, t_i, t_f)$, and plug this in to get $p_f = p_f(p_i, q_i, t_i, t_f)$. That is, given an initial condition (q_i, p_i) at $t = t_i$, we can find (q_f, p_f) at $t = t_f$ given S.
- Henceforth we take q_i and t_i as fixed and implicit, and rename q_f and t_f to q and t. Then we have $S(q, t)$ with
$$dS = -H\,dt + p\,dq$$
where q_i and t_i simply provide the integration constants. The signs here are natural if one imagines them descending from special relativity.
- To evaluate S, we use our result for $\partial S/\partial t$, called the Hamilton-Jacobi equation,
$$H(q, \partial S/\partial q, t) + \frac{\partial S}{\partial t} = 0.$$
That is, S can be determined by solving a PDE. The utility of this method is that the PDE can be separated whenever the problem has symmetry, reducing the problem to a set of *independent* ODEs. We can also run the Hamilton-Jacobi equation in reverse to solve PDEs by identifying them with mechanical systems.
- For a time-independent Hamiltonian, the value of the Hamiltonian is just the conserved energy, so the quantity $S^0 = S + Et$ is time-independent and satisfies the time-independent Hamilton-Jacobi equation
$$H(q, \partial S^0/\partial q) = E.$$
The function S^0 can be used to find the paths of particles of energy E.

We now connect Hamilton's principal function to semiclassical mechanics.

- We can easily find the paths by solving the first-order equation
$$\dot{q} = \left.\frac{\partial H}{\partial p}\right|_{p=\partial S/\partial q}.$$
That is, Hamilton's principal function can reduce the equations of motion to first-order equations on configuration space.
- As a check, we verify that Hamilton's second equation is satisfied. We have
$$\dot{p} = \frac{d}{dt}\frac{\partial S}{\partial q} = \frac{\partial^2 S}{\partial t \partial q} + \frac{\partial^2 S}{\partial q^2}\dot{q}$$
where the partial derivative $\partial/\partial q$ keeps t constant, and
$$\frac{\partial^2 S}{\partial t \partial q} = -\frac{\partial}{\partial q} H(q, \partial S/\partial q, t) = -\frac{\partial H}{\partial q} - \frac{\partial^2 S}{\partial q^2}\dot{q}.$$
Hence combining these results gives $\dot{p} = -\partial H/\partial q$ as desired.

- The quantity $S(q,t)$ acts like a real-valued 'classical wavefunction'. Given a position, its gradient specifies the momentum. To see the connection with quantum mechanics, let

$$\psi(q,t) = R(q,t)e^{iW(q,t)/\hbar}.$$

  We assume the wavefunction varies slowly, in the sense that

$$\hbar \left| \frac{\partial^2 W}{\partial q^2} \right| \ll \left| \frac{\partial W}{\partial q} \right|.$$

  Some care needs to be taken here. We assume R and W are analytic in $\hbar$, but this implies that ψ is not.

- Expanding the Schrodinger equation to lowest order in $\hbar$ gives

$$\frac{\partial W}{\partial t} + \frac{1}{2m}\left(\frac{\partial W}{\partial q}\right)^2 + V(q) = O(\hbar).$$

  Then in the semiclassical limit, W obeys the Hamilton-Jacobi equation. The action $S(q,t)$ is the semiclassical phase of the quantum wavefunction. This result anticipates the de Broglie relations $\mathbf{p} = \hbar\mathbf{k}$ and $E = \hbar\omega$ classically, and inspires the path integral formulation.

- With this intuition, we can read off the Hamilton-Jacobi equation from a dispersion relation. For example, a free relativistic particle has $p_\mu p^\mu = m^2$, which means the Hamilton-Jacobi equation is

$$\eta^{\mu\nu}\partial_\mu S \partial_\nu S = m^2.$$

  This generalizes immediately to curved spacetime by using a general metric.

- To see how classical paths emerge in one dimension, consider forming a wavepacket by superposing solutions with the same phase at time $t_i = 0$ but slightly different energies. The solutions constructively interfere when $\partial S/\partial E = 0$, because

$$\frac{\partial S}{\partial E} = -t + \int \frac{\partial p}{\partial E}\, dq = -t + \int \frac{dq}{\partial H/\partial p} = -t + \int \frac{dq}{\dot{q}} = 0$$

  where we used Hamilton's equations.

There is also a useful analogy with optics.

- Fermat's principle of least time states that light travels between two points in the shortest possible time. We consider an inhomogeneous anisotropic medium. Consider the set of all points that can be reached from point $\mathbf{q}_0$ within time t. The boundary of this set is the wavefront $\Phi_{\mathbf{q}_0}(t)$.

- Huygen's theorem states that

$$\Phi_{\mathbf{q}_0}(s+t) \text{ is the envelope of the fronts } \Phi_{\mathbf{q}}(s) \text{ for } \mathbf{q} \in \Phi_{\mathbf{q}_0}(t).$$

  This follows because $\Phi_{\mathbf{q}_0}(s+t)$ is the set of points we need time $s+t$ to reach, and an optimal path to one of these points should be locally optimal as well. In particular, note that each of the fronts $\Phi_{\mathbf{q}}(s)$ is tangent to $\Phi_{\mathbf{q}_0}(s+t)$.

- Let $S_{\mathbf{q}_0}(\mathbf{q})$ be the minimum time needed to reach point $\mathbf{q}$ from $\mathbf{q}_0$. We define

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{q}}$$

  to be the vector of normal slowness of the front. It describes the motion of wavefronts, while $\dot{\mathbf{q}}$ describes the motion of rays of light. We thus have $dS = \mathbf{p}\, d\mathbf{q}$.

- The quantities $\mathbf{p}$ and $\dot{\mathbf{q}}$ can be related geometrically. Let the indicatrix at a point be the surface defined by the possible velocity vectors; it is essentially the wavefront at that point for infinitesimal time. Define the conjugate of $\dot{\mathbf{q}}$ to be the plane tangent to the indicatrix at $\dot{\mathbf{q}}$.

- The wave front $\Phi_{\mathbf{q}_0}(t)$ at the point $\mathbf{q}(t)$ is conjugate to $\dot{\mathbf{q}}(t)$. By decomposing $t = (t - \epsilon) + \epsilon$ and applying the definition of an indicatrix, this follows from Huygen's theorem.

- Everything we have said here is perfectly analogous to mechanics; we simply replace the total time with the action, and hence the indicatrix with the Lagrangian. The rays correspond to trajectories. The main difference is that the speed the rays are traversed is fixed in optics but variable in mechanics, so our space is $(\mathbf{q}, t)$ rather than just $\mathbf{q}$, and $dS = \mathbf{p}\, d\mathbf{q} - H\, dt$ instead.

2 Electromagnetism

2.1 Electrostatics

The fundamental equations of electrostatics are

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad \nabla \times \mathbf{E} = 0.$$

The latter equation allows us to introduce the potential $\mathbf{E} = -\nabla\phi$, giving Poisson's equation

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}.$$

The case $\rho = 0$ is Laplace's equation and the solutions are harmonic functions.

Example. The field of a point charge is spherically symmetric with $\nabla^2\phi = 0$ except at the origin. Guessing the form $\phi \propto 1/r$, we have

$$\nabla\left(\frac{1}{r}\right) = \frac{-\nabla r}{r^2} = -\frac{\mathbf{r}}{r^3}.$$

Next, we can take the divergence by the product rule,

$$\nabla^2\left(\frac{1}{r}\right) = -\left(\frac{\nabla\cdot\mathbf{r}}{r^3} - \frac{3\hat{\mathbf{r}}\cdot\mathbf{r}}{r^4}\right) = -\left(\frac{3}{r^3} - \frac{3}{r^3}\right) = 0$$

as desired. To get the overall constant, we use Gauss's law, for $\phi = q/(4\pi\epsilon_0 r)$.

Example. The electric dipole has

$$\phi = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{1}{|\mathbf{r}+\mathbf{d}|}\right).$$

To approximate this, we use the Taylor expansion

$$f(\mathbf{r}+\mathbf{d}) = \sum_n \frac{(\mathbf{d}\cdot\nabla)^n}{n!} f(\mathbf{r})$$

which can be understood by expanding in components with $\mathbf{d}\cdot\nabla = d_i\partial_i$. Then

$$\phi \approx \frac{Q}{4\pi\epsilon_0}\left(-\mathbf{d}\cdot\nabla\frac{1}{r}\right) = \frac{Q}{4\pi\epsilon_0}\frac{\mathbf{d}\cdot\mathbf{r}}{r^3}.$$

We see the potential falls off as $1/r^2$, and at large distances only depends on the dipole moment $\mathbf{p} = Q\mathbf{d}$. Differentiating using the usual quotient rule,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}}{r^3}.$$

Taking only the first term of the Taylor series is justified if $r \gg d$. More generally, for an arbitrary charge distribution

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int d\mathbf{r}'\,\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$$

and approximating the integrand with Taylor series gives the multipole expansion.

Note. Electromagnetic field energy. The energy needed to assemble a set of particles is

$$U = \frac{1}{2} \sum_i q_i \phi(\mathbf{r}_i).$$

This generalizes naturally to the energy to assemble a continuous charge distribution,

$$U = \frac{1}{2} \int d\mathbf{r}\, \rho(\mathbf{r}) \phi(\mathbf{r}).$$

Integrating by parts, we conclude that

$$U = \frac{\epsilon_0}{2} \int d\mathbf{r}\, E^2$$

where we tossed away a surface term. However, there's a subtlety when we go back to considering point charges, where these results no longer agree. The first equation explicitly doesn't include a charge's self-interaction, as the potential $\phi(\mathbf{r}_i)$ is supposed to be determined by all other charges. The second equation does, and hence the final result is positive definite. It can be thought of as additionally including the energy needed to assemble each point charge from scratch.

Example. Dipole-dipole interactions. Consider a dipole moment $\mathbf{p}_1$ at the origin, and a second dipole with charge Q at $\mathbf{r}$ and $-Q$ at $\mathbf{r} - \mathbf{d}$, with dipole moment $\mathbf{p}_2 = Q\mathbf{d}$. The potential energy is

$$U = \frac{Q}{2}(\phi(\mathbf{r}) - \phi(\mathbf{r} - \mathbf{d})) = \frac{1}{8\pi\epsilon_0} (\mathbf{d} \cdot \nabla) \frac{\mathbf{p}_1 \cdot \mathbf{r}}{r^3} = \frac{1}{8\pi\epsilon_0} \frac{\mathbf{p}_1 \cdot \mathbf{p}_2 - 3(\mathbf{p}_1 \cdot \hat{\mathbf{r}})(\mathbf{p}_2 \cdot \hat{\mathbf{r}})}{r^3}$$

where we used our dipole potential and the product rule. Then the interaction energy between permanent dipoles falls off as $1/r^3$.

Example. Boundary value problems. Consider a volume bounded by surfaces S_i, which could include a surface at infinity. Then Laplace's equation $\nabla^2 \phi = 0$ has a unique solution (up to constants) if we fix ϕ or $\nabla\phi \cdot \hat{\mathbf{n}} \propto \mathbf{E}_\perp$ on each surface. These are called Dirichlet and Neumann boundary conditions respectively. To see this, let f be the difference of two solutions. Then

$$\int dV\, (\nabla f)^2 = \int dV\, \nabla \cdot (f \nabla f) = \int f \nabla f \cdot d\mathbf{S}$$

where we used $\nabla^2 f = 0$ in the first equality. However, boundary conditions force the right-hand side to be zero, so the left-hand side is zero, which requires f to be constant.

In the case where the surfaces are conductors, it also suffices to specify the charge on each surface. To see this, note that potential is constant on a surface, so

$$\int f \nabla f \cdot d\mathbf{S} = f \int \nabla f \cdot d\mathbf{S} = 0$$

because the total charge on a surface is zero if we subtract two solutions. Then $\nabla f = 0$ as before, giving the same conclusion.

2.2 Magnetostatics

- The fundamental equations of magnetostatics are

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \quad \nabla \cdot \mathbf{B} = 0.$$

- Since the divergence of a curl is zero, we must have $\nabla \cdot \mathbf{J} = 0$. This is simply a consequence of the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

  and the fact that we're doing statics.

- Integrating Ampere's law yields

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I.$$

  This shows that the magnetic field of an infinite wire is $B_\theta = \mu_0 I / 2\pi r$.

- A uniform surface current $\mathbf{K}$ produces discontinuities in the field,

$$\Delta B_{\parallel} = \mu_0 K, \quad \Delta B_{\perp} = 0.$$

  This is similar to the case of a surface charge, except there $E_\perp$ is discontinuous instead.

- Consider an infinite cylindrical solenoid. Then $\mathbf{B} = B(r)\hat{\mathbf{z}}$ by symmetry. Both inside and outside the solenoid, we have $\nabla \times \mathbf{B} = 0$ which implies $\partial B / \partial r = 0$. Since fields vanish at infinity, the field outside must be zero, and by Ampere's law, the field inside is

$$B = \mu_0 K$$

  where K is the surface current density, equal to nI where n is the number of turns per length.

- Define the vector potential as

$$\mathbf{B} = \nabla \times \mathbf{A}.$$

  The vector potential is ambiguous up to the addition of a gradient $\nabla \chi$.

- By adding such a gradient, the divergence of $\mathbf{A}$ is changed by $\nabla^2 \chi$. Then by the existence theorem for Poisson's equation, we can choose any desired $\nabla \cdot \mathbf{A}$ by gauge transformations.

- One useful choice is Coulomb gauge $\nabla \cdot \mathbf{A} = 0$. As a result, Ampere's law becomes

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J}$$

  where we used the curl-of-curl identity,

$$\nabla^2 \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}).$$

Note. What is the vector Laplacian? Formally, the Laplacian of any tensor is defined as

$$\nabla^2 T = \nabla \cdot (\nabla T).$$

In a general manifold with metric, the operations on the right-hand side are defined through covariant derivatives, and depend on a connection. Going to the other extreme of generality, it can be defined in Cartesian components in $\mathbb{R}^n$ as the tensor whose components are the scalar Laplacians of those of T; we can then generalize to, e.g. spherical coordinates by a change of coordinates.

In the case of the vector Laplacian, the most practical definition for curvilinear coordinates on $\mathbb{R}^n$ is to use the curl-of-curl identity in reverse, then plug in the known expressions for divergence, gradient, and curl. This route doesn't require any tensor operations.

We now use our mathematical tools to derive the Biot-Savart law.

- By analogy with the solution to Poisson's equation by Green's functions,

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \, \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}.$$

  We can explicitly prove this by working in components in Cartesian coordinates. This equation also shows a shortcoming of vector notation: read literally, it is ambiguous what the indices on the vectors should be.

- To check whether the Coulomb gauge condition is satisfied, note that

$$\nabla \cdot \mathbf{A}(\mathbf{x}) \propto \int d\mathbf{x}' \, \nabla \cdot \left(\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) = \int d\mathbf{x}' \, \mathbf{J}(\mathbf{x}') \cdot \nabla \frac{1}{|\mathbf{x} - \mathbf{x}'|} = -\int d\mathbf{x}' \, \mathbf{J}(\mathbf{x}') \cdot \nabla' \frac{1}{|\mathbf{x} - \mathbf{x}'|}.$$

  The vector notation has some problems: it's ambiguous what index the divergence acts on (so we try to keep it linked to $\mathbf{J}$ with dots), and it's ambiguous what coordinate it differentiates (so we mark this with primes). In the final step, we used antisymmetry to turn ∇ into $-\nabla'$.

  This expression can be integrated by parts (clearer in index notation) to yield a surface term and a term proportional to $\nabla \cdot \mathbf{J} = 0$, giving $\nabla \cdot \mathbf{A} = 0$ as desired.

- Taking the curl and using the product rule,

$$\mathbf{B}(\mathbf{x}) = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \, \nabla \times \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \left(\nabla \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \times \mathbf{J}(\mathbf{x}') = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \, \frac{\mathbf{J}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}$$

  which is the Biot-Savart law.

Next, we investigate magnetic dipoles and multipoles.

- A current loop tracing out the curve C has vector potential

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \oint_C \frac{d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}$$

  by the Biot-Savart law.

- Just as for electric dipoles, we can expand

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{r} + \frac{\mathbf{r} \cdot \mathbf{r}'}{r^3} + \cdots$$

  for small $\mathbf{r}'$. The first term always integrates to zero about a closed loop, as there are no magnetic monopoles, while the next term gives

$$\mathbf{A}(\mathbf{r}) \approx \frac{\mu_0 I}{4\pi} \oint_C d\mathbf{r}' \frac{\mathbf{r} \cdot \mathbf{r}'}{r^3}.$$

- To simplify, pull the $1/r^3$ out of the integral, then dot the integral with $\mathbf{g}$ for

$$\oint_C g_i r_j r'_j \, dr'_i = \int_S \epsilon_{ijk} \partial'_i (g_j r_\ell r'_\ell) \, dS'_k = \int_S \epsilon_{ijk} r_i g_j \, dS'_k = \mathbf{g} \cdot \int d\mathbf{S}' \times \mathbf{r}$$

by Stokes' theorem. Since both $\mathbf{g}$ and $\mathbf{r}$ are constants, we conclude

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}, \quad \mathbf{m} = I\mathbf{S}, \quad \mathbf{S} = \int_S d\mathbf{S}.$$

Here, $\mathbf{S}$ is the vector area, and $\mathbf{m}$ is the magnetic dipole moment.

- Taking the curl straightforwardly gives the magnetic field,

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{3(\mathbf{m} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}}{r^3}$$

which is the same as the far-field of an electric dipole. (Near the dipoles, the fields differ because the electric and magnetic fields are curlless and divergenceless, respectively).

Example. We can do more complicated variants of these tricks for a general current distribution,

$$A_i(\mathbf{r}) = \frac{\mu_0}{4\pi} \int d\mathbf{r}' \left(\frac{J_i(\mathbf{r}')}{r} + \frac{J_i(\mathbf{r}')(\mathbf{r} \cdot \mathbf{r}')}{r^3} + \dots \right).$$

To simplify the first term, note that

$$\partial_j (J_j r_i) = (\partial_j J_j) r_i + J_i = J_i$$

where we used $\nabla \cdot \mathbf{J} = 0$. Then the monopole term is a total derivative and hence vanishes. The intuitive interpretation is that currents must go around in loops, with no net motion; our identity then says something like 'the center of charge doesn't move'.

To simplify the second term, note that

$$\partial_j (J_j r_i r_k) = J_i r_k + J_k r_i.$$

We can thus use this to 'antisymmetrize' the integrand,

$$\int d\mathbf{r}' \, J_i r_j r'_j = \int d\mathbf{r}' \frac{r_j}{2} (J_i r'_j - J_j r'_i) = \left(\frac{\mathbf{r}}{2} \times \int d\mathbf{r}' \, \mathbf{J} \times \mathbf{r}' \right)_i$$

where we used the double cross product identity. Then we conclude the dipole field has the same form as before, with the more general dipole moment

$$\mathbf{m} = \frac{1}{2} \int d\mathbf{r}' \, \mathbf{r}' \times \mathbf{J}(\mathbf{r}')$$

which is equivalent to our earlier result by the vector identity

$$\frac{1}{2} \int \mathbf{r} \times d\mathbf{s} = \int d\mathbf{S}.$$

Example. The force on a magnetic dipole. The force on a general current distribution is

$$\mathbf{F} = \int d\mathbf{r}\, \mathbf{J}(\mathbf{r}) \times \mathbf{B}(\mathbf{r}).$$

For small distributions localized about $\mathbf{r} = \mathbf{R}$, we can Taylor expand for

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}(\mathbf{R}) + (\mathbf{r} \cdot \nabla')\mathbf{B}(\mathbf{r}')\bigg|_{\mathbf{r}'=\mathbf{R}}.$$

Here, we turned the $\mathbf{R}$ into an $\mathbf{r}'$ evaluated at $\mathbf{R}$ so it's clear what coordinate the derivative is acting on. The first term contributes nothing, by the same logic as the previous example. In indices, the second term is

$$\mathbf{F} = \int d\mathbf{r}\, \mathbf{J}(\mathbf{r}) \times \big((\mathbf{r} \cdot \nabla')\mathbf{B}(\mathbf{r}')\big) = \int d\mathbf{r}\, \epsilon_{ijk} J_i r_\ell \left(\partial'_\ell B_j(\mathbf{r}')\right) \hat{\mathbf{e}}_k.$$

Now we focus on the terms in parentheses. In general, the curl is just the exterior derivative, so if the curl of $\mathbf{B}$ vanishes, then

$$\partial_i B_j - \partial_j B_i = 0.$$

This looks different from the usual (3D) expression for vanishing curl, which contains ϵ_{ijk}, because there we additionally take the Hodge dual. This means that we can swap the indices for

$$\int d\mathbf{r}\, \epsilon_{ijk} J_i r_\ell \left(\partial'_j B_\ell(\mathbf{r}')\right) \hat{\mathbf{e}}_k = -\nabla' \times \int d\mathbf{r}\, (\mathbf{r} \cdot \mathbf{B}(\mathbf{r}'))\mathbf{J}(\mathbf{r}).$$

Now the integral is identical to our magnetic dipole integral from above, with a constant vector of $\mathbf{B}(\mathbf{r}')$ instead. Therefore

$$\mathbf{F} = \nabla \times (\mathbf{B} \times \mathbf{m}) = (\mathbf{m} \cdot \nabla)\mathbf{B} = \nabla(\mathbf{B} \cdot \mathbf{m}), \quad U = -\mathbf{B} \cdot \mathbf{m}.$$

In the first step, we use a product rule along with $\nabla \cdot \mathbf{B} = 0$. For the final step, we again use the 'derivative index swapping' trick which works because the curl of $\mathbf{B}$ vanishes. The resulting potential energy can also be used to find the torque on a dipole.

2.3 Electrodynamics

The first fundamental equation of electrodynamics is Faraday's law,

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0.$$

In particular, defining the emf as

$$\mathcal{E} = \frac{1}{q} \int_C \mathbf{F} \cdot d\mathbf{r}$$

where $\mathbf{F}$ is the Lorentz force on a charge q, we have

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

where Φ is the flux through a surface with boundary C.

- For conducting loops, the resulting emf will create a current that creates a field that opposes the change in flux; this is Lenz's law. This is simply a consequence of energy conservation; if the sign were flipped, we would get runaway positive feedback.

- The integrated form of Faraday's law still holds for moving wires. Consider a loop C with surface S whose points have velocity $\mathbf{v}(\mathbf{r})$ in a static field. After a small time dt, the surface becomes S'. Since the flux through any closed surface is zero,

$$d\Phi = \int_{S'} \mathbf{B} \cdot d\mathbf{S} - \int_S \mathbf{B} \cdot d\mathbf{S} = -\int_{S_c} \mathbf{B} \cdot d\mathbf{S}$$

  where S_c is the surface with boundary C and C'. Choosing this surface to be straight gives $d\mathbf{S} = (d\mathbf{r} \times \mathbf{v})\, dt$, so

$$\frac{d\Phi}{dt} = -\int_C \mathbf{B} \cdot (d\mathbf{r} \times \mathbf{v}) = -\int_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{r}.$$

  Then Faraday's law holds as before, though the emf is now supplied by a magnetic force.

- Define the self-inductance of a curve C with surface S to be

$$L = \frac{\Phi}{I}$$

  where Φ is the flux through S when current I flows through C. Then

$$\mathcal{E} = -L\frac{dI}{dt}, \quad U = \frac{1}{2}LI^2 = \frac{1}{2}I\Phi.$$

  Inductors thus store energy when a current flows through them.

- As an example, a solenoid has $B = \mu_0 nI$ with total flux $\Phi = BAn\ell$ where ℓ is the total length. Therefore $L = \mu_0 n^2 V$ where $V = A\ell$ is the total volume.

- We can use our inductor energy expression to get the magnetic field energy density,

$$U = \frac{1}{2}I\int_S \mathbf{B} \cdot d\mathbf{S} = \frac{1}{2}I\int_C \mathbf{A} \cdot d\mathbf{r} = \frac{1}{2}\int d\mathbf{x}\, \mathbf{J} \cdot \mathbf{A}$$

  where we turned the line integral into a volume integral.

- Using $\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$ and integrating by parts gives

$$U = \frac{1}{2\mu_0}\int d\mathbf{x}\, \mathbf{B} \cdot \mathbf{B}.$$

  This does *not* prove the total energy density of an electromagnetic field is $u \sim E^2 + B^2$ because there can be $\mathbf{E} \cdot \mathbf{B}$ terms, and we've only worked with static fields. Later, we'll derive the energy density properly by starting from a Lagrangian.

Finally, we return to Ampere's law,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}.$$

As noted earlier, this forces $\nabla \cdot \mathbf{J} = 0$, so it must fail in general. The true equation is

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right)$$

so that taking the divergence now gives the full continuity equation. We see a changing electric field behaves like a current; it is a 'displacement current'.

- In vacuum, we thus have

$$\nabla \cdot \mathbf{E} = 0, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}.$$

  Combining these equations, we find

$$\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\nabla \times (\nabla \times \mathbf{E}) = \nabla^2 \mathbf{E}$$

  with a similar equation for $\mathbf{B}$, so electromagnetic waves propagate at speed $c = 1/\sqrt{\mu_0 \epsilon_0}$.

- Taking plane waves with amplitudes $\mathbf{E}_0$ and $\mathbf{B}_0$, we read off from Maxwell's equations

$$\mathbf{k} \cdot \mathbf{E}_0 = \mathbf{k} \cdot \mathbf{B}_0 = 0, \quad \mathbf{k} \times \mathbf{E}_0 = \omega \mathbf{B}_0$$

  using the correspondence $\nabla \sim i\mathbf{k}$. In particular, $E_0 = cB_0$.

- The rate of change of the field energy is

$$\dot{U} = \int d\mathbf{x} \left(\epsilon_0 \mathbf{E} \cdot \dot{\mathbf{E}} + \frac{1}{\mu_0} \mathbf{B} \cdot \dot{\mathbf{B}} \right) = \int d\mathbf{x} \left(\frac{1}{\mu_0} \mathbf{E} \cdot (\nabla \times \mathbf{B}) - \mathbf{E} \cdot \mathbf{J} - \frac{1}{\mu_0} \mathbf{B} \cdot (\nabla \times \mathbf{E}) \right).$$

  Using a product rule, we have

$$\dot{U} = -\int d\mathbf{x}\, \mathbf{J} \cdot \mathbf{E} - \frac{1}{\mu_0} \int (\mathbf{E} \times \mathbf{B}) \cdot d\mathbf{S}.$$

  This is a continuity equation for field energy; the first term is the rate work is done on charges, while the second describes the flow of energy along the boundary. In particular, the energy flow at each point in space is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}.$$

- In an electromagnetic wave, the average field energy density is $\overline{u} = \epsilon_0 E^2/2$, where we get a factor of $1/2$ from averaging a square trigonometric function and a factor of 2 from the magnetic field. As expected, the Poynting vector obeys $S = c\overline{u}$.

- Electromagnetic waves can also be written in terms of potentials, though these have gauge freedom. A common choice for plane waves is to set the electric potential ϕ to zero.

2.4 Relativity

Next, we rewrite our results relativistically. Conservation of charge is specified by the continuity equation

$$\partial_\mu J^\mu = 0, \quad J^\mu = (\rho, \mathbf{J}).$$

For example, transforming an initially stationary charge distribution gives

$$\rho' = \gamma \rho_0, \quad \mathbf{J}' = -\gamma \rho \mathbf{v}.$$

Note. Though the charge density is not invariant, one can show by the continuity equation that the total charge is. The total charge is

$$Q = \int d^3x\, J^0(x) = \int d^4x\, J^\mu(x) n_\mu \delta(n \cdot x).$$

Taking a Lorentz transform, we have

$$Q' = \int d^4x\, \Lambda^\mu{}_\nu J^\nu(\Lambda^{-1}x) n_\mu \delta(n \cdot x).$$

To make progress, we define $n' = \Lambda^{-1}n$ and $x' = \Lambda^{-1}x$. Changing variables to x',

$$Q' = \int d^4x'\, J^\nu(x') n'_\nu \delta(n' \cdot x').$$

This is identical to the expression for Q, except that n has been replaced with n'. Said another way, we can compute the total charge measured in another frame by doing an integral over a tilted spacelike surface in our original frame. Then by the continuity equation, we must have $Q = Q'$. More formally, we can use $n_\mu \delta(n \cdot x) = \partial_\mu \theta(n \cdot x)$ to show the difference is a total derivative.

Example. Deriving magnetism. Consider a wire with positive charges q moving with velocity v and negative charges $-q$ moving with velocity $-v$. Then

$$I = 2nAqv.$$

Now consider a particle moving in the same direction with velocity u, who measures the velocities of the charges to be $v_\pm = u \oplus (\mp v)$. Let n_0 be the number density in the rest frame of each kind of charge, so that $n = \gamma(v) n_0$. Using the property

$$\gamma(u \oplus v) = \gamma(u)\gamma(v)(1 + uv)$$

we can show the particle sees a total charge density of

$$\rho' = q(n_+ - n_-) = -q(uv\gamma(u))n$$

in its rest frame. It thus experiences an electric force of magnitude $F' \sim uv\gamma(u)$. Transforming back to the original frame gives $F \sim uv$, in agreement with our results from magnetostatics.

We now consider gauge transformations and the Faraday tensor.

- The fields are defined in terms of potentials as

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}.$$

- Gauge transformations are of the form

$$\phi \to \phi - \frac{\partial \chi}{\partial t}, \quad \mathbf{A} \to \mathbf{A} + \nabla\chi$$

  and leave the fields invariant.

- In relativistic notation, gauge transformations are

$$A_\mu \to A_\mu - \partial_\mu \chi, \quad A^\mu = (\phi, \mathbf{A}).$$

- The Faraday tensor is defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

and is gauge invariant. It contains the electric and magnetic fields in its components,

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}.$$

- In terms of indices or matrix multiplications,

$$F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma} \quad F' = \Lambda F \Lambda^T.$$

In the latter, F has both indices up, and Λ is the matrix that transforms vectors, $v \to \Lambda v$.

- Under rotations, we can verify that $\mathbf{E}$ and $\mathbf{B}$ also rotate. Under boosts along the x direction, we instead have

$$E'_x = E_x, \quad E'_y = \gamma(E_y - vB_z), \quad E'_z = \gamma(E_z + vB_y)$$

and

$$B'_x = B_x, \quad B'_y = \gamma(B_y + vE_z), \quad B'_z = \gamma(B_z - vE_y).$$

- In the limit $v \ll c$, we have

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}, \quad \mathbf{B}' = \mathbf{B} - \mathbf{v} \times \mathbf{E}.$$

Note that the former quantity is what appears in the Lorentz force law.

- We can construct the Lorentz scalars

$$F_{\mu\nu} F^{\mu\nu} \propto \mathbf{E}^2 - \mathbf{B}^2, \quad F_{\mu\nu} \widetilde{F}^{\mu\nu} \propto \mathbf{E} \cdot \mathbf{B}.$$

The intuition for the latter is that taking the dual simply swaps $\mathbf{E}$ and $\mathbf{B}$ (with some signs, i.e. $\mathbf{E} \to \mathbf{B} \to -\mathbf{E}$), so we can read off the answer.

Example. Slightly boosting the field of a line charge at rest gives a magnetic field $-\mathbf{v} \times \mathbf{E}$ which wraps around the wire, thus yielding Ampere's law. For larger boosts, we pick up a Lorentz contraction factor γ due to the contraction of the charge density.

Example. A boosted point charge. Ignoring constants, the field is

$$\mathbf{E} \sim \frac{\mathbf{r}}{r^3} = \frac{1}{(x^2 + y^2 + z^2)^{3/2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Now consider a frame moving with velocity $\mathbf{v} = v\hat{\mathbf{i}}$. Then the boosted field is

$$\mathbf{E}' \sim \frac{1}{(x^2+y^2+z^2)^{3/2}} \begin{pmatrix} x \\ \gamma y \\ \gamma z \end{pmatrix}$$

using the coordinates in the original field. Switching the coordinates to the boosted ones,

$$\mathbf{E}' \sim \frac{\gamma}{(\gamma^2(x'+vt')^2+y'^2+z'^2)^{3/2}} \begin{pmatrix} x'+vt' \\ y' \\ z' \end{pmatrix}$$

where we used $x = \gamma(x' + vt')$. Interestingly, the field remains radial. However, the x' coordinate in the denominator is effectively $\gamma x'$, so it's as if electric field lines have been length contracted. By charge invariance and Gauss's law, the total flux remains constant, so the field is stronger than usual along the perpendicular direction and weaker than usual along the parallel direction.

We conclude by rewriting Maxwell's equations and the Lorentz force law relativistically.

- Maxwell's equations are
$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu, \quad \partial_\mu \widetilde{F}^{\mu\nu} = 0.$$
Note that this automatically implies current conservation. Also note that the second one holds automatically given $F = dA$.
- The relativistic generalization of the Lorentz force law is
$$\frac{dp^\mu}{d\tau} = qF^{\mu\nu}u_\nu$$
where u is velocity. The spatial part is the usual Lorentz force, while the temporal part is
$$\frac{dE}{d\tau} = q\gamma \mathbf{E}\cdot\mathbf{u}.$$
This simply says that electric fields do work, while magnetic fields don't.
- One neat trick is that whenever $\mathbf{E}\cdot\mathbf{B} = 0$, we can boost to get either zero electric or zero magnetic field. For example, a particle in crossed fields either goes a cycloid-like motion, or falls arbitrarily far; the sign of $E^2 - B^2$ separates the two cases.

2.5 Radiation

In this section, we show how radiation is produced by accelerating charges.

- Expanding the equation of motion, we have
$$\partial_\nu F^{\nu\mu} = \mu_0 J^\mu, \quad \Box A^\mu - \partial^\mu \partial_\nu A^\nu = \mu_0 J^\mu.$$
- To simplify, we work on Lorenz gauge $\partial_\mu A^\mu = 0$, so $\Box A^\mu = \mu_0 J^\mu$. That is, the potential solves the wave equation, and its source is the current.
- Lorenz gauge exists if we can always pick a gauge transformation χ so that $\Box\chi = -\partial_\mu A^\mu$. Thus solving the wave equation will also show us how to get to Lorenz gauge in the first place.

We thus focus on finding the Green's function for $\Box$.

* Our first approach is to perform a Fourier transform in time only, for

$$(\nabla^2 + \omega^2) A_\mu = -\mu_0 J_\mu.$$

This is called the Helmholtz equation; the Poisson equation is the limit $\omega \to 0$. The function $J_\mu(\mathbf{x}, \omega)$ is the time Fourier transform of $J_\mu(\mathbf{x}, t)$ at every point $\mathbf{x}$.

* Define the Green's function for the Helmholtz equation as

$$(\nabla^2 + \omega^2) G_\omega(\mathbf{x}, \mathbf{x}') = \delta^3(\mathbf{x} - \mathbf{x}').$$

Translational and rotational symmetry mean $G_\omega(\mathbf{x}, \mathbf{x}') = G_\omega(r)$ where $r = |\mathbf{x} - \mathbf{x}'|$. We can think of $G_\omega(r)$ as the spatial response to a sinusoidal source of frequency ω at the origin.

* In spherical coordinates,

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dG_\omega}{dr} \right) + \omega^2 G_\omega = \delta^3(r).$$

This equation has solutions

$$G_\omega(r) = -\frac{1}{4\pi} \frac{e^{\pm i\omega r}}{r}.$$

One can arrive at this result by guessing that amplitudes fall as $1/r$, and hence working in terms of rG instead of G. The constant is found by integrating in a ball around $r = 0$.

* Plugging this result in, we have

$$A_\mu(\mathbf{x}, \omega) = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \, \frac{e^{\pm i\omega |\mathbf{x} - \mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|} J_\mu(\mathbf{x}', \omega).$$

Therefore, taking the inverse Fourier transform,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d\!\!\bar{}\omega \int d\mathbf{x}' \, \frac{e^{-i\omega(t \mp |\mathbf{x} - \mathbf{x}'|)}}{|\mathbf{x} - \mathbf{x}'|} J_\mu(\mathbf{x}', \omega) = \frac{\mu_0}{4\pi} \int d\mathbf{x}' \, \frac{J_\mu(\mathbf{x}', t \mp |\mathbf{x} - \mathbf{x}'|)}{|\mathbf{x} - \mathbf{x}'|}.$$

* The result is like the solution to the Poisson equation, except that the current must be evaluated at the retarded or advanced time; we take the retarded time as physical, defining

$$t_{\text{ret}} = t - |\mathbf{x} - \mathbf{x}'|.$$

We see that the Helmholtz equation contains the correct speed of light travel delay.

* Warning: while the potentials just depend on the current in the usual way, just evaluated at the retarded time, the same is *not* true of the fields! When we differentiate the potentials, we pick up extra terms from differentiating t_{ret}. These extra terms are crucial because they provide the radiation fields which fall off as $1/r$, rather than $1/r^2$.

We can also take the Fourier transform in both time and space.

* The Green's function for the wave equation satisfies

$$\Box G(\mathbf{x}, t, \mathbf{x}', t') = \delta^3(\mathbf{x} - \mathbf{x}') \delta(t - t').$$

By translational symmetry in both space and time, $G = G(\mathbf{r}, t)$.

- Taking a Fourier transform and solving, we have

$$G(\mathbf{k},\omega) = -\frac{1}{k^2-\omega^2/c^2}.$$

- Inverting the Fourier transform gives

$$G(\mathbf{r},t) = -\int đ^4k\,\frac{e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}}{k^2-\omega^2/c^2}.$$

  Switching to spherical coordinates with $\hat{\mathbf{z}} \parallel \mathbf{k}$ and doing the angular integration,

$$G(\mathbf{r},t) = \frac{1}{4\pi^3}\int_0^\infty dk\, c^2k^2\frac{\sin kr}{kr}\int_{-\infty}^\infty d\omega\,\frac{e^{-i\omega t}}{(\omega-ck)(\omega+ck)}.$$

- In order to perform the $d\omega$ integration, we need to deal with the poles. By adding an infinitesimal damping forward in time, we can push the poles below the real axis. Now, when $t<0$, the integration contour can be closed in the upper-half plane, giving zero. When $t>0$, we close in the lower-half plane, picking up both poles, so

$$\int_C d\omega\,\frac{e^{-i\omega t}}{(\omega-ck)(\omega+ck)} = -\frac{2\pi}{ck}\theta(t)\sin(ckt).$$

  Finally, doing the dk integral gives some delta functions, for

$$G_{\text{ret}}(\mathbf{r},t) = -\frac{\theta(t)}{4\pi r}\delta(t_{\text{ret}}).$$

  This is the retarded Green's function; plugging it into the wave equation gives us the same expression for the retarded potential as derived earlier.

- We can also apply antidamping, getting the advanced Green's function

$$G_{\text{adv}}(\mathbf{r},t) = -\frac{\theta(-t)}{4\pi r}\delta(t_{\text{adv}}).$$

- Both of these conventions can be visualized by pushing the integration contour above or below the real axis. If we instead tilt it about the origin, we get the Feynman propagator.

Note. Checking Lorenz gauge. Our retarded potential solution has the form

$$A_\mu(x) \sim \int d^4x'\, G(x,x')J_\mu(x').$$

Now consider computing $\partial_\mu A^\mu$. Since the Green's function only depends on $x-x'$, we have

$$\partial_\mu A^\mu \sim \int d^4x'\,\partial_\mu G(x,x')J_\mu(x') = -\int d^4x'\,(\partial'_\mu G(x,x'))J_\mu(x').$$

We can then integrate by parts; since $\partial_\mu J^\mu = 0$, Lorenz gauge holds.

We now use our results to analyze radiation from small objects.

- Consider an object centered on the origin with lengthscale d, with potential

$$A_\mu(\mathbf{x},t) = \frac{\mu_0}{4\pi}\int d\mathbf{x}'\,\frac{J_\mu(\mathbf{x}',t_{\text{ret}})}{|\mathbf{x}-\mathbf{x}'|}.$$

We would like to compute the field at a distance $r = |\mathbf{x}| \gg d$. Taylor expanding,

$$\frac{1}{|\mathbf{x}-\mathbf{x}'|} = \frac{1}{r} + \frac{\mathbf{x}\cdot\mathbf{x}'}{r^3} + \ldots, \qquad J_\mu(\mathbf{x}',t_{\text{ret}}) = J_\mu(\mathbf{x}', t - r/c + \mathbf{x}\cdot\mathbf{x}'/rc + \ldots).$$

- Going to leading order in d/r gives the electric dipole approximation,

$$A_\mu(\mathbf{x},t) \approx \frac{\mu_0}{4\pi r}\int d\mathbf{x}'\,J_\mu(\mathbf{x}', t - r/c).$$

This approximation only makes sense if the motion is nonrelativistic: the next correction term to t_{ret} is of order d/c, which is only small if the characteristic timescale of changes in the current is much greater than d/c.

- It's easiest to compute the field starting with the vector potential. We use the identity

$$\partial_j(J_j x_i) = -\dot{\rho} x_i + J_i, \qquad \int d\mathbf{x}'\,J(\mathbf{x}') = \dot{\mathbf{p}}$$

which is like our results in magnetostatics, but allowing for a varying dipole moment $\mathbf{p}$. Evaluating this at the time $t - r/c$,

$$\mathbf{A}(\mathbf{x},t) \approx \frac{\mu_0}{4\pi r}\dot{\mathbf{p}}(t-r/c).$$

- Applying the product rule, we have

$$\mathbf{B} \approx \frac{\mu_0}{4\pi}\left(-\frac{\hat{\mathbf{x}}\times\dot{\mathbf{p}}(t-r/c)}{r^2} - \frac{\hat{\mathbf{x}}\times\ddot{\mathbf{p}}(t-r/c)}{rc}\right).$$

The former is just the usual magnetic field but time-delayed, and the latter is the $1/r$ radiation field. If the dipole has characteristic frequency ω, then the latter dominates if $r \gg \lambda = c/\omega$, the far-field/radiation zone.

- In the radiation zone, the fields look like plane waves, with $\mathbf{E} = -c\hat{\mathbf{x}}\times\mathbf{B}$. Then

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} = \frac{c}{\mu_0}B^2\hat{\mathbf{x}} = \frac{\mu_0}{16\pi^2 r^2 c}|\hat{\mathbf{x}}\times\ddot{\mathbf{p}}|^2\hat{\mathbf{x}}$$

where we used the triple cross product rule.

- The total instantaneous power is thus

$$\mathcal{P} = \frac{\mu_0}{16\pi^2 c}\int \sin^2\theta\,d\Omega = \frac{\mu_0}{6\pi c}|\ddot{\mathbf{p}}|^2.$$

- Consider a particle of charge Q oscillating in the $\hat{\mathbf{z}}$ direction with frequency ω and amplitude d, and hence dipole moment $p = Qz$. Expanding and time averaging,

$$\mathcal{P}_{\text{av}} = \frac{\mu_0 p^2\omega^4}{12\pi c} = \frac{Q^2 a^2}{12\pi\epsilon_0 c^3}.$$

This is the Larmor formula; note that it is quadratic in charge and acceleration (the field is linear, but energy is bilinear). Since we used the electric dipole approximation, it only applies for nonrelativistic motion.

- Note that the radiation fields are zero along the $\hat{\mathbf{z}}$ axis. This is related to the hairy ball theorem: since the radiation fields are everywhere tangent to spheres about the charge, they must vanish somewhere.
- By taking higher-order terms in our Taylor series, we can get magnetic dipole and electric quadrupole terms, and so on. The magnetic dipole term is dominant in situations where there is no electric dipole moment (e.g. a current loop), but for moving charges its power is suppressed by v^2/c^2 and hence is much smaller in the nonrelativistic limit.

We can apply our results to scattering.

- As a warmup, we consider Thomson scattering. Consider a free particle in light, and assume that it never moves far compared to the wavelength of the light. Equivalently, we assume it never moves relativistically fast. Then
$$m\ddot{\mathbf{x}}(t) \approx q\mathbf{E}(\mathbf{x}=0,t), \quad \mathbf{x}(t) = -\frac{q\mathbf{E}_0}{m\omega^2}\sin(\omega t).$$
Applying the Larmor formula,
$$\mathcal{P}_{\text{av}} = \frac{\mu_0 q^4 E_0^2}{12\pi m^2 c}.$$
- The averaged Poynting vector for the light is
$$\mathcal{S}_{\text{av}} = \frac{cE_0^2}{2\mu_0}.$$
Therefore, Thomson scattering has a 'cross section' of
$$\sigma = \frac{\mathcal{P}_{\text{av}}}{\mathcal{S}_{\text{av}}} = \frac{8\pi}{3}r_q^2, \quad \frac{q^2}{4\pi\epsilon_0 r_q} = mc^2.$$
Here, r_q is called the classical electron radius. Note that it is independent of frequency.
- Thomson scattering is elastic, but if the particle moves relativistically fast, the scattered light can be redshifted by radiation recoil effects.
- Experimentally, it was found that the scattered light had a shifted wavelength for high frequencies and arbitrarily low intensities (Compton scattering), which provided support for the particle nature of light.
- Rayleigh scattering describes the scattering of light off a neutral but polarizable atom or molecule. We effectively add a spring and damping to the model of Thomson scattering, so
$$\mathbf{x}(t) = -\frac{q\mathbf{E}(t)/m}{\omega^2 - \omega_0^2 + i\gamma\omega}.$$
- In the limit $\omega \ll \omega_0$, which is a good approximation for visible light and molecules in the atmosphere, the amplitude is constant (rather than the $1/\omega^2$ for Thomson scattering), giving
$$\sigma = \frac{8\pi r_q^2}{3}\left(\frac{\omega}{\omega_0}\right)^4.$$
The fact that $\sigma \propto \omega^4$ explains why the sky is blue. Intuitively, scattering of low frequency light is suppressed because the 'molecular springs' limit how far the electrons can go.

- Rayleigh scattering holds when the size of the molecules involved is much smaller than the wavelength of the light. In the case where they are comparable, we get Mie scattering, which preferentially scatters longer wavelengths. The reason is that nearby molecules oscillate in phase, so their amplitudes superpose, giving a quadratic increase in power. Mie scattering applies for water droplets in the atmosphere, explaining why clouds are visible, and white. In the case where the scattering particles are much larger, we simply use geometric optics.

Note. As a final note, we can generalize our results to a relativistically moving charge. Suppose a point charge has position $\mathbf{r}(t)$. Then its retarded potential is

$$\phi(\mathbf{x},t) \propto \int d\mathbf{x}' \, \frac{\delta^3(\mathbf{x}' - r(t_{\text{ret}}))}{|\mathbf{x}-\mathbf{x}'|}.$$

The tricky part is that t_{ret} depends on $\mathbf{x}'$ nontrivially. Instead, it's easier to switch the delta function to be over time,

$$\phi(\mathbf{x},t) \propto \int d\mathbf{x}' \, dt \, \frac{\delta^3(\mathbf{x}' - r(t))\delta(t - t_{\text{ret}})}{|\mathbf{x}-\mathbf{x}'|} = \int dt' \, \frac{\delta(t - t' - |\mathbf{x} - \mathbf{r}(t')|/c)}{|\mathbf{x} - \mathbf{r}(t')|}.$$

The argument of the delta function changes both because of the t' and because of the velocity of the particle towards the point $\mathbf{x}$, giving an extra contribution akin to a Doppler shift. Then

$$\phi(\mathbf{x},t) = \frac{q}{4\pi\epsilon_0} \frac{1}{R(t')(1 - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')/c)}, \quad \mathbf{A}(\mathbf{x},t) = \frac{q\mu_0}{4\pi} \frac{\mathbf{v}(t')}{R(t')(1 - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')/c)}, \quad t' + R(t')/c = t$$

where $\mathbf{R}$ is the separation vector $\mathbf{R}(t) = \mathbf{x} - \mathbf{r}(t)$. These are the Lienard-Wierchert potentials. Carrying through the analysis, we can find the fields of a relativistic particle and the relativistic analogue of the Larmor formula. The result is that the radiation rate is greatly enhanced, and concentrated along the direction of motion of the particle.

3 Statistical Mechanics

3.1 Ensembles

First, we define the microcanonical ensemble.

- The fundamental postulate of statistical mechanics is that, for an isolated system in equilibrium, all accessible microstate are equally likely. Here, accessible means 'reachable due to small fluctuations'. For example, such fluctuations cannot modify conserved quantities.

- The fundamental postulate can be justified by Liouville's theorem: if $\rho(t)$ is constant, then it must be constant over all phase space trajectories, and generically we expect trajectories to fully explore the available phase space, i.e. we assume there are no hidden conserved quantities that get in the way.

- For now, we suppose that energy is the only conserved quantity. Then the probability of occupying state $|n\rangle$ is
$$p(n) = \frac{1}{\Omega(E)}$$
where $\Omega(E)$ is the number of states with energy E.

- We know that for a quantum system the energy levels can be discrete, but for a thermodynamically large system they form a continuum. Then what we really mean by $\Omega(E)$ is the number of states with energy in $[E, E + \delta E]$ where δE specifies how well we know the energy.

- We define the entropy of the system to be
$$S(E) = k_B \log \Omega(E).$$
For two non-interacting systems, Ω multiplies, so S adds. That is, entropy is extensive.

- Now suppose we allow the systems to weakly interact, so they can exchange energy, but the energy levels of the states aren't significantly shifted. Then the number of states is
$$\Omega(E_{\text{total}}) = \prod_{E_i} \Omega_1(E_i)\Omega_2(E_{\text{total}} - E_i) = \sum_{E_i} \exp\left(\frac{S_1(E_i) + S_2(E_{\text{total}} - E_i)}{k_B}\right).$$
After allowing the systems to come to equilibrium, so that the new system is described by a microcanonical ensemble, we find the entropy has increased. This is an example of the Second Law of Thermodynamics.

- Since S is extensive, the argument of the exponential above is huge in the thermodynamic limit, so we can approximate the sum by its maximum summand. (This is just the discrete saddle point method.) Then the final entropy is approximately $S_{\text{total}} = S_1(E_*) + S_2(E_{\text{total}} - E_*)$ where E_* is chosen to maximize S_{total}.

Next, we define temperature.

- We define the temperature T as
$$\frac{1}{T} = \frac{\partial S}{\partial E}.$$
Comparing this with our previous result, we find that in thermal equilibrium, the temperatures of the two systems are equal. Moreover, in the approach to equilibrium, energy flows from the hotter system to the colder one.

- The heat capacity is defined as
$$C = \frac{\partial E}{\partial T}, \quad \Delta S = \int \frac{C(T)}{T} dT.$$
Hence measuring the heat capacity allows us to measure the entropy.
- Above, we are only guaranteed that E_* maximizes S_{total} if, for each of the two systems,
$$\partial^2 S_i / \partial E^2 < 0.$$
If a system does not satisfy this condition, it is thermodynamically unstable. Placed in contact with a reservoir, it would never reach thermal equilibrium, instead emitting or absorbing as much energy as possible. In terms of the heat capacity, stability requires $C > 0$.
- For example, black holes are hotter than the CMB, and so emit energy by Hawking radiation. Since they get hotter as they lose energy, they continue emitting energy until they disappear.
- Given the above definitions, it is still perfectly consistent for a system to have negative temperature. Such a system gets more ordered as it absorbs energy.

We now add pressure and volume as thermodynamic variables.

- We now let Ω, and hence S, depend on volume. Define the pressure p as
$$p = T \frac{\partial S}{\partial V}.$$
Then by similar arguments as above, the pressures of systems are equal in thermal equilibrium.
- Consider two gases separated by a movable piston. The pressure difference will cause the piston to move; when it comes to rest, the entropy has increased and the pressures are equal. On a longer timescale, small vibrations of the piston transfer heat, setting the temperatures equal. This is because two interacting gases will have equal $\langle mv^2 \rangle$ after many collisions, and we can think of the piston as a one-particle gas.
- Rearranging the total differential of entropy, we find
$$dE = T dS - p dV.$$
We call 'work' the energy transferred by exchange of volume; the rest is 'heat'. More generally, we can write the work as a sum $\sum J_i dx_i$ where the x_i are generalized displacements and the J_i are their conjugate generalized forces.
- In general, the (x_i, J_i) behave similarly to (S, T). In equilibrium, the J_i are equal. For stability, we must have $\partial^2 E / \partial x^2 > 0$, which implies that the matrix $\partial J_i / \partial x_j$ is positive definite. For example, a gas with $(\partial p / \partial V)|_T > 0$ is unstable to expansion or collapse.

Next, we define the canonical ensemble.

- Consider a system S in thermal equilibrium with a large reservoir R. Then the number of microstates associated with a state where the system has energy E_n is
$$\Omega = \Omega_R(E_{\text{total}} - E_n) = \exp\left(\frac{S_R(E_{\text{total}} - E_n)}{k_B}\right) \approx \exp\left(\frac{S_R(E_{\text{total}})}{k_B} - \frac{\partial S_R}{\partial E_{\text{total}}} \frac{E_n}{k_B}\right)$$
where the approximation holds because the reservoir is very large.

- We conclude $\Omega \propto e^{-E_n/k_BT}$, so the probability of occupancy of a state of energy n is
$$p(n) = \frac{e^{-E_n/k_BT}}{Z}, \quad Z = \sum_n e^{-E_n/k_BT}.$$
For convenience, we define $\beta = 1/k_BT$.

Example. For noninteracting systems, the partition functions multiply. Another useful property is that the partition function is similar to the cumulant generating function for the energy,
$$f(\gamma) = \log\langle e^{\gamma E}\rangle = \log \sum_n \frac{e^{-(\beta-\gamma)E_n}}{Z}.$$
The cumulants are the derivatives of f evaluated at $\gamma = 0$. Only the numerator contributes to this term, and since it contains only $(\beta - \gamma)$ we can differentiate with respect to β instead,
$$f^{(n)}(\gamma)|_{\gamma=0} = (-1)^n \frac{\partial^n (\log Z)}{\partial \beta^n}.$$
As an explicit example,
$$\langle E\rangle = -\frac{\partial \log Z}{\partial \beta}, \quad \operatorname{var} E = \frac{\partial^2 \log Z}{\partial \beta^2}.$$
However, since $\operatorname{var} E = -\partial\langle E\rangle/\partial\beta$, we have
$$\operatorname{var} E = k_BT^2 C_V$$
which is a relative of the fluctuation-dissipation theorem. Moreover, all cumulants of the energy can be found by differentiating $\langle E\rangle$, so they are all extensive. Then in the thermodynamic limit the system has a definite energy and the canonical and microcanonical ensembles coincide. (This doesn't hold when we're applying the canonical ensemble to a small system, like a single atom.)

To see this another way, note that
$$Z = \sum_{E_i} \Omega(E_i) e^{-\beta E_i}$$
where we are now summing over energies instead of states. But in the thermodynamic limit, the two factors in the sum are rapidly rising and falling, so they are dominated by the maximum term, which has fixed energy.

Example. We now compute the entropy of the canonical ensemble. Suppose we had W copies of the canonical ensemble; then there will be $p(n)W$ systems in state $|n\rangle$. Since W is large, we can consider all the copies to lie in the microcanonical ensemble, for which the entropy is
$$S = k_B \log \Omega = k_B \log \frac{W!}{\prod_n (p(n)W)!} = -k_B W \sum_n p(n) \log p(n).$$
Since entropy is extensive, the entropy of one copy is
$$S = -k_B \sum_n p(n) \log p(n)$$
and this expression is called the Gibbs entropy. It is proportional to the Shannon entropy of information theory; it is the amount of information we gain if we learn what the microstate is, given knowledge of the macrostate.

Next, we define the free energy and other potentials.

- We define the free energy in the canonical ensemble as
$$F = E - TS.$$
We have tacitly taken the thermodynamic limit, defining E as $\langle E \rangle$.

- The differential of F is
$$dF = -SdT - pdV, \quad S = -\left.\frac{\partial F}{\partial T}\right|_V, \quad p = -\left.\frac{\partial F}{\partial V}\right|_T.$$

- To relate F to Z, use our expression for the Gibbs entropy for
$$S/k_B = -\sum_n \frac{e^{-\beta E_n}}{Z} \log \frac{e^{-\beta E_n}}{Z} = \log Z + \langle \beta E \rangle.$$
Rearranging, we find that
$$F = -k_B T \log Z.$$

- Next, we can allow the particle number N to vary, and define the chemical potential
$$\mu = -T \left.\frac{\partial S}{\partial N}\right|_{E,V}.$$
The total differential of energy becomes
$$dE = TdS - pdV + \mu dN, \quad \mu = \left.\frac{\partial E}{\partial N}\right|_{S,V}.$$

- Note that the chemical potential for an classical gas is negative, because it is the energy cost of a particle at fixed S. To keep the entropy the same, we typically have to remove more energy than the particle's presence added. By contrast, for the Fermi gas at zero temperature, $\mu = E_F > 0$ because the entropy is exactly zero.

- We may similarly define the grand canonical ensemble by allowing N to vary. Then
$$p(n) = \frac{e^{-\beta(E_n - \mu N_n)}}{\mathcal{Z}}, \quad \mathcal{Z}(T, \mu, V) = \sum_n e^{-\beta(E_n - \mu N_n)}$$
where $\mathcal{Z}$ is the grand canonical partition function.

- We can extract information about the distribution of N by differentiating $\mathcal{Z}$. The cumulant generating function argument goes through as before, giving
$$\langle N \rangle = \frac{\partial \log \mathcal{Z}}{\partial(\beta\mu)}, \quad \operatorname{var} N = \frac{\partial^2 \log \mathcal{Z}}{\partial(\beta\mu)^2}.$$
In particular, as with energy, we see that variance is extensive, so fluctuations disappear in the thermodynamic limit.

- Similarly, we define the grand canonical potential $\Phi = F - \mu N$, so that
$$d\Phi = -SdT - pdV - Nd\mu, \quad \Phi = -k_B T \log \mathcal{Z}$$
by analogous arguments to before.

Example. In most cases, the energy and entropy are extensive. This implies that
$$E(\lambda S, \lambda V, \lambda N) = \lambda E(S, V, N).$$
Differentiating at $\lambda = 1$, we find
$$E = TS - pV + \mu N.$$
Taking the total differential, we have the constraint
$$SdT - Vdp + Nd\mu = 0.$$
We also see that the grand canonical potential is $\Phi = -pV$, which provides an easy way to calculate the pressure. Note that if we performed one further Legendre transformation from V to p, we would get a potential that is identically zero! This makes sense, as with no extensive variables left, our "system" would have no characteristics independent of the bath. As such, the potential $\Phi + pV$ is not useful. Another useful insight is that $\mu = G/N$, so the chemical potential measures the Gibbs free energy per molecule.

3.2 Thermodynamics

At this point, we start over with thermodynamics. We ignore microscopic details, instead taking the Laws of Thermodynamics as phenomenological axioms.

- We describe systems only by macroscopically observable quantities (e.g. p and V for a gas). For simplicity, we only consider gases below.
- The Zeroth Law states that thermal equilibrium between systems exists, and is transitive. This means that we can assign systems a temperature $T(p, V)$ so that systems with the same temperature are in equilibrium. The equation $T = T(p, V)$ is an equation of state. At this stage, T can be replaced by $f(T)$ for any monotonic f.
- The First Law tells us that energy is a state function. Defining work to be a special subset of energy transfer (e.g. from macroscopic changes in macroscopic quantities like volume) and heat to be everything else, we have
$$dE = đQ + đW$$
where the $đ$ indicates an inexact differential. (Note that here we use 'exact' in the same sense as in the theory of differential forms.)
- The Second Law tells us that it's impossible to transfer heat from a colder body to a warmer body without any other effects.
- A Carnot cycle is a process involving an ideal gas that extracts heat Q_H from a hot reservoir and performs work W and dumps heat Q_L to a cold reservoir. We define the efficiency
$$\eta = \frac{W}{Q_H}.$$
By construction, the Carnot cycle is reversible. Then by the Second Law, no cycle can have greater efficiency.

- By composing two Carnot cycles, we have the constraint

$$(1 - \eta(T_1, T_3)) = (1 - \eta(T_1, T_2))(1 - \eta(T_2, T_3))$$

  where T is the temperature. Therefore

$$1 - \eta(T_1, T_2) = \frac{f(T_2)}{f(T_1)}.$$

  For simplicity, we make the *choice* $f(T) = T$. (In statistical mechanics, this choice is forced by the definition $S = k_B \log \Omega$.)

- Under this choice, the Carnot cycle satisfies $Q_H/T_H + Q_C/T_C = 0$. Since any reversible process can be decomposed into infinitesimal Carnot cycles,

$$\oint \frac{đQ}{T} = 0$$

  for any reversible cycle. This implies that $\int đQ/T$ is independent of path, as long as we only use reversible paths, so we can define a state function

$$S(A) = \int_0^A \frac{đQ}{T}.$$

- Again using the Second Law, we have the Clausius inequality

$$\oint \frac{đQ}{T} \leq 0$$

  for any cycle. In particular, suppose we have an irreversible adiabatic path from A to B and a reversible path back. Then the Clausius inequality says $S(B) \geq S(A)$, which is the usual statement of the Second Law.

- The Third Law tells us that S/N goes to zero as T goes to zero; this means that heat capacities must go to zero. Another statement is that it takes infinitely many steps to get to $T = 0$ via isothermal and adiabatic processes.

- In statistical mechanics, the Third Law simply says that the log-degeneracy of the ground state can't be extensive. For example, in a system of N spins in zero field, one might think that the ground state has degeneracy 2^N. But in reality, arbitrarily weak interactions always break the degeneracy.

Note. Reversible and irreversible proceses. For reversible processes only, we have

$$đQ_{\text{rev}} = TdS, \quad đW_{\text{rev}} = -pdV.$$

For example, in the process of free expansion, the volume and entropy change, even though there is no heat or work. Now, for a reversible process the First Law gives $dE = TdS - pdV$. Since both sides are state functions, this must be true for *all* processes, though the individual terms will no longer describe heat or work! We'll ignore this subtlety below and think of all changes as reversible.

Example. We define the enthalpy, Helmholtz free energy, and Gibbs free energy as

$$H = U + PV, \quad F = U - TS, \quad G = U + PV - TS.$$

Then we have

$$dH = TdS + Vdp, \quad dF = -SdT - pdV, \quad dG = -SdT + Vdp.$$

From these differentials, we can read off the natural variables of these functions; when we change a state function by changing its natural variables, the change in the state function is easy to compute. To convert between the quantities, we use the Gibbs–Helmholtz equations

$$U = F - T\left(\frac{\partial F}{\partial T}\right)_V = -T^2\left(\frac{\partial (F/T)}{\partial T}\right)_V, \quad H = -T^2\left(\frac{\partial (G/T)}{\partial T}\right)_p.$$

Note. The potentials defined above have direct physical interpretations. Consider a system with $đW = -pdV + đW'$, where $đW'$ contains other types of work, such as electrical work supplied by a battery. Since $đQ \leq TdS$, the First Law gives

$$-pdV + đW' \geq dU - TdS.$$

If the process is carried out at constant volume, then $dF = dU - TdS$, so $đW' \geq dF$. Then the Helmholtz free energy represents the maximum amount of work that can be extracted at fixed temperature. If instead we fix the pressure, then $đW' \geq dG$, so the Gibbs free energy represents the maximum amount of non-pdV work that can be extracted.

The interpretation of enthalpy is different; at constant pressure, we have $dH = TdS = đQ_{\text{rev}}$, so changes in enthalpy tell us whether a chemical reaction is endothermic or exothermic.

Note. Deriving the Maxwell relations. Recall that area in the TS plane is heat and area in the pV plane is work. In a closed cycle, the change in U is zero, so the heat and work are equal,

$$\int dp\, dV = \int dT\, dS.$$

Since the cycle is arbitrary, this means that $dpdV = dTdS$, which can be thought of as an equality of differential forms; in terms of calculus, it means the Jacobian for changing variables from (p, V) to (T, S) is one. This trick can be used to reproduce all the Maxwell relations.

We now give some examples of problems using the Maxwell relations and partial derivative rules.

Example. The natural variables of U are S and V, because its partial derivatives with respect to these variables are simple. Other derivatives, such as $\partial U/\partial V|_T$, are complicated, though one can be deceived because it is simple (i.e. zero) for ideal gases. But generally we have

$$\left.\frac{\partial U}{\partial V}\right|_T = \left.\frac{\partial U}{\partial V}\right|_S + \left.\frac{\partial U}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T = -p + T\left.\frac{\partial p}{\partial T}\right|_V = \left.\frac{\partial (p/T)}{\partial T}\right|_V$$

where we used a Maxwell relation in the second equality. This is the simplest way of writing $\partial U/\partial V|_T$ in terms of thermodymamic variables.

Example. The ratio of isothermal and adiabatic compressibilities is

$$\frac{\kappa_T}{\kappa_S} = \frac{(\partial V/\partial p)|_T}{(\partial V/\partial p)|_S} = \frac{(\partial V/\partial T)|_p(\partial T/\partial p)|_V}{(\partial V/\partial S)|_p(\partial S/\partial p)||_V} = \frac{(\partial V/\partial T)|_p(\partial S/\partial V)|_p}{(\partial p/\partial T)|_V(\partial S/\partial p)||_V} = \frac{(\partial S/\partial T)|_p}{(\partial S/\partial T)|_V} = \gamma$$

where we used the triple product rule, the reciprocal rule, and the regular chain rule.

Example. The entropy for one mole of an ideal gas. We have

$$dS = \left(\frac{\partial S}{\partial T}\right)_V dT + \left(\frac{\partial S}{\partial V}\right)_T dV = \frac{C_V}{T} dT + \left(\frac{\partial p}{\partial T}\right)_V dV.$$

Using the ideal gas law, $(\partial p/\partial T)|_V = R/V$, and integrating gives

$$S = \int \frac{C_V}{T} dT + \int \frac{R}{V} dV = C_V \log T + R \log V + \text{const.}$$

where we can do the integration easily since the coefficient of dT doesn't depend on V, and vice versa. The singular behavior for $T \to 0$ is incomptible with the Third Law, as is the result $C_P = C_V + R$, as all heat capacities must vanish for $T \to 0$. These tensions are because Third Law is quantum mechanical, and they indicate the classical model of the ideal gas must break down.

Example. Work for a rubber band. Instead of $dW = -pdV$, we have $dW = fdL$, where f is the tension. Now, we have

$$\left(\frac{\partial S}{\partial L}\right)_T = -\left(\frac{\partial f}{\partial T}\right)_L = -\left(\frac{\partial f}{\partial L}\right)_T \left(\frac{\partial L}{\partial T}\right)_f$$

where we used a Maxwell relation, and both of the terms on the right are positive (rubber bands act like springs, and contract when cold). The sign can be understood microscopically: an expanding gas has more position phase space, but if we model a rubber band as a chain of molecules taking a random walk with a constrained total length, there are fewer microstates if the length is longer.

Next, using the triple product rule gives

$$\left(\frac{\partial S}{\partial T}\right)_L \left(\frac{\partial T}{\partial L}\right)_S > 0$$

and the first term must be positive by thermodynamic stability; therefore a rubber band heats up if it is quickly stretched, just the opposite of the result for a gas.

Example. Work for electric dipoles. The potential energy of a dipole in a field is

$$U_{\text{pot}} = -\mathbf{p} \cdot \mathbf{E}, \quad dU_{\text{pot}} = -\mathbf{p} \cdot d\mathbf{E} - \mathbf{E} \cdot d\mathbf{p}.$$

However, the dipole also has an internal energy (e.g. the masses can be connected by a spring), and it increases by $\mathbf{E} \cdot d\mathbf{p}$ when the dipole is stretched. Adding this contribution, we have

$$đW = -\mathbf{p} \cdot d\mathbf{E}$$

and similarly $đW = -\mathbf{m} \cdot d\mathbf{B}$ for magnetic dipoles. This 'extra' term is not entirely unfamiliar. For example, the 'potential energy' of a box of gas at pressure p and volume V is $U_{\text{pot}} = pV$. This is the energy associated with 'hollowing out' a space of volume V in an atmosphere with pressure p. Yet the increment of work is only pdV, not $pdV + Vdp$. The main difference between magnets and gases is that m decreases with temperature, while p increases; then cycles involving magnets in (m, B) space run opposite the analogous direction for gases.

Note. Chemical reactions. For multiple reactions, we get a contribution $\sum_i \mu_i dN_i$ to the energy. Now, consider an isolated system where some particle has no conservation law; then the amount N_i of that particle is achieved by minimizing the free energy, which sets $\mu = 0$. This is the case for photons in most situations. More generally, if chemical reactions can occur, then minimizing the free energy means that chemical potentials are balanced on both sides of the reaction.

Note. The Clausius-Clapeyron equation. At a phase transition, the chemical potentials of the two phases (per molecule) are equal. Now consider two nearby points on a coexistence curve in (p, T) space. If we connect these points by a path in the region with phase i, then

$$\Delta\mu_i = -s_i dT + v_i dP$$

where we used $\mu = G/N$, and s_i and v_i are the entropy and volume divided by the total particle nunmber N. Since we must have $\Delta\mu_1 = \Delta\mu_2$,

$$\frac{dP}{dT} = \frac{s_2 - s_1}{v_2 - v_1} = \frac{L}{T(V_2 - V_1)}.$$

This can also be derived by demanding that a heat engine running on a phase transition doesn't violate the Second Law.

Note. Insight into the Legendre transform. The Legendre transform of a function $F(x)$ is the function $G(s)$ satisfying

$$G(s) + F(x) = sx, \quad s = \frac{dF}{dx}$$

from which one may show that $x = dG/ds$. The symmetry of the above equation makes it clear that the Legendre transform is its own inverse. Moreover, the Legendre transform crucially requires $F(x)$ to be convex, in order to make the function $s(x)$ single-valued. It is useful whenever s is an easier parameter to control or measure than x.

However, the Legendre transforms in thermodynamics seem to come with some extra minus signs. The reason is that the fundamental quantity is entropy, not energy. Specifically, we have

$$F(\beta) + S(E) = \beta E, \quad \beta = \frac{\partial S}{\partial E}, \quad E = \frac{\partial F}{\partial \beta}.$$

That is, we are using β and E as conjugate variables, not T and S! Another hint of this comes from the definition of the partition function,

$$Z(\beta) = \int \Omega(E) e^{-\beta E}\, dE, \quad F(\beta) = -\log Z(\beta), \quad S(E) = \log \Omega(E)$$

from which we recover the above result by the saddle point approximation.

3.3 Classical Gases

We first derive the partition function by taking the classical limit of a noninteracting quantum gas.

Example. For each particle, we have the Hamiltonian $\hat{H} = \hat{p}^2/2m + V(\hat{q})$, where the potential confines the particle to a box. The partition function is defined as $Z = \operatorname{tr} e^{-\beta \hat{H}}$. In the classical limit, we neglect commutators,

$$e^{-\beta\hat{H}} = e^{-\beta\hat{p}^2/2m} e^{-\beta V(\hat{q})} + O(\hbar).$$

Taking the trace over the position degrees of freedom,

$$Z \approx \int dq\, e^{-\beta V(q)} \langle q | e^{-\beta\hat{p}^2/2m} | q\rangle = \int dq\, dp\, dp'\, e^{-\beta V(q)} \langle q|p\rangle \langle p| e^{-\beta\hat{p}^2/2m} |p'\rangle \langle p'|q\rangle.$$

Evaluating the p' integral, and using $\langle q|p\rangle = e^{ipq/\hbar}/\sqrt{2\pi\hbar}$, we find

$$Z = \frac{1}{h}\int dq\, dp\, e^{-\beta H(p,q)}$$

in the classical limit. Generically, we get integrals of $e^{-\beta H}$ over phase space, where h is the unit of phase space volume. The value of h won't affect our classical calculation, as it only affects Z by a multiplicative constant.

Next, we recover the properties of the classical ideal gas.

- For a particle in an ideal gas, the position integral gives a volume factor V. Performing the Gaussian momentum integrals,

$$Z = \frac{V}{\lambda^3}, \quad \lambda = \sqrt{\frac{2\pi\hbar^2}{mk_BT}}.$$

  The thermal de Broglie wavelength λ is the typical de Broglie wavelength of a particle. Then our expression for Z makes sense if we think of Z as the 'number of thermally accessible states', each of which could be a wavepacket of volume λ^3.

- For N particles, we have

$$Z = \frac{1}{N!}\frac{V^N}{\lambda^{3N}}.$$

  The factor of $N!$ comes from particle indistinguishability, which we discuss below. Using this, we can recover the ideal gas law and the internal energy, which obeys equipartition.

- The entropy of the ideal gas is

$$S = -\frac{\partial F}{\partial T} = \frac{\partial}{\partial T}(k_BT\log Z) = Nk_B\left[\log\frac{V}{N\lambda^3} + \frac{5}{2}\right]$$

  where we used Stirling's approximation and dropped sub-extensive terms. This is the Sackur-Tetrode equation. Note that while the entropy depends explicitly on h, the value of h is not detectable since only entropy differences can be measured classically.

- In the grand canonical ensemble, we have

$$\mathcal{Z} = \sum_N e^{\beta\mu N} Z(N) = \exp\left(\frac{e^{\beta\mu}V}{\lambda^3}\right).$$

  Then the particle number is

$$N = \frac{1}{\beta}\frac{\partial}{\partial\mu}\log\mathcal{Z} = \frac{e^{\beta\mu}V}{\lambda^3}, \quad \mu = k_BT\log\frac{\lambda^3 N}{V}.$$

  The chemical potential is thus negative, as the classical limit is valid for $\lambda^3 \ll V/N$.

- We can easily derive the velocity distribution F and speed distribution f,

$$F(\mathbf{v}) \propto e^{-mv^2/2k_BT} \quad f(v) \propto v^2 e^{-mv^2/2k_BT}.$$

One slick derivation is to assume that velocity in orthogonal directions should be iid, and F can only depend on the speed by rotational symmetry. Then

$$F(v) = \phi(v_x)\phi(v_y)\phi(v_z)$$

which has only one solution, $F(\mathbf{v}) \propto e^{-Av^2}$. However, this derivation is a *fake*, as it only works for a nonrelativistic gas.

Example. Gaseous reactions. At constant temperature, the chemical potential per mole of gas is

$$\mu(p) = \mu^\circ + RT\log(p/p^\circ)$$

where μ° is the chemical potential at standard pressure p°. For a reaction $A \leftrightarrow B$, we define the equilibrium constant as $K = p_B/p_A$. Then

$$\Delta G = \Delta G^\circ + RT\log\frac{p_B}{p_A} \quad \log K = -\frac{\Delta G^\circ}{RT}.$$

This result also holds for arbitrarily complicated reactions. Applying the Gibbs-Helmholtz relation,

$$\frac{d\log K}{dT} = \frac{\Delta H^\circ}{RT^2}$$

which is an example of Le Chatelier's principle.

Example. Counting degrees of freedom. A monatomic gas has three degrees of freedom; the atom has kinetic energy $(3/2)k_BT$. The diatomic gas has seven: the three translational degrees of freedom of the center of mass, the two rotations, and the vibrational mode, which counts twice due to the potential energy of the bond, but is frozen out at room temperature.

An alternating counting method is to simply assign $(3/2)k_BT$ kinetic energy to every atom; this is correct because the derivation of the monatomic gas's energy holds for each atom separately, in the moment it collides with another. The potential energy then adds $(1/2)k_BT$.

We now consider the effects of weak interactions.

- Corrections to the ideal gas law are often expressed in terms of a density expansion,

$$\frac{p}{k_BT} = \frac{N}{V} + B_2(T)\frac{N^2}{V^2} + B_3(T)\frac{N^3}{V^3} + \cdots$$

  where the $B_i(T)$ are called the virial coefficients.

- To calculate the coefficients, we need an ansatz for the interaction potential. We suppose the density is relatively low, so only pairwise interactions matter, so

$$H_{\text{int}} = \sum_{i<j} U(r_{ij}).$$

- If the atoms are neutral with no permanent dipole moment, they will have an attractive $1/r^6$ van der Waals interactions. Atoms will also have a strong repulsion at short distances; in the Lennard-Jones potential, we take it to be $1/r^{12}$ for convenience. In our case, we will take the even simpler choice of a hard core repulsion,

$$U(r) = \begin{cases} \infty & r < r_0 \\ -U_0(r_0/r)^6 & r \geq r_0. \end{cases}$$

- Performing the momentum integral as usual, the partition function is

$$Z(N,V,T) = \frac{1}{N!\lambda^{3N}} \int \prod_i d\mathbf{r}_i \, e^{-\beta \sum_{j<k} U(r_{jk})}.$$

It is tempting to expand in βU, but this doesn't work because U is large (infinite!). Instead we define the Mayer f function

$$f(r) = e^{-\beta U(r)} - 1$$

which is bounded here between -1 and 0. Then

$$Z(N,V,T) = \frac{1}{N!\lambda^{3N}} \int \prod_i d\mathbf{r}_i \prod_{j>k} (1 + f_{jk}).$$

An expansion of powers in f is thus more sensible. This is an expansion in 'perturbations to occupancy probabilities/densities' rather than perturbations to energies.

- The zeroth order term recovers V^N. The first order term gives

$$\int \prod_i d\mathbf{r}_i \sum_{j>k} f_{jk} \approx \frac{N^2}{2} V^{N-2} \int d\mathbf{r}_1 d\mathbf{r}_2 \, f(r_{12}) \approx \frac{N^2}{2} V^{N-1} \int d\mathbf{r} f(\mathbf{r})$$

where we integrated out the center of mass coordinate. We don't have to worry about bounds of integration on the $\mathbf{r}$ integral, as most of its contribution comes from atomic-scale r.

- Since $\overline{f} \sim r_0^3$, the ratio of the first and zeroth order terms goes as Nr_0^3/V, giving us a measure of what 'low density' means.

- Denoting the integral as $\overline{f}$, we find that to first order in f,

$$Z = \frac{V^N}{N!\lambda^{3N}} \left(1 + \frac{N^2 \overline{f}}{2V}\right) \approx \frac{V^N}{N!\lambda^{3N}} \left(1 + \frac{N\overline{f}}{2V}\right)^N$$

so that

$$F = F_{\text{ideal}} - Nk_BT \log(1 + N\overline{f}/2V) \approx F_{\text{ideal}} - \frac{N^2 k_B T}{2V} \overline{f}.$$

- Calculating the pressure as $p = -\partial F/\partial V$, we find

$$\frac{pV}{Nk_BT} = 1 - \frac{N\overline{f}}{2V}.$$

Evidently, we have computed the virial coefficient $B_2(T)$. Finding $\overline{f}$ explicitly yields the van der Waals equations of state.

- Note that the integral for $\overline{f}$ diverges if the potential falls off as $1/r^3$ or slower. These potentials are 'long-ranged'.

Higher order corrections can be found efficiently using the cluster expansion.

- Consider a generic term $O(f^E)$ term in the full expansion of Z above. Such a term can be represented by a graph G with N vertices and E edges, with no edges repeated. Denoting the value of a graph by $W[G]$, we have

$$Z = \frac{1}{N!\lambda^{3N}} \sum_G W[G].$$

- Each graph G factors into connected components called clusters, each of which contributes an independent multiplicative factor to $W[G]$.

- The most convenient way to organize the expansion is by the number and sizes of the clusters. Let U_l denote the contribution from all l-clusters,

$$U_l = \int \prod_{i=1}^{l} d\mathbf{r}_i \sum_{G \text{ is } l-\text{cluster}} W[G].$$

  Now consider the contributions of all graphs with m_l l-clusters, so that $\sum m_l l = N$. They have the value

$$\prod_l \frac{N!}{(l!)^m_l} \frac{U_l^{m_l}}{m_l!}.$$

  where the various factorials prevent overcounting within and between l-clusters.

- Summing over $\{m_l\}$, the partition function is

$$Z = \frac{1}{\lambda^{3N}} \sum_{\{m_l\}} \prod_l \frac{U_l^{m_l}}{(l!)^{m_l} m_l!}$$

  where the $N!$ factor has been canceled.

- The annoying part is the restriction $\sum m_l l = N$, which we eliminate by going to the grand canonical ensemble. Defining the fugacity $z = e^{\beta\mu}$, we have

$$\mathcal{Z}(\mu) = \sum_N z^n Z(N) = \sum_{\{m_l\}} \prod_l \frac{1}{m_l!} \left(\frac{z^l U_l}{\lambda^{3l} l!}\right)^{m_l} = \prod_l \exp\left(\frac{U_l}{z^l} \lambda^{3l} l!\right).$$

- Defining $b_l = (\lambda^3/V)(U_l/l!\lambda^{3l})$, our expression reduces to

$$\mathcal{Z}(\mu) = \exp\left(\frac{V}{\lambda^3} \sum_l b_l z^l\right).$$

  We see that if we take the log to get the free energy, only b_l appears, not higher powers of b_l. This reduces a sum over all diagrams to a sum over only connected diagrams. Expanding in powers of z allows us to find the virial coefficients.

3.4 Bose-Einstein Statistics

We now turn to bosonic quantum gases.

Note. Before continuing, it's useful to get the density of states. For independent particles in a box with periodic boundary conditions, the states are plane waves, leading to the usual $1/h^3$ density of states in phase space. Integrating out position and momentum angle, we have

$$g(k) = \frac{4\pi V}{(2\pi)^3} k^2.$$

Changing variables to energy using $dk = (dk/dE)dE$, for a nonrelativistic particle we find

$$g(E) = \frac{V}{4\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} E^{1/2}.$$

For a relativistic particle, the same procedure gives

$$g(E) = \frac{VE}{2\pi^2 \hbar^3 c^3} \sqrt{E^2 - m^2 c^4}.$$

In particular, for massless particles, we get

$$g(E) = \frac{VE^2}{2\pi^2 \hbar^3 c^3}.$$

In general, we should also multiply by the number of spin states/polarizations.

Example. Photons in blackbody radiation. Using $E = \hbar\omega$ and the fact that photons are bosons with two polarizations, the partition function for photons in a mode of frequency ω is

$$Z_\omega = 1 + e^{-\beta\hbar\omega} + e^{-2\beta\hbar\omega} + \ldots = \frac{1}{1 - e^{-\beta\hbar\omega}}.$$

Note that the number of photons is not fixed. We can imagine we're working in the canonical ensemble, but summing over states of the quantum field. Alternatively, we can imagine we're working in the grand canonical ensemble, where $\mu = 0$ since photon number is not conserved; instead the photon number sits at a minimum of the Gibbs free energy. In this case there are no extra combinatoric factors because photons are identical.

In any case, the entire partition function is

$$\log Z = \int_0^\infty d\omega\, g(\omega) \log Z_\omega = -\frac{V}{\pi^2 c^3} \int_0^\infty d\omega\, \omega^2 \log(1 - e^{-\beta\hbar\omega}).$$

The energy density is

$$E = -\frac{\partial}{\partial \beta} \log Z = \frac{V\hbar}{\pi^2 c^3} \int_0^\infty d\omega\, \frac{\omega^3}{e^{\beta\hbar\omega} - 1}$$

where the integrand is the Planck distribution. Taking the high T limit then recovers the Rayleigh-Jeans law, from equipartition. Now, to evaluate the integral, note that it has dimensions ω^4, so it must produce $1/(\beta\hbar)^4$. Then

$$E \propto V(k_B T)^4$$

which recovers the Stefan-Boltzmann law. To get other quantities, we can differentiate the free energy. One particularly important result is

$$p = \frac{E}{3V}$$

which is important in cosmology. One way to derive the constant is to note that the pressure from kinetic theory depends on pv, and pv is twice the kinetic energy for a nonrelativistic gas, but equal to the kinetic energy for a photon gas. Thus $pV = (1/2)(2E/3)$ for a photon gas.

Note. More on the thermodynamics of the photon gas. By considering an isochoric change,

$$dS = \frac{dE}{T} \propto VT^2 dT \quad S \propto VT^3$$

where the constant is zero by the Third Law. We also know that the typical energy of a photon is proportional to T, so $N \propto VT^3$. As a result, adiabatic expansion keeps particle number constant, and conserves pV^γ for $\gamma = 4/3$. Additionally, since $VT^3 \propto (LT)^3$ is conserved, the temperature is inversely proportional to the length; this appears as the cosmological redshift.

Note. Since $\gamma = 4/3$ for the photon gas, it seems we have six degrees of freedom. The catch is that the equipartition theorem only works for quadratic degrees of freedom; for a linear degree of freedom, as in $E = pc$, the contribution is twice as much, giving twice as many effective 'degrees of freedom' as for a monatomic gas. However, this analogy has limits: while a classical ultrarelativistic gas has energy $3Nk_BT$, this is not true for a quantum gas, and a photon gas is always quantum. We cannot hide the discreteness of the quantum states by raising the temperature, because the energy will always mostly come from photons with energy of order k_BT, so the mode occupancy numbers will always be order 1.

Note. Above, we've thought of every photon mode as a harmonic oscillator. To see this microscopically, note that $\mathbf{A}$ is the conjugate momentum to $\mathbf{E}$ and the energy is

$$H \sim \frac{1}{2}(E^2 + B^2) \sim \frac{1}{2}(E^2 + \omega^2 A^2)$$

where we worked in Coulomb gauge. This is then formally identical to a harmonic oscillator. The reason that $\mathbf{E}$ and $\mathbf{B}$ are in phase, rather than the usual 90° out of phase, is that $\mathbf{B}$ is a derivative of the true canonical variable $\mathbf{A}$.

Example. Phonons. The exact same logic applies for phonons in a solid, except that there are three polarization states, and the speed of light c is replaced with the speed of sound c_s. (That is, we are assuming the dispersion relation remains linear.) There is also a high-frequency cutoff ω_D imposed by the lattice.

To get a reasonable number for ω_D, note that the number of normal modes is equal to the number of degrees of freedom, so

$$\int_0^{\omega_D} d\omega\, g(\omega) = 3N$$

where N is the number of lattice ions. The partition function is very similar to the blackbody case. At low temperatures, the cutoff ω_D doesn't matter, so the integral is identical, and

$$E \propto T^4 \quad C \propto T^3.$$

At high temperatures, one can show that with the choice of ω_D above, we simply reproduce the Dulong-Petit law. The only problem with the Debye model is that the phonon dispersion relation isn't actually linear. This doesn't matter at very high or low temperatures, but yields slight deviations at intermediate ones.

Now we formally introduce the Bose-Einstein distribution. For convenience, we work in the grand canonical ensemble.

- Consider a configuration of particles where n_i particles are in state i, and $\sum_i n_i = N$. In the Maxwell-Boltzmann distribution, we treat the particles as distinguishable, then divide by $1/N!$ at the end, so the 'weight' of this configuration is

$$\frac{1}{N!}\binom{N}{n_1}\binom{N-n_1}{n_2}\cdots = \prod_i \frac{1}{n_i!}.$$

  In the Bose-Einstein distribution, we treat each configuration as one state of the quantum field, so all states have weight 1.

- As long as all of the n_i are zero or one (the classical limit), the two methods agree. However, once we introduce discrete quantum states, simply dividing by $1/N!$ no longer 'takes us from distinguishable to indistinguishable'. States in which some energy levels have multiple occupancy aren't weighted enough.

- Similarly, the Fermi-Dirac distribution also agrees with the classical result, as long as $\langle n_i \rangle \ll 1$.

- Another way of saying this is that in the classical case, we're imagining we can paint labels on all the particles; at the end we divide by $1/N!$ because the labels are arbitrary. This is an imperfect approximation to true indistinguishability, because when two particles get into the same state, we must lose track of the labels!

- For one single-particle quantum state $|r\rangle$, the Bose-Einstein partition function is

$$\mathcal{Z}_r = \sum_{n_r} e^{-\beta n_r (E_r - \mu)} = \frac{1}{1 - e^{-\beta(E_r - \mu)}}.$$

  Note that in the classical case, we would have also multiplied by $1/n_r!$. Without this factor, the sum might not converge, so we also demand $E_r > \mu$ for all E_r. Setting the ground state energy E_0 to zero, we require $\mu < 0$.

- The expected occupancy can be found by summing an arithmetic-geometric series, or noting

$$\langle n_r \rangle = \frac{1}{\beta}\frac{\partial}{\partial \mu} \log \mathcal{Z}_r = \frac{1}{e^{\beta(E_r - \mu)} - 1}.$$

  This equation is sometimes called the 'Bose-Einstein distribution'.

- Taking the product, the grand partition function is

$$\mathcal{Z} = \prod_r \frac{1}{1 - e^{\beta(E_r - \mu)}}$$

  where the product ranges over all single-particle states.

- Using the Bose-Einstein distribution, we can compute properties of the Bose gas,

$$N = \int dE \, \frac{g(E)}{z^{-1}e^{\beta E} - 1}, \quad E = \int dE \, \frac{Eg(E)}{z^{-1}e^{\beta E} - 1}$$

  where $z = e^{\beta\mu}$ is the fugacity. The stability requirement $\mu < 0$ means $z < 1$.

- To compute the pressure, note that

$$pV = \frac{1}{\beta} \log \mathcal{Z} = -\frac{1}{\beta} \int dE \, g(E) \log(1 - ze^{-\beta E}).$$

  In the nonrelativistic case, $g(E) \sim E^{1/2}$. Integrating by parts then shows

$$pV = \frac{2}{3}E$$

  which matches that of a classical monatomic gas. For comparison, we saw above that in the ultrarelativistic case we get $1/3$ instead.

- At high temperatures, we can compute the corrections to the ideal gas law by expanding in $z \ll 1$, finding

$$\frac{N}{V} = \frac{z}{\lambda^3} \left(1 + \frac{z}{2\sqrt{2}} + \cdots \right)$$

  To see why $z \ll 1$ is a high-temperature expansion, note that $z \sim \lambda^3 \sim T^{-3/2}$ here. Next, we can similarly expand the energy for

$$\frac{E}{V} = \frac{3z}{2\lambda^3\beta} \left(1 + \frac{z}{4\sqrt{2}} + \cdots \right).$$

  Combining these equations, we find the first correction to the ideal gas law,

$$pV = Nk_BT \left(1 - \frac{\lambda^3 N}{4\sqrt{2}V} + \ldots \right).$$

  The pressure is less; the physical intuition is that bosons 'like to clump up', since they're missing the $1/n_r!$ weights that a classical gas has.

Note. To get more explicit results, it's useful to define the functions

$$g_n(z) = \frac{1}{\Gamma(n)} \int_0^\infty dx \, \frac{x^{n-1}}{z^{-1}e^x - 1}.$$

To simplify this, expand the denominator as a geometric series for

$$g_n(z) = \frac{1}{\Gamma(n)} \sum_{m=1}^\infty \int dx \, x^{n-1} e^{-mx} z^m = \frac{1}{\Gamma(n)} \sum_{m=1}^\infty \frac{z^m}{m^n} \int_0^\infty du \, u^{n-1} e^{-u} = \sum_{m=1}^\infty \frac{z^m}{m^n}.$$

The $g_n(z)$ are monotonic in z, and we have

$$\frac{N}{V} = \frac{g_{3/2}(z)}{\lambda^3}, \quad \frac{E}{V} = \frac{3}{2} \frac{k_BT}{\lambda^3} g_{5/2}(z)$$

for the ideal Bose gas. Finally, for photon gases where $\mu = 0$ we use

$$g_n(1) = \zeta(n).$$

Useful particular values of the zeta function include

$$\zeta(2) = \frac{\pi^2}{6}, \quad \zeta(4) = \frac{\pi^4}{90}.$$

These results may be derived by evaluating

$$\int_{-\pi}^{\pi} dx\, |f(x)|^2$$

for $f(x) = x$ and $f(x) = x^2$, respectively, using direct integration and Fourier series.

We now use these results to investigate Bose-Einstein condensation.

- Consider low temperatures, which correspond to high z, and fix N. Since we have

$$\frac{N}{V} = \frac{g_{3/2}(z)}{\lambda^3}$$

  the quantity $g_{3/2}(z)$ must increase as λ^3 increases. However, we know that the maximum value of $g_{3/2}(z)$ is $g_{3/2}(1) = \zeta(3/2)$, so this is impossible below some critical temperature T_c.

- The problem is that, early on, we took the continuum limit and turned sums over states into integrals; this is a good approximation whenever the occupancy of any state is small. But for $T < T_c$, the occupancy of the ground state becomes macroscopically large!

- The ground state isn't counted in the integral because $g(0) = 0$, so we manually add it, for

$$\frac{N}{V} = \frac{g_{3/2}(z)}{\lambda^3} + n_0, \quad n_0 = \frac{1}{z^{-1} - 1}.$$

  Then for $T < T_c$, z becomes extremely close to one ($z \sim 1 - 1/N$), and the second term makes up for the first. In the limit $T \to 0$, all particles sit in the ground state.

- We say that for $T < T_c$, the system forms a Bose-Einstein condensate (BEC). Since the number of uncondensed particles in a BEC at fixed temperature is independent of the density, the equation of state of a BEC doesn't depend on the density.

- To explicitly see the phase transition behavior, note that for $z \to 1$, one can show

$$g_{3/2}(z) \approx \zeta(3/2) + A\sqrt{1-z} + \ldots.$$

  Applying the definition of T_c, we have

$$\left(\frac{T}{T_c}\right)^{3/2} - 1 \sim A\sqrt{1-z} - \frac{1}{N}\frac{1}{1-z}.$$

  Dropping all constants, switching to reduced temperature t, and letting $x = 1 - z$,

$$t \sim \sqrt{x} - \frac{1}{Nx}.$$

Since x is never zero, the function $t(x)$ (and hence $x(t)$) is perfectly analytic, and there is no phase transition. However, in the thermodynamic limit, we instead have

$$\lim_{N\to\infty} x(t) = \begin{cases} t^2 & t > 0 \\ 0 & t < 0 \end{cases}$$

which is nonanalytic, as it has a discontinuous second derivative.

- Differentiating the energy, we find the heat capacity is

$$C_V = \frac{dE}{dT} \sim \frac{g_{5/2}(z)}{\lambda^3} + \frac{1}{\lambda^3}\frac{dg_{5/2}}{dz}\frac{dz}{dT}.$$

  Then the derivative of the heat capacity depends on d^2z/dT^2, and is discontinuous at $t = 0$.

- Another way of characterizing the BEC transition is that it occurs when the chemical potential increases to the ground state energy, creating a formally divergent number of particles in it.

Note. In a gas where the particle number N is not conserved, particles are created or destroyed freely to maximize the entropy, setting the chemical potential μ to zero. For such a gas, Bose-Einstein condensation cannot occur. Instead, as the temperature is lowered, N goes to zero.

Note that if N is almost conserved, with N changing on a timescale T much greater than the thermalization time, then for times much less than T we can see a quasiequilibrium with nonzero μ. Also note that setting $\mu = 0$ formally makes N diverge if there are zero energy states. This infrared divergence is actually perfectly correct; for instance, a formally infinite number of photons are created in every single scattering event. This is not problematic since these photons cannot be detected.

3.5 Fermi-Dirac Statistics

Now we turn to fermions, which obey Fermi-Dirac statistics.

- Each single-particle quantum state $|r\rangle$ can be occupied by one or two particles, so

$$\mathcal{Z}_r = 1 + e^{-\beta(E_r-\mu)} \quad \langle n_r\rangle = \frac{1}{e^{\beta(E_r-\mu)}+1}.$$

  Our expression for n_r is called the Fermi-Dirac distribution; it differs from the Bose-Einstein distribution by only a sign. Since there are no convergence issues, μ can be positive.

- Our expression for N, E, and pV are almost identical to the Bose gas case, again differing by a few signs. As before, we have $pV = (2/3)E$. The extra minus signs result in a first-order increase in pressure over that of a classical gas at high temperatures.

- In the low-temperature limit, the Fermi-Dirac distribution becomes

$$n(E) = \theta(\mu - E).$$

  All states with energies up to the Fermi energy E_F are filled, where in this case E_F is just equal to the chemical potential. These filled states form the 'Fermi sea' or 'Fermi sphere', and its boundary is the Fermi surface. The quantity E_F can be quite high, with the corresponding temperature $T_F = E_F/k_B$ at around 10^4 K for metals and 10^7 K for white dwarfs.

- The total energy is
$$E = \int_0^{E_f} dE\, Eg(E) = \frac{3}{5} N E_F$$
and the pressure is
$$pV = \frac{1}{\beta}\log \mathcal{Z} = \frac{1}{\beta}\int dE\, g(E) \log(1+e^{-\beta(E-\mu)}) = \int_0^{E_f} dE\,(\mu - E) g(E) = \frac{2}{5} N E_F.$$
This zero-temperature pressure is called the degeneracy pressure.

- Next, consider the particle number and energy density near zero temperature,
$$N = \int_0^\infty dE\, \frac{g(E)}{z^{-1}e^{\beta E}+1}, \quad E = \int_0^\infty dE\, \frac{Eg(E)}{z^{-1}e^{\beta E}+1}$$
where $g(E)$ is the density of states. We look at how E and μ depend on T, holding N fixed.

- First we claim that $d\mu/dT = 0$ at $T = 0$. We know that if μ is fixed, $\Delta N \sim T^2$, as the Fermi-Dirac distribution spreads out symmetrically about $E = E_F$. But if $d\mu/dT \neq 0$, then $\Delta N \sim T$ as the Fermi surface shifts outward, so we cannot have $\Delta N = 0$.

- For higher temperatures, μ should decrease, as we know it becomes negative as we approach the ideal gas. In $d = 2$, μ is exponentially rather than quadratically suppressed because the density of states is constant.

- Next, consider the change in energy. Since $dN/dT = 0$, the only effect is that k_BT/E_F of the particles are excited by energy on the order of k_BT. Then $\Delta E \sim T^2$, so $C_V \sim T$.

- Therefore, the low-temperature specific heat of a metal goes as
$$C_V = \gamma T + \alpha T^3$$
where the second term is from phonons. We can test this by plotting C_V/T against T^2. The linear contribution is only visible at very low temperatures.

Note. The classical limit. Formally, both the Fermi-Dirac and Bose-Einstein distributions approach the Maxwell-Boltzmann distribution in the limit of low occupancy numbers,
$$\frac{E-\mu}{T} \ll 1.$$
Since this is equivalent to $T \gg E - \mu$, it is sometimes called the low temperature limit, but this is deceptive; it would be better to call it the 'high energy limit'. Specifically, the high energy tail of a Bose or Fermi gas always behaves classically. But at low temperature Bose and Fermi gases look 'more quantum' as a whole.

Note. The chemical potential is a bit trickier when the energy levels are discrete, since it can't be defined by a derivative; it is instead defined by fixing N. It can be shown that in the zero temperature limit, the chemical potential is the average of the energies of the highest occupied state and the lowest unoccupied state. This ensures that N is fixed upon turning in a small T. In particular, it holds even if these two states have different degeneracies, because the adjustment in μ needed to cancel this effect is exponentially small.

Note. We can establish the above results quantitatively with the Sommerfeld expansion. Define

$$f_n(z) = \frac{1}{\Gamma(n)} \int_0^\infty dx\, \frac{x^{n-1}}{z^{-1}e^x + 1}$$

which are the fermionic equivalent of the g_n functions. Then

$$\frac{N}{V} = \frac{g_s}{\lambda^3} f_{3/2}(z), \quad \frac{E}{V} = \frac{3}{2}\frac{g_s}{\lambda^3} k_B T f_{5/2}(z)$$

where we plugged in the form of $g(E)$, and g_s is the number of spin states. We want to expand the $f_n(z)$ at high z. At infinite z, the integrands are just $x^{n-1}\theta(\beta\mu - x)$, so the integral is $(\beta\mu)^n/n$.

For high z, the integrands still contain an approximate step function. Then it's convenient to peel off the difference from the step function by splitting the integral into two pieces,

$$\Gamma(n) f_n(z) = \int_0^{\beta\mu} dx\, x^{n-1}\left(1 - \frac{1}{1 + ze^{-x}}\right) + \int_{\beta\mu}^\infty dx\, \frac{x^{n-1}}{z^{-1}e^x + 1}.$$

The first term simply reproduces the infinite temperature result. Now, the deviations above and below $\beta\mu$ tend to cancel each other, as we saw for dN/dT above. Then it's useful to subtract them against each other; defining $\eta = \beta\mu - x$ and $\eta = x - \beta\mu$ respectively, we get

$$\Gamma(n) f_n(z) = \frac{(\log z)^n}{n} + \int_0^\infty d\eta\, \frac{(\beta\mu + \eta)^{n-1} - (\beta\mu - \eta)^{n-1}}{1 + e^\eta}$$

where we extended a limit of integration from $\beta\mu$ to ∞, incurring an exponentially small $O(z^{-1})$ error. Taylor expanding to lowest order in $\beta\mu$ gives

$$\Gamma(n) f_n(z) = \frac{(\log z)^n}{n} + 2(n-1)(\log z)^{n-2} \int_0^\infty d\eta \frac{\eta}{e^\eta + 1}.$$

This integral can be done by expanding the denominator as a geometric series in $e^{-\eta}$. Termwise integration gives the series $\sum (-1)^{m+1}/m^2 = (1/2)\sum 1/m^2 = \pi^2/12$, giving the final result

$$f_n(z) = \frac{(\log z)^n}{\Gamma(n+1)}\left(1 + \frac{\pi^2}{6}\frac{n(n-1)}{(\log z)^2} + \cdots\right).$$

By keeping more terms in the Taylor expansion, we get a systematic expansion in $1/\log z = 1/\beta\mu$.

Applying the expansion to N/V, we immediately find

$$\Delta N \sim \left(\frac{k_B T}{\mu}\right)^2$$

which shows that, to keep N constant,

$$\Delta\mu \sim \left(\frac{k_B T}{E_F}\right)^2$$

as expected earlier. Similarly, the first term in ΔE goes as T^2, giving a linear heat capacity.

Example. Pauli paramagnetism. Paramagnetism results from dipoles aligning with an external field, and Pauli paramagnetism is the alignment of spin. In a field B, electrons have energy

$$E = \mu_B B s, \quad s = \pm 1, \quad \mu_B = \frac{|e|\hbar}{2mc}$$

where μ_B is the Bohr magneton. Then the occupancy numbers are

$$\frac{N_\uparrow}{V} = \frac{1}{\lambda^3} f_{3/2}(z e^{\beta \mu_B B}), \quad \frac{N_\downarrow}{V} = \frac{1}{\lambda^3} f_{3/2}(z e^{-\beta \mu_B B}).$$

The resulting magnetization is

$$M = \mu_B (N_\uparrow - N_\downarrow).$$

In the high-temperature limit, z is small and $f_{3/2}(z) \approx z$, so

$$M = \frac{2\mu_B V z}{\lambda^3} \sinh(\beta \mu_B B) = \mu_B N \tanh(\beta \mu_B B)$$

where $N = N_\uparrow + N_\downarrow$. This is simply the classical result, as given by Maxwell-Boltzmann statistics. One important feature is that the susceptibility $\chi = \partial M / \partial B$ goes as $1/T$, i.e. Curie's law.

In the low-temperature limit, we take the leading term in the Sommerfeld expansion, then expand to first order in B, for

$$M = \mu_B^2 g(E_F) B.$$

Then at low temperatures, the susceptibility no longer obeys Curie's law, but instead saturates to a constant. To understand this result, note that only $g(E_F)\Delta E = g(E_F)\mu_B B$ electrons are close enough to the Fermi surface to participate, and they each contribute magnetization μ_B.

Note. Landau diamagnetism. Classically, charged particles exhibit diamagnetism because they begin moving in circles when a magnetic field is turned on, creating an opposing field. However, this explanation isn't completely right because of the Bohr-van Leeuwen theorem; the canonical partition function Z does not depend on the external field, as can be seen by shifting $\mathbf{p} - e\mathbf{A}$ to $\mathbf{p}$ in the integral.

Physically, this is because the particles must be in a finite box, say with reflecting walls. Then the particles whose orbits hit the walls effectively orbit backwards. Since the magnetic moment is proportional to the area, this cancels the magnetic moment of the bulk exactly. This is a significantly trickier argument. It is much easier to simply calculate Z and then differentiate the free energy, since Z itself is less sensitive to the boundary conditions.

In quantum mechanics, this argument does not hold. The partition function isn't an integral, so the first argument fails; we will instead find nontrivial dependence of Z on the field. In terms of the energy levels, electron states near the boundary are much higher energy due to the repulsive potential, so they are less relevant, though this is a bit difficult to see.

Note. The Euler summation formula is

$$\sum_{n=0}^{\infty} h(n + 1/2) = \int_0^\infty h(x)\, dx + \frac{1}{24} h'(0) + \ldots.$$

The idea behind the Euler summation formula is that one can approximate a smooth function by a low-order polynomial (or a Taylor series with decreasing coefficients). To see the origin of the first term, consider the formula for a unit interval,

$$h(1/2) \approx \int_0^1 h(x)\, dx + \ldots.$$

There is no correction term if $h(x)$ is a first-order polynomial. The correction due to second-degree terms in $h(x)$ can be found by subtracting $h'(x)$ at the endpoints,

$$h(1/2) \approx \int_0^1 h(x)\,dx + c(h'(0) - h'(1)) + \ldots.$$

To find the value of c, consider $h(x) = (x - 1/2)^2$, which fixes $c = 1/24$. Telescoping the sum gives the $h'(0)/24$ term in the formula above. Generally, all higher correction terms will have odd derivatives, because terms like $(x - 1/2)^{2n+1}$ don't contribute to the area.

Example. An explicit calculation of Landau diamagnetism. When the electrons are constrained to the xy plane, they occupy Landau levels with

$$E = \left(n + \frac{1}{2}\right)\hbar\omega_c, \quad \omega_c = \frac{eB}{m}$$

with degeneracy

$$N = \frac{\Phi}{\Phi_0}, \quad \Phi = L^2 B, \quad \Phi_0 = \frac{2\pi\hbar c}{e}.$$

Allowing the electrons to move in the third dimension gives an energy contribution $\hbar^2 k_z^2/2m$. Then the grand partition function is

$$\log \mathcal{Z} = \frac{L}{2\pi}\int dk_z \sum_{n=0}^{\infty} \frac{2L^2 B}{\Phi_0} \log\left[1 + z\exp\left(-\frac{\beta\hbar^2 k_z^2}{2m} - \beta\hbar\omega_c(n + 1/2)\right)\right]$$

where we added a factor of 2 to account for the spin sum, and converted the k_z momentum sum into an integral. Now we apply the Euler summation formula with the choice

$$h(x) = \int dk_z\, \log\left[1 + \exp\left(-\frac{\beta\hbar^2 k_z^2}{2m} + \beta x\right)\right].$$

Then our grand partition function becomes

$$\log \mathcal{Z} = \frac{VB}{\pi\Phi_0}\sum_{n=0}^{\infty} h(\mu - \hbar\omega_c(n + 1/2)).$$

The magnetization response comes from the first correction term, and we find

$$M = \frac{1}{\beta}\frac{\partial(\log \mathcal{Z})}{\partial B} = -\frac{\mu_B^2}{3} g(E_f) B$$

showing a diamagnetic effect comparable to the size of the Pauli paramagnetism. Since the paramagnetic effect is larger, one might expect that every solid is paramagnetic. However, there are also diamagnetic effects from electrons in core orbitals, so that in many heavy elements diamagnetism wins out.

4 Fundamentals of Quantum Mechanics

4.1 Mathematical Preliminaries

We postulate the state of a system at a given time is given by a ray in a Hilbert space $\mathcal{H}$.

- A Hilbert space is complex vector spaces with a positive-definite sesquilinear form $\langle\alpha|\beta\rangle$. Elements of $\mathcal{H}$ are called kets, while elements of the dual space $\mathcal{H}^*$ are called bras. Using the form, we can canonically identify $|\alpha\rangle$ with the bra $\langle\alpha|$, analogously to raising and lowering indices. This is an antilinear map, i.e. $c|\alpha\rangle \leftrightarrow \bar{c}\langle\alpha|$, since the form is sesquilinear.

- A ray is a ket up to the equivalence relation $|\psi\rangle \sim c|\psi\rangle$ for any nonzero complex number c, indicating that global phases in quantum mechanics at not important.

- Hilbert spaces are also complete, i.e. every Cauchy sequence of kets converges in $\mathcal{H}$.

- A Hilbert space V is separable if it has a countable subset D so that $\overline{D} = V$, which turns out to be equivalent to having a countable orthonormal basis. Hilbert spaces that aren't separable are mathematically problematic, so we'll usually assume this separability.

- If $\{|\phi_i\rangle\}$ is an orthonormal basis, then we have the completeness relation

  $$\sum_i |\phi_i\rangle\langle\phi_i| = 1.$$

  We also have the Schwartz equality,

  $$\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2.$$

  The trick to the proof is to use $\langle\gamma|\gamma\rangle \geq 0$ for $|\gamma\rangle = |\alpha\rangle + \lambda|\beta\rangle$, with $\lambda = -\langle\beta|\alpha\rangle/\langle\beta|\beta\rangle$.

Example. Finite-dimensional Hilbert spaces are all complete and separable. We will mostly deal with countably infinite-dimensional Hilbert spaces, like the QHO basis $|n\rangle$. Such spaces are separable, though uncountably infinite-dimensional spaces are not.

Example. Not all countably infinite-dimensional spaces are complete: consider the space V of infinite vectors with a finite number of nonzero entries. Then the sequence

$$(1, 0, \ldots), (1, 1/2, 0, \ldots), (1, 1/2, 1/3, 0, \ldots), \ldots$$

is Cauchy but does not converge in V.

Next, we consider operators on the Hilbert space.

- Given an operator $A\colon V \to W$, we may define the pullback operator $A^*\colon W^* \to V^*$ by

  $$A^*(\langle\beta|)|\alpha\rangle = \langle\beta|(A|\alpha\rangle).$$

  Since we can always do this, we'll write A^* as A, and write both of the above quantities as $\langle\beta|A|\alpha\rangle$ by letting A^* act to the left on bras. Note that this definition is independent of the existence of an inner product.

- Take $V = W$, and define the Hermitian adjoint $A^\dagger$ of A by
$$A^\dagger|\alpha\rangle \leftrightarrow \langle\alpha|A$$
where the $\leftrightarrow$ means the bra/ket correspondence from the inner product. Writing the bra associated with $A|\alpha\rangle$ as $\langle A\alpha|$, the above condition says $\langle A^\dagger\alpha| = \langle\alpha|A$. In more standard linear algebra notation, it means $\langle A^\dagger w, v\rangle = \langle w, Av\rangle$.

- Applying the definition of the adjoint,
$$\langle\alpha|A^\dagger|\beta\rangle = (\langle\beta|A|\alpha\rangle)^*.$$
Since the adjoint of a number is its complex conjugate, this suggests how to take the adjoint of a general expression: conjugate all numbers, flip bras and kets, and take adjoints of operators.

- We say A is Hermitian if $A = A^\dagger$, plus certain technical conditions which we'll tacitly ignore. Observables are Hermitian operators on $\mathcal{H}$, though not all Hermitian operators are observables, as we'll see below.

- We say A is an isometry if $A^\dagger A = 1$, so that A preserves inner products, and A is unitary if it is an invertible isometry, $A^\dagger = A^{-1}$, which implies $AA^\dagger = 1$ as well.

Example. A subtle example in an uncountably infinite-dimensional space. Consider functions on $[a, b]$ with basis $|x\rangle$. Then $K = -i(d/dx)$ appears Hermitian based on matrix elements:
$$K_{xx'} = -i\delta'(x - x'), \quad K^*_{x'x} = (-i\delta'(x - x'))^* = +i\delta'(x' - x) = -i\delta'(x - x').$$
However, K is not Hermitian! It does not satisfy the condition
$$\langle g|K|f\rangle = \langle f|K|g\rangle^*.$$
If we try to verify this equation, by inserting $1 = \int dx|x\rangle\langle x|$ and integrating, we see it's off by a surface term. To see why, consider using the finite difference, so that $f'(x) \sim f(x + 1) - f(x - 1)$. Then the matrix elements of K look like
$$K \propto \begin{pmatrix} & 1 & & & \\ 0 & & 1 & & \\ & -1 & & \ldots & \\ & & \ldots & & 0 \\ & & & -1 & \end{pmatrix}$$
At the endpoints, K is "missing" matrix elements because derivatives there are only defined from one side. However, K is Hermitian if we restrict to the subspace of functions that vanish at a and b, chopping off the last rows and columns of the matrix.

Example. Not all isometries are unitary: if $|n\rangle$ is an orthonormal basis with $n \in \mathbb{Z}$, the shift operator $A|n\rangle = |n + 1\rangle$ is an isometry, but $AA^\dagger \neq 1$.

- The spectral theorem states that if $A = A^\dagger$, then all eigenvalues of A are real, and all eigenspaces with distinct a_i are orthogonal. If the space is separable, every eigenspace has finite dimension, so we can construct an orthonormal eigenbasis by Gram-Schmidt.

- An operator A is a projection if $A^2 = A$. For example, $A = |\alpha\rangle\langle\alpha|$ is a projection if $\langle\alpha|\alpha\rangle = 1$.
- A basis $|\phi_i\rangle$ is complete if $\sum_i |\phi_i\rangle\langle\phi_i| = 1$. The sum of projections is the identity.
- Given a complete orthonormal basis, we can decompose operators and vectors into matrix elements. For example,

$$A = \sum_{i,j} |\phi_i\rangle\langle\phi_i|A|\phi_j\rangle\langle\phi_j| \sim \begin{pmatrix} \langle\phi_1|A|\phi_1\rangle & \langle\phi_1|A|\phi_2\rangle & \dots \\ \langle\phi_2|A|\phi_1\rangle & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix}.$$

  In this matrix notation, $A^\dagger$ is the conjugate transpose of A.

Example. If we consider infinite-dimensional spaces, not all Hermitian operators have a complete eigenbasis. Let $\mathcal{H} = \mathcal{L}^2([0,1])$ and let $A = \hat{x}$. Then A has no eigenvectors in $\mathcal{H}$!

This is worrying, because we physically prefer observables with a complete eigenbasis.

- We say A is bounded if

$$\sup_{|\alpha\rangle \in \mathcal{H}/\{|0\rangle\}} \frac{\langle\alpha|A|\alpha\rangle}{\langle\alpha|\alpha\rangle} < \infty.$$

  We say A is compact if every bounded sequence $\{|\alpha_n\rangle\}$ (with $\langle\alpha_n|\alpha_n\rangle < \beta$ for some fixed β) has a subsequence $\{|\alpha_{n_k}\rangle\}$ so that $\{A|\alpha_{n_k}\rangle\}$ is norm-convergent in $\mathcal{H}$.
- One can show that if A is compact, A is bounded. Compactness is sufficient for a Hermitian operator to be complete, but boundedness is neither necessary not sufficient. However, we will still consider observables that are neither bounded nor compact, when it turns out to be useful.
- If $|a_i\rangle$ and $|b_i\rangle$ are two complete orthonormal bases, then U defined by $U|a_i\rangle = |b_i\rangle$ is unitary. This yields the change of basis formula,

$$X = X_{ij}|a_i\rangle\langle a_j| = Y_{kl}|b_k\rangle\langle b_l|, \quad X_{ij} = U_{ik}Y_{kl}U^\dagger_{lj}.$$

- Using the above formula, a finite-dimensional Hermitian matrix can always be diagonalized by a unitary, i.e. a matrix that changes basis to an orthonormal eigenbasis.
- If A and B are diagonalizable, they are simultaneously diagonalizable iff $[A, B] = 0$, in which case we say A and B are compatible. The forward direction is easy. For the converse, let $A|\alpha_i\rangle = a_i|\alpha_i\rangle$. Then $AB|\alpha_i\rangle = a_iB|\alpha_i\rangle$ so B preserves A's eigenspaces. Therefore when A is diagonalized, B is block diagonal, and we can make B diagonal by diagonalizing within each eigenspace of A.

We gather here for later reference a few useful commutator identities.

Prop. The Hadamard lemma. For operators A and B, we have

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots.$$

Intuitively, this is simply the adjoint action of A on B, which infinitesimally is the commutator $[A, B]$. Therefore the operation of e^A on B must be the exponential of the commutator operation. Defining $\mathrm{ad}_A(B) = [A, B]$, this means

$$e^A B e^{-A} = e^{\mathrm{ad}_A} B$$

which is exactly the desired identity. The more straightforward way of proving this is to define

$$F(\lambda) = e^{\lambda A} B e^{-\lambda A}$$

and finding a differential equation for F; this is fundamentally the same idea in different notation.

Prop. Glauber's theorem. If $[A, B]$ commutes with both A and B, then

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A, B]\right).$$

To see this, define

$$F(\lambda) = e^{\lambda A} e^{\lambda B}, \quad F'(\lambda) = (A + e^{\lambda A} B e^{-\lambda A}) F(\lambda).$$

However, using the previous theorem, we have

$$F'(\lambda) = (A + B + \lambda[A, B]) F(\lambda).$$

We therefore guess the solution

$$F(\lambda) = \exp\left(\lambda(A + B) + \frac{\lambda^2}{2}[A, B]\right)$$

This solution satisfies the differential equation as long as the argument of the exponential commutes with its derivative, which we can quickly verify. Setting $\lambda = 1$ gives the result.

Prop. A special case of Baker-Campbell-Haussdorf. If $[A, B] = cI$, then

$$e^{A+B} = e^B e^A e^{c/2} = e^A e^B e^{-c/2}.$$

This tells us how to multiply things that 'almost commute'.

4.2 The Postulates of Quantum Mechanics

We are now ready to state the postulates of quantum mechanics.

1. The state of a system at time t is given by a ray in a Hilbert space $\mathcal{H}$. By convention, we normalize states to unit norm.

2. Observable quantities correspond to Hermitian operators whose eigenstates are complete. These quantities may be measured in experiments.

3. A observable H called the Hamiltonian defines time evolution by

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = H|\psi(t)\rangle.$$

4. If an observable A is measured when the system is in a state $|\alpha\rangle$, where A has an orthonormal basis of eigenvectors $|\alpha_i\rangle$ with eigenvalues a_i, the probability of observing $A = a$ is

$$\sum_{a_j = a} |\langle a_j|\alpha\rangle|^2 = \langle\alpha|P_a|\alpha\rangle, \quad P_a = \sum_{a_j = a} |a_j\rangle\langle a_j|.$$

   After this occurs, the (unnormalized) state of the system is $P_a|\alpha\rangle$.

5. If two individual systems have Hilbert spaces $\mathcal{H}^{(i)}$ with orthonormal bases $|\phi_n^{(i)}\rangle$, then the composite system describing both of them has Hilbert space $\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$, with orthonormal basis $|\phi_{ij}\rangle = |\phi_i^{(1)}\rangle \otimes |\phi_j^{(2)}\rangle$. An operator A on $\mathcal{H}^{(1)}$ is promoted to $A \otimes I$, and so on.

The fourth postulate implies the state of a system can change in an irreversible, discontinuous way. There are other formalisms that do not have this feature, though we'll take it as truth here.

Example. Let all eigenvalues of A be nondegenerate. Then if $|\alpha\rangle = \sum c_i |a_i\rangle$, the probability of observing $A = a_i$ is $|c_i|^2$, and the resulting state is $|a_i\rangle$. The expectation value of A is

$$\langle A\rangle = \sum |c_i|^2 a_i = \langle \alpha|A|\alpha\rangle.$$

Example. Spin 1/2. The Hilbert space is two-dimensional, and the operators that measure spin about each axis are

$$S_i = \frac{\hbar}{2}\sigma_i, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -1 \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

For a general axis $\hat{n}$, the operator $\vec{S}\cdot\hat{n}$ has eigenvalues $\pm\hbar/2$, and measures spin along the $\hat{n}$ direction. The commutation relations are

$$[S_i, S_j] = i\hbar\epsilon_{ijk}S_k, \quad \{S_i, S_j\} = \frac{1}{2}\hbar^2\partial_{ij}, \quad S^2 = \frac{3\hbar^2}{4}$$

so that S^2 commutes with S_i.

Example. The uncertainty principle. For an observable A and state $|\alpha\rangle$, define $\Delta A = A - \langle A\rangle$. Then the dispersion (variance) of A is

$$\langle \Delta A^2\rangle = \langle A^2\rangle - \langle A\rangle^2.$$

Now note that for observables A and B,

$$\langle\alpha|\Delta A^2|\alpha\rangle\langle\alpha|\Delta B^2|\alpha\rangle \geq |\langle\alpha|\Delta A \Delta B|\alpha\rangle|^2$$

by the Schwartz inequality. Note that we can write

$$\Delta A \Delta B = \frac{1}{2}\left([\Delta A, \Delta B], \{\Delta A, \Delta B\}\right).$$

These two terms are skew-Hermitian and Hermitian, so their expectation values are imaginary and real, respectively. Then we have

$$\langle \Delta A^2\rangle\langle \Delta B^2\rangle \geq \frac{1}{4}\left(|\langle[A, B]\rangle|^2 + |\langle\{\Delta A, \Delta B\}\rangle|^2\right).$$

Ignoring the second term gives

$$\sigma_A\sigma_B \geq \frac{1}{2}|\langle[A, B]\rangle|$$

where σ_X is the standard deviation. This is the uncertainty principle.

We now review position and momentum operators for particles on a line.

- The state of a particle on a line is an element of the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$, the set of square integrable functions on $\mathbb{R}$. This space is separable, and hence has a countable basis.

- Typical observables in this space include the projections,

$$(P_{[a,b]}f)(x) = f(x) \text{ for } a \leq x \leq b, 0 \text{ otherwise.}$$

However, this approach is physically inconvenient because most operators of interest (e.g. $\hat{x}$, $\hat{p} = -i\hbar\partial_x$) cannot be diagonalized in $\mathcal{H}$, as their eigenfunctions would not be normalizable.

- We will treat all of these operators as acceptable, and formally include their eigenvectors, even if they are not in $\mathcal{H}$. This severely enlarges the space under consideration, because x and p have uncountable eigenbases while the original space had a countable basis. Physically, this is not be a problem because all physical measurements of x are 'smeared out' and not infinitely precise. Thus the observables we actually measure do live in $\mathcal{H}$, and x is just a convenient formal tool.

- To begin, let $|x\rangle$ with $x \in \mathbb{R}$ be a complete orthonormal eigenbasis for x, with

$$\hat{x}|x\rangle = x|x\rangle, \quad \langle x'|x\rangle = \delta(x' - x), \quad \int dx|x\rangle\langle x| = 1.$$

Using completeness,

$$|\psi\rangle = \int dx\,|x\rangle\langle x|\psi\rangle = \int dx\,\psi(x)|x\rangle.$$

The quantity $\psi(x) = \langle x|\psi\rangle$ is called the wavefunction.

- If we take $[\hat{x}, \hat{p}] = i\hbar$, motivated by classical Poisson brackets, we uniquely specify $\hat{p}$ up to isomorphism as

$$\hat{p} = -i\hbar\frac{\partial}{\partial x}$$

by the Stone-von Neumann theorem. Let $|p\rangle$ be an orthonormal basis for $\hat{p}$,

$$\hat{p}|p\rangle = p|p\rangle.$$

Hence we may define a momentum space wavefunction, and the commutation relation immediately yields the Heisenberg uncertainty principle $\sigma_x\sigma_p \geq \frac{\hbar}{2}$.

- We can relate the $|x\rangle$ and $|p\rangle$ bases by noting that

$$-i\hbar\partial_x\langle x|p\rangle = p\langle x|p\rangle, \quad \langle x|p\rangle = Ne^{ipx/\hbar}.$$

Here, we acted with p to the left on $\langle x|$. To normalize, note that

$$\langle p|p'\rangle = \int dx\,\langle p|x\rangle\langle x|p'\rangle = |N|^2\int dx\,e^{ix(p'-p)/\hbar} = |N|^2(2\pi\hbar)\delta(p - p').$$

Therefore, we conclude

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar}$$

where we set an arbitrary phase to one.

Example. The momentum basis is complete if the position basis is. Insert the identity twice for

$$\int dp\,|p\rangle\langle p| = \int dx dx' dp\,|x\rangle\langle x|p\rangle\langle p|x'\rangle\langle x'| = \int dx dx' dp\,|x\rangle\frac{e^{ip(x'-x)/\hbar}}{2\pi\hbar}\langle x'| = \int dx\,|x\rangle\langle x|.$$

Then if one side is the identity, so is the other.

Example. The momentum-space wavefunction $\phi(p) = \langle p|\psi\rangle$ is related to $\psi(x)$ by Fourier transform,

$$\phi(p) = \int dx\,\langle p|x\rangle\langle x|\psi\rangle = \frac{1}{\sqrt{2\pi\hbar}}\int dx\,e^{-ipx/\hbar}\psi(x), \quad \psi(x) = \frac{1}{\sqrt{2\pi\hbar}}\int dp\,e^{ipx/\hbar}\phi(p).$$

This is the main place where conventions may differ. The original factor of 2π comes from the representation of the delta function

$$\delta(x) = \int d\xi\,e^{2\pi i x\xi}.$$

When defining the momentum eigenstates, we have freedom in choosing the scale of p, which can change the $\langle x'|p'\rangle$ expression above. This allows us to move the 2π factor around. In field theory texts, we prefer define momentum integrals to have a differential of the form $d^k p/(2\pi)^k$.

We now cover some facts about one-dimensional wave mechanics.

- Setting $\hbar = 2m = 1$, the Schrodinger equation is

  $$-\psi'' + V\psi = E\psi.$$

  Consider two degenerate solutions ψ and ϕ. Then combining the equations gives

  $$\phi\psi'' - \psi\phi'' = 0 = \frac{dW}{dx}$$

  where W is the Wronskian of the solutions,

  $$W = \phi\psi' - \psi\phi' = \det\begin{pmatrix}\phi & \psi \\ \phi' & \psi'\end{pmatrix}.$$

  In general, the Wronskian determines the independence of a set of solutions of a differential equation; if it is zero the solutions are linearly dependent.

- In this case, if both ψ and ϕ vanish at some point, then $W = 0$ so the solutions are simply multiples of each other. In particular, bound state wavefunctions vanish at infinity, so bound states are not degenerate. Unbound states can be two-fold degenerate, such as $e^{\pm ikx}$ for the free particle.

- Since the Schrodinger equation is real, if ψ is a solution with energy E, then ψ^* is a solution with energy E. If the solution ψ is not degenerate, then we must have $\psi = c\psi^*$, which means ψ is real up to a constant phase. Hence bound state wavefunctions can be chosen real. It turns out nonbound state wavefunctions can also be chosen real. These arguments are really just time-reversal invariant arguments in disguise, since we are conjugating the wavefunction.

- For bound states, the bound state with the n^{th} lowest energy has $n-1$ nodes. Moreover, the nodes interleave as n is increased.

- As an application, consider a double well potential. The symmetric combination of ground states will have lower energy than the antisymmetric combination, because its wavefunction has no nodes.

Note. The probability density and current are

$$\rho = |\psi|^2, \quad \mathbf{J} = \frac{1}{2}(\psi^* \mathbf{v}\psi + \psi \mathbf{v}\psi^*) = \mathrm{Re}(\psi^* \mathbf{v}\psi)$$

where the velocity operator is defined in general by Hamilton's equations,

$$\mathbf{v} = \frac{\partial H}{\partial \mathbf{p}} = \frac{\mathbf{p}}{m} = -\frac{i\hbar}{m}\nabla.$$

They satisfy the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0.$$

In particular, note that for an energy eigenfunction, $\mathbf{J} = 0$ identically since it can be chosen real. Also note that with a magnetic field, we would have $\mathbf{v} = (\mathbf{p} - q\mathbf{A})/m$ instead.

However, physically interpreting ρ and $\mathbf{J}$ is subtle. For example, consider multiplying by the particle charge q, so we have formal charge densities and currents. It is not true that a particle sources an electromagnetic field with charge density $e\rho$ and current density $e\mathbf{J}$. The electric field of a particle at $\mathbf{x}$ is

$$\mathbf{E}_{\mathbf{x}}(\mathbf{r}) = \frac{q(\mathbf{r} - \mathbf{x})}{|\mathbf{r} - \mathbf{x}|^3}.$$

Hence a perfect measurement of $\mathbf{E}$ is a measurement of the particle position $\mathbf{x}$. Thus for the hydrogen atom, we would not measure an exponentially small electric field at large distances, but a dipole field! The state of the system is not $|\psi\rangle \otimes \mathbf{E}_\rho$, but rather an entangled state like

$$\int d\mathbf{x}\, |\mathbf{x}\rangle \otimes |\mathbf{E}_{\mathbf{x}}\rangle$$

where we consider only the electrostatic field. To avoid these errors, it's better to think of the wavefunction as describing an ensemble of particles, rather than a single "spread out" particle. Note if the measurement takes longer than the characteristic orbit time of the electron, then we will only see the averaged field due to $q\mathbf{J}$.

Note. The Ehrenfest relations. The Heisenberg equations of motion are

$$\dot{\mathbf{x}} = -\frac{i}{\hbar}[\mathbf{x}, H] = \frac{\mathbf{p}}{m}, \quad \dot{\mathbf{p}} = -\frac{i}{\hbar}[\mathbf{p}, H].$$

Since expectation values are the same in all pictures, this implies in Schrodinger picture

$$\frac{d\langle \mathbf{x}\rangle}{dt} = \frac{\langle \mathbf{p}\rangle}{m}, \quad \frac{d\langle \mathbf{p}\rangle}{dt} = -\langle \nabla V\rangle, \quad m\frac{d^2\langle \mathbf{x}\rangle}{dt^2} = -\langle \nabla V\rangle$$

which holds exactly. When the particle is well-localized, we can replace $\langle \nabla V\rangle$ with $\nabla V(\langle \mathbf{x}\rangle, t)$, which implies that $\langle \mathbf{x}\rangle$ obeys the classical equations of motion. On the other hand, the error in making this approximation comes in only through the second derivative of $V(\mathbf{x})$, so this statement always holds for quadratic potentials, which include the harmonic oscillator and a uniform electric or gravitational field. More generally, it holds whenever the Hamiltonian is quadratic in $\mathbf{p}$ and $\mathbf{x}$, which includes a uniform magnetic field.

4.3 The Adiabatic Theorem

We now review the adiabatic theorem.

- Suppose we have a Hamiltonian $H(x^a, \lambda^i)$ with control parameters λ^i. If the energies never cross, we can index the eigenstates as a function of λ as $|n(\lambda)\rangle$. If the space of control parameters is contractible, the $|n(\lambda)\rangle$ can be taken to be smooth, though we will see cases where they cannot.
- The adiabatic theorem states that if the λ^i are changed sufficiently slowly, a state initially in $|n(\lambda(t_i))\rangle$ will end up in the state $|n(\lambda(t_f))\rangle$, up to an extra phase called the Berry phase. This is essentially because the rapid phase oscillations of the coefficients prevent transition amplitudes from accumulating, as we've seen in time-dependent perturbation theory.
- The phase oscillations between two energy levels have timescale $\hbar/\Delta E$, so the adiabatic theorem holds if the timescale of the change in the Hamiltonian is much greater than this.
- The quantum adiabatic theorem implies that quantum numbers n are conserved, and in the semiclassical limit
$$\oint p\, dq = nh$$
which implies the classical adiabatic theorem.
- To parametrize the error in the adiabatic theorem, we could write the time dependence as $H = H(\tau)$ with $\tau = \epsilon t$ and take $\epsilon \to 0$ and $t \to \infty$, holding τ fixed. We can then expand the coefficients in a power series in ϵ.
- When this is done carefully, we find that as long as the energy levels are nondegenerate, the adiabatic theorem holds to all orders in ϵ. To see why, note that the error terms will look like
$$\int_{\tau_i}^{\tau_f} d\tau\, e^{i\omega\tau/\epsilon} f(\tau)$$
If the levels are nondegenerate, then the integral must be evaluated by the saddle point approximation, giving a result of the form $e^{-\omega\tau/\epsilon}$, which vanishes faster than any power of ϵ.
- For comparison, note that for a constant perturbation, time-dependent perturbation theory gives a transition amplitude that goes as ϵ, rather than $e^{-1/\epsilon}$. This discrepancy is because the constant perturbation is suddenly added, rather than adiabatically turned on; if all time derivatives of the Hamiltonian are smooth, we get $e^{-1/\epsilon}$.

We now turn to Berry's phase.

- We assume the adiabatic theorem holds and plug the ansatz
$$|\psi(t)\rangle = e^{i\gamma(t)}|n(\lambda(t))\rangle$$
into the Schrodinger equation,
$$i\frac{\partial|\psi\rangle}{\partial t} = H(\lambda(t))|\psi\rangle$$
where $\gamma(t)$ is a phase to be determined. For simplicity we ignore all other states, and set the energy of the current state to zero at all times to ignore the dynamical phase.

- Plugging in the ansatz and operating with $\langle\psi|$, we find
$$i\dot{\gamma} + \langle n|\dot{n}\rangle = 0.$$
The quantity γ is real because
$$0 = \frac{d}{dt}\langle n|n\rangle = \langle \dot{n}|n\rangle + \langle n|\dot{n}\rangle = 2\,\mathrm{Re}\langle n|\dot{n}\rangle.$$

- Using the chain rule, we find
$$\gamma(t) = \int \mathcal{A}_i(\lambda)\, d\lambda^i, \quad \mathcal{A}_i(\lambda) = i\langle n|\frac{\partial}{\partial\lambda^i}|n\rangle$$
where $\mathcal{A}$ is called the Berry connection, and implicitly depends on n. However, this phase is only meaningful for a closed path in parameter space, because the Berry connection has a gauge redundancy from the fact that we can redefine the states $|n(\lambda)\rangle$ by phase factors.

- More explicitly, we may redefine the states by the 'gauge transformation'
$$|n'(\lambda)\rangle = e^{i\omega(\lambda)} n(\lambda)$$
in which case the Berry connection is changed to
$$\mathcal{A}_i' = \mathcal{A}_i + \partial_i \omega.$$
This is just like a gauge transformation in electromagnetism, except there, the parameters λ^i are replaced by spatial coordinates. Geometrically, $\mathcal{A}_i$ is a one-form over the space of parameters, like A_i is a one-form over Minkowski space.

- Hence we can define a gauge-invariant curvature
$$\mathcal{F}_{ij}(\lambda) = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i$$
called the Berry curvature. Using Stokes' theorem, we may write the Berry phase as
$$\gamma = \int_C \mathcal{A}_i\, d\lambda^i = \int_S \mathcal{F}_{ij}\, dS^{ij}$$
where S is a surface bounding the closed curve C.

- Geometrically, we can describe this situation using a $U(1)$ bundle over M, the parameter space. The Berry connection is simply a connection on this bundle; picking a phase convention amounts to choosing a section.

Next, we consider the Born-Oppenheimer approximation, an important application.

- In the theory of molecules, the basic Hamiltonian includes the kinetic energies of the nuclei and electrons, as well as Coulomb interactions between them. We have a small parameter $\kappa \sim (m/M)^{1/4}$ where m is the electron mass and M is the mass of the nuclei.

- In a precise treatment, we would expand in orders of κ. For example, for diatomic molecules we can directly show that electronic excitations have energies of order $E_0 = e^2/a_0$, where a_0 is the Bohr radius, vibrational modes have energies of order $\kappa^2 E_0$, and rotational modes have energies of order $\kappa^4 E_0$. These features generalize to all molecules.

- A simpler approximation is to simply note that if the electrons and nuclei have about the same kinetic energy, the nuclei move much slower. Moreover, the uncertainty principle places weaker constraints on their positions and momenta. Hence we could treat the positions of the nuclei as classical, giving a Hamiltonian $H_{\text{elec}}(\mathbf{r}, \mathbf{p}; \mathbf{R})$ for the electrons, then apply the adiabatic theorem to variations of $\mathbf{R}$. In particular, we include the nucleon-nucleon interaction energy.
- Applying the adiabatic theorem, we find eigenfunctions and energies

$$\phi_n(\mathbf{r}; \mathbf{R}), \quad E_n(\mathbf{R}).$$

  We can also compute a Berry connection $\mathbf{A}(\mathbf{R})$ in the configuration space of the nuclei.
- On the other hand, we know the total energy is conserved, which means the nuclei move in the classical effective potential

$$H_{\text{cl}}(\mathbf{R}, \mathbf{P}) = \frac{\mathbf{P}^2}{2M} + E_n(\mathbf{R}).$$

  More rigorously, this is the effective potential felt by the nuclei at lowest nontrivial order in κ. Note that it still includes the Coulomb interaction energy, and it depends on the electronic state n. This may be used to explain the chemical bond. At the next order, the Berry connection yields an effective gauge potential for the nuclei.

Example. A particle with spin s in a magnetic field of fixed magnitude. The parameter space S^2 is in magnetic field space. We may define states in this space as

$$|\theta, \phi, m\rangle = e^{i\phi m} e^{-i\phi S_z} e^{-i\theta S_y} |0, 0, m\rangle.$$

This is potentially singular at $\theta = 0$ and $\theta = \pi$, and the extra phase factor ensures there is no singularity at $\theta = 0$. The Berry connection is

$$A^{(m)} = m(\cos\theta - 1)\, d\phi$$

by direct differentiation, which gives a field strength

$$F^{(m)}_{\phi\theta} = m \sin\theta, \quad \int_{S^2} F = 4\pi m.$$

Hence we have a magnetic monopole in $\mathbf{B}$-space of strength proportional to m, and the singularity in the states and in $A^{(m)}$ is due to the Dirac string.

4.4 Particles in Electromagnetic Fields

Next, we set up the quantum mechanics of a particle in an electromagnetic field.

- The Hamiltonian for a particle in an electromagnetic field is

$$H = \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + q\phi$$

  as in classical mechanics. Here, $\mathbf{p}$ is the canonical momentum, so it corresponds to $-i\hbar\nabla$.
- There is an ordering ambiguity, since $\mathbf{A}$ and $\mathbf{p}$ do not commute at the quantum level. We will set the term linear in $\mathbf{A}$ to $\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}$, as this is the only combination that makes H Hermitian, as one can check by demanding $\langle\psi|H|\psi\rangle$ to be real.

- The kinetic momentum is $\boldsymbol{\pi} = \mathbf{p} - q\mathbf{A}$ and the velocity operator is $\mathbf{v} = \boldsymbol{\pi}/m$. The velocity operator is the operator that should appear in the continuity equation for probability, as it corresponds to the classical velocity.
- Under a gauge transformation specified by an arbitrary function α, called the gauge scalar,

$$\phi \to \phi - \partial_t \alpha, \quad \mathbf{A} \to \mathbf{A} + \nabla \alpha.$$

  As a result, the Hamiltonian is not gauge invariant.
- In order to make the Schrodinger equation gauge invariant, we need to allow the wavefunction to transform as well, by

$$\psi \to e^{iq\alpha/\hbar}\psi.$$

  If the Schrodinger equation holds for the old potential and wavefunction, then it also holds for the gauge-transformed potential and wavefunction. Roughly speaking, the extra $e^{iq\alpha/\hbar}$ factor can be 'pulled through' the time and space derivatives, leaving behind extra $\partial_\mu \alpha$ factors that exactly cancel the additional terms from the gauge transformation.
- In the context of gauge theories, the reasoning goes the other way. Given that we want to make $\psi \to e^{iq\alpha/\hbar}\psi$ a symmetry of the theory, we conclude that the derivative (here, $\mathbf{p}$) must be converted into a covariant derivative (here, $\boldsymbol{\pi}$).
- The phase of the wavefunction has no direct physical meaning, since it isn't gauge invariant. Similarly, the canonical momentum isn't gauge invariant, but the kinetic momentum $\boldsymbol{\pi}$ is. The particle satisfies the Lorentz force law in Heisenberg picture if we work in terms of $\boldsymbol{\pi}$.
- The fact that the components of velocity $\mathbf{v}$ don't commute can be understood directly from our intuition for Poisson brackets; in the presence of a magnetic field parallel to $\hat{\mathbf{z}}$, a particle moving in the $\hat{\mathbf{x}}$ direction is deflected in the $\hat{\mathbf{y}}$ direction.

Electromagnetic fields lead to some interesting topological phenomena.

Example. A particle around a flux tube. Consider a particle constrained to lie on a ring of radius r, through which a magnetic flux Φ passes. Then we can take

$$A_\phi = \frac{\Phi}{2\pi r}$$

and the Hamiltonian is

$$H = \frac{(p_\phi - qA_\phi)^2}{2m} = \frac{1}{2mr^2}\left(-i\hbar\partial_\phi - \frac{q\Phi}{2\pi}\right)^2.$$

The energy eigenstates are still exponentials, of the form

$$\psi = \frac{1}{\sqrt{2\pi r}} e^{in\phi}$$

where $n \in \mathbb{Z}$ since the wavefunction is single-valued. Plugging this in, the energy is

$$E = \frac{\hbar^2}{2mr^2}\left(n - \frac{\Phi}{\Phi_0}\right)^2$$

where $\Phi_0 = 2\pi\hbar/q$ is the quantum of flux. Since generally Φ/Φ_0 is not an integer, the presence of the magnetic field affects the spectrum even though the magnetic field is zero everywhere the wavefunction is nonzero!

We can also look at this phenomenon in a slightly different way. Suppose we were to try to gauge away the vector potential. Since

$$\mathbf{A} = \nabla\alpha, \quad \alpha = \frac{\Phi\phi}{2\pi}$$

we might try a gauge transformation with gauge scalar α. Then the wavefunction transforms as

$$\psi \to \exp\left(\frac{iq\alpha}{\hbar}\right)\psi = \exp\left(\frac{\Phi}{\Phi_0}i\phi\right)\psi.$$

However, the wavefunction only remains single-valued if Φ is a multiple of Φ_0! This reflects the fact that the spectrum really changes when Φ/Φ_0 is not an integer; it is a physical effect that can't be gauged away. The constraint that ψ is single-valued is perfectly physical; it's just what we used to get the energy eigenstates when A is zero. The reason it restricts the gauge transformations allowed is because the wavefunction wraps around the flux tube. This is a first look at how topology appears in quantum mechanics.

Example. Aharanov-Bohm effect. Consider the double slit experiment, but with a solenoid hidden behind the wall between the slits. Then the presence of the solenoid affects the interference pattern! To see this, note that a path from the starting point to a point $\mathbf{x}$ picks up a phase

$$\Delta\theta = \frac{q}{\hbar}\int^{\mathbf{x}} \mathbf{A}(\mathbf{x}')\cdot d\mathbf{x}'.$$

Then the two possible paths through the slits pick up a relative phase

$$\Delta\theta = \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{x} = \frac{q}{\hbar}\int \mathbf{B}\cdot d\mathbf{S} = \frac{q\Phi}{\hbar}$$

which shifts the interference pattern. Again, we see that if Φ is a multiple of Φ_0, the effect vanishes, but in general there is a physically observable effect.

Note. There are two ways to justify the phases. In the path integral formulation, we sum over all classical paths with phase $e^{iS/\hbar}$. The dominant contribution comes from the two classical paths, so we can ignore all others; the phase shift for each path is just $e^{i\Delta S/\hbar}$.

Alternatively, we can use the adiabatic theorem. Suppose that we have a well-localized, slowly-moving particle in a vector potential $\mathbf{A}(\mathbf{x})$. Then we can apply the adiabatic theorem, where the parameter is the particle's position, the Berry connection is $\mathbf{A}$, and the Berry curvature is $\mathbf{B}$, giving the same conclusion. In the path integral method, the adiabatic assumption manifests as ignoring the $\mathbf{p}\cdot d\mathbf{x}$ phase.

Note. Another way to derive the above effects is to use the fiber bundle formalism. This isn't necessary because all $U(1)$ bundles over S^1 are trivial, but suppose we cover S^1 with two patches. Then we can gauge away $\mathbf{A}$ within each patch, and the physical phases in both examples above arise just from transition functions. This can be a bit more convenient, since we don't have to bother with explicitly writing $\mathbf{A}$ and solving the Schrodinger equation with nonzero $\mathbf{A}$.

Example. Dirac quantization of magnetic monopoles. A magnetic monopole has a magnetic field

$$\mathbf{B} = \frac{g\hat{\mathbf{r}}}{4\pi r^2}$$

where the magnetic charge g is its total flux. To get around Gauss's law (i.e. writing $\mathbf{B} = \nabla \times \mathbf{A}$), we must use a singular vector potential. Two possible examples are

$$A_\phi^N = \frac{g}{4\pi r}\frac{1-\cos\theta}{\sin\theta}, \quad A_\phi^S = -\frac{g}{4\pi r}\frac{1+\cos\theta}{\sin\theta}.$$

These vector potentials are singular along the lines $\theta = \pi$ and $\theta = 0$, respectively, which we call 'Dirac strings'. Physically, we can think of a magnetic monopole as one end of a solenoid that extends off to infinity that's too thin to detect; the solenoid then lies on the Dirac string. Note that there is only one Dirac string, not two, but where it is depends on whether we use A_ϕ^N or A_ϕ^S.

To solve the Schrodinger equation for a particle in this field, we must solve it separately in the Northern hemisphere (where A_ϕ^N is nonsingular) and the Southern hemisphere, giving wavefunctions ψ_N and ψ_S. On the equator, where they overlap, they must differ by a gauge transformation

$$\psi_N = e^{iq\alpha/\hbar}\psi_S, \quad \alpha = \frac{g\phi}{2\pi}.$$

But since the wavefunction must be single-valued, g must be a multiple of Φ_0, giving the Dirac quantization condition

$$qg = 2\pi\hbar n.$$

Therefore, if even a single magnetic monopole exists, charge is quantized!

Alternatively, going in the opposite direction, the experimental observation of quantization of charge tells us that the gauge group of electromagnetism should be $U(1)$ rather than $\mathbb{R}$, and magnetic monopoles can only exist in the former. Hence the quantization of charge gives us a good theoretical reason to expect monopoles to exist.

Note. An alternate derivation of the Dirac quantization condition. Consider a particle that moves in the field of a monopole, in a closed path that subtends a magnetic flux Φ. As we know already, the resulting phase shift is $\Delta\theta = q\Phi/\hbar$. But we could also have taken a surface that wrapped about the monopole the other way, with a flux $\Phi - g$ and phase shift $\Delta\theta' = q(\Phi - g)/\hbar$.

Since we consider the exact same path in both situations (and the phase shift is observable, as we could interfere it with a state that didn't move at all), the phase shifts must differ by a multiple of 2π for consistency. This recovers the Dirac quantization condition.

Note. In general, the wavefunction is single-valued. Formally, it is a section of a complex line bundle associated with the $U(1)$ gauge bundle. However, it is important to note that in the case of a nontrivial bundle, the wavefunction can only be defined on patches; attempting to define it globally will give a multivalued or singular wavefunction. Sometimes one says that because the wavefunction differs between patches, it is multivalued, but this is incorrect. It would be better to say that its description depends on the patch.

Note. We can also derive the Dirac quantization condition without referring to matter. The point is that $A^N - A^S = d\lambda$ for a (single-valued) function λ defined on the equator S^1. Then

$$\int_{S^2} F = \int_N dA^N + \int_S dA^S = \int_{S^1} (A^N - A^S) = \int_{S^1} d\lambda$$

which is an integer. This quantity is called the first Chern class of the $U(1)$ bundle.

4.5 The Harmonic Oscillator

We consider the model system of the harmonic oscillator.

- The Hamiltonian
$$H = \frac{\hat{p}^2}{2m} + \frac{m\omega^2\hat{x}^2}{2}$$
has a characteristic length $\sqrt{\hbar/m\omega}$, characteristic momentum $\sqrt{m\hbar\omega}$, and characteristic energy $\hbar\omega$. Setting all of these quantities to one, or equivalently setting $\omega = \hbar = m = 1$,
$$H = \frac{\hat{p}^2}{2} + \frac{\hat{x}^2}{2}, \quad [\hat{x}, \hat{p}] = i.$$
We can later recover all units by dimensional analysis.

- Since the potential goes to infinity at infinity, there are only bound states, and hence the spectrum of H is discrete. Moreover, since we are working in one dimension, the eigenfunctions of H are nondegenerate.

- Classically, the Hamiltonian may be factored as
$$\frac{1}{2}(x^2 + p^2) = \frac{x + ip}{\sqrt{2}} \frac{x - ip}{\sqrt{2}}.$$
This motivates the definitions
$$a = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{p}), \quad a^\dagger = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{p}).$$
However, these two operators have the nontrivial commutation relation
$$[a, a^\dagger] = 1, \quad H = a^\dagger a + \frac{1}{2} = N + \frac{1}{2}.$$
The addition of the 1/2 is thus an inherently quantum effect.

- We note that the operator N is positive, because
$$\langle \phi | N | \phi \rangle = \| a | \phi \rangle \|^2 \geq 0.$$
Therefore, N only has nonnegative eigenvalues; we let the eigenvectors be
$$N|\nu\rangle = \nu|\nu\rangle, \quad \nu \geq 0.$$

- Applying the commutation relations, we find
$$Na = a(N - 1), \quad Na^\dagger = a^\dagger(N + 1).$$
This implies that $a|\nu\rangle$ is an eigenket of N with eigenvalue $\nu - 1$, and similarly $a^\dagger|\nu\rangle$ has eigenvalue $\nu + 1$. Therefore, starting with a single eigenket, we can get a ladder of eigenstates.

- This ladder terminates if $a|\nu\rangle$ or $a^\dagger|\nu\rangle$ vanishes. But note that

$$\|a|\nu\rangle\|^2 = \langle\nu|a^\dagger a|\nu\rangle = \nu, \quad \|a^\dagger|\nu\rangle\| = \nu + 1.$$

  Therefore, the ladder terminates on the bottom with $\nu = 0$ and doesn't terminate on the top. Moreover, all eigenvalues ν must be integers; if not, we could lower until the eigenvalue was negative, contradicting the positive definiteness of N. We can show there aren't multiple copies of the ladder by switching to wavefunctions and using uniqueness, as shown below.

- Therefore, the eigenstates of the harmonic oscillator are indexed by integers,

$$H|n\rangle = E_n|n\rangle, \quad E_n = n + \frac{1}{2}.$$

- Using the equations above, we find that for the $|n\rangle$ to be normalized, we have

$$a|n\rangle = \sqrt{n}|n-1\rangle, \quad a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle.$$

  There can in principle be a phase factor, but we use our phase freedom in the eigenkets to rotate it to zero. Repeating this, we find

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle.$$

Example. Explicit wavefunctions. The ground state wavefunction satisfies $a|0\rangle = 0$, so

$$\frac{1}{\sqrt{2}}(x + \partial_x)\psi_0(x) = 0, \quad \psi_0(x) = \frac{1}{\pi^{1/4}}e^{-x^2/2}.$$

Similarly, the excited states satisfy

$$\psi_n(x) = \frac{1}{\pi^{1/4}}\frac{1}{\sqrt{n!2^n}}(x - \partial_x)^n e^{-x^2/2}$$

To simplify the derivative factor, we 'commute past the exponential', using the identity

$$(x - \partial_x)e^{x^2/2}f = e^{-x^2/2}\partial_x f.$$

Therefore we find

$$\psi_n(x) = \frac{1}{\pi^{1/4}}\frac{(-1)^n}{\sqrt{n!2^n}}e^{x^2/2}\partial_x^n e^{-x^2}.$$

In terms of the Hermite polynomials, we have

$$\psi_n(x) = \frac{1}{\pi^{1/4}}\frac{1}{\sqrt{n!2^n}}H_n(x)e^{-x^2/2}, \quad H_n(x) = (-1)^x e^{x^2}\partial_x^n e^{-x^2}.$$

Generally the n^{th} state is an n^{th} degree polynomial times a Gaussian.

Note. Similarly, we can find the momentum space wavefunction $\widetilde{\psi}_n(p)$ by writing $a^\dagger$ in momentum space. The result turns out to be identical up to phase factors and scaling; this is because unitary evolution with the harmonic oscillator potential for time $\pi/2$ Fourier transforms the wavefunction (as shown in the next section), and this evolution leaves $\psi_n(x)$ unchanged up to a phase factor.

Next we turn to coherent states, where it's easiest to work in Heisenberg picture.

- The Hamiltonian is still $H = (\hat{x}^2 + \hat{p}^2)/2$, but the operators have time-dependence equivalent to the classical equations of motion,
$$\frac{d\hat{x}}{dt} = \hat{p}, \quad \frac{d\hat{p}}{dt} = -\hat{x}.$$
The solution to this is simply clockwise circular motion in phase space, as it is classically,
$$\begin{pmatrix} \hat{x}(t) \\ \hat{p}(t) \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} \hat{x}_0 \\ \hat{p}_0 \end{pmatrix}.$$
Then the expectation values of position and momentum behave as they do classically.

- Moreover, the time evolution for $\pi/2$ turns position eigenstates into momentum eigenstates. To see this, let $U = e^{-iH(\pi/2)}$ and let $x_0|x\rangle = x|x\rangle$. Then
$$Ux_0U^{-1}U|x\rangle = Ux|x\rangle$$
which implies that
$$p_0(U|x\rangle) = x(U|x\rangle).$$
Hence $U|x\rangle$ is a momentum eigenstate with (dimensionless) momentum x. A corollary is that time evolution for $\pi/2$ applies a Fourier transform to the wavefunction in Schrodinger picture.

- Classically, it is convenient to consider the complex variable
$$z = \frac{1}{\sqrt{2}}(x + ip), \quad \overline{z} = \frac{1}{\sqrt{2}}(x - ip).$$
Expressing the Hamiltonian in terms of these new degrees of freedom gives $H = \overline{z}z$, so $\dot{z} = -iz$ and $\dot{\overline{z}} = i\overline{z}$. As a result, the variable z rotates clockwise in the complex plane.

- The quantum analogues of z and $\overline{z}$ are a and $a^\dagger$, satisfying
$$\dot{a} = -ia, \quad \dot{a}^\dagger = ia^\dagger, \quad a(t) = e^{-it}a(0), \quad a^\dagger(t) = e^{it}a^\dagger(0).$$

- We define a coherent state as one satisfying
$$\Delta x = \Delta p = \frac{1}{\sqrt{2}}$$
which saturates the uncertainty relation. These states as 'are classical as possible', in the sense that they have maximally well defined position and momentum. Semiclassically, thinking of a quantum state as a phase space distribution, a coherent state is a circle in phase space with the minimum area h. In addition, there are 'squeezed states' that saturate the uncertainty relation but are ellipses in phase space.

- The state $|0\rangle$ is a coherent state, and we can generate others by applying the position and momentum translation operators
$$T(a) = e^{-ia\hat{p}}, \quad S(b) = e^{ib\hat{x}}$$

- By expanding in a Taylor series, or applying the Hadamard lemma,

  $$(T(a)\psi)(x) = \psi(x-a), \quad (T(a)\phi)(p) = e^{-iap}\phi(p), \quad (S(b)\psi)(x) = e^{ibx}\psi(x), \quad (S(b)\phi)(p) = \phi(p-b).$$

  Therefore the translation operators shift expectation values and keep dispersions constant. Moreover, they don't commute; using the above relations, we instead have

  $$S(b)T(a) = e^{iab}T(a)S(b)$$

  so we pick up a phase factor unless $ab = nh$.

- Due to the noncommutativity, the order of the position and momentum translations matters. To put them on an equal footing, we define the Heisenberg operators

  $$W(a,b) = e^{i(b\hat{x} - a\hat{p})}.$$

  By Glauber's theorem, we have

  $$W(a,b) = e^{iab/2}T(a)S(b) = e^{-iab/2}S(b)T(a),$$

  so it simply 'splits the phase' between the two orders.

- We define coherent states by

  $$|a,b\rangle = W(a,b)|0\rangle.$$

  We visualize this state as a circle centered at $(x,p) = (a,b)$ in phase space; the position space and momentum space wavefunctions are Gaussians.

With this setup, it's easy to show some important properties of coherent states.

- From our Heisenberg picture results, we know that the expectation values of $|a,b\rangle$ will evolve classically. To show that the dispersions are constant over time, it's convenient to switch to raising and lowering operators. Defining the complex variable z as before, we have

  $$W(x,p) = \exp\left(i(p\hat{x} - x\hat{p})\right) = \exp(za^\dagger - \overline{z}a) = W(z)$$

  Applying Glauber's theorem implies

  $$W(z) = e^{-|z|^2/2}\exp(za^\dagger)\exp(-\overline{z}a).$$

- Therefore, the coherent state $|z\rangle = W(z)|0\rangle$ is

  $$|z\rangle = e^{-|z|^2/2}\exp(za^\dagger)|0\rangle = e^{-|z|^2/2}\sum_{n=0}^{\infty}\frac{z^n}{\sqrt{n!}}|n\rangle.$$

  Then $|z\rangle$ is an eigenstate of the lowering operator with eigenvalue z.

- The expectation value of the energy is $|z|^2 = (x^2 + p^2 + 1)/2$, matching the classical value in the limit $\hbar \to 0$. The number distribution is Poisson with mean $|z|^2$.

- The time evolution of the coherent state is

$$U(t)|z\rangle = e^{-it/2}|e^{-it}z\rangle$$

in accordance with the classical $z(t)$ evolution we saw before. This implies the coherent state remains coherent. We can also see this result from the Heisenberg time evolution of a and $a^\dagger$.

- In the $z/\overline{z}$ variables, the uncertainty relation is $\Delta n \Delta\varphi \gtrsim 1$, where φ is the uncertainty on the phase of z. Physically, if we consider the quantum electromagnetic field, this relation bounds the uncertainty on the number of photons and the phase of the corresponding classical wave.

Example. Since a is not Hermitian, its eigenvectors are not a complete set. However, they are an 'overcomplete' set, in that

$$\int \frac{dxdp}{2\pi}|z\rangle\langle z| = 1.$$

To see this, act with $\langle m|$ on the left and $|n\rangle$ on the right for

$$\frac{1}{2\pi}\int dxdp\, e^{-|z|^2}\frac{z^n(z^*)^m}{\sqrt{n!m!}} = \int d|z|^2 e^{-|z|^2}\int \frac{d\varphi}{2\pi}\frac{z^n(z^*)^m}{\sqrt{n!m!}}.$$

The phase integral is zero unless $n = m$. When $n = m$, the phase integral is 1, and the $d|z|^2$ integral also gives 1, showing the result.

Note. Coherent states are ubiquitous in nature, because they are produced by forcing a harmonic oscillator. For a harmonic oscillator experiencing force $f(t)$, we have

$$x(t) = x_0(t) + \int dt'\, \sin(t-t')\theta(t-t')f(t')$$

by Green's functions, where $x_0(t)$ is a homogeneous solution. Then in Heisenberg picture,

$$\hat{x}(t) = \frac{\hat{a}e^{-it} + \hat{a}^\dagger e^{it}}{\sqrt{2}} + \int dt'\, \frac{i\theta(t-t')f(t')}{2}(e^{-i(t-t')} - e^{-i(t-t')})$$

where we fix $\hat{a}$ and $\hat{a}^\dagger$ to be the Heisenberg operators at time $t = 0$. Now we focus on times t after the driving ends. The step function is just 1, so

$$\hat{x}(t) = \frac{1}{\sqrt{2}}\left(\left(\hat{a} + \frac{i}{\sqrt{2}}\tilde{f}(1)\right)e^{-it} + \left(\hat{a}^\dagger - \frac{i}{\sqrt{2}}\tilde{f}(-1)\right)e^{it}\right)$$

where the expressions look a little strange because we have set $\omega = 1$. However, for all times,

$$\hat{x}(t) = \frac{\hat{a}(t) + \hat{a}^\dagger(t)}{\sqrt{2}}$$

so the final expressions for $\hat{a}(t)$ and $\hat{a}^\dagger(t)$ must be the factors above. The ground state evolves into a state annihilated by $\hat{a}(t)$, which is precisely a coherent state. The other states evolve into this state raised by $\hat{a}^\dagger(t)$.

Note. The classical electromagnetic field in a laser is really a coherent state of the quantum electromagnetic field; in general classical fields emerge from quantum ones by stacking many quanta together. A more exotic example occurs for superfluids, where the excitations are bosons which form a coherent field state, $\hat{\psi}(\mathbf{x})|\psi\rangle = \psi(\mathbf{x})|\psi\rangle$. In the limit of large occupancies, we may treat the state as a classical field $\psi(\mathbf{x})$, which is often called a "macroscopic wavefunction".

4.6 The WKB Approximation

In this section, we introduce the WKB approximation and connect it to classical mechanics.

- We consider the standard 'kinetic-plus-potential' Hamiltonian, and attempt to solve the time-independent Schrodinger equation. For a constant potential, the solutions are plane waves,

$$\psi(\mathbf{x}) = Ae^{iS(\mathbf{x})/\hbar}, \quad S(\mathbf{x}) = \mathbf{p} \cdot \mathbf{x}.$$

  The length scale here is the de Broglie wavelength $\lambda = h/p$.

- Now consider a potential that varies on scales $L \gg \lambda$. Then we have

$$\psi(\mathbf{x}) = A(\mathbf{x})e^{iS(\mathbf{x})/\hbar}$$

  where we expect $A(\mathbf{x})$ varies slowly, on the scale L, while $S(\mathbf{x})$ still varies rapidly, on the scale λ. Then the solution locally looks like a plane wave with momentum

$$\mathbf{p}(\mathbf{x}) = \nabla S(\mathbf{x}).$$

  Hence $S(\mathbf{x})$ is analogous to Hamilton's principal function.

- Our approximation may also be thought of as an expansion in $\hbar$, because $L \gg \lambda$ is equivalent to $pL \gg \hbar$. However, the WKB approximation is fundamentally about widely separated length scales; it is also useful in classical mechanics.

- To make this more quantitative, we write the logarithm of the wavefunction as a series in $\hbar$,

$$\psi(\mathbf{x}) = \exp\left(\frac{i}{\hbar}W(\mathbf{x})\right), \quad W(\mathbf{x}) = W_0(\mathbf{x}) + \hbar W_1(\mathbf{x}) + \hbar^2 W_2(\mathbf{x}) + \ldots.$$

  Comparing this to our earlier ansatz, we identify W_0 with S and W_1 with $-i\log A$, though the true S and A receive higher-order corrections.

- Plugging this into the Schrodinger equation gives

$$\frac{1}{2m}(\nabla W)^2 - \frac{i\hbar}{2m}\nabla^2 W + V = E.$$

  At lowest order in $\hbar$, this gives the time-independent Hamilton-Jacobi equation

$$\frac{1}{2m}(\nabla S)^2 + V(\mathbf{x}) = E$$

  which describes particles of energy E.

- At the next order,

$$\frac{1}{m}\nabla W_0 \cdot \nabla W_1 - \frac{i}{2m}\nabla^2 W_0 = 0, \quad \nabla S \cdot \nabla \log A + \frac{1}{2}\nabla^2 S = 0$$

  which is equivalent to

$$\nabla \cdot (A^2 \nabla S) = 0.$$

  This is called the amplitude transport equation.

- To see the meaning of this result, define a velocity field and density

$$\mathbf{v}(\mathbf{x}) = \frac{\partial H}{\partial \mathbf{p}} = \frac{\mathbf{p}(\mathbf{x})}{m}, \quad \rho(\mathbf{x}) = A(\mathbf{x})^2.$$

  Then the amplitude transport equation says

$$\nabla \cdot \mathbf{J} = 0, \quad \mathbf{J}(\mathbf{x}) = \rho(\mathbf{x})\mathbf{v}(\mathbf{x})$$

  which is simply conservation of probability in a static situation.

- Semiclassically, we can think of a stationary state as an ensemble of classical particles with momentum field $\mathbf{p}(\mathbf{x})$, where $\nabla \times \mathbf{p} = 0$, and the particle density is constant in time. This picture is correct up to $O(\hbar^2)$ corrections.

- The same reasoning can be applied to the time-dependent Schrodinger equation with a time-dependent Hamiltonian, giving

$$\frac{1}{2m}(\nabla S)^2 + V(\mathbf{x}, t) + \frac{\partial S}{\partial t} = 0.$$

  This is simply the time-dependent Hamilton-Jacobi equation.

Note. The general definition of a quantum velocity operator is

$$\mathbf{v}(\mathbf{x}) = \frac{\partial H}{\partial \mathbf{p}} = \frac{\partial \omega}{\partial \mathbf{k}}.$$

Therefore, the velocity operator gives the group velocity of a wavepacket. This makes sense, since we also know that the velocity operator appears in the probability flux.

We now specialize to one-dimensional problems.

- In the one-dimensional case, we have, at lowest order,

$$\psi(x) = A(x)e^{iS(x)/\hbar}, \quad \frac{1}{2m}\left(\frac{dS}{dx}\right)^2 + V(x) = E, \quad \frac{d}{dx}\left(A^2 \frac{dS}{dx}\right) = 0.$$

  The solutions are

$$\frac{dS}{dx} = p(x) = \pm\sqrt{2m(E - V(x))}, \quad A(x) = \frac{\text{const}}{\sqrt{p(x)}}.$$

  Since S is the integral of $p(x)$, it is simply the phase space area swept out by the classical particle's path.

- Note that in classically forbidden regions, S becomes imaginary, turning oscillation into exponential decay. In classically allowed regions, the two signs of S are simply interpreted as whether the particle is moving left or right. For concreteness we choose

$$p(x) = \begin{cases} \sqrt{2m(E - V(x))} & E > V(x), \\ i\sqrt{2m(V(x) - E)} & E < V(x). \end{cases}$$

- The result $A \propto 1/\sqrt{p}$ has a simple classical interpretation. Consider a classical particle oscillating in a potential well. Then the amount of time it spends at a point is inversely proportional to the velocity at that point, and indeed $A^2 \propto 1/p \propto 1/v$. Then the semiclassical swarm of particles modeling a stationary state should be uniformly distributed in time.
- This semiclassical picture also applies to time-independent scattering states, which can be interpreted as a semiclassical stream of particles entering and disappearing at infinity.
- Note that the WKB approximation breaks down for classical turning points (where $V(x) = E$) since the de Broglie wavelength diverges.

We now derive the connection formulas, which deal with turning points.

- Suppose the classically allowed region is $x < x_r$. In this region, we define

$$S(x) = \int_{x_r}^{x} p(x')\, dx'.$$

Then the WKB solution for $x < x_r$ is

$$\psi_{\mathbf{I}}(x) = \frac{1}{\sqrt{p(x)}} \left(c_r e^{iS(x)/\hbar + i\pi/4} + c_\ell e^{-iS(x)/\hbar - i\pi/4} \right)$$

where c_r and c_ℓ represent the right-moving and left-moving waves.

- For the classically forbidden region, we define

$$K(x) = \int_{x_r}^{x} |p(x')|\, dx'$$

to deal with only real quantities. Then the general WKB solution is

$$\psi_{\mathbf{II}}(x) = \frac{1}{\sqrt{|p(x)|}} \left(c_g e^{K(x)/\hbar} + c_d e^{-K(x)/\hbar} \right)$$

where the solutions grow and decay exponentially, respectively, as we go rightward.

- The connection formulas relate c_r and c_ℓ with c_g and c_d. Taylor expanding near the turning point, the Schrodinger equation is

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V'(x_r)(x - x_r)\psi = 0.$$

To nondimensionalize, we switch to the shifted and scaled variable z defined by

$$x = x_r + az, \quad a = \left(\frac{\hbar^2}{2mV'(x_r)} \right)^{1/3}, \quad \frac{d^2\psi}{dz^2} - z\psi = 0.$$

This differential equation is called Airy's equation.

- The two independent solutions to Airy's equation are $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$. They are the exact solutions of Schrodinger's equation for a particle in a uniform field, such a gravitational or electric field. Both oscillate for $z \ll 0$, and exponentially decay and grow for $z \gg 0$,

$$\mathrm{Ai}(x) = \begin{cases} \dfrac{\cos\alpha(z)}{\sqrt{\pi}(-z)^{1/4}} & z \ll 0 \\ \\ \dfrac{e^{-\beta(z)}}{2\sqrt{\pi}z^{1/4}} & z \gg 0, \end{cases} \qquad \mathrm{Bi}(x) = \begin{cases} \dfrac{\sin\alpha(z)}{\sqrt{\pi}(-z)^{1/4}} & z \ll 0 \\ \\ \dfrac{e^{\beta(z)}}{\sqrt{\pi}z^{1/4}} & z \gg 0, \end{cases}$$

where

$$\alpha(z) = -\frac{2}{3}(-z)^{3/2} + \frac{\pi}{4}, \quad \beta(z) = \frac{2}{3}z^{3/2}$$

as can be shown by the saddle point approximation.

- Let the solution near the turning point be

$$\psi_{\mathrm{tp}}(x) = c_a\,\mathrm{Ai}(z) + c_b\,\mathrm{Bi}(z).$$

We first match this with the solution on the left. Writing the solution in terms of complex exponentials,

$$\psi_{\mathrm{tp}}(z) = \frac{1}{2\sqrt{\pi}(-z)^{1/4}}((c_a - ic_b)e^{i\alpha(z)} + (c_a + ic_b)e^{-i\alpha(z)}), \quad z \ll 0.$$

On the other hand, the phase factors have been chosen so that in the linear approximation, the WKB solution is

$$\psi_{\mathbf{I}}(x) = \frac{1}{\sqrt{p(x)}}(c_r e^{i\alpha(z)} + c_\ell e^{-i\alpha(z)}).$$

Thus we read off the simple result

$$\frac{c_a - ic_b}{2\sqrt{\pi}} = \sqrt{\frac{a}{\hbar}}c_r, \quad \frac{c_a + ic_b}{2\sqrt{\pi}}\sqrt{\frac{a}{\hbar}}c_\ell.$$

- In the classically forbidden region, similar reasoning gives

$$\frac{c_a}{2\sqrt{\pi}} = \sqrt{\frac{a}{\hbar}}c_d, \quad \frac{c_b}{\sqrt{\pi}} = \sqrt{\frac{a}{\hbar}}c_g.$$

Combining these results gives the connection formulas

$$\begin{pmatrix} c_g \\ c_d \end{pmatrix} = \begin{pmatrix} i & -i \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} c_r \\ c_\ell \end{pmatrix}.$$

- The analysis for a classically forbidden region on the left is very similar. On the left,

$$\psi_{\mathbf{III}}(x) = \frac{1}{\sqrt{|p(x)|}}\left(c_g e^{K(x)/\hbar} + c_d e^{-K(x)/\hbar}\right), \quad K(x) = \int_{x_\ell}^{x} |p(x')|\,dx'$$

and on the right,

$$\psi_{\mathbf{IV}}(x) = \frac{1}{\sqrt{p(x)}}\left(c_r e^{iS(x) - i\pi/4} + c_\ell e^{-iS(x) - i\pi/4}\right), \quad S(x) = \int_{x_\ell}^{x} p(x')\,dx'$$

where the phase factors are again chosen for convenience. Then we find

$$\begin{pmatrix} c_g \\ c_d \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -i & i \end{pmatrix} \begin{pmatrix} c_r \\ c_\ell \end{pmatrix}.$$

We now apply the connection formulas to some simple problems.

- First, consider a classically forbidden region for $x > x_r$ that is impenetrable. Then we must have $c_g = 0$ in this region, so $c_r = c_\ell$ and the wavefunction on the left is
$$\psi_{\mathbf{I}}(x) = \frac{1}{\sqrt{p(x)}}(e^{iS(x)+i\pi/4} + e^{-iS(x)-i\pi/4}).$$
Another way to write this is to match the phases at the turning point,
$$\psi_{\mathbf{I}}(x) = \frac{1}{\sqrt{p(x)}}(e^{iS(x)} + re^{-iS(x)}), \quad r = -i.$$
To interpret this, we picture the wave as accumulating phase $d\theta = p\,dx/\hbar$ as it moves. Then the reflection coefficient tells us the 'extra' phase accumulated due to the turning point, $-\pi/2$.
- Next, consider a oscillator with turning points x_ℓ and x_r. This problem can be solved by demanding exponential decay on both sides. Intuitively, the particle picks up a phase of
$$\frac{1}{\hbar}\oint p\,dx - \pi$$
through one oscillation, so demanding the wavefunction be single-valued gives
$$2\pi I = \oint p\,dx = (n + 1/2)h, \quad n = 0, 1, 2, \ldots$$
which is the Bohr-Sommerfeld quantization rule. The quantity I is proportional to the phase space area of the orbit, and called the action in classical mechanics. The semiclassical estimate for the energy of the state is just the energy of the classical solution with action I.
- In the case of the simple harmonic oscillator, we have
$$\oint p\,dx = \pi\sqrt{2mE}\sqrt{\frac{2E}{m\omega^2}} = \frac{2\pi E}{\omega}$$
which yields
$$E_n = (n + 1/2)\hbar\omega$$
which are the exact energy eigenvalues; however, the energy eigenstates are not exact.
- We can also consider reflection from a hard wall, i.e. an infinite potential. In this case the right-moving and left-moving waves must cancel exactly at the wall, $c_\ell = -ic_r$, which implies that the reflected wave picks up a phase of $-\pi$.
- For example, the quantization condition for a particle in a box is
$$\oint p\,dx = (n + 1)h, \quad n = 0, 1, 2, \ldots$$
and if the box has length L, then
$$E_n = \frac{(n+1)^2\hbar^2\pi^2}{2mL^2}$$
which is the exact answer.

- Finally, we can have periodic boundary conditions, such as when a particle moves on a ring. Then there are no phase shifts at all, and the quantization condition is just $\oint p\,dx = nh$.
- Generally, we find that for a system with an n-dimensional configuration space, each stationary state occupies a phase space volume of h^n. This provides a quick way to calculate the density of states.

Note. Classical and quantum frequencies. The classical frequency ω_c is the frequency of the classical oscillation, and obeys $\omega_c = dE/dI$. The quantum frequency ω_q is the rate of change of the quantum phase. These are different; for the harmonic oscillator ω_c does not depend on n but ω_q does.

Now, when a quantum oscillator transitions between states with difference $\Delta\omega_q$ in quantum frequencies, it releases radiation of frequency $\Delta\omega_q$. On the other hand, we know that a classical particle oscillating at frequency ω_c radiates at frequency ω_c. To link these together, suppose a quantum oscillator has $n \gg 1$ and transitions with $\Delta n = -1$. Then

$$\omega_q = \frac{\Delta E}{\hbar} \approx \frac{\Delta E}{\Delta I} \approx \frac{dE}{dI} = \omega_c$$

which recovers the classical expectation. For higher Δn, radiation is released at multiples of ω_c. This also fits with the classical expectation, where these harmonics come from the higher Fourier components of the motion.

5 Path Integrals

5.1 Formulation

- Define the propagator as the position-space matrix elements of the time evolution operator,
$$K(x,t;x_0,t_0) = \langle x|U(t,t_0)|x_0\rangle.$$
Then we automatically have $K(x,t_0;x_0,t_0) = \delta(x-x_0)$.

- Since we often work in the position basis, we distinguish the Hamiltonian operator acting on kets, $|H\rangle$ and the differential operator acting on wavefunctions, H. They are related by
$$\langle x|\hat{H}|\psi\rangle = H\langle x|\psi\rangle.$$

- Using the above, the time evolution of the propagator is
$$i\hbar\frac{\partial K(x,t;x_0,t_0)}{\partial t} = H(t)K(x,t;x_0,t_0)$$
which means that $K(x,t)$ is just a solution to the Schrodinger equation with initial condition $\psi(x,t_0) = \delta(x-x_0)$. However, $K(x,t)$ is not a valid wave function; its initial conditions are quite singular, and non-normalizable. Since a delta function contains all momenta, $K(x,t)$ is typically nonzero for all x, for any $t > t_0$.

- Given the propagator, we can compute general time evolution as
$$\psi(x,t) = \int dx_0\, K(x,t;x_0,t_0)\psi(x_0,t_0).$$

Example. The propagator for the free particle. Since the problem is time-independent, we set $t_0 = 0$ and drop it. Then
$$\begin{aligned}
K(x,x_0,t) &= \langle x|\exp(-it\hat{p}^2/2m\hbar)|x_0\rangle \\
&= \int dp\,\langle x|\exp(-it\hat{p}^2/2m\hbar)|p\rangle\langle p|x_0\rangle \\
&= \int \frac{dp}{2\pi\hbar}\exp\left[\frac{i}{\hbar}\left(p(x-x_0) - \frac{p^2t}{2m}\right)\right] \\
&= \sqrt{\frac{m}{2\pi i\hbar t}}\exp\left(\frac{i}{\hbar}\frac{m(x-x_0)^2}{2t}\right)
\end{aligned}$$
where we performed a Gaussian integral. The limit $t \to 0$ is somewhat singular; we expect it is a delta function, yet the magnitude of the propagator is equal for all x. The resolution is that the phase oscillations in x get faster and faster, so that $K(x,t)$ behaves like a delta function when integrated against a test function.

The path integral is an approach for calculating the propagator in more complicated settings. We work with the Hamiltonian $H = T + V = p^2/2m + V(x)$, as more general Hamiltonians with higher powers of p are more difficult to handle.

- The time evolution for a small time ϵ is

$$U(\epsilon) = 1 - \frac{i\epsilon}{\hbar}(T+V) + O(\epsilon^2) = e^{-i\epsilon T/\hbar} e^{-i\epsilon V/\hbar} + O(\epsilon^2).$$

  Therefore the time evolution for a time $t = N\epsilon$ is

$$U(t) = \left(e^{-i\epsilon T/\hbar} e^{-i\epsilon V/\hbar}\right)^N + O(1/N).$$

  This is a special case of the Lie product formula; the error vanishes as $N \to \infty$.

- Using this decomposition, we insert the identity $N-1$ times for

$$K(x, x_0, t) = \lim_{N\to\infty} \int dx_1 \ldots dx_{N-1} \prod_{j=0}^{N-1} \langle x_{j+1}|e^{-i\epsilon T/\hbar} e^{-i\epsilon V/\hbar}|x_j\rangle, \quad x = x_N.$$

  Within each factor, we insert a resolution of the identity in momentum space for

$$\int dp\, \langle x_{j+1}|e^{-i\epsilon \hat{p}^2/2m\hbar}|p\rangle\langle p|e^{-i\epsilon V(\hat{x})/\hbar}|x_j\rangle = \sqrt{\frac{m}{2\pi i\epsilon\hbar}} \exp\left[\frac{i}{\hbar}\left(m\frac{(x_{j+1}-x_j)^2}{2\epsilon} - \epsilon V(x_j)\right)\right]$$

  where we performed a Gaussian integral almost identical to the free particle case. Then

$$K(x, x_0, t) = \lim_{N\to\infty} \left(\frac{m}{2\pi i\hbar\epsilon}\right)^{N/2} \int dx_1 \ldots dx_{N-1} \exp\left[\frac{i\epsilon}{\hbar}\sum_{j=0}^{N-1}\left(m\frac{(x_{j+1}-x_j)^2}{2\epsilon^2} - \epsilon V(x_j)\right)\right]$$

- Recognizing a Riemann sum, the above formula shows that

$$K(x, x_0, t) = C \int d[x(\tau)] \exp\left(\frac{i}{\hbar}\int_0^t L\, d\tau\right)$$

  where C is a normalization constant and $d[x(\tau)]$ is the volume element in 'path space'.

- For each time interval Δt, the range of positions that contributes significantly to the amplitude is $\Delta x \sim \sqrt{\Delta t}$, since rapid oscillations cancel the contribution outside this range. This implies that typical path integral paths are continuous but not differentiable. This is problematic for the compact action integral notation above, since the Lagrangian formalism assumes differentiable paths, but we ignore it for now.

- If we don't perform the momentum integration, we get the phase space path integral,

$$K(x, x_0, t) = C \int d[x(\tau)]d[p(\tau)] \exp\left(\frac{i}{\hbar}\int_0^t (p\dot{x} - H)\, d\tau\right)$$

  where x is constrained at the endpoints but p is not. This form is less common, but more general, as it applies even when the kinetic energy is not quadratic in momentum. In such cases the momentum integrals are not Gaussian and cannot be performed. Luckily, the usual path integral will work in all the cases we care about.

- The usual path integral can also accommodate terms linear in p, as these are shifted Gaussians; for example, they arise when coupling to a magnetic field, $p^2 \to (p - eA)^2$.

5.2 Gaussian Integrals

One of the strengths of the path integral is that it keeps classical paths in view; this makes it well-suited for semiclassical approximations. We first review some facts about Gaussian integration.

- The fundamental result for Gaussian integration is
$$\int dx\, e^{-ax^2/2} = \sqrt{\frac{2\pi}{a}}, \quad \operatorname{Re} a > 0.$$
All bounds of integration are implicitly from $-\infty$ to ∞.
- By differentiating the above, we find
$$\int dx\, e^{-ax^2/2} x^2 = \sqrt{\frac{2\pi}{a^3}}.$$
- By completing the square and shifting, we have
$$\int dx\, e^{-ax^2/2+bx} = \sqrt{\frac{2\pi}{a}} e^{b^2/2a}.$$
In particular, this result holds even for complex b, as we can shift the integration contour in the complex plane; this is legal since there are no singularities to hit.
- Later, we will have to generalize the Gaussian integral to complex arguments; the fundamental result is
$$\int d(\overline{z}, z)\, e^{-\overline{z}wz} = \frac{\pi}{w}, \quad \operatorname{Re} w > 0.$$
Here, the notation $\int d(\overline{z}, z)$ is a formal notation that stands for $\int dx\, dy$ where $z = x + iy$ and $\overline{z} = x - iy$, and in practice, we always evaluate these integrals by breaking z into real and imaginary parts and doing the dx and dy integrals instead. (Note that, in terms of differential forms, $dzd\overline{z} = dxdy$ up a constant.)
- Similarly, by taking real/imaginary parts, we find
$$\int d(\overline{z}, z)\, e^{-\overline{z}wz+\overline{u}z+\overline{z}v} = \frac{\pi}{w} e^{\overline{u}v/w}, \quad \operatorname{Re} w > 0.$$
- The multidimensional generalization of the real Gaussian integral is
$$\int d\mathbf{v}\, e^{-\mathbf{v}^T A\mathbf{v}/2} = \sqrt{\frac{(2\pi)^N}{\det A}}$$
where A must be positive definite and real. Then A is symmetric and can be diagonalized, separating the integral into N standard Gaussian integrals; the positive definiteness ensures that these integrals converge.
- Similarly, with a linear term in $\mathbf{v}$, we have
$$\int d\mathbf{v}\, e^{-\mathbf{v}^T A\mathbf{v}/2+\mathbf{j}^T\mathbf{v}} = \sqrt{\frac{(2\pi)^N}{\det A}} e^{\mathbf{j}^T A^{-1}\mathbf{j}/2}.$$
This can be shown using the shift $\mathbf{v} \to \mathbf{v} + A^{-1}\mathbf{j}$.

- Next, we can differentiate the above identity with respect to j at $\mathbf{j} = 0$. But since
$$\partial_{j_m} e^{\mathbf{j}^T A^{-1} \mathbf{j}/2} = (A^{-1}\mathbf{j})_m e^{\mathbf{j}^T A^{-1} \mathbf{j}/2}$$
the result vanishes for a single derivative when valuated at $\mathbf{j} = 0$. However, for two derivatives, we can get a nonzero result by differentiating the $A^{-1}\mathbf{j}$ term, giving
$$\int d\mathbf{v}\, e^{-\mathbf{v}^T A \mathbf{v}/2} v_m v_n = \sqrt{\frac{(2\pi)^N}{\det A}} A^{-1}_{mn}.$$
Interpreting the Gaussian as a probability distribution, this implies
$$\langle v_m v_n \rangle = A^{-1}_{mn}.$$
Similarly, for any even number of derivatives, we get a sum over all pairings,
$$\langle v_{i_1} \cdots v_{i_{2n}} \rangle = \sum_{\text{pairings}} A^{-1}_{i_{k_1} i_{k_2}} \ldots A^{-1}_{i_{k_{2n-1}} i_{k_{2n}}}.$$
This is known as Wick's theorem.

- In the complex case, we have
$$\int d(\mathbf{v}^\dagger, \mathbf{v})\, e^{-\mathbf{v}^\dagger A \mathbf{v}} = \frac{\pi^N}{\det A}$$
where A must be positive definite. (The conclusion also holds if A only has positive definite Hermitian part.) With a linear term, we have
$$\int d(\mathbf{v}^\dagger, \mathbf{v})\, e^{-\mathbf{v}^\dagger A \mathbf{v} + \mathbf{w}^\dagger \mathbf{v} + \mathbf{v}^\dagger \mathbf{w}'} = \frac{\pi^N}{\det A} e^{\mathbf{w}^\dagger A^{-1} \mathbf{w}'}.$$
Similarly, we can take derivatives; to get nonzero results, we must pair derivatives with respect to v with derivatives with respect to $\overline{v}$. Then Wick's theorem is
$$\langle \overline{v}_{i_1} \cdots \overline{v}_{i_n} v_{j_1} \cdots v_{j_n} \rangle = \sum_{\text{perms}} A^{-1}_{j_1 i_{P_1}} \cdots A^{-1}_{j_n i_{P_n}}$$
where the sum is over permutations of N integers.

- In the continuum limit, the vectors and matrices above become functions and operators, and the integral becomes a path integral, giving
$$\int Dv(x) \exp\left(-\frac{1}{2}\int dx\, dx'\, v(x) A(x,x') v(x') + \int dx\, j(x) v(x)\right)$$
$$\propto \frac{1}{\sqrt{\det A}} \exp\left(\frac{1}{2}\int dx\, dx'\, j(x) A^{-1}(x,x') j(x')\right).$$
Here, A^{-1} is the Green's function for A, satisfying
$$\int dx'\, A(x,x') A^{-1}(x',x'') = \delta(x - x'')$$
and we have thrown away some normalization factors, which drop out of averages. Wick's theorem generalizes to this case straightforwardly.

Note. We now review the stationary phase approximation. We consider the integral

$$\int dx\, e^{i\varphi(x)/\kappa}$$

for small κ. Then the integrand oscillates wildly except at points of stationary phase $\overline{x}$. Approximating the exponent as a quadratic there, we have a Gaussian integral, giving

$$\int dx\, e^{i\varphi(x)/\kappa} \approx \sqrt{\frac{2\pi i\kappa}{\varphi''(\overline{x})}} e^{i\varphi(\overline{x})/\kappa} = e^{i\nu\pi/4}\sqrt{\frac{2\pi\kappa}{|\varphi''(\overline{x})|}} e^{i\varphi(\overline{x})/\kappa}, \quad \nu = \operatorname{sign}(\varphi''(\overline{x}))$$

If there are multiple points of stationary phase, we must sum over each such point. Similarly, we can consider the multidimensional integral

$$\int d\mathbf{x}\, e^{i\varphi(\mathbf{x})/\kappa}$$

for small κ. Then the stationary points are where $\nabla\varphi = 0$. Expanding about these points and applying our multidimensional Gaussian formula,

$$\int d\mathbf{x}\, e^{i\varphi(\mathbf{x})/\kappa} = e^{i\nu\pi/4}(2\pi\kappa)^{n/2}\left|\det \frac{\partial^2\varphi(\overline{x})}{\partial x_k \partial x_l}\right|^{-1/2} e^{i\varphi(\overline{x})/\kappa}, \quad \nu = \sum_i \operatorname{sign}(\lambda_i).$$

To get the full result, we sum over all stationary points.

5.3 Semiclassical Approximation

Given this setup, we now apply the stationary phase approximation to the path integral.

- In this case, the large parameter is $\kappa = \hbar$ and the function is the discretized Lagrangian

  $$\varphi(x_1, \ldots, x_{N-1}) = \epsilon \sum_{j=0}^{N-1} \left(\frac{m}{2}\frac{(x_{j+1}-x_j)^2}{\epsilon^2} - V(x_j) \right).$$

  Differentiating, we have

  $$\frac{\partial\varphi}{\partial x_k} = \epsilon\left(\frac{m}{\epsilon^2}(2x_k - x_{k+1} - x_{k-1}) - V'(x_k)\right), \quad \frac{\partial^2\varphi}{\partial x_k \partial x_\ell} = \frac{m}{\epsilon} Q_{k\ell}$$

  where the matrix $Q_{k\ell}$ is tridiagonal,

  $$Q_{k\ell} = \begin{pmatrix} 2-c_1 & -1 & 0 & 0 & \ldots \\ -1 & 2-c_2 & -1 & 0 & \ldots \\ 0 & -1 & 2-c_3 & -1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad c_k = \frac{\epsilon^2}{m}V''(x_k).$$

- In the limit $N \to \infty$, the stationary points are simply the classical paths $\overline{x}(\tau)$, so

  $$\lim_{N\to\infty} \varphi(\overline{x}) = S(x, x_0, t).$$

  In the case of multiple stationary paths, we add a branch index.

- Next, we must evaluate $\det Q$. This must combine with the path integral prefactor, which is proportional to $\epsilon^{-N/2}$, to give a finite result, so we expect $\det Q \propto 1/\epsilon$. The straightforward way to do this would be to diagonalize Q, finding eigenfunctions of the second variation of the action. However, we can do the whole computation in one go by a slick method.

- Letting D_k be the determinant of the upper-left $k \times k$ block, we have

$$D_{k+1} = (2 - c_{k+1})D_k - D_{k-1}.$$

This may be rearranged into a difference equation, which becomes, in the continuum limit

$$m\frac{d^2F(\tau)}{d\tau^2} = -V''(\overline{x}(\tau))F(\tau), \quad F_k = \epsilon D_k.$$

We pulled out a factor of ϵ to make $F(\tau)$ regular, with initial conditions

$$F(0) = \lim_{\epsilon\to 0} \epsilon D_0 = \lim_{\epsilon\to 0} \epsilon = 0, \quad F'(0) = \lim_{\epsilon\to 0}(D_1 - D_0) = 1.$$

- The equation of motion for F is the equation of motion for a small deviation about the classical path, $x(\tau) = \overline{x}(\tau) + F(\tau)$, as the right-hand side is the linearized change in force. Thus $F(t)$ is the change in position at time t per unit change in velocity at $t = 0$, so

$$F(t) = \frac{\partial x}{\partial v_i} = m\left(\frac{\partial p_i}{\partial x}\right)^{-1} = -m\left(\frac{\partial^2 S}{\partial x_0 \partial x}\right)^{-1}.$$

This is regular, as expected, and we switch back to $D(t)$ by dividing by ϵ. Intuitively, this factor tells us how many paths near the original classical path contribute. In the case where $V''(\tau) < 0$, nearby paths rapidly diverge away, while for $V''(\tau) < 0$ a restoring force pushes them back, enhancing the contribution.

- Finally, we need the number of negative eigenvalues, which we call μ. It will turn out that μ approaches a definite limit as $N \to \infty$. In that limit, it is the number of perturbations of the classical path that further decrease the action, which is typically small.

- Putting everything together and restoring the branch index gives the Van Vleck formula

$$K(x, x_0, t) \approx \sum_b \frac{e^{-i\mu_b \pi/2}}{\sqrt{2\pi i\hbar}} \left|\frac{\partial^2 S_b}{\partial x \partial x_0}\right|^{1/2} \exp\left(\frac{i}{\hbar} S_b(x, x_0, t)\right).$$

The Van Vleck formula expands the action to second order about stationary paths. It is exact when the potential energy is at most quadratic, i.e. for a particle that is free, in a uniform electric or gravitational field, or in a harmonic oscillator. It is also exact for a particle in a magnetic field, since the Lagrangian remains at most quadratic in velocity.

Example. The free particle. In this case the classical paths are straight lines and

$$S = \frac{m\dot{x}^2 t}{2} = \frac{m(x - x_0)^2}{2t}.$$

The determinant factor is

$$\left|\frac{\partial^2 S}{\partial x \partial x_0}\right|^{1/2} = \sqrt{\frac{m}{t}}.$$

The second-order change in action would be the integral of $m(\delta\dot{x})^2/2$ which is positive definite, so $\mu = 0$. Putting everything together gives

$$K(x, x_0, t) = \sqrt{\frac{m}{2\pi i\hbar t}} \exp\left(\frac{i}{\hbar}\frac{m(x-x_0)^2}{2t}\right)$$

as we found earlier.

Example. Recovering the Schrodinger equation. For a small time $t = \epsilon$, we have

$$\psi(\mathbf{x}, \epsilon) = \psi(\mathbf{x}, 0) - \frac{i\epsilon}{\hbar}\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right)\psi(\mathbf{x}, 0) + O(\epsilon^2).$$

Now we compare this to the path integral. Here we use a single timestep, so

$$\psi(\mathbf{x}, \epsilon) = \int d\mathbf{y}\, K(\mathbf{x}, \mathbf{y}, \epsilon)\psi(\mathbf{y}, 0), \quad K(\mathbf{x}, \mathbf{y}, 0) = \left(\frac{m}{2\pi i\hbar\epsilon}\right)^{3/2} \exp\left(\frac{i\epsilon}{\hbar}\left(\frac{m(\mathbf{x}-\mathbf{y})^2}{2\epsilon^2} - V(\mathbf{y})\right)\right).$$

The expansion is a little delicate because of the strange dependence on ϵ. The key is to note that by the stationary phase approximation, most of the contribution comes from $\boldsymbol{\xi} = \mathbf{x} - \mathbf{y} = O(\epsilon^{1/2})$. We then expand everything to first order in ϵ, treating $\boldsymbol{\xi} = O(\epsilon^{1/2})$, for

$$\psi(\mathbf{x}, \epsilon) = \left(\frac{m}{2\pi i\hbar\epsilon}\right)^{3/2} \int d\boldsymbol{\xi}\, \exp\left(\frac{im\xi^2}{2\epsilon\hbar}\right)\left(1 - \frac{i\epsilon}{\hbar}V(\mathbf{x}+\boldsymbol{\xi}) + \ldots\right) \\ \times \left(\psi(\mathbf{x}, 0) + \xi^i\partial_i\psi(\mathbf{x}, 0) + \frac{1}{2}\xi^i\xi^j\partial_i\partial_j\psi(\mathbf{x}, 0) + \ldots\right).$$

where we cannot expand the remaining exponential since its argument is $O(1)$. Now we consider the terms in the products of the two expansions. The $O(1)$ term gives $\psi(\mathbf{x}, 0)$, as expected. The $O(\epsilon^{1/2})$ term gives zero because it is odd in ξ. The $O(\epsilon)$ term is

$$-\frac{i\epsilon}{\hbar}V(\mathbf{x})\psi(\mathbf{x}, 0) + \frac{1}{2}\xi^i\xi^j\partial_i\partial_j\psi(\mathbf{x}, 0).$$

The first of these terms is the potential term. The second term integrates to give the kinetic term. Finally, the $O(\epsilon^{3/2})$ term vanishes by symmetry, proving the result.

Example. Path integrals in quantum statistical mechanics. Since the density matrix is $\rho = e^{-\beta H}/Z$, we would like to compute the matrix elements of $e^{-\beta H}$. This is formally identical to what we've done before if we set $t = -i\hbar\beta$. Substituting this in, we have

$$\langle x|e^{-\beta H}|x_0\rangle = \lim_{N\to\infty}\left(\frac{m}{2\pi\hbar\eta}\right)^{N/2}\int dx_1\ldots dx_{N-1}\, \exp\left(-\frac{\eta}{\hbar}\sum_{j=0}^{N-1}\left(\frac{m(x_{j+1}-x_j)^2}{2\eta^2} + V(x_j)\right)\right)$$

where we have defined $\eta = \hbar\beta/N$, and $\epsilon = -i\eta$. The relative sign between the kinetic and potential terms has changed, so we have an integral for the Hamiltonian instead, and the integral is now damped rather than oscillatory. Taking the continuum limit, the partition function is

$$Z = C\int dx_0 \int d[x(u)]\, \exp\left(-\frac{1}{\hbar}\int_0^{\beta\hbar} H\, du\right)$$

where the path integral is taken over paths with $x(0) = x(\beta\hbar) = x_0$. As a simple example, suppose that the temperature is high, so $\beta\hbar$ is small. Then the particle can't move too far from $x(0)$ in the short 'time' $u = \beta\hbar$, so we can approximate the potential as constant,

$$Z \approx C \int dx_0 \, e^{-\beta V(x_0)} \int d[x(u)] \, \exp\left(-\frac{1}{\hbar}\int_0^{\beta\hbar} \frac{m}{2}\left(\frac{dx}{du}\right)^2 du\right) = \sqrt{\frac{m}{2\pi\beta\hbar^2}} \int dx_0 \, e^{-\beta V(x_0)}$$

where the last step used the analytically continued free particle propagator. This is the result from classical statistical mechanics, where Z is simply an integral of $e^{-\beta H}$ over phase space, but we can now find corrections order by order in $\beta\hbar$.

Example. The harmonic oscillator with frequency ω. This is somewhat delicate since some choices of (x_0, x, t) give infinitely many branches, or no branches at all. However, assuming we have chosen a set with exactly one branch, we can show

$$S(x, x_0, t) = \frac{m\omega}{2\sin(\omega t)}((x_0^2 + x^2)\cos(\omega t) - 2xx_0).$$

To find μ, note that we may write the second variation as

$$\delta S = \int d\tau \, \delta x(\tau) \left(-\frac{m}{2}\left(\frac{d^2}{d\tau^2} + \omega^2\right)\right) \delta x(\tau)$$

by integration by parts; hence we just need the number of negative eigenvalues of the operator above, where the boundary conditions are $\delta x(0) = \delta x(t) = 0$. The eigenfunctions are of the form $\sin(n\pi\tau/t)$ for positive integer n with eigenvalue $(n\pi/t)^2 - \omega^2$. Therefore the number of negative eigenvalues depends on the value of t, but for sufficiently small t there are none.

Applying the Van Vleck formula gives the exact propagator,

$$K(x, x_0, t) = \sqrt{\frac{m\omega}{2\pi i\hbar\sin(\omega t)}} \exp(iS(x, x_0, t)/\hbar), \quad t < \pi/\omega.$$

Setting $t = -i\hbar\beta$ and simplifying gives the partition function

$$Z = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}}$$

which matches the results from standard statistical mechanics.

Example. Operator ordering in the path integral. At the quantum level, operators generally do not commute, and their ordering affects the physics. But all the variables in the path integral appear to commute. It turns out that the operator ordering is determined by the discretization procedure. For example, for a particle in an electromagnetic field, the phase factor is

$$\exp\left(\frac{i\epsilon}{\hbar}\sum_{j=0}^{N-1}\left(\frac{m(\mathbf{x}_{j+1} - \mathbf{x}_j)^2}{2\epsilon^2} + \frac{q}{c}\frac{\mathbf{x}_{j+1} - \mathbf{x}_j}{\epsilon} \cdot \mathbf{A}\left(\frac{\mathbf{x}_{j+1} + \mathbf{x}_j}{2}\right) - V(\mathbf{x}_j)\right)\right)$$

where V is evaluated as usual at the initial point, but $\mathbf{A}$ is evaluated at the midpoint. One can show this is the right choice by expanding order by order in ϵ as we did before. While the evaluation point of V doesn't matter, the evaluation point of $\mathbf{A}$ ensures that the path integral describes a Hamiltonian with term $\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}$.

Naively, the evaluation point can't matter because it makes no difference in the continuum limit. The issue is that the path integral paths are not differentiable, as we saw earlier, with $\boldsymbol{\xi} = O(\epsilon^{1/2})$ instead of $\boldsymbol{\xi} = O(\epsilon)$. The midpoint evaluation makes a difference at order $O(\xi^2) = O(\epsilon)$, which is exactly the term that matters. This subtlety is swept under the rug in the casual, continuum notation for path integrals.

6 Angular Momentum

6.1 Classical Rotations

First, we consider rotations classically.

- Physical rotations are operators $\mathcal{R}$ that take spatial points to spatial points in an inertial coordinate system, preserving lengths and the origin.
- By taking coordinates, $\mathbf{r} = x_i \hat{\mathbf{e}}_i$, we can identify every spatial point with a 3-vector. As a result, we can identify rotation operators $\mathcal{R}$ with 3×3 rotation matrices R_{ij}. Under a rotation $\mathbf{r}' = R\mathbf{r}$, we have $x'_i = R_{ij} x_j$.
- We distinguish the physical rotations $\mathcal{R}$ and the rotation matrices R. The latter provide a representation of the former.
- It's also important to distinguish active/passive transformations. We prefer the active viewpoint; the passive viewpoint is tied to coordinate systems, so we can't abstract out to the geometric rotations $\mathcal{R}$.
- Using the length-preserving property shows $R^t = R^{-1}$, so the group of rotations is isomorphic to $O(3)$. From now on we specialize to proper rotations, with group $SO(3)$. The matrices R acting on $\mathbb{R}^3$ form the fundamental representation of $SO(3)$.
- Every proper rotation can be written as a rotation of an angle θ about an axis $\hat{\mathbf{n}}$, $\mathcal{R}(\hat{\mathbf{n}}, \theta)$.

  Proof: every rotation has a unit eigenvalue because $\prod \lambda_i = 1$ and $|\lambda_i| = 1$. The corresponding eigenvalue is the axis. (Note that this argument fails in higher dimensions.)
- Working in the fundamental representation, we consider the infinitesimal elements $R = I + \epsilon A$. Then we require $A + A^t = 0$, so the (fundamental representation of the) Lie algebra $\mathfrak{so}(3)$ contains antisymmetric matrices. One convenient basis is

  $$(J_i)_{jk} = -\epsilon_{ijk}$$

  and we write an algebra element as $A = \mathbf{a} \cdot \mathbf{J}$.
- Using the above definition, we immediately find

  $$(J_i J_j)_{jk} = \delta_{il}\delta_{kj} - \delta_{ij}\delta_{kl}$$

  which gives the commutation relations

  $$[J_i, J_j] = \epsilon_{ijk} J_k, \quad [\mathbf{a} \cdot \mathbf{J}, \mathbf{b} \cdot \mathbf{J}] = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{J}.$$
- We also immediately find that for an arbitrary vector $\mathbf{u}$,

  $$A\mathbf{u} = \mathbf{a} \times \mathbf{u}$$

  Physically, we can picture $\mathbf{a}$ as specifying an angular velocity and $A\mathbf{u}$ as the resulting velocity of $\mathbf{u}$. This also shows that an infinitesimal axis-angle rotation is

  $$R(\hat{\mathbf{n}}, \theta) = I + \theta \hat{\mathbf{n}} \cdot \mathbf{J}, \quad \theta \ll 1.$$

  Exponentiating gives the result

  $$R(\hat{\mathbf{n}}, \theta) = \exp(\theta \hat{\mathbf{n}} \cdot \mathbf{J}).$$

- More generally, the set of infinitesimal elements of a Lie group is a Lie algebra, and we go between the two by taking exponentials, or differentiating paths through the origin (to get tangent vectors).

A group acts on itself by conjugation; this is called the adjoint action. The Lie algebra is closed under this operation, giving an action of the group on the algebra. Viewing the algebra as a vector space, this gives a representation of the Lie group on $V = \mathfrak{g}$ called the adjoint representation.

Example. In the case of $SO(3)$, the fundamental representation happens to coincide with the adjoint representation. To see this, note that

$$R(\mathbf{a} \times \mathbf{u}) = (R\mathbf{a}) \times (R\mathbf{u})$$

which simply states that the cross product transforms as a vector under rotations (it's actually a pseudovector). Then we find

$$R(\mathbf{a} \cdot \mathbf{J})\mathbf{u} = ((R\mathbf{a}) \cdot \mathbf{J})R\mathbf{u}, \quad R(\mathbf{a} \cdot \mathbf{J})R^{-1} = (R\mathbf{a}) \cdot \mathbf{J}.$$

This provides a representation of the Lie group, representing $\mathcal{R}$ as the operator that takes the vector $\mathbf{a}$ to $R\mathbf{a}$. This is just the fundamental representation, but viewed in a more abstract way – the vector space now contains infinitesimal rotations rather than spatial vectors.

Another statement of the above is that 'angular velocity is a vector'. This is not generally true; in $SO(2)$, it is a scalar and the adjoint representation is trivial; in $SO(4)$, the Lie group is six-dimensional, and the angular velocity is more properly a two-form.

Example. Variants of the adjoint representation. Exponentiating the above gives the formula for the adjoint action on the group,

$$R_0 R(\hat{\mathbf{n}}, \theta) R_0^{-1} = R(R_0(\hat{\mathbf{n}}), \theta).$$

We can also derive the adjoint action of an algebra on itself, which yields a representation of the Lie algebra. First consider conjugation acting on an infinitesimal group element,

$$A(1 + \epsilon h)A^{-1} = 1 + \epsilon A h A^{-1}, \quad A \in G, \quad h \in \mathfrak{g}.$$

This shows that the adjoint action also conjugates algebra elements. Then if $A = 1 + \epsilon g$ with $g \in \mathfrak{g}$,

$$h \to (1 + \epsilon g)h(1 - \epsilon g) = h + \epsilon(gh - hg).$$

Taking the derivative with respect to ϵ to define the algebra's adjoint action, we find that g acts on h by sending it to $[g, h]$. Incidentally, this is also a proof that the Lie algebra is closed under commutators, since we know the algebra is closed under the adjoint action.

As a direct example, consider the matrix Lie group $SO(3)$. Since the operation is matrix multiplication, the commutator above is just the matrix commutator. Our above calculations shows that the adjoint action of the Lie algebra $\mathfrak{so}(3)$ on itself is the cross product.

Example. Noncommutativity in the Lie group reflects a nontrivial Lie bracket. The first manifestation of this is the fact that

$$e^{tg} e^{th} e^{-tg} e^{-th} = 1 + t^2 [g, h] + \dots$$

This tells us that a nonzero Lie bracket causes the corresponding group elements to not commute; as a simple example, the commutator of small rotations about $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ is a rotation about $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}$. Conversely, the Lie bracket is zero, the commutator is zero.

Another form of the above statement is the Baker-Campbell-Hausdorff theorem, which is the matrix identity

$$e^X e^Y = e^Z, \quad Z = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] + \frac{1}{12}[Y,[Y,X]] + \ldots$$

where all the following terms are built solely out of commutators of X and Y. Therefore, if we can compute the commutator in the algebra, we can in principle compute multiplication in the group.

The group $SO(3)$ is a compact connected three-dimensional manifold; it is also the configuration space for a rigid body, so wavefunctions for rigid bodies are defined on the $SO(3)$ manifold. As such, it's useful to have coordinates for it; one set is the Euler angles.

Example. The Euler angles. A rotation corresponds to an orientation of a coordinate system; therefore, we can specify a rotation uniquely by defining axes $\hat{\mathbf{x}'}$, $\hat{\mathbf{y}'}$, $\hat{\mathbf{z}'}$ that we would like to rotate our original axes into. Suppose the spherical coordinates of $\hat{\mathbf{z}'}$ in the original frame are α and β. Then the rotation

$$R(\hat{\mathbf{z}}, \alpha) R(\hat{\mathbf{y}}, \beta)$$

will put a vector originally pointing along $\hat{\mathbf{z}}$ along $\hat{\mathbf{z}'}$. However, the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ axes won't be in the right place. To fix this, we can perform a pre-rotation about $\hat{\mathbf{z}}$ before any of the other rotations; therefore, any rotation may be written as

$$R(\alpha, \beta, \gamma) = R(\hat{\mathbf{z}}, \alpha) R(\hat{\mathbf{y}}, \beta) R(\hat{\mathbf{z}}, \gamma).$$

This is the zyz convention for the Euler angles. We see that α and γ range from 0 to 2π, while β ranges from 0 to π. The group manifold $SO(3)$, however, is not $S^1 \times S^1 \times [0, \pi]$. This is reflected in the fact that for extremal values of the angles, the Euler angle parametrization is not unique.

6.2 Representations of $\mathfrak{su}(2)$

Next we consider quantum spin, focusing on the case of spin 1/2.

- Given a quantum mechanical system with an associated Hilbert space, we expect rotations R are realized by unitary operators $U(R)$ on the space. It is reasonable to expect that $R \to U(R)$ is a group homomorphism, so we have a representation of $SO(3)$ on the Hilbert space.
- Given a representation of a Lie group, we automatically have a representation of the Lie algebra. Specifically, we define

  $$J_k = i\hbar \left. \frac{\partial U(\boldsymbol{\theta})}{\partial \theta_k} \right|_{\boldsymbol{\theta}=0}$$

  where $U(\boldsymbol{\theta})$ is the rotation with axis $\hat{\boldsymbol{\theta}}$ and angle θ. Then we must have

  $$[J_i, J_j] = i\hbar \epsilon_{ijk} J_k.$$

  This can be shown directly by considering the commutator of infinitesimal rotations.

- The operators $\mathbf{J}$ generate rotations, the factor of i makes them Hermitian, and the factor of $\hbar$ makes them have dimensions of angular momentum. We hence define $\mathbf{J}$ to be the angular momentum operator of the system.

- With this definition, near-identity rotations take the form

$$U(\hat{\mathbf{n}}, \theta) = 1 - \frac{i}{\hbar}\theta \hat{\mathbf{n}} \cdot \mathbf{J} + \ldots, \quad U(\hat{\mathbf{n}}, \theta) = \exp\left(-\frac{i}{\hbar}\theta \hat{\mathbf{n}} \cdot \mathbf{J}\right).$$

  Since we can recover a representation of the group by exponentiation, it suffices to find representations of the algebra, i.e. triplets of matrices that satisfy the above commutation relations.

- One possible representation is

$$\mathbf{J} = \frac{\hbar}{2}\boldsymbol{\sigma}$$

  in which case

$$U(\hat{\mathbf{n}}, \theta) = e^{-i\theta \hat{\mathbf{n}} \cdot \boldsymbol{\sigma}/2} = \cos\frac{\theta}{2} - i(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}) \sin\frac{\theta}{2}.$$

  This gives the spin 1/2 representation; it tells us how states transform under rotations.

- Even though the angular momentum of a spin 1/2 particle is not a vector, we still expect that angular momentum behaves like a vector under rotations, in the sense that the expectation value $\langle \mathbf{J} \rangle$ transforms as a vector. Then we require

$$\langle U\psi | \boldsymbol{\sigma} | U\psi \rangle = \langle \psi | U^\dagger \boldsymbol{\sigma} U | \psi \rangle = R \langle \psi | \boldsymbol{\sigma} | \psi \rangle$$

  which implies that

$$U^\dagger \boldsymbol{\sigma} U = R\boldsymbol{\sigma}.$$

  This may be verified directly using our explicit formula for U above.

- The above formula is equivalent to our earlier adjoint formula. Inverting and dotting with $\mathbf{a}$, we find

$$U(\mathbf{a} \cdot \boldsymbol{\sigma}) U^\dagger = (R\mathbf{a}) \cdot \boldsymbol{\sigma}.$$

  This is just another formula for the adjoint action; conjugation by the group takes $\mathbf{a}$ to $R\mathbf{a}$.

- Using our explicit formula above, we notice that

$$U(\hat{\mathbf{n}}, 2\pi) = -1.$$

  This phase is physically observable; in neutron inferferometry, we may observe it by splitting a beam, rotating by a relative 2π, and recombining it. Then our representation is actually one-to-two. Mathematically, this tells us we actually want projective representations of $SO(3)$, which turns out to be equivalent to representations of $SU(2)$, the double cover of $SO(3)$. In the case of spin 1/2, we're simply working with the fundamental representation of $SU(2)$.

- Using the definition of $SU(2)$, we find that for any $U \in SU(2)$,

$$U = x_0 + i\mathbf{x} \cdot \boldsymbol{\sigma}, \quad \sum x_i^2 = 1$$

  so $SU(2)$ is topologically S^3. The x_i are called the Cayley-Klein parameters.

Example. Euler angle decomposition also works for spinor rotations, with

$$U(\hat{\mathbf{x}},\theta) = \begin{pmatrix} \cos\theta/2 & -i\sin\theta/2 \\ -i\sin\theta/2 & \cos\theta/2 \end{pmatrix}, U(\hat{\mathbf{y}},\theta) = \begin{pmatrix} \cos\theta/2 & -\sin\theta/2 \\ \sin\theta/2 & \cos\theta/2 \end{pmatrix}, U(\hat{\mathbf{z}},\theta) = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}.$$

Then a general rotation may be written as

$$U(\alpha,\beta,\gamma) = U(\hat{\mathbf{z}},\alpha)U(\hat{\mathbf{y}},\beta)U(\hat{\mathbf{z}},\gamma)$$

where $\alpha \in [0,2\pi]$, $\beta \in [0,\pi]$, $\gamma \in [0,4\pi]$. The extended range of γ accounts for the double cover. To see that this gives all rotations, note that classical rotations R are a representation of spinor rotations U with kernel $\pm I$. Then with the extended range of γ, which provides the -1, we get everything.

Example. The ket $|+\rangle = (1,0)$ points in the $+\hat{\mathbf{z}}$ direction, since $\langle +|\boldsymbol{\sigma}|+\rangle = \hat{\mathbf{z}}$ and $\sigma_z|+\rangle = |+\rangle$. Similarly, we can define the kets pointing in arbitrary directions as

$$|\hat{\mathbf{n}},+\rangle = U|+\rangle.$$

Writing $\hat{\mathbf{n}}$ in spherical coordinates and applying the Euler angle decomposition,

$$U = U(\hat{\mathbf{z}},\alpha)U(\hat{\mathbf{y}},\beta), \quad |\hat{\mathbf{n}},+\rangle = \begin{pmatrix} e^{-i\alpha/2}\cos\beta/2 \\ e^{i\alpha/2}\sin\beta/2 \end{pmatrix}.$$

Applying the adjoint formula, we have

$$\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}|\hat{\mathbf{n}},+\rangle = |\hat{\mathbf{n}},+\rangle, \quad \langle\hat{\mathbf{n}},+|\boldsymbol{\sigma}|\hat{\mathbf{n}},+\rangle = \hat{\mathbf{n}}.$$

Then the expectation value of the spin along any direction perpendicular to $\hat{\mathbf{n}}$ vanishes.

Note. The above reasoning doesn't work for higher spin. For example, using the notation in the next section, for a spin 1 particle, the state $(0,1,0)$ has $\langle\boldsymbol{\sigma}\rangle = 0$, so it's not 'pointing' in any direction. For spin higher than $1/2$, the action of the rotation operators $U(R)$ on the states $|\psi\rangle$ isn't even transitive, since the dimension of $SU(2)$ is less than the (real) dimension of the state space.

We now consider general representations of $\mathfrak{su}(2)$ on a Hilbert space. That is, we are looking for triplets of operators $\mathbf{J}$ satisfying the angular momentum commutation relations. Given these operators, we can recover the rotation operators by exponentiation; conversely, we can get back to the angular momentum operators by differentiation at $\boldsymbol{\theta} = 0$.

- Begin by constructing the operator

  $$J^2 = J_1^2 + J_2^2 + J_3^2.$$

  which commutes with $\mathbf{J}$; such an operator is called a Casimir operator. As a result, J^2 commutes with any function of $\mathbf{J}$, including the rotation operators.

- Given the above structure, we consider simultaneous eigenkets $|am\rangle$ of J^2 and J_3, with eigenvalues $\hbar^2 a$ and $\hbar m$. Since J^2 and J_3 are Hermitian, a and m are real, and since J^2 is nonnegative definite, $a \geq 0$. For simplicity, we assume we are dealing with an irrep; physically, we can guarantee this by postulating that J^2 and J_3 form a CSCO.

- We introduce the ladder operators

$$J_\pm = J_1 \pm iJ_2, \quad [J_3, J_\pm] = \pm\hbar J_\pm, \quad [J_+, J_-] = 2\hbar J_3, \quad [J^2, J_\pm] = 0.$$

  They satisfy the relations

$$J^2 = \frac{1}{2}(J_+J_- + J_-J_+) + J_3^2, \quad J_-J_+ = J^2 - J_3(J_3 + \hbar), \quad J_+J_- = J^2 - J_3(J_3 - \hbar).$$

  In this setting, $J_\pm$ play a very similar formal role to a and $a^\dagger$ for the QHO.

- Next, as in the QHO, we investigate norms. We have

$$\langle am|J_-J_+|am\rangle = \hbar^2(a - m(m+1)) \geq 0$$

  and similarly

$$\hbar^2(a - m(m-1)) \geq 0.$$

  Therefore, we require $a \geq \max(m(m+1), m(m-1))$. If the maximum value of $|m|$ is j, the corresponding value of a is $j(j+1)$. For convenience, we switch to labeling the states by j and m values.

- Then our first equation above becomes

$$\langle jm|J_-J_+|jm\rangle = \hbar^2(j-m)(j+m+1) \geq 0$$

  where we have equality if $j = m$. (The other case is forbidden by our second equation.) Doing a similar analysis on the second equation, we conclude

$$J_+|jm\rangle = 0 \text{ iff } m = j, \quad J_-|jm\rangle = 0 \text{ iff } m = -j.$$

- Finally, using the commutation relations, we see that acting with $J_\pm$ doesn't change the j value, but raises/lowers m by 1.

  As a result, we conclude that $m - j$ is an integer; if not, we can keep applying the raising operator until our inequalities above are broken. Similarly, $m - (-j)$ is an integer. Therefore, $2j$ is an integer and $m = -j, \ldots, +j$. These are all of the irreps of $\mathfrak{su}(2)$.

Now that we've found all of the irreps, we turn to calculations and applications.

- Using our norm calculation above, we find

$$J_+|jm\rangle = \hbar\sqrt{(j-m)(j+m+1)}|j, m+1\rangle, \quad J_-|jm\rangle = \hbar\sqrt{(j+m)(j-m+1)}|j, m-1\rangle.$$

  Above we used the phase freedom in the $|jm\rangle$ to set all possible phase factors to zero. Then

$$|jm\rangle = \sqrt{\frac{(j+m)!}{(2j)!(j-m)!}} \left(\frac{J_-}{\hbar}\right)^{j-m} |jj\rangle.$$

- Given the above, we know the matrix elements of $J_\pm$, as well as the matrix elements of J_3,

$$\langle j'm'|J_3|jm\rangle = \hbar\delta_{j'j}\delta_{m'm}m.$$

  Then we can simply write down the matrix elements of all of the $\mathbf{J}$, and hence the matrix of any function of $\mathbf{J}$, including the rotation operators.

Example. What j values apply to what physical situations? If we're considering central force motion of a particle, it turns out that only integral j matter. If we consider p-wave scattering, $j = 1$ appears. The spin state of a photon is described by $j = 1$, but the spin state of two electrons is described by $j = 0, 1$. Which j values appear must be analyzed separately for each physical situation.

Example. In the case $j = 1$, we have

$$J_3 = \hbar \begin{pmatrix} 1 & & \\ & & \\ & & -1 \end{pmatrix}, \quad J_+ = \hbar \begin{pmatrix} & \sqrt{2} & \\ & & \sqrt{2} \\ & & \end{pmatrix}.$$

Evaluating $e^{-2\pi i J_3/\hbar}$, we find that a rotation by 2π is the identity. In general, for integer j, we end up with a normal representation of $SO(3)$, rather than a projective one.

Example. Reading a table of rotation matrices. The operator $U(\hat{\mathbf{n}}, \theta)$ has matrix elements

$$D^j_{m'm}(U) = \langle jm'|U|jm\rangle.$$

Note that U must be diagonal in j-space, so we aren't missing any information here. We think of the $D^j_{m'm}$ as a set of matrices indexed by j. Parametrizing a rotation by Euler angles as above,

$$D^j_{mm'}(\alpha, \beta, \gamma) = \langle jm|e^{-i\alpha J_z/\hbar} e^{-i\beta J_y/\hbar} e^{-i\gamma J_z/\hbar}|jm'\rangle.$$

It is straightforward to expand this since J_z is diagonal, giving

$$D^j_{mm'}(\alpha, \beta, \gamma) = e^{-i\alpha m - i\gamma m'} d^j_{mm'}(\beta), \quad d^j_{mm'}(\beta) = \langle jm|e^{-i\beta J_y/\hbar}|jm\rangle.$$

Here, $d^j_{mm'}(\beta)$ is the reduced rotation matrix. Using tables of $d^j_{mm'}$ values, we may construct rotation matrices for arbitrary spin.

The D^j matrices have numerous properties which aid calculation. We can view them as a representation of the U operators; the distinction is that while the U operators act on a Hilbert space, the D^j matrices are just numbers acting on vectors of numbers. Since U is unitary, and we are using the orthonormal basis $|jm\rangle$, the D^j are also unitary. These two properties imply

$$D^j_{mm'}(U^{-1}) = D^{j*}_{m'm}(U).$$

This is one of several symmetries of the D matrices.

Example. Multiple copies of the same irrep. If we have just one copy of an irrep, we construct an orthonormal basis for it by starting with $|jm\rangle$ for $m = j$ and acting with J_-. Similarly, if there are many copies, we may pick an orthonormal basis for the $m = j$ subspace, labeling the vectors by γ, and carry them down with J_-. We write the resulting basis vectors as $|\gamma jm\rangle$, and all matrix elements defined above are identical except for a factor of $\delta_{\gamma'\gamma}$. In particular, the $D^j_{mm'}$ matrices still suffice to calculate everything we need.

Example. The adjoint formula carries over, becoming

$$U\mathbf{J}U^\dagger = R^{-1}\mathbf{J}.$$

Here, $\mathbf{J} = \hat{J}_i \mathbf{e}_i$ is a vector of operators; the U operates on the $\hat{J}_i$ and the R operates on the $\mathbf{e}_i$. The formula can be proven by considering infinitesimal rotations and building them up; for an infinitesimal rotation $U(\hat{n}, \theta)$ with $\theta \ll 1$, the left-hand side is

$$\mathbf{J} - \frac{i\theta}{\hbar}[\hat{\mathbf{n}} \cdot \mathbf{J}, \mathbf{J}].$$

The commutator is equal to

$$[n_i J_i, J_j \hat{\mathbf{e}}_j] = i\hbar \epsilon_{ijk} n_i \hat{\mathbf{e}}_j J_k = -i\hbar \hat{\mathbf{n}} \times \mathbf{J}.$$

Therefore, the left-hand side is $\mathbf{J} - \theta \hat{\mathbf{n}} \times \mathbf{J}$, which is simply the infinitesimal spatial rotation R^{-1}.

6.3 Spin and Orbital Angular Momentum

Next, we turn to physical realizations of angular momentum in quantum mechanics. We first consider the case of spins in a magnetic field.

- The Hamiltonian of a magnetic moment $\boldsymbol{\mu}$ in a magnetic field $\mathbf{B}$ is $H = -\boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{x})$, both classically and in quantum mechanics. Magnetic moments obey

$$F = -\nabla U = \nabla(\boldsymbol{\mu} \cdot \mathbf{B}), \quad \boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}.$$

- Experimentally, we find that for nuclei and elementary particles, $\boldsymbol{\mu} \propto \mathbf{J}$, and the relevant state space is just a single copy of a single irrep of $\mathfrak{su}(2)$.

- For a classical current loop with total mass m and charge q, we can show that

$$\boldsymbol{\mu} = \frac{q}{2mc}\mathbf{L}.$$

  For general configurations, $\boldsymbol{\mu}$ and $\mathbf{L}$ need not be proportional, since the former depends only on the current distribution while the latter depends only on the mass distribution. However, the relation turns out to hold for orbital angular momentum in quantum mechanics.

- For spin, the relation above fails, but since the proportionality remains, we can fix it with a fudge factor,

$$\boldsymbol{\mu} = g\frac{q}{2mc}\mathbf{S}.$$

  For example, for the electron, we find $g \approx 2$. For nuclei, the number must be determined experimentally. For the deuteron (one proton and one nucleus), the magnetic moment comes from a combination of the magnetic moments of the proton and neutron, and the orbital motion of the proton. Similarly, the magnetic moment of the proton comes from a mix of the spin and orbital motion of the quarks and gluons. For spin zero particles, like the α particle, $\mathbf{S} = 0$, so $\boldsymbol{\mu} = 0$.

We now show why the experimental facts above make sense.

- Assuming rotational invariance, the spectrum of the Hamiltonian is split into irreps each containing $2j + 1$ degenerate states. Now, since accidental degeneracies are very unlikely, the irreps won't be degenerate; instead, they will be separated by energies on the nuclear energy scale. This energy scale is much larger than the splitting within each irrep induced by an external field; therefore, if the nucleus starts in the ground state, it suffices to only consider the lowest-energy irrep.

- While additional symmetries can cause more degeneracies, these symmetries typically don't appear in real systems.
- The above argument explains the situation for nuclei; for fundamental particles, it turns out that only a single j value is allowed at all. Explaining this requires relativistic quantum mechanics, which we consider much later.
- Supposing that a single irrep is relevant, one can show that every vector operator (i.e. triplet of operators transforming as a vector) is a multiple of $\mathbf{J}$. Since $\boldsymbol{\mu}$ is a vector, this forces $\boldsymbol{\mu} \propto \mathbf{J}$.
- In the case of atoms, the irreps are much closer together, as the atomic energy scale is much smaller than the nuclear energy scale. In this case we do see mixing of irreps for sufficiently strong fields, such as in the strong field Zeeman effect. Each irrep has its own g-factor, so that the total $\boldsymbol{\mu}$ is no longer proportional to the total angular momentum, recovering the classical behavior.

We now consider the example of a spinless particle in three-dimensional space. We again assume rotational symmetry, which in this case means $V = V(r)$.

- We can define angular momentum as $\mathbf{r} \times \mathbf{p}$, but instead we define it as the generator of rotations, which is more fundamental. Let
$$U(R)|\mathbf{x}\rangle = |R\mathbf{x}\rangle.$$
Then it's straightforward to check the $U(R)$ are a unitary representation of $SO(3)$.
- Wavefunctions transform as
$$\psi'(\mathbf{x}) = \psi(R^{-1}\mathbf{x}) \text{ where } |\psi'\rangle = U(R)|\psi\rangle.$$
One way of remembering this rule is to note that if the rotation takes $\mathbf{x}$ to $\mathbf{x'}$, then we must have $\psi'(\mathbf{x'}) = \psi(\mathbf{x})$. This rule is necessary in the active point of view, which we take throughout these notes.
- To find the form of the $\mathbf{L}$ operators, we substitute infinitesimal rotations
$$R(\hat{\mathbf{n}}, \theta) = 1 + \theta \hat{\mathbf{n}} \cdot \mathbf{J}, \quad U(\hat{\mathbf{n}}, \theta) = 1 - \frac{i}{\hbar} \theta \hat{\mathbf{n}} \cdot \mathbf{L}$$
into the above relation, where $\mathbf{J}$ contains the generators of the fundamental representation of $\mathfrak{so}(3)$, as defined earlier. Equating first-order terms in θ, we have
$$\left(-\frac{i}{\hbar}\theta \hat{\mathbf{n}} \cdot \mathbf{L}\right) \psi(\mathbf{x}) = -\theta(\hat{\mathbf{n}} \times \mathbf{x}) \cdot \nabla \psi$$
where we used the property $(\mathbf{a} \cdot \mathbf{J})\mathbf{u} = \mathbf{a} \times \mathbf{u}$. Simplifying,
$$(\hat{\mathbf{n}} \cdot \mathbf{L})\psi = (\hat{\mathbf{n}} \times \mathbf{x}) \cdot \mathbf{p}\psi = \hat{\mathbf{n}} \cdot (\mathbf{x} \times \mathbf{p})$$
which implies $\mathbf{L} = \mathbf{x} \times \mathbf{p}$ as expected.

- We now find the standard angular momentum basis $|lm\rangle$ in the position basis. That is, we are looking for wavefunctions $\psi_{lm}(\mathbf{x})$ such that

$$L^2\psi_{lm} = \hbar^2 l(l+1)\psi_{lm}, \quad L_z\psi_{lm} = \hbar m\psi_{lm}.$$

The easiest way to start is with the stretched state $m = l$, satisfying

$$L_z\psi_{ll} = l\hbar\psi_{ll}, \quad L_+\psi_{ll} = 0.$$

- We know the L_i in Cartesian coordinates; switching to spherical coordinates, we find

$$L_z = -i\hbar\partial_\phi, \quad L_\pm = -i\hbar e^{\pm i\phi}\left(\pm i\partial_\theta - \cot\theta\partial_\phi\right), L^2 = -\hbar^2\left(\frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta) + \frac{1}{\sin^2\theta}\partial_\phi^2\right).$$

- We notice that ∂_r appears nowhere above, which makes sense since angular momentum generates rotations, which keep r constant. Therefore, it suffices to find wavefunctions on the unit sphere, $f(\theta,\phi) = f(\hat{\mathbf{r}})$. We define their inner product by

$$\langle f|g\rangle = \int d\Omega\, f(\theta,\phi)^* g(\theta,\phi), \quad d\Omega = \sin\theta\, d\theta d\phi.$$

As an example, the state $|\mathbf{r}\rangle$ has wavefunction $\delta(\theta-\theta_0)\delta(\phi-\phi_0)/\sin\theta$, where the sine cancels the Jacobian factor in $d\Omega$.

- The solutions for the ψ_{lm} on the sphere are the spherical harmonics Y_{lm}. Using the definition of L_z, we have $Y_{lm} \propto e^{im\phi}$. After solving for Y_{ll}, we apply the lowering operator to find

$$Y_{lm}(\theta,\phi) = \frac{(-1)^l}{2^l l!}\sqrt{\frac{2l+1}{4\pi}\frac{(l+m)!}{(l-m)!}}\frac{e^{im\phi}}{\sin^m\theta}\left(\frac{d}{d(\cos\theta)}\right)^{l-m}\sin^{2l}\theta.$$

Here, the choice of phase factor $(-1)^l$ is conventional and makes Y_{l0} real and positive at the North pole. The $(l+m)!/(l-m)!$ normalization factor comes from the application of L_-.

- We may also write the θ dependence in terms of the Legendre polynomials

$$P_l(x) = \frac{(-1)^l}{2^l l!}\frac{d^l}{dx^l}(1-x^2)^l$$

and the associated Legendre functions

$$P_{lm}(x) = (1-x^2)^{m/2}\frac{d^m P_l(x)}{dx^m}$$

which yields

$$Y_{lm}(\theta,\phi) = (-1)^m\sqrt{\frac{2l+1}{4\pi}\frac{(l+m)!}{(l-m)!}}e^{im\phi}P_{lm}(\cos\theta), \quad m \geq 0$$

where the $m < 0$ spherical harmonics are related by

$$Y_{l,-m} = (-1)^m Y_{lm}^*.$$

- In the above analysis, we have found that precisely one copy of each integer irrep appears, since the solution to $L_+\psi_{ll} = 0$ is unique for each l.

  For a particle in three-dimensional space, the Y_{lm} will be multiplied by a function $u(r)$. Then multiple copies of each irrep may appear, depending on how many solutions there are for $u(r)$, and we must index the states by a third quantum number (e.g. n for the hydrogen atom).

- The spherical harmonics are then our standard angular momentum basis $|lm\rangle$. We can find an identity by computing $\langle \hat{\mathbf{r}}|U(R)|lm\rangle$ in two different ways. Acting on the right, we have $Y_{lm}(R^{-1}\hat{\mathbf{r}})$. Alternatively, we may insert an identity for

$$\sum_{m'} \langle \hat{\mathbf{r}}|lm'\rangle\langle lm'|U(R)|lm\rangle = \sum_{m'} Y_{lm'}(\hat{\mathbf{r}}) D^l_{m'm}(R).$$

  Here, we only needed to insert states with the same l since they form an irrep. Then

$$Y_{lm}(R^{-1}\hat{\mathbf{r}}) = \sum_{m'} Y_{lm'}(\hat{\mathbf{r}}) D^l_{m'm}(R).$$

- One useful special case of the above is to choose $\hat{\mathbf{r}} = \hat{\mathbf{z}}$ and replace R with R^{-1}, for

$$Y_{lm}(\hat{\mathbf{r}}) = \sum_{m'} Y_{lm'}(\hat{\mathbf{z}}) D^l_{m'm}(R^{-1})$$

  where R is the rotation that maps $\hat{\mathbf{z}}$ to $\hat{\mathbf{r}}$, i.e. the one with Euler angles $\alpha = \phi$ and $\beta = \theta$. Moreover, only the $m = 0$ spherical harmonic is nonzero at $\hat{\mathbf{z}}$ (because of the centrifugal force), and plugging it in gives

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}} D^{l*}_{m0}(\phi, \theta, 0)$$

  where we applied the unitarity of the D matrices.

- For a multiparticle system, with state space $|\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$, the angular momentum operator is $\mathbf{L} = \sum \mathbf{x}_i \times \mathbf{p}_i$. To construct the angular momentum basis, we use addition of angular momentum techniques, as discussed later.

6.4 Central Force Motion

We now apply the results of the previous section to central force motion.

- Consider a spinless particle moving in a central potential. Since L^2 and L_z commute with H, the eigenstates are of the form

$$\psi(r, \theta, \phi) = R(r) Y_{lm}(\theta, \phi).$$

  Substituting this into the Schrodinger equation, and noting that L^2 is $-\hbar^2/r^2$ times the angular part of the Laplacian, we have

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\partial_r(r^2\partial_r R) + UR = ER, \quad U(r) = V(r) + \frac{l(l+1)\hbar^2}{2mr^2}$$

  where the extra contribution to the effective potential $U(r)$ is equal to $L^2/2mr^2$. As in the classical case, this is the angular part of the kinetic energy.

- Next, we let $f(r) = rR(r)$. This is reasonable, because then $|f|^2$ gives the radial probability density, so we expect this should simplify the radial kinetic energy term. Indeed we have

$$-\frac{\hbar^2}{2m}\frac{d^2 f(r)}{dr^2} + U(r)f(r) = Ef(r), \quad \int_0^\infty dr\, |f(r)|^2 = 1.$$

  The resulting equation looks just like the regular 1D Schrodinger equation, but on $(0, \infty)$.

- We could also have arrived at this conclusion using separation of variables. Generally, this technique works when there is a continuous symmetry. Then the (differential) operator that generates this symmetry commutes with the Hamiltonian, and we can take the eigenfunctions to be eigenfunctions of that operator. In an appropriate coordinate system (i.e. when fixing some of the coordinates gives an orbit of the symmetry) this automatically gives separation of variables; for example, L_z generates rotations which change only ϕ, so diagonalizing L_z separates out the coordinate ϕ.

- As another example, the free particle separates in Cartesian coordinates by conservation of linear momentum. The hydrogen atom has a hidden $SO(4)$ symmetry, so it can be separated in confocal parabolic coordinates, in addition to spherical coordinates.

- We index the radial solutions for a given l by n, giving

$$\psi_{nlm}(r, \theta, \phi) = R_{nl}(l)Y_{lm}(\theta, \phi).$$

  These account for the bound states; there also may be unbound states with a continuous spectrum. Focusing on just the bound states, the irreps are indexed by n and l and each contain $2l + 1$ states.

- There generally is no degeneracy in l unless there is additional symmetry; this occurs for the hydrogen atom (hidden $SO(4)$ symmetry) and the 3D harmonic oscillator ($SU(3)$ symmetry).

- Next, we consider degeneracy in n, i.e. degenerate eigenfunctions $f(r)$ of the same effective potential. These eigenfunctions satisfy the same Schrodinger equation (with the same energy E and effective potential $U(r)$), so there can be at most two of them, as the Schrodinger equation is second-order. However, as we'll show below, we must have $f(0) = 0$, which effectively removes one degree of freedom – eigenfunctions are solely determined by $f'(0)$. Therefore there is only one independent solution for each energy, bound or not, so different values of n are nondegenerate. (In the bound case, we can also appeal to the fact that f vanishes at infinity.)

  Therefore we conclude irreps are generically nondegenerate.

- We now consider the behavior of $R(r)$ for small r. If $R(r) \sim ar^k$ for small r, then the terms in the reduced (1D) Schrodinger equation scale as:
  - Radial kinetic energy: $-a(\hbar^2/2m)k(k+1)r^{k-2}$.
  - Centrifugal potential: $a(\hbar^2/2m)l(l+1)r^{k-2}$.
  - Potential energy: $aV(r)r^k$.
  - Right-hand side: aEr^k.

If we suppose the potential is regular at the origin and diverges no faster than $1/r$, then the last two terms are negligible. Then for the equation to remain true, the first two terms must cancel, so

$$k(k+1) = l(l+1), \quad k = l \text{ or } k = -l-1.$$

The second solution is nonnormalizable for $l \geq 1$, so we ignore it. For $l = 0$, it gives $R(r) \propto 1/r$, which is the solution for the delta function potential, which we have ruled out by regularity. Therefore the first solution is physical,

$$R(r) \sim r^l \text{ for small r}$$

and hence $f(0) = 0$ in general.

Now we consider some important examples of central force motion.

Example. Two-body interactions. Suppose that two massive bodies interact with Hamiltonian

$$H = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + V(|\mathbf{x}_1 - \mathbf{x}_2|).$$

In this case it's convenient to switch to the coordinates

$$\mathbf{R} = \frac{m_1\mathbf{x}_1 + m_2\mathbf{x}_2}{M}, \quad \mathbf{r} = \mathbf{x}_2 - \mathbf{x}_1$$

where $M = m_1 + m_2$. Defining the conjugate momenta $\mathbf{P} = -i\hbar\partial_{\mathbf{R}}$ and $\mathbf{p} = -i\hbar\partial_{\mathbf{r}}$, we have

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2, \quad \mathbf{p} = \frac{m_1\mathbf{p}_2 - m_2\mathbf{p}_1}{M}.$$

This transformation is an example of a canonical transformation, as it preserves the canonical commutation relations. The Hamiltonian becomes

$$H = \frac{P^2}{2M} + \frac{p^2}{2\mu} + V(r), \quad \frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}.$$

We see that $P^2/2M$ commutes with H, so we can separate out the variable R, giving the overall center-of-mass motion. We then focus on the wavefunction of the relative coordinate, $\psi(\mathbf{r})$. This satisfies the same equation as a single particle in a central force, with m replaced with μ.

Finally, we may decompose the total angular momentum $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$ into

$$\mathbf{L} = \mathbf{R} \times \mathbf{P} + \mathbf{r} \times \mathbf{p}$$

which is a 'orbit' plus 'spin' (really, 'relative') contribution, just as in classical mechanics. The relative contribution commutes with the relative-coordinate Hamiltonian $p^2/2\mu + V(r)$, so the quantum numbers l and m in the solution for $\psi(\mathbf{r})$ refer to the angular momentum of the particles in their CM frame.

Example. The rigid rotor. Consider two masses m_1 and m_2 connected with a massless, rigid rod of length r_0. The Hamiltonian is

$$H = \frac{L^2}{2I}, \quad I = \mu r_0^2.$$

Since the length r_0 is fixed, there is no radial dependence; the solution is just

$$E_l = \frac{l(l+1)\hbar^2}{2\mu r_0^2}, \quad \psi_{lm}(\theta, \phi) = Y_{lm}(\theta, \phi).$$

This can also be viewed as a special case of the central force problem, with a singular potential.

Example. Diatomic molecules. For a typical diatomic molecule, such as CO, the reduced mass is on the order of several times the atomic mass, so the rotational energy levels are much more closely spaced than the atomic levels. (Here, we treat the two atoms as point particles; this is justified by the Bohr-Oppenheimer approximation, which works because the atomic degrees of freedom are faster, i.e. higher energy.) There are also vibrational degrees of freedom due to oscillations in the separation distance between the atoms.

To estimate the energy levels of the vibrational motion, we use dimensional analysis on the parameters m, e, and $\hbar$, where m and e are the mass and charge of the electron; this is reasonable because valence electrons are responsible for bonding. We don't use c, as the situation is nonrelativistic. We find the following units:

- Distance: $a_0 = \hbar^2/me^2 \approx 0.5\,\text{Å}$, the Bohr radius.
- Energy: $K_0 = e^2/a_0 = me^4/\hbar^2 \approx 27\,\text{eV}$, twice the Rydberg constant.
- Velocity: $v_0 = e^2/\hbar = \alpha c$, which confirms the motion is nonrelativistic.

Now, we estimate the diatomic bond as a harmonic oscillator near its minimum. Assuming that the 'spring constant' of the bond is about the same as the 'spring constant' of the bond between the valence electrons and their own atoms (which makes sense since the bond is covalent), and using $\omega \propto 1/\sqrt{m}$, we have

$$\omega_{\text{vib}} = \sqrt{\frac{m}{M}}\omega_0, \quad \omega_0 = K_0/\hbar$$

where M is the reduced mass, on the order of $10,000m$. Therefore the vibrational energy level spacing is about 100 times closer than the electronic energy level spacing, or equivalently the bond dissociation energy. The rotational energy levels have a different dependence, as

$$\Delta E_{\text{rot}} = \frac{\hbar^2}{2I} \sim \frac{\hbar^2}{Ma_0^2} = \frac{m}{M}K_0 \sim 10^{-4}K_0.$$

That is, the rotational levels are another factor of 100 times closer spaced than the vibrational ones. At room temperature, the rotational levels are active, and the rotational levels are partially or completely frozen out, depending on the mass of the atoms involved.

Example. Hydrogen. We consider a spinless, electrostatic, nonrelativistic model, with potential

$$V(r) = -\frac{Ze^2}{r}.$$

The radial Schrodinger equation is

$$-\frac{\hbar^2}{2\mu}\frac{d^2 f}{dr^2} + \left(\frac{l(l+1)\hbar^2}{2\mu r^2} - \frac{Ze^2}{r}\right) f = Ef$$

where μ is the reduced mass. In this potential, the atomic units above are slightly modified.

- The characteristic distance is $a = \hbar^2/me_{\text{el}}e_{\text{nuc}} = a_0/Z$, so the electrons orbit closer.
- The characteristic energy is $K = e_{\text{el}}e_{\text{nuc}}/a = Z^2K_0$, so the energies are higher.
- The characteristic velocity is $v = e_{\text{el}}e_{\text{nuc}}/\hbar = Zv_0 = (Z\alpha)c$, so for heavy nuclei, the nonrelativistic approximation breaks down.

In particular, taking distance and energy in units of a and K, we have

$$\frac{d^2 f}{dr^2} + \left(-\frac{l(l+1)}{r^2} + \frac{2}{r} + 2E\right) f = 0.$$

There are both bound states and free states in the spectrum. Searching for bound states, we change radial variable to

$$\nu = \frac{1}{\sqrt{-2E}}, \quad \rho = \frac{2r}{\nu}$$

which reduces the equation to

$$\frac{d^2 f}{d\rho^2} + \left(-\frac{l(l+1)}{\rho^2} + \frac{\nu}{\rho} - \frac{1}{4}\right) f = 0.$$

We now solve the equation by standard methods. First, we take the high ρ limit to find the asymptotic behavior for normalizable solutions, $f \propto \rho^{l+1} e^{-\rho/2}$. Next, we peel off the asymptotic behavior and express the error term in a power series in ρ. For the overall solution to remain normalizable, the power series must terminate, and the condition is

$$\nu = n \in \mathbb{Z}, \quad l < n.$$

We call n the principal quantum number. The solutions for f are polynomials times the exponential $e^{-\rho/2}$, and the energies are $E_n = -1/2n^2$ independent of l. Therefore we have n^2 degeneracy for each value of n.

The bound $l < n$ can be understood classically. For a planet orbiting a star with a fixed energy (and hence fixed semimajor axis), there is a highest possible angular momentum corresponding to $l \approx n$ (in the appropriate units), corresponding to a circular orbit. The analogous quantum states have $f(\rho)$ peaked around a single value. The low angular momentum states correspond to long, thin ellipses, and indeed the analogous $f(\rho)$ extend further out with multiple nodes.

Many perturbations break the degeneracy in l. For example, consider the case of an alkali atom with one valence electron; then the potential interpolates between $-e^2/r$ at long distances and $-Ze^2/r$ at short distances. Then orbits which approach the core are lowered in energy, and this happens more for low values of l. For example, in sodium, this effect makes the $3s$ state significantly lower in energy than the $3p$ state.

For reference, we summarize facts about special functions and the contexts in which they appear.

- The most general equation we consider is the time-independent Schrodinger equation,

  $$-\nabla^2\psi + V\psi = E\psi$$

  which comes from separating the ordinary Schrodinger equation. We only consider the rotationally symmetric case $V = V(r)$.

- If we separate the wave equation, the spatial part is the Helmholtz equation, which is the special case $V = 0$ above. If we further set $E = 0$ above, we get Laplace's equation, whose solutions are harmonic functions. These represent static solutions of the wave equation.

- It only makes sense to add source terms to full PDEs, not separated ones, so we shouldn't add sources to the time-independent Schrodinger equation or the Helmholtz equation. By contrast, Laplace's equation is purely spatial, and adding a source term gives Poisson's equation.

* By rotational symmetry, the time-independent Schrodinger equation separates into a radial and angular part. The angular solutions are the eigenfunctions of L^2, the angular part of the Laplacian, and are called spherical harmonics.
  - The spherical harmonics $Y_{\ell m}(\theta, \phi)$ form a complete basis for functions on the sphere. The quantity ℓ can take on nonnegative integer values.
  - They are proportional to $e^{im\phi}$ times an associated Legendre function $P_\ell^m(\cos\theta)$.
  - Setting $m = 0$ gives the Legendre polynomials, which are orthonormal on $[-1, 1]$.
  - More generally, the associated Legendre functions satisfy orthogonality relations which, combined with those for $e^{im\phi}$, ensure that the spherical harmonics are orthogonal.
  - Spherical harmonics are *not* harmonic functions on the sphere, which have zero L^2 eigenvalue – the only such function is the constant function Y_{00}.
  - If we were working in two dimensions, we'd just get $e^{im\theta}$.
* The radial equation depends on the potential $V(r)$ and the total angular momentum ℓ, which contributes a centrifugal force term.
  - For $V = 0$, the solutions are spherical Bessel functions, $j_\ell(r)$ and $y_\ell(r)$. They are called Bessel functions of the first and second kind; the latter are singular at $r = 0$.
  - For high r, the Bessel functions asymptote to sinusoids with amplitude $1/r$. (As a special case, setting $\ell = 0$ gives $j_0(r) = \sin(r)/r$, $y_0(r) = \cos(r)/r$, recovering the familiar form of an isotropic spherical wave.)
  - If we were working in two dimensions, we would instead get the ordinary, or cylindrical Bessel functions. We also define the Hankel functions in terms of linear combinations of Bessel functions to correspond to incoming and outgoing waves at infinity.
  - For a Coulomb field, the solutions are exponentials times associated Laguerre polynomials. Again, there are two solutions, with exponential growth and decay, but only the decaying solution is relevant for bound states.
* Our results also apply to Laplace's equation, in which case the radial equation yields solutions r^ℓ and $1/r^{\ell+1}$. These are the small-r limits of the spherical Bessel functions, because near the origin the energy term $E\psi$ is negligible compared to the centrifugal term.
* As an application, applying this decomposition to the potential created by a charge distribution near the origin yields the multipole expansion, with $\ell = 0$ giving the monopole contribution, and so on.

6.5 Addition of Angular Momentum

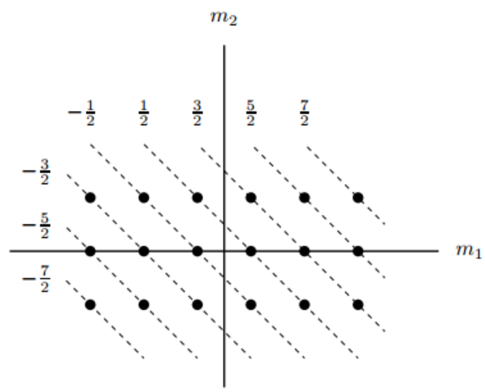
We now discuss addition of angular momentum.

* Consider two Hilbert spaces with angular momentum operators $\mathbf{J}_1$ and $\mathbf{J}_2$. Then the tensor product space has angular momentum operator

$$\mathbf{J} = \mathbf{J}_1 \otimes 1 + 1 \otimes \mathbf{J}_2 = \mathbf{J}_1 + \mathbf{J}_2.$$

  The goal is to relate the angular momentum basis of the joint system $|jm\rangle$ in terms of the uncoupled angular momentum basis $|j_1 m_1\rangle \otimes |j_2 m_2\rangle = |j_1 m_1 j_2 m_2\rangle$.

- It suffices to consider the tensor product of two irreps; for concreteness, we consider $\frac{5}{2} \otimes 1$. The J_z eigenvalue is just $m_1 + m_2$, so the m eigenvalues of the uncoupled basis states are:



- To find the coupled angular momentum basis, we first consider the state $|\frac{5}{2}\frac{5}{2}\rangle \otimes |11\rangle$, which has $m = 7/2$. This state is a one-dimensional eigenspace of J_z. However, since J_z commutes with J^2, it must also be a one-dimensional eigenspace of J^2, so it has a definite j value. Since there are no states with higher m, we must have $j = 7/2$, so $|\frac{5}{2}\frac{5}{2}11\rangle = |\frac{7}{2}\frac{7}{2}\rangle$.

- Next, we may apply the total lowering operator to give $|\frac{7}{2}\frac{5}{2}\rangle$. There are two states with $m = 5/2$, and hence by similar reasoning, the orthogonal state with $m = 5/2$ must be an eigenstate of J^2, so it is $|\frac{5}{2}\frac{5}{2}\rangle$.

- Continuing this process, lowering our basis vectors and finding new irreps by orthogonality, we conclude that $\frac{5}{2} \otimes 1 = \frac{3}{2} \oplus \frac{5}{2} \oplus \frac{7}{2}$. By very similar reasoning, we generally have

$$j_1 \otimes j_2 = |j_1 - j_2| \oplus |j_1 - j_2| + 1 \oplus \cdots \oplus j_1 + j_2.$$

- We define the Clebsch-Gordan coefficients as the overlaps $\langle j_1 j_2 m_1 m_2 | jm \rangle$. These coefficients satisfy the relations

$$\sum_{m_1 m_2} \langle jm | j_1 j_2 m_1 m_2 \rangle \langle j_1 j_2 m_1 m_2 | j' m' \rangle = \delta_{jj'} \delta_{mm'},$$

$$\sum_{jm} \langle j_1 j_2 m_1 m_2 | jm \rangle \langle jm | j_1 j_2 m_1' m_2' \rangle = \delta_{m_1 m_1'} \delta_{m_2 m_2'}$$

which simply follow from completeness of the coupled and uncoupled bases. In addition we have the selection rule

$$\langle jm | j_1 j_2 m_1 m_1 \rangle \propto \delta_{m, m_1 + m_2}.$$

We may also obtain recurrence relations for the Clebsch-Gordan coefficients by applying J_- in both the coupled and uncoupled bases.

- Next, we consider the operation of rotations. Since $\mathbf{J}_1$ and $\mathbf{J}_2$ commute,

$$U(\hat{\mathbf{n}}, \theta) = e^{-i\theta \hat{\mathbf{n}} \cdot (\mathbf{J}_1 + \mathbf{J}_2)/\hbar} = U_1(\hat{\mathbf{n}}, \theta) U_2(\hat{\mathbf{n}}, \theta)$$

where the U_i are the individual rotation operators. Then

$$U|j_1j_2m_1m_2\rangle = \sum_{jm}\sum_{m'}|jm'\rangle D^j_{m'm}\langle jm|j_1j_2m_1'm_2'\rangle$$

in the coupled basis, and

$$U|j_1j_2m_1m_2\rangle = U_1|j_1m_1\rangle U_2|j_2m_2\rangle = \sum_{m_1'm_2'}|j_1j_2m_1'm_2'\rangle D^{j_1}_{m_1'm_1}D^{j_2}_{m_2'm_2}$$

in the uncoupled basis. Combining these and relabeling indices, we have

$$D^{j_1}_{m_1m_1'}D^{j_2}_{m_2m_2'} = \sum_{jmm'}\langle j_1j_2m_1m_2|jm\rangle D^j_{mm'}\langle jm'|j_1j_2m_1'm_2'\rangle$$

which allows products of D matrices to be reduced.

Example. Combining spin and spatial degrees of freedom for the electron. We must work in the tensor product space with basis $|\mathbf{r}, m\rangle$. Wavefunctions are of the form

$$\psi(\mathbf{r}, m) = \langle\mathbf{r}, m|\psi\rangle$$

which is often written in the notation

$$\psi(\mathbf{r}) = \begin{pmatrix} \psi_s(\mathbf{r}) \\ \psi_{s-1}(\mathbf{r}) \\ \vdots \\ \psi_{-s}(\mathbf{r}) \end{pmatrix}$$

which has a separate wavefunction for each spin component, or equivalently, a spinor for every position in space. The inner product is

$$\langle\phi|\psi\rangle = \sum_m \int d^3\mathbf{r}\,\phi^*(\mathbf{r}, m)\psi(\mathbf{r}, m).$$

In the case of the electron, the Hamiltonian is the sum of the spatial and spin Hamiltonians we have considered before,

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{q}{c}\mathbf{A}\right)^2 + q\phi - \boldsymbol{\mu}\cdot\mathbf{B}, \quad \boldsymbol{\mu} = \frac{g}{2}\mu\boldsymbol{\sigma}.$$

This is called the Pauli Hamiltonian and the resulting evolution equation is the Pauli equation. Note that the only part that couples different spins is the $\boldsymbol{\mu}\cdot\mathbf{B}$ term.

6.6 Tensor Operators

Classically, we say the position $\mathbf{x}$ is a vector because of how it transforms under rotations. In quantum mechanics, observables correspond to operators, motivating us to consider how operators transform under rotations.

- States transform as $|\psi\rangle \to |\psi'\rangle = U(R)|\psi\rangle$. Under a rotation, an operator A becomes

  $$A' = U(R)AU(R)^\dagger$$

  so that $\langle\psi'|A'|\psi'\rangle = \langle\psi|A|\psi\rangle$.

- A scalar operator K is any operator invariant under rotations, $K' = K$. Therefore K commutes with all rotations, or, taking the case of an infinitesimal rotation, K commutes with $\mathbf{J}$. One important example is the Hamiltonian is a central force problem.
- A vector operator $\mathbf{V}$ is a triplet of operators satisfying

$$\langle \psi' | \mathbf{V} | \psi' \rangle = R \langle \psi | \mathbf{V} | \psi \rangle.$$

  That is, $\mathbf{V}$ corresponds to a classical vector quantity. Expanding in components yields

$$U(R) V_i U(R)^\dagger = V_j R_{ji}.$$

  Taking infinitesimal rotations on both sides gives the commutation relations

$$[J_i, V_j] = i\hbar \epsilon_{ijk} V_k$$

  which serves as an alternate definition of a vector operator.
- Similarly, we may show that the dot product of vector operators is a scalar operator, the cross product is a vector operator, and so on. For example, for orbital angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, both $\mathbf{r}$ and $\mathbf{p}$ are vector operators. Hence $\mathbf{L}$ itself is a vector operator, while p^2 is a scalar operator. More generally, the adjoint formula shows that angular momentum is always a vector operator.
- Similarly, we define a rank-2 tensor operator as one that transforms by

$$U(R) T_{ij} U(R)^\dagger = T_{kl} R_{ki} R_{lj}.$$

  For example, the outer product of vector operators $T_{ij} = V_i W_j$ is a tensor operator. A physical example of a rank-2 tensor operator is the quadrupole moment.

Next, we turn to an apparently unrelated subject: the spherical basis of $\mathbb{R}^3$.

- Starting with the Cartesian basis $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$, we define the spherical basis vectors

$$\hat{\mathbf{e}}_1 = -\frac{\hat{\mathbf{x}} + i\hat{\mathbf{y}}}{\sqrt{2}}, \quad \hat{\mathbf{e}}_0 = \hat{\mathbf{z}}, \quad \hat{\mathbf{e}}_{-1} = \frac{\hat{\mathbf{x}} - i\hat{\mathbf{y}}}{\sqrt{2}}.$$

  We may expand vectors in this basis (or technically, the basis $\hat{\mathbf{e}}_q^*$) by

$$\mathbf{X} = \hat{\mathbf{e}}_q^* X_q, \quad X_q = \hat{\mathbf{e}}_q \cdot \mathbf{X}$$

- As an example application, consider calculating the dipole transition rate, which is proportional to $\langle n'\ell'm' | \mathbf{x} | n\ell m \rangle$. This is messy, but a simplification occurs if we expand $\mathbf{x}$ in the spherical basis, because

$$r Y_{1q}(\Omega) = \sqrt{\frac{3}{4\pi}} x_q.$$

  Then the matrix element factors into an angular and radial part,

$$\langle n'\ell'm' | x_q | n\ell m \rangle = \int_0^\infty r^2 dr\, R^*_{n'\ell'}(r) r R_{n\ell}(r) \times \sqrt{\frac{4\pi}{3}} \int d\Omega\, Y^*_{\ell'm'}(\Omega) Y_{1q}(\Omega) Y_{\ell m}(\Omega).$$

  This is a substantial improvement: we see that n and n' only appear in the first factor, while m and m' only appear in the second.

- The 'coincidence' arises because both the spherical harmonics and spherical basis arise out of the representation theory of $SU(2)$. The Y_{lm}'s are the standard angular momentum basis for the action of rotations on functions on the sphere. Similarly, the spherical basis is the standard angular momentum basis for the action of rotations in space, which carries the representation $j = 1$.

- More generally, tensor quantities carry representations of $SO(3)$ classically, and hence tensor operators carry representations of $SU(2)$ in quantum mechanics. Hence it is natural for the photon, which is represented by the vector $\mathbf{A}$ classically, to have spin 1.

- Tensor operators can be broken down into irreps. Scalar and vector operators are already irreps, but the tensor operator $T_{ij} = V_i W_j$ contains the scalar and vector irreps

  $$\operatorname{tr} T = \mathbf{V} \cdot \mathbf{W}, \quad \mathbf{X} = \mathbf{V} \times \mathbf{W}.$$

  The remaining degrees of freedom form a five-dimensional irrep, the symmetric traceless part of T_{ij}. This is in accordance with the Clebsch-Gordan decomposition $1 \otimes 1 = 0 \oplus 1 \oplus 2$. The same decomposition holds for arbitrary T_{ij} by linearity.

- Irreps in the standard basis transform by the same D matrices that we introduced earlier. For example, an irreducible tensor operator of order k is a set of $2k+1$ operators T_q^k satisfying

  $$U T_q^k U^\dagger = T_{q'}^k D_{q'q}^k(U).$$

  An irreducible tensor operator of order k transforms like a spin j particle. In our new language, writing $\mathbf{x}$ in terms of the x_q is just writing it as an irreducible tensor operator of order 1.

- Rotations act on kets by multiplication by $U(R)$, while rotation act on operators by conjugation, which turns into commutation for infinitesimal rotations. Therefore the angular momentum operators affect the irreducible tensor operator T_q^k exactly as they affect the kets $|kq\rangle$, but with commutators,

  $$[J_z, T_q^k] = \hbar k T_q^k, \quad [J_i, [J_i, T_q^k]] = \hbar^2 k(k+1) T_q^k.$$

  We don't even have to prove this independently; it just carries over from our previous work.

- In the case of operators, there's no simple 'angular momentum operator' as in the other cases, because it would have to be a superoperator, i.e. a linear map of operators.

Note. The ideas above can be used to understand higher spherical harmonics as well. The functions x, y, and z form an irrep under rotations, and hence the set of homogeneous second-order polynomials forms a representation as well. Using the decomposition $1 \otimes 1 = 0 \oplus 1 \oplus 2$ yields a five-dimensional irrep, and dividing these functions by r^2 yields the $\ell = 2$ spherical harmonics.

This explains the naming of chemical orbitals. The p orbitals are p_x, p_y, and p_z, corresponding to angular parts x/r, y/r, and z/r. Note that this is not the standard angular momentum basis; the functions are instead chosen to be real and somewhat symmetrical. The names of the d orbitals are similar, though d_{z^2} should actually be called $d_{3z^2-r^2}$.

We now introduce the Wigner-Eckart theorem, which simplifies matrix elements of irreducible tensor operators.

- Consider a setup with rotational symmetry, and work in the basis $|\gamma j m\rangle$. A scalar operator K commutes with both J_z and J^2, and hence preserves j and m. Moreover, since it commutes with $J_\pm$, its matrix elements do not depend on m,

$$\langle \gamma' j' m' | K | \gamma j m \rangle = \delta_{j'j} \delta_{m'm} C^j_{\gamma'\gamma}.$$

  This implies, for instance, that the eigenvalues come in multiplets of degeneracy $2j+1$. We've already seen this reasoning before, for the special case $K = H$, but the result applies for any scalar operator in any rotationally symmetric system.

- More generally, the Wigner-Eckart theorem states that

$$\langle \gamma' j' m' | T^k_q | \gamma j m \rangle = \langle \gamma' j' || T^k || \gamma j \rangle \langle j' m' | j k m q \rangle$$

  where the first factor is called a reduced matrix element, and the second is a Clebsch-Gordan coefficient. The reduced matrix element is not a literal matrix element, but just stands in for a quantity that only depends on k and the γs and js.

- The Wigner-Eckart theorem factors the matrix element into a part that depends only on the irreps, and a part that depends on the m's that label states inside the irreps. This simplifies the computation of transition rates, as we saw earlier. Fixing the γ's and j's, there are generally $(2j+1)(2j'+1)(2k+1)$ matrix elements to compute, but we can just compute one (to get the reduced matrix element) and use Clebsch-Gordan coefficients to find the rest.

- The intuition for the Clebsch-Gordan coefficient is that $T^k_q |jm\rangle$ transforms like the ket $|kq\rangle|jm\rangle$. The Clebsch-Gordan factor also provides several selection rules,

$$m' = m + q, \quad j' \in \{|j-k|, \ldots, j+k\}.$$

- If there is only one irrep, then all irreducible tensor operators of order k must be proportional to each other. This provides a proof for why we must have $\boldsymbol{\mu} \propto \mathbf{S}$ for spins.

Example. An alpha particle is a nucleus whose ground state has spin zero. Restricting our Hilbert space to this irrep, the selection rules show that every irreducible tensor operator with $k > 0$ must be zero. In particular, this means that alpha particles cannot have a magnetic dipole moment.

7 Discrete Symmetries

7.1 Parity

In the previous section, we studied proper rotations. We now add in parity, an improper rotation, and consider its representations.

- In classical mechanics, the parity operator P inverts all spatial components. It has matrix representation $-I$, satisfies $P^2 = I$, and commutes with all proper rotations, $PRP^{-1} = R$.

- In quantum mechanics, we look for a parity operator $\pi = U(P)$ which satisfies

$$\pi^\dagger \pi = 1, \quad \pi^2 = 1, \quad \pi U(R) \pi^\dagger = U(R).$$

  Mathematically, these conditions mean that we are looking for unitary representations of $O(3)$. Combining the first postulates show that π is Hermitian, so the parity is observable. The third postulate is equivalent to $[\pi, \mathbf{J}] = 0$, i.e. that π is a scalar operator.

- The above postulates rule out projective representations. These are allowed in principle, but won't be necessary for any of our applications.

- For a spinless particle, we have previously defined $U(R)|\mathbf{x}\rangle = |R\mathbf{x}\rangle$. Similarly, we may define $\pi|\mathbf{x}\rangle = -|\mathbf{x}\rangle$, which obeys all of the postulates above. We may also explicitly compute

$$\pi \mathbf{x} \pi^\dagger = -\mathbf{x}, \quad \pi \mathbf{p} \pi^\dagger = -\mathbf{p}, \quad \pi \mathbf{L} \pi^\dagger = \mathbf{L}$$

  where $\mathbf{L}$ is the orbital angular momentum $\mathbf{r} \times \mathbf{p}$. the parity of the state $|lm\rangle$ is $(-1)^l$.

- Another example is a spin-s particle with no spatial wavefunction. The states are $|sm\rangle$ for $m = -s, \ldots, s$. Since π is a scalar operator, we must have

$$\pi|sm\rangle = \eta|sm\rangle$$

  for some constant $\eta = \pm 1$. In nonrelativistic quantum mechanics, the sign has no physical consequences, so we choose $\eta = 1$ so that parity does nothing to the spin state. Adding back the spatial degrees of freedom gives $\pi|\mathbf{x}, m\rangle = |-\mathbf{x}, m\rangle$.

- In relativistic quantum mechanics, the sign of η makes a physical difference because particle number can change, but the overall parity must be conserved; this provides some selection rules. For example, the fact that the photon has negative parity is related to the fact that the parity of an atom flips during an electric dipole transition, which involves one photon.

- Given a vector operator $\mathbf{V}$, if

$$\pi \mathbf{V} \pi^\dagger = \pm \mathbf{V}$$

  then we say $\mathbf{V}$ is a true/polar vector if the sign is -1, and a pseudovector/axial vector if the sign is $+1$. For example, $\mathbf{x}$ and $\mathbf{p}$ are polar vectors but $\mathbf{L}$ is an axial vector.

- Similarly, for a scalar operator K, if

$$\pi K \pi^\dagger = \pm K$$

  then K is a true scalar if the sign is $+1$ and a pseudoscalar if the sign is -1. For example, $\mathbf{p} \cdot \mathbf{S}$ is a pseudoscalar.

- Note that $\mathbf{E}$ is a polar vector while $\mathbf{B}$ is an axial vector. In particular, adding an external magnetic field does not break parity symmetry.

Next, we consider the consequences of parity symmetry of the Hamiltonian.

- Parity is conserved if $[\pi, H] = 0$. This is satisfied by the central force Hamiltonian, and more generally to any system of particles interacting by pairwise forces of the form $V(|\mathbf{r}_i - \mathbf{r}_j|)$.

- Parity remains conserved when we account for relativistic effects. For example, such effects lead to a spin-orbit coupling $\mathbf{L} \cdot \mathbf{S}$, but this term is a true scalar. Parity can appear to be violated when photons are emitted (or generally when a system is placed in an external field), but remains conserved as long as we account for the parity of the electromagnetic field.

- Parity is also conserved by the strong interaction, but not by the weak interaction. The weak interaction is extremely weak at atomic energy scales, so parity symmetry is extremely accurate in atomic physics.

- Just like rotational symmetry, parity symmetry can lower the dimensionality of a system. If $[\pi, H] = 0$, then we can split the Hilbert space into representations with $+1$ and -1 parity and diagonalize H within them separately, which is more computationally efficient.

- In the case of rotational symmetry, every rotational irrep has definite parity since π is a scalar operator. In particular, if there is no degeneracy of irreps, then every energy eigenstate is automatically a parity eigenstate.

  For example, in hydrogen, the $2s$ and $2p$ irreps are degenerate, with even and odd parity. A linear combination of these states gives an energy eigenstate without definite parity.

- As another example, consider one-dimensional motion in an even potential. Assuming no degeneracy (which is true for bound states), every eigenstate is either even or odd. However, for the free particle we have the eigenstates $e^{\pm ikx}$ which do not have definite parity; the combinations $\sin(kx)$ and $\cos(kx)$ do.

Example. Selection rules for electric dipole transitions. Such a transition is determined by the matrix element $\langle n'\ell'm'|\mathbf{x}|n\ell m\rangle$. It must be parity invariant, but under parity it picks up a factor of $(-1)^{\ell+\ell'+1}$, giving the selection rule $\Delta\ell = \text{odd}$. The Wigner-Eckart theorem rules out $|\Delta\ell| > 1$, so we must have $\Delta\ell = \pm 1$. The Wigner-Eckart theorem also gives $|\Delta m| \leq 1$.

Example. A spin-orbit coupling. Consider a particle with spatial state $|n\ell m_\ell\rangle$, which separates into a radial and angular part $|n\ell\rangle|\ell m_\ell\rangle$, and a spin state $|sm_s\rangle$. Ignoring the radial part, which separates out, we consider the total spin states

$$|\ell j m_j\rangle = \sum_{m_\ell, m_s} |\ell m_\ell\rangle|sm_s\rangle\langle \ell s m_\ell m_s|jm_j\rangle.$$

The wavefunction of such a state takes in an angular coordinate and outputs a spinor. A spin-orbit coupling is of the form $\boldsymbol{\sigma} \cdot \mathbf{x}$. Since this term is rotationally invariant, it conserves j and m_j. From the standpoint of the spatial part, it's like an electric dipole transition, so $\Delta\ell = \pm 1$. Thus the interaction can transfer angular momentum between the spin and orbit, one unit at a time.

7.2 Time Reversal

Next, we consider time reversal, which is more subtle because it is realized by an antilinear operator. We begin with the classical case.

- In Newtonian mechanics, if $\mathbf{x}(t)$ is a valid trajectory for a particle in a potential (such as an external electric field), $\mathbf{x}(-t)$ is a valid trajectory as well. Since the velocity is flipped, the momentum is also flipped, $\mathbf{p}(t) \to -\mathbf{p}(-t)$.
- This reasoning fails in the case of an external magnetic field. However, if we consider the field to be internally generated by charges in the system, then time reversal takes

$$\rho \to \rho, \ \mathbf{J} \to -\mathbf{J}, \quad \mathbf{E} \to \mathbf{E}, \ \mathbf{B} \to -\mathbf{B},$$

  where we suppress time coordinates. This gives an extra sign flip that restores the symmetry.
- Note that this is the opposite of the situation with parity. In this case, $\mathbf{J}$ is flipped as well, but $\mathbf{E}$ is flipped while $\mathbf{B}$ isn't.
- In the case of quantum mechanics, we have the Schrodinger equation

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

  It is tempting to implement time reversal by taking $\psi(\mathbf{x}, t) \to \psi(\mathbf{x}, -t)$, but this doesn't work because only the left-hand changes sign. However, if we take

$$\psi_r(\mathbf{x}, t) = \psi^*(\mathbf{x}, -t)$$

  then we do get a solution, as we can conjugate both sides. Since position information is in the magnitude of ψ and momentum information is in the phase, this is simply performing the flip $\mathbf{p} \to -\mathbf{p}$ we already did in the classical case.
- In the case of an external magnetic field, we have

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\left[-i\hbar\nabla - \frac{q}{c}\mathbf{A}(\mathbf{x})\right]^2\psi(\mathbf{x}, t)$$

  and we again have a problem, as the terms linear in $\mathbf{A}$ are imaginary. As in the classical case, the fix is to reverse the magnetic field, $\mathbf{A} \to -\mathbf{A}$.

We now define and investigate the time reversal operator.

- We define the time reversal operator Θ as

$$|\psi_r(t)\rangle = \Theta|\psi(-t)\rangle.$$

  Setting $t = 0$, the time reversal operator takes the initial condition $|\psi(0)\rangle$ to the initial condition for the reversed motion $|\psi_r(0)\rangle$.

- Since probabilities should be conserved under time reversal, we postulate

$$\Theta^\dagger \Theta = 1.$$

By analogy with classical mechanics, we require

$$\Theta \mathbf{x} \Theta^\dagger = \mathbf{x}, \quad \Theta \mathbf{p} \Theta^\dagger = -\mathbf{p}.$$

Then the orbital angular momentum also flips, $\Theta \mathbf{L} \Theta^\dagger = -\mathbf{L}$.

- We postulate that spin angular momentum flips as well. This can be understood classically by thinking of spin as just internal rotation. Since $\boldsymbol{\mu} \propto \mathbf{S}$, the magnetic moment also flips.

- The above postulates cannot be satisfied by any unitary operator, because

$$\Theta[x,p]\Theta^\dagger = i\hbar \Theta\Theta^\dagger = i\hbar = [x,p]$$

but we must get $[x,-p]$. However, we can construct Θ if we let it be an *antilinear* operator, i.e. an operator that complex conjugates everything to its right. We have already seen this in the conjugated wavefunction above.

- Generally, Wigner's theorem states that any map that preserves probabilities,

$$|\langle \psi' | \phi' \rangle| = |\langle \psi | \phi \rangle|$$

must be either unitary or antiunitary. Time reversal is the only symmetry that requires the latter option.

Working with antilinear operators is delicate, because Dirac notation is made for linear operators.

- Let A be antilinear and let L be linear. The usual rule $Lc = c^* L$ becomes

$$Ac = c^* A$$

for any scalar c. Note that the product of antilinear operators is linear.

- We defined the action of L on bras by the rule

$$(\langle \phi | L) | \psi \rangle = (\langle \phi |)(L | \psi \rangle).$$

However, if we naively extend this to antilinear operators, then $\langle \phi | A$ would be an antilinear functional, while bras must be linear functionals. Thus we add a complex conjugation,

$$(\langle \phi | A) | \psi \rangle = [(\langle \phi |)(A | \psi \rangle)]^* .$$

It matters which way an antilinear operator acts, and switching it gives a complex conjugate.

- Next, we define the Hermitian conjugate. For linear operators, we let

$$\langle \phi | L^\dagger | \psi \rangle = [\langle \psi | L | \phi \rangle]^* .$$

To extend this to antilinear operators, we need to find which way A and $A^\dagger$ act. The correct rule is to flip the direction of action,

$$\left(\langle \phi | A^\dagger \right) | \psi \rangle = [\langle \psi | \left(A | \phi \rangle \right)]^* .$$

One can check that this behaves correctly when $|\psi\rangle$ and $|\phi\rangle$ are multiplied by scalars. The rule can be remembered by simply 'flipping everything' when taking the Hermitian conjugate. Equivalently, we simply maintain the rule

$$(A^\dagger|\psi\rangle) = (\langle\psi|A)^\dagger$$

as for linear operators.

- An antiunitary operator is an antilinear operator satisfying

$$A^\dagger A = AA^\dagger = 1.$$

Antiunitary operators preserve probabilities, because

$$\langle\psi'|\phi'\rangle = (\langle\psi|A^\dagger)(A|\phi\rangle) = \Big[\langle\psi|(A^\dagger A|\phi\rangle)\Big]^* = \langle\psi|\phi\rangle^*.$$

- It is useful to factor an antilinear operator as $A = LK$ where L is linear and K is a standard antilinear operator. For example, given a basis $|n\rangle$ (say, an eigenbasis of the CSCO Q), we may define $K_Q|n\rangle = |n\rangle$. Then $K_{\mathbf{x}}$ maps the wavefunction $\psi(\mathbf{x})$ into $\psi(\mathbf{x})^*$.

Next, we apply time reversal symmetry to specific situations.

- In the case of a spinless system, it can be verified that $K_{\mathbf{x}}$ conjugates $\mathbf{x}$ and $\mathbf{p}$ in the appropriate manner, so this is the time reversal operator; as we've seen, it conjugates wavefunctions.

- Now consider a particle of spin s, ignoring spatial degrees of freedom. Since $\Theta\mathbf{S}\Theta^\dagger = -\mathbf{S}$, we have $\Theta S_z\Theta^\dagger = -S_z$, so Θ flips m, $\Theta|sm\rangle = c_m|s,-m\rangle$. On the other hand, we have

$$\Theta S_+\Theta^\dagger = \Theta(S_x + iS_y)\Theta^\dagger = -S_x + iS_y = -S_-$$

whch yields $c_{m_1} = -c_m$, so that $c_m = \eta(-1)^{s-m}$. We can absorb an arbitrary phase into η. The common choice is

$$\Theta|sm\rangle = i^{2m}|s,-m\rangle.$$

- An alternate way to derive this result is to set $K = K_{S_z}$, so that K is conjugation in the standard angular momentum basis, then choose L to fix up the commutation relations,

$$\Theta = e^{-i\pi S_y/\hbar}K = Ke^{-i\pi S_y/\hbar}$$

where the exponential commutes with K because its matrix elements are real.

- Restoring the spatial degrees of freedom,

$$\Theta = K_{\mathbf{x},S_z}e^{-i\pi S_y/\hbar}.$$

One might wonder why S_y appears. Choosing the S_z basis already breaks rotational invariance; the remaining symmetry between x and y is broken by phase conventions. Defining Θ is one of the first places where phase conventions have a nontrivial effect.

- In the case of many particles with spin, we may either multiply the individual Θ's or replace S_y and S_z above with the total angular momenta. These give the same result because the Clebsch-Gordan coefficients are real.

- Time reversal invariance holds for any Hamiltonian of the form $H = p^2/2m + V(\mathbf{x})$. It is broken by an external magnetic field, but not by internal fields. For example, the spin-orbit coupling $\mathbf{L} \cdot \mathbf{S}$ is time-reversal invariant because both the angular momenta flip.

- Time reversal invariance implies that fundamental particles should have no electric dipole moment. By the Wigner-Eckart theorem, $\boldsymbol{\mu}_e \propto \mathbf{S}$ for an irrep, so the electric dipole moment flips under time reversal. But $H_{\text{int}} \propto \boldsymbol{\mu}_e \cdot \mathbf{E}$, so $\boldsymbol{\mu}_e$ cannot flip under time reversal if the Hamiltonian is to be time reversal invariant.

- Note that this argument only applies to particles described by a single irrep. That is, it applies to neutrons because we are assuming the irreps of nuclei are spaced far apart; there's no symmetry that would make the lowest irrep degenerate. But a typical molecule in typical laboratory conditions has enough energy to enter many irreps.

Finally, we apply time reversal to dynamics.

- First, we verify the time-reversed state obeys the Schrodinger equation. Setting $\hbar = 1$,

$$i\partial_t |\psi_r(t)\rangle = i\partial_t \Theta |\psi(-t)\rangle = \Theta \left[-i\partial_t |\psi(-t)\rangle \right].$$

  Writing $\tau = -t$, we have

$$i\partial_t |\psi_r(t)\rangle = \Theta \left[i\partial_\tau |\psi(\tau)\rangle \right] = \Theta H |\psi(\tau)\rangle = (\Theta H \Theta^\dagger) |\psi_r(t)\rangle.$$

  Hence the time-reversed state satisfies the Schrodinger equation under the time-reversed Hamiltonian. The Hamiltonian itself is invariant under time reversal if $[\Theta, H] = 0$.

- If the Hamiltonian is invariant under time reversal and $|\psi\rangle$ is a nondegenerate energy eigenstate, we must have $\Theta|\psi\rangle = e^{i\theta}|\psi\rangle$, where the eigenvalue is a phase because Θ preserves norms. Then the state $e^{i\theta/2}|\psi\rangle$ has Θ eigenvalue 1.

  For the case of spatial degrees of freedom, this implies that the wavefunctions of nondegenerate states can be chosen real.

- More generally, Θ can link pairs of degenerate energy eigenstates. One can show that we can always change basis in this subspace so that both have Θ eigenvalue 1.

  For example, for the free particle, $e^{\pm ikx}$ can be combined into $\sin(kx)$ and $\cos(kx)$. Orbitals in chemistry, which are real, and chosen from combinations of the $Y_{\ell,\pm m}$.

- In general, we have

$$\Theta^2 = K e^{-i\pi S_y/\hbar} K e^{-i\pi S_y/\hbar} = e^{-i(2\pi)S_y/\hbar} = \begin{cases} 1 & \text{bosons} \\ -1 & \text{fermions} \end{cases}.$$

  This does not depend on phase conventions, as any phase adjustment cancels itself out.

- When there are an odd number of fermions, $\Theta^2 = -1$. Then energy levels must be twofold degenerate, because if they were not, we would have $\Theta^2|\psi\rangle = \Theta e^{i\theta}|\psi\rangle = |\psi\rangle$, which contradicts $\Theta^2 = -1$. This result is called Kramer's degeneracy. For example, $|l, m\rangle$ pairs with $|l, -m\rangle$, except when $m = 0$, which only occurs for integer spin.

8 Atomic Physics

8.1 Chemistry

8.2 Time Independent Perturbation Theory

8.3 The Stark and Zeeman Effects

8.4 Fine and Hyperfine Structure

8.5 The Thomas-Fermi Model

8.6 The Hartree-Fock Method

9 Time Dependent Perturbation Theory

9.1 Interaction Picture

In time-dependent perturbation theory, we consider the Hamiltonian

$$H(t) = H_0 + H_1(t)$$

where H_0 is solvable and H_1 is treated as a perturbation. We are interested in calculating the transition amplitudes

$$\langle f|U(t)|i\rangle$$

where typically $|i\rangle$ and $|f\rangle$ are two eigenstates of the unperturbed Hamiltonian, and $U(t)$ is the time evolution operator. To set up the perturbation theory, it's useful to go to the interaction picture; we begin by reviewing Heisenberg picture.

- In Heisenberg picture, we transfer all time-dependence to the operators, so

  $$A_H(t) = U^\dagger(t) A_S(t) U(t)$$

  where $U(t)$ is the time evolution operator for $H(t)$ from 0 to t. The states are 'frozen' at time $t = 0$. Then all expectation values come out the same as in Schrodinger picture.
- If $C_S = A_S B_S$, then $C_H = A_H B_H$. As a result, all operator identities remain true in Heisenberg picture, including commutation relations.
- In particular, the expression for the Hamiltonian remains valid, so

  $$H_H(t) = H_S(p_H(t), x_H(t), t).$$

  In the special case $[H_S(t), H_S(t')] = 0$ for all times (e.g. when it is time-independent) we find

  $$H_H(t) = H_S(t).$$
- Differentiating the Heisenberg operator definition and using

  $$i\hbar \frac{\partial U(t)}{\partial t} = H_S(t) U(t)$$

  we find the Heisenberg equation of motion,

  $$i\hbar \frac{dA_H(t)}{dt} = [A_H(t), H_H(t)] + i\hbar \left(\frac{\partial A_S(t)}{\partial t}\right)_H.$$
- Time-independent Schrodinger operators that always commute with the Hamiltonian are said to be 'conserved' in Schrodinger picture; in Heisenberg picture, they have no time evolution.

Example. The harmonic oscillator. Setting all constants to one,

$$H_S = \frac{p_S^2 + x_S^2}{2}.$$

Since the Hamiltonian is time-independent, $H_S = H_H$. To check this, note that

$$H_H = \frac{p_H^2 + x_H^2}{2} = \frac{(p_S^2 + x_S^2)(\cos^2 t + \sin^2 t)}{2} = H_S$$

where we plugged in the known time dependence of p_H and x_H.

Now we turn to the interaction picture. We leave S subscripts implicit.

- Let $U_0(t)$ be the time evolution due to just H_0, so

$$i\hbar \frac{\partial U_0(t)}{\partial t} = H_0 U_0(t)$$

  In the interaction picture, we 'cancel out' the state evolution due to H_0, defining

$$|\psi_I(t)\rangle = U_0^\dagger(t)|\psi_S(t)\rangle, \quad A_I(t) = U_0^\dagger(t) A_S(t) U_0(t)$$

  with the operator evolution chosen to preserve expectation values.

- Define the time evolution operator in the interaction picture as

$$|\psi_I(t)\rangle = W(t)|\psi_I(0)\rangle.$$

  Combining the above results, we find

$$W(t) = U_0(t)^\dagger U(t).$$

  That is, we evolve forward in time according to the exact Hamiltonian, then evolve backward under the unperturbed Hamiltonian.

- Differentiating and simplifying gives

$$i\hbar \frac{\partial W(t)}{\partial t} = H_{1I}(t) W(t)$$

  where $H_{1I}(t)$ is the perturbation term in the interaction picture. Integrating this gives

$$W(t) = 1 + \frac{1}{i\hbar} \int_0^t dt'\, H_{1I}(t') W(t')$$

  and plugging this equation into itself gives a series solution for $W(t)$, the Dyson series.

- A succinct way to write the full result is by a time-ordered exponential,

$$W(t) = T \exp\left(\frac{1}{i\hbar} \int_0^t dt'\, H_{1I}(t') \right).$$

  This is the generic solution to a Schrodinger equation with time-dependent Hamiltonian.

- In general, we can always split the Hamiltonian so that one piece contributes to the time evolution of the operators (by the Heisenberg equation) and the other contributes to the time evolution of the states (by the Schrodinger equation). Interaction picture is just the particular splitting into H_0 and $H_1(t)$.

9.2 Fermi's Golden Rule

For simplicity, we begin with the case where H_0 has a discrete spectrum, $H_0|n\rangle = E_n|n\rangle$, with initial state $|i\rangle$.

- Applying the Dyson series, the interaction picture state at a later time is

$$|\psi_I(t)\rangle = |i\rangle + \frac{1}{i\hbar}\int_0^t dt'\, H_{1I}(t')|i\rangle + \frac{1}{(i\hbar)^2}\int_0^t dt' \int_0^{t'} dt''\, H_{1I}(t')H_{1I}(t'')|i\rangle + \cdots .$$

- Our goal is to calculate the coefficients

$$|\psi_I(t)\rangle = \sum_n c_n(t)|n\rangle.$$

  The $c_n(t)$ differ from the transition amplitudes mentioned earlier because they lack the rapidly oscillating phase factors $e^{iE_nt/\hbar}$; such factors don't affect transition probabilities. (Note that the eigenstates $|n\rangle$ are the same in all pictures; states evolve in time but eigenstates don't.)

- Using the Dyson series, we can expand each coefficient in a power series

$$c_n(t) = \delta_{ni} + c_n^{(1)}(t) + c_n^{(2)}(t) + \ldots .$$

  The first term is

$$c_n^{(1)}(t) = \frac{1}{i\hbar}\int_0^t dt'\, \langle n|H_{1I}(t')|i\rangle = \frac{1}{i\hbar}\int_0^t dt'\, e^{i\omega_{ni}t'}\langle n|H_1(t')|i\rangle, \quad \omega_{ni} = \frac{E_n - E_i}{\hbar}$$

  where we converted H_1 back to Schrodinger picture.

- Similarly, the second-order term is

$$c_n^{(2)}(t) = \frac{1}{(i\hbar)^2}\int_0^t dt' \int_0^{t'} dt'' \sum_k e^{i\omega_{nk}t' + i\omega_{ki}t''}\langle n|H_1(t')|k\rangle\langle k|H_1(t'')|i\rangle.$$

  Here, we added a resolution of the identity; the second-order term evidently accounts for transitions through one intermediate state.

- To make further progress, we need to specify more about the perturbation term H_1. For example, for a constant perturbation, the phase factors come out the integral, giving

$$c_n^{(1)}(t) = \frac{2}{i\hbar}e^{i\omega_{ni}t/2}\frac{\sin\omega_{ni}t/2}{\omega_{ni}}\langle n|H_1|i\rangle.$$

  The corresponding transition frequency, to first order, is

$$P_n(t) = \frac{4}{\hbar^2}\frac{\sin^2\omega_{ni}t/2}{\omega_{ni}^2}|\langle n|H_I|i\rangle|^2.$$

  We see the probability oscillates sinusoidally in time, to first order.

Example. The next simplest example is sinusoidal driving. The most general example is

$$H_1(t) = Ke^{-i\omega_0 t} + K^\dagger e^{i\omega_0 t}$$

where K need not be Hermitian. As a result, the expression for $c_n^{(1)}$ has two terms, with denominators of $\omega_{ni} \pm \omega_0$. Therefore, the effect of a sinusoidal driving can be very large when it is 'on resonance' with a transition. Supposing that $\omega_{ni} \approx \omega_0$, so that only the K term counts, we have

$$P_n(t) = \frac{4}{\hbar^2} \frac{\sin^2(\omega_{ni} - \omega_0)t/2}{(\omega_{ni} - \omega_0)^2} |\langle n|K|i\rangle|^2.$$

Physically, this could translate to absorption of light, where a sinusoidal electromagnetic field is the driving; the response is Lorentzian. Since the $K^\dagger$ term must be there as well, we also get resonance for $\omega_{ni} \approx -\omega_0$. Physically, that process corresponds to stimulated emission.

Generally, the probability is proportional to $1/(\Delta\omega)^2$ and initially grows as t^2. The probability can exceed unity close to resonance, signaling that first-order perturbation theory breaks down.

Next, we consider the case of a continuum of final states, which yields Fermi's golden rule.

- Shifting the frequency to be zero on resonance, the total transition probability to all states near resonance is
$$P(t) \approx \frac{4}{\hbar^2} \int_{-\infty}^{\infty} d\omega \, \frac{\sin^2 \omega t/2}{\omega^2} g(\omega) |\langle f_\omega|K|i\rangle|^2$$
where $g(\omega)$ is the density of states.
- The function $\sin^2(\omega t/2)/\omega^2$ is peaked around $|\omega| \lesssim 1/t$ to a height of $t^2/4$, so area of the central lobe is $O(t)$. Away from the lobe, for $|\omega| \gtrsim 1/t$, we have oscillations of amplitude $1/\omega^2$. Integrating, the total area of the side lobes also grows as t. We thus expect the total area to grow as t, and contour integrating shows
$$\int_{-\infty}^{\infty} d\omega \, \frac{\sin^2 \omega t/2}{\omega^2} = \frac{\pi t}{2}.$$
- As $t \to \infty$, the integral's contribution becomes concentrated about $\omega = 0$, so
$$\lim_{t\to\infty} \frac{1}{t} \frac{\sin^2 \omega t/2}{\omega^2} = \frac{\pi}{2} \delta(\omega).$$
More generally, for arbitrary t, we can define
$$\frac{1}{t} \frac{\sin^2 \omega t/2}{\omega^2} = \frac{\pi}{2} \Delta_t(\omega).$$
Plugging this into our integral and taking the long time limit gives
$$P(t) \approx \frac{2\pi t}{\hbar^2} g(\omega_{ni}) |\langle f|K|i\rangle|^2$$
where f is a representative final state. This is called Fermi's golden rule.
- The transition probability grows linearly in time, which fits with our classical intuition (i.e. for absorption of light), as the system has a constant 'cross section'. For long times, the probability exceeds unity, again signaling that first-order perturbation theory breaks down.

- For very early times, the rule also fails, and we recover the t^2 dependence. To do this, note that $\lim_{t\to 0} \Delta_t(\omega) = t^2/4$. Therefore, we can pull $\Delta_t(\omega)$ out of the integral to get

$$P(t) \propto t^2 \int d\omega\, g(\omega) |\langle f_\omega|K|i\rangle|^2 \propto t^2.$$

  Fermi's golden rule becomes valid once the variation of $g(\omega)|\langle f_\omega|K|i\rangle|^2$ is slow compared to the variation of $\Delta_t(\omega)$, and we can pull the former out of the integral instead.

Note. It is sometimes said that for finite times, transitions can violate energy conservation, because $\Delta_t(\omega)$ has support for $\omega \neq 0$, so we can have transitions of energy greater than or lesser than $\hbar\omega$. However, what's really going on is that for finite times, the energy of the *photons* we're sending into the system aren't definite to begin with, since they must be a finite-time wavepacket. Energy is always conserved, even in quantum mechanics.

Note. On the other hand, this thinking explains why the probability can go as t rather than t^2. Roughly speaking, the amplitude to go into each decay state scales as t, giving a probability t^2 for each state, but the number of accessible states has $\Delta E \sim \hbar/t$ by the energy-time uncertainty principle, giving the observed linear dependence and expected exponential decay. For very early times, we see deviations because ΔE is so large we can hit all the states.

9.3 The Born Approximation

We apply time-dependent perturbation theory to scattering, first reviewing classical scattering.

- In classical scattering, we consider a collimated beam of particles with momentum $\mathbf{p}$ parametrized by impact parameter $\mathbf{b}$ which hits a localized potential $U(\mathbf{x})$ centered at the origin. The particles are scattered in the asymptotic direction $\hat{\mathbf{n}}(\mathbf{b})$.
- We define the differential cross section $d\sigma/d\Omega$ by

$$\text{area} = \frac{d\sigma}{d\Omega} d\Omega$$

  where the left-hand side is an area in impact-parameter space; it is a function of θ and ϕ.
- To convert a cross section to a count rate, we let $\mathbf{J}$ be the flux of incident particles and w be the total count rate of scattered particles. Then

$$w = J\sigma, \quad \sigma = \int d\Omega\, \frac{d\sigma}{d\Omega}$$

  where σ is the total cross section, and the integral omits the forward direction.
- For example, for hard-sphere scattering off an obstacle of radius r, the cross section is $\sigma = \pi r^2$. However, classically the total cross section is often infinite, as we count particles that are scattered even a tiny amount.
- In the case of two-body scattering, we switch to the center-of-mass frame, with variables

$$\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2, \quad \mathbf{p} = \frac{m_1\mathbf{p}_2 - m_2\mathbf{p}_1}{m_1 + m_2}.$$

  The momentum $\mathbf{p}$ is simply chosen to be the conjugate momentum to $\mathbf{r}$. It is the momentum of one of the particles in the center-of-mass frame.

- In the case of two beams scattering off each other, with number density n_1 and n_2 and relative velocity v,
$$\frac{dw}{d\Omega} = \frac{d\sigma}{d\Omega} v \int d\mathbf{x}\, n_1 n_2.$$

We now set up the same situation in quantum mechanics.

- We split the Hamiltonian as $H_0 = p^2/2m$ and $H_1 = V(\mathbf{x})$. The perturbation is not time-dependent, but the results above hold just as well.

- We take periodic boundary conditions in a cube of volume $V = L^3$ with plane wave states $|\mathbf{k}\rangle$ with wavefunctions
$$\psi_{\mathbf{k}}(\mathbf{x}) = \langle \mathbf{x}|\mathbf{k}\rangle = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\sqrt{V}}.$$
These are the eigenstates of H_0. We take the initial state to be $|\mathbf{k}_i\rangle$.

- The first-order transition amplitude to $|\mathbf{k}\rangle$ is
$$c_{\mathbf{k}}^{(1)}(t) = \frac{2}{i\hbar} e^{i\omega t/2} \left(\frac{\sin(\omega t/2)}{\omega}\right) \langle \mathbf{k}|U(\mathbf{x})|\mathbf{k}_i\rangle, \quad \omega = \frac{\hbar}{2m}(k^2 - k_i^2).$$
To make contact with our classical theory, we consider the rate of scattering into a cone of solid angle $\Delta\Omega$,
$$\frac{dw}{d\Omega}\Delta\Omega = \sum_{\mathbf{k}\in\text{cone}} \frac{2\pi}{\hbar^2} \Delta_t(\omega) |\langle \mathbf{k}|U(\mathbf{x})|\mathbf{k}_i\rangle|^2,$$
where w is now interpreted as probability per time, corresponding to a classical count rate. The incident flux is also interpreted as a probability flux, $J = n_i v_i = \hbar k_i/mV$.

- For sufficiently long times t, we have
$$\Delta_t(\omega) \approx \delta(\omega) = \frac{m}{\hbar k_i}\delta(k - k_i).$$
Moreover, in the limit $V \to \infty$, we have
$$\sum_{\mathbf{k}\in\text{cone}} \quad \to \quad \frac{V}{(2\pi)^3}\Delta\Omega \int_0^\infty k^2\, dk.$$

- Plugging everything in and using the symmetric convention for the Fourier transform,
$$\frac{d\sigma}{d\Omega} = \frac{2\pi}{\hbar^2}\left(\frac{m}{\hbar k_i}\right)^2 \int_0^\infty dk\, k^2 \delta(k - k_i) |\widetilde{U}(\mathbf{k} - \mathbf{k}_i)|^2 = \frac{2\pi m^2}{\hbar^4} |\widetilde{U}(\mathbf{k}_f - \mathbf{k}_i)|^2$$
where $\mathbf{k}_f$ is parallel to $\mathbf{k}$ with $k_f = k_i$ by energy conservation. This is the first Born approximation.

- If the potential $U(\mathbf{x})$ has lengthscale a, then $\widetilde{U}(\mathbf{k})$ has scale $1/a$. Hence Fermi's golden rule applies for times $t \gg a/v$ where v is the velocity.

  Physically, we can understand this by looking at the initial state $|\mathbf{k}\rangle$. This state is unphysical because it has uniform momentum everywhere; this is okay far from the potential, where it represents a uniform beam, but not in the potential itself. Thus the time a/v can be interpreted as the time needed for these transient, 'unphysical' particles to get out of the way.

- After a time $t \gg a/v$, the evolved wavefunction $U(t)|\mathbf{k}\rangle$ will look like an energy eigenstate in a region of radius about tv about the origin, as we have reached a 'steady state' of particles coming in and being scattered out. This lends some intuition for why scattering rates can be computed using energy eigenstates alone.

Example. We consider scattering off the Yukawa potential

$$U(r) = A\frac{e^{-\kappa r}}{r}, \quad \widetilde{U}(\mathbf{q}) = \frac{2A}{(2\pi)^{1/2}}\frac{1}{\kappa^2 + q^2}$$

which arises in nuclear physics because it is the Green's function for the Klein-Gordan equation. Applying our scattering formula, $\mathbf{q} = \mathbf{k} - \mathbf{k}_i$ and hence $q^2 = 4k^2 \sin^2(\theta/2)$, giving

$$\frac{d\sigma}{d\Omega} = \frac{4A^2 m^2}{\hbar^4}\frac{1}{(4k^2\sin^2(\theta/2) + \kappa^2)^2}.$$

In particular, in the case of Coulomb scattering, $\kappa \to 0$ and $A = Z_1 Z_2 e^2$, giving

$$\frac{d\sigma}{d\Omega} = \frac{Z_1^2 Z_2^2 e^4 m^2}{4\hbar^4 k^4}\frac{1}{\sin^4(\theta/2)}.$$

This is the Rutherford cross section, the exact result for classical nonrelativistic Coulomb scattering. It is also the exact result in nonrelativistic quantum mechanics if the particles are distinguishable, though we couldn't have known this as we only computed the first term in a perturbation series.

However, the scattering *amplitude* for the Coulomb potential turns out to be incorrect by phase factors, because the Coulomb potential doesn't fall off quickly enough. This doesn't matter for distinguishable particles, but for identical particles it renders our answer incorrect because we must combine distinct scattering amplitudes with phases intact. The correct answer for two electrons is called the Mott cross section.

9.4 The Photoelectric Effect

We now consider the photoelectric effect as an extended example.

- We consider photons of energy $E_0 = \hbar\omega_0$ and momentum $\mathbf{p}_0 = \hbar\mathbf{k}_0$ incident on a single-electron atom in the ground state $|g\rangle$ with energy E_g, and compute the rate at which the electron is ejected into a plane-wave final state $|\mathbf{k}\rangle$.

- By conservation of energy, we must have $\hbar\omega_0 > |E_g|$, and we further assume that

  $$\hbar\omega_0 \gg |E_g|.$$

  This is necessary because of the long-range Coulomb field of the nucleus; by assuming this, we can ignore the field and consider the ejected electron to be approximately free.

- We also require that the electron be nonrelativistic, with final energy

  $$E = \hbar\omega_0 + E_g \ll mc^2.$$

  For hydrogen, these constraints imply $100\,\mathrm{eV} \lesssim \hbar\omega_0 \lesssim 100\,\mathrm{keV}$, which contains the far UV and X-ray ranges.

- We model the light wave classically, with potentials
$$\phi = 0, \quad \mathbf{A}(\mathbf{x}, t) = A_0 \boldsymbol{\epsilon} e^{i(\mathbf{k}_0 \cdot \mathbf{x} - \omega t)}.$$
This is a common choice for treating plane waves in a nonrelativistic context. Using the transversality condition $\boldsymbol{\epsilon} \cdot \mathbf{k}_0 = 0$ shows that the vector potential is in Coulomb gauge, $\nabla \cdot \mathbf{A} = 0$, and hence as operators, $\mathbf{p} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{p}$. (Note that at the quantum level, $\mathbf{k}_0$ is not an operator but $\mathbf{x}$ and $\mathbf{p}$ above both are.)

- We use the standard replacement $\mathbf{p} \to \mathbf{p} + e\mathbf{A}/c$, which gives perturbing Hamiltonian
$$H_1 = \frac{e}{mc} \mathbf{p} \cdot \mathbf{A}.$$
Since we are working to first order, we neglect the A^2 term.

- In particular, this is of a sinudoidal form with
$$K = \frac{eA_0}{mc} (\boldsymbol{\epsilon} \cdot \mathbf{p}) e^{i\mathbf{k}_0 \cdot \mathbf{x}}.$$
Hence the transition rate is
$$\frac{dw}{d\Omega} \Delta\Omega = \frac{2\pi}{\hbar^2} \sum_{\mathbf{k} \in \text{cone}} |\langle \mathbf{k} | K | g \rangle|^2 \Delta_t(\omega), \quad \omega = \frac{E - \hbar\omega_0 - E_g}{\hbar},$$
where we take the sum over final states in a cone of solid angle $\Delta\Omega$.

- We next convert from $dw/d\Omega$ to a cross-section $d\sigma/d\Omega$ using
$$\frac{dw}{d\Omega} = n_i v_i \frac{d\sigma}{d\Omega}.$$
Now, the velocity is simply $v_i = c$, while the number density can be found by computing the energy in two different ways,
$$u = n_i \hbar\omega_0, \quad u = \frac{E^2 + B^2}{8\pi} = \frac{\omega_0^2 A_0^2}{2\pi c^2}.$$

- Next, we compute the matrix element. We have
$$\langle \mathbf{k} | (\boldsymbol{\epsilon} \cdot \mathbf{p}) e^{i\mathbf{k}_0 \cdot \mathbf{x}} | g \rangle = \hbar (\boldsymbol{\epsilon} \cdot \mathbf{k}) \langle \mathbf{k} | e^{i\mathbf{k}_0 \cdot \mathbf{x}} | g \rangle.$$
The remaining factor is proportional to $\widetilde{\psi}_g(\mathbf{q})$ where $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$ by logic we've seen before. Note that for typical optics applications, where k_0 is in the visible range and hence $e^{i\mathbf{k}_0 \cdot \mathbf{x}}$ varies slowly, we often expand the exponential instead, yielding a multipole expansion.

- Putting everything together, taking $\Delta_t(\omega) \to \delta(\omega)$, and simplifying gives
$$\frac{d\sigma}{d\Omega} = (2\pi)^2 \frac{e^2}{mc^2} \frac{k_f}{k_0} (\boldsymbol{\epsilon} \cdot \mathbf{k}_f)^2 |\widetilde{\psi}_g(\mathbf{q})|^2$$
where the magnitude of the final momentum $\mathbf{k}_f$ is set by energy conservation. We can then proceed further with an explicit form for $|g\rangle$, which would show that harder (higher energy) X-rays penetrate further, and that larger atoms are more effective at stopping them.

- We might wonder why momentum isn't conserved here, while energy is. The reason is that momentum is absorbed by the nucleus, which we have implicitly assumed to be infinitely heavy by taking the potential as static. Without the nucleus present, the reaction $\gamma + e \to e$ is forbidden. The same effect is observed in Bremsstrahlung, $e \to e + \gamma$, which only occurs when matter is nearby to absorb the momentum.

10 Scattering

10.1 Introduction

In the previous section, we considered scattering from a time-dependent point of view. In this section, we instead solve the time-independent Schrodinger equation.

- We consider scattering off a potential $V(\mathbf{x})$ which goes to zero outside a cutoff radius $r > r_{\text{co}}$. Outside this radius, energy eigenstates obey the free Schrodinger equation.

- As argued earlier, if we feed in an incident plane wave, the wavefunction will approach a steady state after a long time, with constant probability density and current; hence it approach an energy eigenstate. Thus we can also compute scattering rates by directly looking at energy eigenstates; such eigenstates are all nonnormalizable.

- We look for energy eigenstates $\psi(\mathbf{x})$ which contain an incoming plane wave, i.e.

  $$\psi(\mathbf{x}) = \psi_{\text{inc}}(\mathbf{x}) + \psi_{\text{scat}}(\mathbf{x}), \quad \psi_{\text{inc}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}.$$

  For large r, the scattered wave must be a spherical wave with the same energy as the original wave (i.e. same magnitude of momentum),

  $$\psi_{\text{scat}}(\mathbf{x}) \sim \frac{e^{ikr}}{r} f(\theta, \phi).$$

  The function $f(\theta, \phi)$ is called the scattering amplitude.

- Now, if we wanted ψ_{scat} to be an exact eigenstate for $r > r_{\text{co}}$, then f would have to be constant, yielding an isotropic spherical wave. However, the correction terms for arbitrary f are subleading in r, and we only care about the large r behavior.

  Similarly, the incoming plane wave $e^{i\mathbf{k}\cdot\mathbf{x}}$ isn't an eigenstate; the correction terms are included in $\psi_{\text{inc}}(\mathbf{x})$ and are subleading.

- Next, we convert the scattering amplitude to a cross section. The probability current is

  $$\mathbf{J} = \frac{\hbar}{m} \text{Im}(\psi^* \nabla \psi).$$

  For the incident wave, $\mathbf{J}_{\text{inc}} = \hbar\mathbf{k}/m$. For the outgoing wave,

  $$\mathbf{J}_{\text{scat}} \sim \frac{\hbar k}{m} \frac{|f(\theta, \phi)|^2}{r^2} \hat{\mathbf{r}}.$$

  The area of a cone of solid angle $\Delta\Omega$ at radius r is $r^2 \Delta\Omega$, and hence

  $$\frac{d\sigma}{d\Omega} = \frac{r^2 J_{\text{scat}}(\Omega)}{J_{\text{inc}}} = |f(\theta, \phi)|^2$$

  which is a very simple result.

- We've ignored a subtlety above: the currents for the incident and scattered waves should interfere because $\mathbf{J}$ is bilinear. We ignore this because the incident wave has a finite area in reality, so it is zero for all angles except the forward direction. In the forward direction, the incident and scattered waves interfere destructively, as required by conservation of probability. Applying this quantitatively yields the optical theorem.

- The total cross section almost always diverges classically, because we count any particle scattered by an arbitrarily small amount. By contrast, in quantum mechanics we can get finite cross sections because an 'arbitrarily small push' can instead become an arbitrarily small scattering amplitude, plus a high amplitude for continuing exactly in the forward direction. (However, the cross section can still diverge if $V(r)$ falls slowly enough.)

10.2 Partial Waves

We now focus on the case of a central force potential.

- Solutions to the Schrodinger equation separate,
$$\psi_{k\ell m}(\mathbf{x}) = R_{k\ell}(r) Y_{\ell m}(\theta, \phi).$$
The quantum number k parametrizes the energy by $E = \hbar^2 k^2/2m$. It is the wavenumber of the incident and scattered waves far from the potential, i.e. $R_{kl}(r) \propto e^{ikr}$.

- Defining $u_{k\ell}(r) = rR_{k\ell}(r)$, the radial Schrodinger equation is
$$\frac{1}{r^2}\frac{d}{dr}\left(r^2 \frac{dR_{k\ell}}{dr}\right) + k^2 R_{k\ell}(r) = W(r) R_{k\ell}(r), \quad u''_{k\ell}(r) + k^2 u_{k\ell}(r) = W(r) u_{k\ell}(r)$$
where
$$W(r) = \frac{\ell(\ell+1)}{r^2} + \frac{2m}{\hbar^2} V(r).$$

- Therefore, the general solution of energy E is
$$\psi(\mathbf{x}) = \sum_{\ell m} A_{\ell m} R_{k\ell}(r) Y_{\ell m}(\theta, \phi).$$
Our next task is to find the expansion coefficients $A_{\ell m}$ to get a scattering solution.

- In the case of the free particle, the solutions for the radial wavefunction $R_{k\ell}$ are the spherical Bessel functions $j_\ell(kr)$ and $y_\ell(kr)$, where
$$j_\ell(\rho) \approx \frac{1}{\rho}\sin\left(\rho - \ell\pi/2\right), \quad y_\ell(\rho) \approx -\frac{1}{\rho}\cos(\rho - \ell\pi/2)$$
for $\rho \gg \ell$, and the y-type Bessel functions are singular at $\rho = 0$.

- Since the incident wave $e^{i\mathbf{k}\cdot\mathbf{x}}$ describes a free particle, it must be possible to write in terms of the j-type Bessel functions. One can show
$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kr) Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{r}}).$$
Next, using the addition theorem for spherical harmonics,
$$P_\ell(\cos\gamma) = \frac{4\pi}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{r}})$$
where γ is the angle between $\mathbf{k}$ and $\mathbf{r}$, we have
$$e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_\ell i^\ell (2\ell+1) j_\ell(kr) P_\ell(\cos\gamma).$$

- Next, we find the asymptotic behavior of the radial wavefunction $R_{k\ell}(r)$ for large r. If the potential $V(r)$ cuts off at a finite radius r_0, then the solutions are Bessel functions of both the j and y-type, since we don't care about the region $r < r_0$, giving $u_{k\ell}(r) \sim e^{\pm ikr}$.
- If there is no sharp cutoff, parametrize the error as $u_{k\ell}(r) = e^{g(r) \pm ikr}$, giving
$$g'' + g'^2 \pm 2ikg' = W(r).$$
We already know the centrifugal term alone gives Bessel functions, so we consider the case where the potential dominates for long distances, $V(r) \sim 1/r^p$ where $0 < p < 2$. Taking the leading term on both sides gives $g(r) \sim 1/r^{p-1}$, so the correction factor g goes to zero for large r only if $p > 1$. In particular, the Coulomb potential is ruled out, as it gives logarithmic phase shifts $e^{i\log(kr)}$. This can also be shown using the first-order WKB approximation.
- Assuming that $V(r)$ does fall faster than $1/r$, we may write
$$R_{k\ell} \sim \frac{\sin(kr - l\pi/2 + \delta_\ell)}{kr}$$
for large r. To interpret the phase shift δ_ℓ, note that we would have $\delta_\ell = 0$ in the case of a free particle, by the expansion of $j_\ell(kr)$. Thus the phase shift tells us how the potential asymptotically modifies radial phases.

Finally, we combine these ingredients to get our desired incident-plus-scattering states.

- We write the general solution as
$$\psi(\mathbf{x}) = 4\pi \sum_{\ell m} i^\ell A_{\ell m} R_{k\ell}(r) Y_{\ell m}(\hat{\mathbf{r}}).$$
Subtracting off a plane wave, we have
$$\psi_{\text{scat}}(\mathbf{x}) = 4\pi \sum_{\ell m} i^\ell \left[A_{\ell m} R_{k\ell}(r) - j_\ell(kr) Y^*_{\ell m}(\hat{\mathbf{k}}) \right] Y_{\ell m}(\hat{\mathbf{r}}).$$
- For large r, the quantity in square brackets can be expanded as the sum of incoming and outgoing waves e^{-ikr}/r and e^{ikr}/r, and we only want an outgoing component, which gives
$$A_{\ell m} = e^{i\delta_\ell} Y^*_{\ell m}(\hat{\mathbf{k}}).$$
Substituting this in and simplifying, we have
$$\psi_{\text{scat}}(\mathbf{x}) \sim 4\pi \frac{e^{ikr}}{kr} \sum_{\ell m} e^{i\delta_\ell} \sin(\delta_\ell) Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{r}}) = \frac{e^{ikr}}{kr} \sum_\ell (2\ell+1) e^{i\delta_\ell} \sin(\delta_\ell) P_\ell(\cos\theta)$$
where we used the addition theorem for spherical harmonics and set $\hat{\mathbf{k}} = \hat{\mathbf{z}}$.
- The above result is known as the partial wave expansion. It gives the scattering amplitude
$$f(\theta, \phi) = \frac{1}{k} \sum_\ell (2\ell+1) e^{i\delta_\ell} \sin(\delta_\ell) P_\ell(\cos\theta).$$
There is no dependence on ϕ and hence no angular momentum in the z-direction because the problem is symmetric about rotations about $\hat{\mathbf{z}}$. Instead the scattered waves are parametrized by their total angular momentum ℓ. The individual terms are $m = 0$ spherical harmonics, and are called the s-wave, the p-wave, and so on. Each of these contributions are present in the initial plane wave and scatter independently, since L^2 is conserved.

- The differential cross section has interference terms, but the total cross section does not due to the orthogonality of the Legendre polynomials, giving

$$\sigma = \frac{4\pi}{k^2} \sum_\ell (2\ell+1) \sin^2 \delta_\ell.$$

This is the partial wave expansion of the total cross section.

- For any localized potential with lengthscale a, then when $ka \lesssim 1$, s-wave scattering ($\ell = 0$) dominates and the scattered particles are spherically symmetric. To see this, note that the centrifugal potential is equal to the energy when

$$\frac{\ell(\ell+1)\hbar^2}{2ma^2} = E = \frac{\hbar^2 k^2}{2m}$$

which has solution $\ell \approx ka$. Then for $ka \lesssim 1$ the particle cannot classically reach the potential at all, so it has the same phase as a free particle and hence no phase shift.

- In reality, the phase shift will be small but nonzero for $ka > 1$ because of quantum tunneling, but drops off exponentially to zero. In the case where the potential is a power law (long-ranged), the phase shifts instead drop off as powers.

- In many experimental situations, s-wave scattering dominates (e.g. neutron scattering off nuclei in reactors). In this case we can replace the potential $V(r)$ with any potential with the same δ_0. A common and convenient choice is a δ-function potential.

- We can also import some heuristic results from our knowledge of Fourier transforms, though the partial wave expansions is in Legendre polynomials instead. If the scattering amplitude is dominated by terms up to ℓ_{cutoff}, the maximum angular size of a feature is about $1/\ell_{\text{cutoff}}$. Moreover, if the phase shifts fall off exponentially, then the scattering amplitude will be analytic. Otherwise, we generally get singularities in the forward direction.

- Each scattering term σ_ℓ is bounded by $(4\pi/k^2)(2\ell+1)$. This is called the unitarity bound; it simply says we can't scatter out more than we put in.

Example. Hard sphere scattering. We let

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

The radial wavefunction takes the form

$$R_{k\ell}(r) = \cos(\delta_\ell) j_\ell(kr) - \sin(\delta_\ell) y_\ell(kr)$$

for $r > a$, where δ_ℓ is the phase shift, as can be seen by taking the $r \to \infty$ limit. The boundary condition $R_{k\ell}(a) = 0$ gives

$$\tan(\delta_\ell) = \frac{j_\ell(ka)}{y_\ell(ka)}.$$

First we consider the case $ka \ll 1$. Applying the asymptotic forms of the Bessel functions,

$$\sin(\delta_\ell) \approx \delta_\ell \approx -\frac{(ka)^{2\ell+1}}{(2\ell-1)!!(2\ell+1)!!}.$$

In particular this means the scattering is dominated by the s-wave, giving

$$\sigma = \frac{4\pi}{k^2}(ka)^2 = 4\pi a^2$$

which is several times larger than the classical result $\sigma = \pi a^2$. Next we consider the case $ka \gg 1$. For terms with $\ell \ll ka$, using the asymptotic forms of the Bessel functions (this time for large argument) gives

$$\delta_\ell = -ka + \frac{\ell\pi}{2}.$$

As ℓ approaches ka, the phase shifts go to zero, cutting off the partial wave expansion. Intuitively, this is because when $ka \gg 1$ the scattering is essentially classical, with the incoming wave acting like a discrete particle. If a particle is scattered off a sphere of radius a, its angular momentum is

$$L = pa\sin\theta \leq \hbar ka.$$

The total cross section is

$$\sigma \approx \frac{4\pi}{k^2}\sum_{\ell=0}^{ka}(2\ell+1)(1/2) \approx 2\pi a^2$$

where we replaced the rapidly oscillating factor $\sin^2(\delta_\ell)$ with its average, $1/2$. It is puzzling that we get twice the classical cross section. Physically, the extra πa^2 comes from diffraction around the edge of the sphere which 'fills in' the shadow. This gives a sharp scattering peak in the forward diffraction, formally the same as the central peak in light diffraction with a circular aperture.

Note. The optical theorem relates the total cross section to the forward scattering amplitude. For central force potentials, we simply note that

$$f(0) = \frac{1}{k}\sum_\ell (2\ell+1)e^{i\delta_\ell}\sin(\delta_\ell).$$

Comparing this with the total cross section immediately gives

$$\sigma = \frac{4\pi}{k}\mathrm{Im}(f(0)).$$

If we expand f in a series, the optical theorem relates terms of different orders, since $d\sigma/d\Omega \sim |f|^2$ but $\sigma \sim f$. This makes an appearance in quantum field theory through 'cut' diagrams.

The optical theorem can also be derived more generally by looking at the probability flux. By conversation of probability, we must have

$$\int \mathbf{J}\cdot d\mathbf{S} = 0$$

over a large sphere. The flux $\mathbf{J}$ splits into three terms: the incident wave (which contributes zero flux), the scattered wave (which contributes $v\sigma$), and the interference term,

$$\mathbf{J}_{\text{int}} = \frac{\hbar}{m}\mathrm{Im}\left(\psi^*_{\text{scat}}\nabla\psi_{\text{inc}} + \psi^*_{\text{inc}}\nabla\psi_{\text{scat}}\right) = vr\mathrm{Re}\left(f(\theta,\phi)^* e^{ik(x-r)}\hat{\mathbf{x}} + f(\theta,\phi)e^{ik(r-x)}\hat{\mathbf{r}}\right).$$

Integrating over a sphere of radius r, we must have

$$\sigma = r\,\mathrm{Re}\left[\int d\phi\int \sin\theta d\theta\, e^{ikr(1-\cos\theta)}f(\theta,\phi)(1+\cos\theta)\right]$$

in the limit $r\to\infty$. Then the phase factor is rapidly oscillating, so the only contribution comes from the endpoints $\theta = 0, \pi$ since there are no points of stationary phase. The contribution at $\theta = \pi$ is zero due to the $(1+\cos\theta)$ factor, while the $\theta = 0$ peak gives the desired result.

Example. Resonances. Intuitively, a resonance is a short-lived excitation that is formed in a scattering process. To understand them, we apply the WKB approximation to a potential

$$V_{\text{tot}}(r) = V(r) + \frac{\ell(\ell+1)\hbar^2}{2mr^2}$$

which has a well between the turning points $r = r_0$ and $r = r_1$, and a classically forbidden region between $r = r_1$ and the turning point $r = r_2$. We define

$$p(r) = \sqrt{2m(E - V_{\text{tot}}(r))}, \quad \Phi = \frac{2}{\hbar}\int_{r_0}^{r_1} p(r)\,dr, \quad \kappa = \frac{1}{\hbar}\int_{r_1}^{r_2} |p(r)|\,dr.$$

Note that Φ is the action for an oscillation inside the well, so the bound state energies satisfy

$$\Phi(E_n) = 2\pi(n + 1/2).$$

Starting with an exponentially decaying solution for $r < r_0$, the connection formulas give

$$u(r) = \frac{1}{\sqrt{p(r)}}\left(2e^K \cos\frac{\Phi}{2} + \frac{i}{2}e^{-K}\sin\frac{\Phi}{2}\right)e^{iS(r)/\hbar - i\pi/4} + \text{c.c.}, \quad S(r) = \int_{r_2}^{r} p(r)\,dr$$

in the region $r > r_2$, where $\cos(\Phi/2) = 0$ for a bound state. Suppose the forbidden region is large, so $e^K \gg 1$. Then away from bound states, the e^{-K} term does not contribute; we get the same solution we would get if there were no potential well at all. In particular, assuming $V(r)$ is negligible for $r > r_2$, the particle doesn't feel its effect at all, so $\delta_\ell = 0$.

Now suppose we are near a bound state, $E = E_n + \delta E$. Then

$$\Phi(E) = 2\pi(n + 1/2) + \frac{\delta E}{\hbar\omega_c}$$

according to the theory of action-angle variables, and expanding to lowest order in δE gives

$$e^{2i\delta_\ell} = \frac{-\delta E + i\Gamma/2}{-\delta E - i\Gamma/2}, \quad \Gamma = \hbar\omega_c e^{-2K}.$$

That is, across a resonance, the phase shift rapidly changes by π. Then we have a Lorentzian resonance in the cross-section,

$$\sin^2\delta_\ell = \frac{\Gamma^2/4}{(E - E_n)^2 + \Gamma^2/4}.$$

Since we have assumed K is large, the width Γ is much less than the spacing between energy levels $\hbar\omega_c$, so the cross-section has sharp spikes as a function of E. Such spikes are common in neutron-nucleus scattering. Physically, we imagine that the incoming particle tunnels through the barrier, gets 'stuck inside' bouncing back and forth for a timescale $1/\Gamma$, then exits. This is the physical model for the production of decaying particles in quantum field theory.

10.3 Green's Functions

In this section we make some formal definitions, which will be put to use in the next section. We begin with a heuristic example from electromagnetism.

- Schematically, Maxwell's equation read $\Box A = J$. The corresponding homogeneous equation is $\Box A_h = 0$, and the general solution of the inhomogeneous equation is

$$A(x) = A_h(x) + \int dx'\, G(x,x')J(x'), \quad \Box G(x,x') = \delta(x-x')$$

  where $\Box$ acts on the x coordinate.

- In general, we see that solutions to inhomogeneous equations are ambiguous up to adding a homogeneous solution. In particular, the Green's function is defined by an inhomogeneous equation, so it is ambiguous too; we often specify it with boundary conditions.

- Now we consider the case where the source is determined by A itself, $J = \sigma A$. Then Maxwell's equations read

$$\Box A = \sigma A, \quad (\Box - \sigma)A = 0.$$

  We have arrived at a homogeneous equation, but now A must be determined self-consistently; it will generally be the sum of an incident and scattered term, both sourcing current.

- As a specific example, consider reflection of an incident wave off a mirror, which is a region of high σ. The usual approach is to search for a solution of $\Box A = 0$ containing an incoming wave, satisfying a boundary condition at the mirror. But as shown above, we can also solve self-consistently, letting $A = A_{\text{inc}} + A_{\text{scat}}$ where $\Box A = \sigma A$. We would then find that A_{scat} cancels A_{inc} inside the mirror and also contains a reflected wave.

- Similarly, defining $H_0 = p^2/2m$, the time-independent Schrodinger equation for potential scattering is

$$(H_0 + V)\psi = E\psi, \quad (E - E_0)\psi = V\psi.$$

  The latter equation is formally like the equation $\Box A = \sigma A$. We can think of solving for $\psi = \psi_{\text{inc}} + \psi_{\text{scat}}$ where both terms collectively produce the 'source' term $V(\mathbf{x})\psi(\mathbf{x})$.

- Given a Green's function for ψ, we will not have a closed form for ψ. Instead, we'll get a self-consistent expression for ψ in terms of itself, which we can expand to get a series solution.

We now define time-dependent Green's functions for the Schrodinger equation.

- The inhomogeneous time-dependent Schrodinger equation is

$$\left(i\hbar\frac{\partial}{\partial t} - H(t)\right)\psi(\mathbf{x},t) = S(\mathbf{x},t).$$

  We define a Green's function to satisfy this equation for the source $i\hbar\delta(t-t')\delta^3(\mathbf{x}-\mathbf{x}')$, where the $i\hbar$ is by convention. We always indicate sources by primed coordinates.

- Earlier, we defined the propagator as

$$K(\mathbf{x},t,\mathbf{x}',t') = \langle \mathbf{x}|U(t,t')|\mathbf{x}\rangle.$$

  It is not a Green's function, as it satisfies the homogeneous Schrodinger equation; it instead propagates effects forward and backward in time.

- The outgoing (or retarded) time-dependent Green's function is

$$K_+(\mathbf{x}, t, \mathbf{x}', t') = \Theta(t - t')K(\mathbf{x}, t, \mathbf{x}', t').$$

The additional step function gives the desired δ-function when differentiated. This Green's function is zero for all $t < t'$. In terms of a water wave analogy, it describes the surface of a lake which is previously still, which we poke at $(\mathbf{x}', t')$.

- Using the outgoing Green's function gives the solution

$$\psi(\mathbf{x}, t) = \psi_h(\mathbf{x}, t) + \int_{-\infty}^{\infty} dt' \int d\mathbf{x}' \, K_+(\mathbf{x}, t, \mathbf{x}', t')S(\mathbf{x}', t').$$

If we want a causal solution, then $\psi_h(\mathbf{x}, t)$ must also vanish before the driving starts, but this implies it must vanish for all times. Therefore

$$\psi(\mathbf{x}, t) = \int_{-\infty}^{t} dt' \int d\mathbf{x}' \, K(\mathbf{x}, t, \mathbf{x}', t')S(\mathbf{x}', t')$$

is the unique causal solution.

- Similarly, we have the incoming (or advanced) Green's function

$$K_-(\mathbf{x}, t, \mathbf{x}', t') = -\Theta(t' - t)K(\mathbf{x}, t, \mathbf{x}', t').$$

For $t \to 0^-$, it approaches $-\delta^3(\mathbf{x} - \mathbf{x}')$. In terms of water waves, it describes waves in a lake forming for $t < t'$, then finally coalescing into a spike at $t = t'$ which is absorbed by our finger. For practical problems, we thus prefer the outgoing Green's function.

- We define the Green's operators $\hat{K}_\pm$ to satisfy

$$K_\pm(\mathbf{x}, t, \mathbf{x}', t') = \langle \mathbf{x} | \hat{K}_\pm(t, t') | \mathbf{x}' \rangle$$

which satisfy

$$K_\pm(t, t') = \pm\Theta(\pm(t - t'))U(t, t'), \quad (i\hbar - H(t)) \, \hat{K}_\pm(t, t') = i\hbar\delta(t - t').$$

This form is often more useful it does not privilege the position basis. In particular, Green's operators can be defined for systems with a much broader range of Hilbert spaces, such as spin systems or field theories.

Example. In the case of a time-independent Hamiltonian, we will replace the arguments t and t' with one argument, t for the time difference. For example, for a free particle in three dimensions,

$$K_0(\mathbf{x}, \mathbf{x}', t) = \left(\frac{m}{2\pi i\hbar t}\right)^{3/2} \exp\left(\frac{i}{\hbar} \frac{m(\mathbf{x} - \mathbf{x}')^2}{2t}\right)$$

as we found in the section on path integrals.

Next, we turn to energy-dependent Green's functions, which are essentially the Fourier transforms of time-dependent ones.

- We consider the inhomogeneous time-dependent Schrodinger equation,

$$(E-H)\psi(\mathbf{x}) = S(\mathbf{x})$$

where H is a time-independent Hamiltonian. An energy-dependent Green's function $G(\mathbf{x},\mathbf{x}',E)$ satisfies this equation with energy E and source $\delta(\mathbf{x}-\mathbf{x}')$.

- Given an energy-dependent Green's function, the general solution is

$$\psi(\mathbf{x}) = \psi_h(\mathbf{x}) + \int d\mathbf{x}\, G(\mathbf{x},\mathbf{x}',E)S(\mathbf{x}').$$

Note that the homogeneous solution $\psi_h(\mathbf{x})$ is simply a stationary state with energy E.

- We imagine the energy-dependent Green's functions as follows. We consider a lake with finite area which is quiet for $t<0$. At $t=0$, we begin driving a point $\mathbf{x}'$ sinusoidally with frequency E. After a long time, the initial transients die out by dissipation and the surface approaches a sinusoidally oscillating steady state; this is $G(\mathbf{x},\mathbf{x}',E)$.

- If we drive exactly at an eigenfrequency of a lake, the corresponding eigenmode has a high amplitude which goes to infinity as the dissipation $\epsilon \to 0$, so the Green's function does not exist without dissipation.

- Finally, we can consider driving at an eigenfrequency in a continuous spectrum. This is only realizable in an infinite lake, as the corresponding eigenmodes are unbounded. We find a wave field with size $1/\epsilon$, where energy continually radiates out from the driving point $\mathbf{x}'$. In the limit $\epsilon \to 0$ the wave field becomes infinite, and we see that energy is transported out to infinity. However, this wave pattern is not an eigenfunction because eigenfunctions have zero net energy flux through any closed boundary.

- We can recast the energy-dependent Green's function as an operator,

$$G(\mathbf{x},\mathbf{x}',E) = \langle \mathbf{x}|\hat{G}(E)|\mathbf{x}'\rangle, \quad (E-H)\hat{G}(E) = 1.$$

Then naively we have the solution $\hat{G}(E) = 1/(E-H)$, but this is generally not well defined. As usual, the ambiguity that exists comes from freedom in the boundary conditions.

- Note that we are not explicitly distinguishing the operator H, which acts on the Hilbert space, and the coordinate form of H, which is a differential operator that acts on wavefunctions.

Next, we carefully define energy-dependent Green's operators.

- As a first attempt, we try to define

$$\hat{G}_+(E) = \frac{1}{i\hbar}\int_{-\infty}^{\infty} dt\, e^{iEt/\hbar}\hat{K}_+(t).$$

then we have

$$\hat{G}_+(E) = \frac{1}{i\hbar}\int_0^{\infty} dt\, e^{iEt/\hbar}U(t) = \frac{1}{i\hbar}\int_0^{\infty} dt\, e^{i(E-H)t/\hbar} = -\left.\frac{e^{i(E-H)t/\hbar}}{E-H}\right|_0^{\infty}$$

where all functions of operators are defined by power series. Then $\hat{G}_+(E)$ would be a Green's operator if we could neglect the upper limit of integration.

- The problem above is due to the fact that the Schrodinger equation has no damping, so initial transients never die out. Instead we replace $H \to H - i\epsilon$, giving exponential decay, or equivalently $E \to E + i\epsilon$. Then generally we may define
$$\hat{G}_+(z) = \frac{1}{i\hbar} \int_0^\infty e^{izt/\hbar} U(t) = \frac{1}{z - H}$$
for any $z = E + i\epsilon$ with $\epsilon > 0$.

- For $\operatorname{Im} z > 0$, the Green's operator has a complete set of eigenfunctions (since H does), though it is not Hermitian. Moreover, none of the eigenvalues are vanishing because they all have nonzero imaginary part. Thus the inverse of $z - H$ exists and is unique. (We ignore subtle mathematical issues, such as nonnormalizable eigenfunctions.)

- Suppose that H has a discrete spectrum with negative energies E_n and a continuous spectrum with positive energies E, as is typical for scattering problems,
$$H|n\alpha\rangle = E_n|n\alpha\rangle, \quad H|E\alpha\rangle = E|E\alpha\rangle.$$
Using standard normalization, the resolution of the identity is
$$1 = \sum_{n\alpha} |n\alpha\rangle\langle n\alpha| + \int_0^\infty dE \sum_\alpha |E\alpha\rangle\langle E\alpha|.$$
Therefore the Green's operator can be written as
$$\hat{G}_+(E + i\epsilon) = \frac{1}{E + i\epsilon - H} = \sum_{n\alpha} \frac{|n\alpha\rangle\langle n\alpha|}{E + i\epsilon - E_n} + \int_0^\infty dE' \sum_\alpha \frac{|E'\alpha\rangle\langle E'\alpha|}{E + i\epsilon - E'}.$$

- From the above expression we conclude that $\hat{G}_+(E + i\epsilon)$ is well-defined in the upper-half plane, but may become singular in the limit $\epsilon \to 0$. We define
$$\hat{G}_+(E) = \lim_{\epsilon \to 0} \hat{G}_+(E + i\epsilon)$$
where the right-hand side is often written as $\hat{G}_+(E + i0)$. When E is not an eigenvalue, then the limit exists by the decomposition above. When E is a discrete eigenvalue, the limit is singular and the Green's function fails to exist. Finally, when $E > 0$ the integrand above diverges, though it turns out the limit of the integral exists, as we'll show in an example later. All these results are perfectly analogous to the water waves above.

- When $\hat{G}_+(E)$ is well-defined, it is a Green's operator, because
$$(E - H)\hat{G}_+(E) = \lim_{\epsilon \to 0} (E + i\epsilon - H - i\epsilon) \frac{1}{E + i\epsilon - H} = \lim_{\epsilon \to 0} (1 - i\epsilon \hat{G}(E + i\epsilon)) = 1.$$

- We similarly define the incoming energy-dependent Green's operator
$$\hat{G}_-(z) = -\frac{1}{i\hbar} \int_{-\infty}^0 e^{izt/\hbar} U(t) = \frac{1}{z - H}$$
where now $z = E - i\epsilon$. It is defined in the lower-half plane and limits to $\hat{G}_-(E)$ for $\epsilon \to 0$, where the limit is well defined if E is not equal to any of the E_n.

- In the water wave analogy, we have 'antidamping', and energy is continually absorbed by the drive. In the case $E < 0$, this makes no difference in the limit $\epsilon \to 0$, where the drive absorbs zero energy. But in the case of a continuous eigenfrequency $E > 0$, the drive will continuously absorb energy even for $\epsilon \to 0$ because it 'comes in from infinity', just as it continuously radiates energy out in the outgoing case.
- Note that since everything in the definitions of $\hat{G}_\pm$ is real except for the $i\epsilon$, the $\hat{G}_\pm$ are Hermitian conjugates.

With the above water wave intuition, we can understand the Green's operators analytically.

- Define the difference of the Green's operators by

$$\hat{\Delta}(E) = \lim_{\epsilon\to 0}\left[\hat{G}_+(E+i\epsilon) - \hat{G}_-(E-i\epsilon)\right] = \lim_{\epsilon\to 0}\left(\frac{1}{E+i\epsilon-H} - \frac{1}{E-i\epsilon-H}\right).$$

- This limit is easier to understand in terms of ordinary numbers,

$$\lim_{\epsilon\to 0}\left(\frac{1}{x-x_0+i\epsilon} - \frac{1}{x-x_0-i\epsilon}\right) = \lim_{\epsilon\to 0}\frac{-2i\epsilon}{(x-x_0)^2+\epsilon^2} = -2\pi i\delta(x-x_0).$$

  Therefore we have

$$\hat{\Delta}(E) = -2\pi i\delta(E-H).$$

  The operator on the right-hand side is defined by each eigenvector, i.e. an eigenvector of H with eigenvalue E_0 becomes an eigenvector with eigenvalue $\delta(E-E_0)$. Explicitly,

$$\delta(E-H) = \sum_{n\alpha}|n\alpha\rangle\langle n\alpha|\delta(E-E_n) + \int_0^\infty dE'\sum_\alpha|E'\alpha\rangle\langle E'\alpha|\delta(E-E').$$

  We see that $\hat{\Delta}(E)$ is zero when E is not an eigenvalue, diverges when $E = E_n$, and is finite when $E > 0$ with $\hat{\Delta}(E) = -2\pi i\sum_\alpha|E\alpha\rangle\langle E\alpha|$.

- Therefore $\hat{G}_-(z)$ is the analytic continuation of $\hat{G}_+(z)$ through the gaps between the discrete eigenvalues, so they are both part of the same analytic function called the resolvent,

$$\hat{G}(z) = \frac{1}{z-H}$$

  which is defined for all z that are not eigenvalues of H. The resolvent has poles at every discrete eigenvalue, and a branch cut along the continuous eigenvalues.

- We can analytically continuous $\hat{G}_+(z)$ across the positive real axis, 'pushing aside' the branch cut to reach the second Riemann sheet of the resolvent. In this case we can encounter additional singularities in the lower-half plane, which correspond to resonances (e.g. long-lived bound states). **(need a good example for this!)**

Example. The free particle Green's functions $G_{0\pm}(\mathbf{x},\mathbf{x}',E)$ in three dimensions. Setting $z = E+i\epsilon$,

$$G_{0+}(\mathbf{x},\mathbf{x}',z) = \langle\mathbf{x}|(z-H_0)^{-1}|\mathbf{x}'\rangle = \int d\mathbf{p}\,d\mathbf{p}'\,\langle\mathbf{x}|\mathbf{p}\rangle\langle\mathbf{p}|\frac{1}{z-H_0}|\mathbf{p}'\rangle\langle\mathbf{p}'|\mathbf{x}'\rangle = \int\frac{d\mathbf{p}}{(2\pi\hbar)^3}\frac{e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}')/\hbar}}{z-p^2/2m}.$$

To simplify, we set $\mathbf{x}' = 0$ for simplicity, by translational invariance, let $\mathbf{p} = \hbar\mathbf{q}$, and let $z = E + i\epsilon = \hbar^2 w^2/2m$ for a complex wavenumber w (so that w is the first quadrant), giving

$$G_{0+}(\mathbf{x}, z) = -\frac{1}{(2\pi)^3}\frac{2m}{\hbar^2}\int d\mathbf{q}\,\frac{e^{i\mathbf{q}\cdot\mathbf{x}}}{q^2 - w^2} = \frac{1}{(2\pi)^2}\frac{2m}{\hbar^2}\frac{i}{x}\int_{-\infty}^{\infty} dq\,\frac{qe^{iqx}}{(q-w)(q+w)}$$

where we performed the angular integration. To do the final integral, we close the contour in the upper-half plane, picking up the $q = w$ pole. Then

$$G_{0+}(\mathbf{x}, z) = -\frac{1}{4\pi}\frac{2m}{\hbar^2}\frac{e^{iwx}}{x}.$$

The incoming Green's function is similar, but now we choose the branch of the square root so that w lies in the fourth quadrant, so we pick up the $q = -w$ pole instead, giving e^{-iwx}. Converting back to wavenumbers, we have

$$G_{0\pm}(\mathbf{x}, E) = -\frac{1}{4\pi}\frac{2m}{\hbar^2}\begin{cases} e^{\pm ikx}/x, & E \geq 0, \\ e^{-\kappa x}/x, & E \leq 0 \end{cases}$$

where the quantities $k, \kappa \sim \sqrt{\pm E}$ are all real and positive. By taking this choice of branches, we have ensured that $G_{0\pm}$ is continuous across the negative real axis, but as a result it is discontinuous across the positive real axis, as expected.

10.4 The Lippmann-Schwinger Equation

Green's functions provide a powerful general formalism for scattering problems. Below we focus on potential scattering, though the same techniques work in many contexts, such as field theories.

- We are interested in solutions to the driven time-independent Schrodinger equation

  $$(E - H_0)\psi(\mathbf{x}) = V(\mathbf{x})\psi(\mathbf{x})$$

  where $E > 0$, and have shown that solutions can be written as

  $$\psi(\mathbf{x}) = \phi(\mathbf{x}) + \int d\mathbf{x}'\, G_0(\mathbf{x}, \mathbf{x}', E)V(\mathbf{x}')\psi(\mathbf{x}')$$

  where $\phi(\mathbf{x})$ solves the homogeneous equation (i.e. free particle with energy E).

- Since we are interested in scattering solutions, we take the outgoing Green's function G_{0+} and let the homogeneous solution be an incoming plane wave $|\phi_{\mathbf{k}}\rangle = |\mathbf{k}\rangle$, which satisfies $E = \hbar^2 k^2/2m$. This yields the Lippmann-Schwinger equation. In terms of kets, it reads

  $$|\psi_{\mathbf{k}}\rangle = |\phi_{\mathbf{k}}\rangle + \hat{G}_{0+}(E)V|\psi_{\mathbf{k}}\rangle$$

  We add the subscript $\mathbf{k}$ to emphasize that the solution depends on the choice of $\mathbf{k}$, not just on E, as it tells us which direction the particles are launched in. In terms of wavefunctions,

  $$\psi_{\mathbf{k}}(\mathbf{x}) = \phi_{\mathbf{k}}(\mathbf{x}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\int d\mathbf{x}'\,\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}V(\mathbf{x}')\psi_{\mathbf{k}}(\mathbf{x}').$$

- There are many variations on the Lippmann-Schwinger equation. For example, in proton-proton scattering V is the sum of a Coulomb potential and the nuclear potential. Then we might include the Coulomb term in H_0, so that the incoming wave would be a Coulomb solution of positive energy, and we would use Green's functions for the Coulomb potential.

- Now suppose that the potential cuts off after a finite radius, and we observe the scattering at a much larger radius $r = |\mathbf{x}|$. Then $x' \ll r$ in the integral above, and we may expand in a power series in x'/r, throwing away all terms falling faster than $1/r$, giving

$$\psi_{\mathbf{k}}(\mathbf{x}) \approx \phi_{\mathbf{k}}(\mathbf{x}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\frac{e^{ikr}}{r}\int d\mathbf{x}'\, e^{-i\mathbf{k}'\cdot\mathbf{x}'}V(\mathbf{x}')\psi_{\mathbf{k}}(\mathbf{x}').$$

  In particular, this matches the 'incident plus scattered' form of the wavefunction postulated in the beginning of this section, with scattering amplitude

$$f(\mathbf{k},\mathbf{k}') = -\frac{(2\pi)^{3/2}}{4\pi}\frac{2m}{\hbar^2}\int d\mathbf{x}'\, e^{-i\mathbf{k}'\cdot\mathbf{x}'}V(\mathbf{x}')\psi_{\mathbf{k}}(\mathbf{x}') = -\frac{4\pi^2 m}{\hbar^2}\langle\mathbf{k}'|V|\psi_{\mathbf{k}}\rangle.$$

  Thus we have proven that the wavefunction must have such a form in general. We can also prove a similar statement for rapidly decaying potentials, but it fails for the Coulomb potential.

- We can also use the incoming Green's function; this describes a solution where waves come in from infinity and combine to come out as a plane wave. Since the outgoing solution is much more realistic, we focus on it and may leave the plus sign implicit.

- Finally, when $E < 0$, we get an integral expression for bound states,

$$\psi(\mathbf{x}) = -\frac{1}{4\pi}\frac{2m}{\hbar^2}\int d\mathbf{x}'\,\frac{e^{-\kappa|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}V(\mathbf{x}')\psi(\mathbf{x}')$$

  where there is no homogeneous term, because free particle solutions do not decay at infinity. Solutions only exist for discrete values of E. There is also no choice in Green's function as both agree on the negative real axis.

We can use the Lippmann-Schwinger equation to derive a perturbation series for scattering, called the Born series.

- We may rewrite the Lippmann-Schwinger equation in the form

$$|\mathbf{k}\rangle = (1 - G_{0+}(E)V)|\psi_{\mathbf{k}}\rangle$$

  which has the formal solution

$$|\psi_{\mathbf{k}}\rangle = \Omega_+(E)|\mathbf{k}\rangle, \quad \Omega_+(E) = (1 - G_{0+}(E)V)^{-1}$$

  where $\Omega_+(E)$ is called the Moller scattering operator. Similarly we may define an incoming form $\Omega_-(E)$ and a general operator $\Omega(z)$ with complex energy and

$$\Omega(z) = (1 - G_0(z)V)^{-1}, \quad \Omega_\pm(E) = \lim_{\epsilon\to 0}\Omega(E \pm i\epsilon).$$

- Expanding in a series in V gives the Born series,
$$\Omega(z) = 1 + G_0(z)V + G_0(z)VG_0(z)V + \ldots$$
which explicitly gives
$$|\psi_{\mathbf{k}}\rangle = |\mathbf{k}\rangle + G_{0+}(E)V|\mathbf{k}\rangle + G_{0+}(E)VG_{0+}(E)V|\mathbf{k}\rangle + \ldots.$$
Substituting this into the expression for the scattering amplitude gives
$$f(\mathbf{k}, \mathbf{k}') = -\frac{4\pi^2 m}{\hbar^2}\left[\langle\mathbf{k}'|V|\mathbf{k}\rangle + \langle\mathbf{k}'|VG_{0+}(E)V|\mathbf{k}\rangle + \ldots\right].$$
When we truncate these series at V^n, we get the n^{th} Born approximation. The Born series can also be derived by plugging the Lippmann-Schwinger equation into itself.

- The first Born approximation recovers our first-order result from time-dependent perturbation theory: the scattering amplitude is proportional to the Fourier transform of the potential. In general, the Dyson series (from time-dependent perturbation theory) is very similar to the Born series. They both expand in powers of V, but in the time/energy domain respectively.

- We can also phrase the results in terms of the exact Green's operator
$$G(z) = \frac{1}{z - H}.$$
Playing around and suppressing the z argument, we have
$$G = G_0 + G_0VG = G_0 + GVG_0$$
which are Lippmann-Schwinger equations for G. This gives the exact Green's function as a series in the number of scatterings off the potential.

- By playing around some more, we find
$$\Omega = 1 + GV, \quad |\psi_{\mathbf{k}}\rangle = (1 + GV)|\mathbf{k}\rangle.$$
In this picture, a scattering process occurs through an initial scattering, then propagation by the exact Green's function.

Example. We show the scattering states $|\psi_{\mathbf{k}}\rangle$ are orthonormal using Green's functions. We have
$$\langle\psi_{\mathbf{k'}}|\psi_{\mathbf{k}}\rangle = \langle\psi_{\mathbf{k'}}|\mathbf{k}\rangle + \langle\psi_{\mathbf{k'}}|G_+(E)V|\mathbf{k}\rangle = \langle\psi_{\mathbf{k'}}|\mathbf{k}\rangle + \lim_{\epsilon\to 0}\frac{1}{E + i\epsilon - E'}\langle\psi_{\mathbf{k'}}|V|\mathbf{k}\rangle$$
where $E' = \hbar^2 k'^2/2m$. Next, using the Lippmann-Schwinger equation on the first factor,
$$\langle\psi_{\mathbf{k}'}|\mathbf{k}\rangle = \langle\mathbf{k}'|\mathbf{k}\rangle + \langle\psi_{\mathbf{k}'}|VG_{0-}(E')|\mathbf{k}\rangle = \langle\mathbf{k}'|\mathbf{k}\rangle + \lim_{\epsilon\to 0}\frac{1}{E' - i\epsilon - E}\langle\psi_{\mathbf{k}'}|V|\mathbf{k}\rangle.$$
Then the extra terms cancel, giving $\langle\psi_{\mathbf{k'}}|\psi_{\mathbf{k}}\rangle = \langle\mathbf{k}'|\mathbf{k}\rangle = \delta(\mathbf{k} - \mathbf{k}')$. The completeness relation is
$$\sum_{n\alpha}|n\alpha\rangle\langle n\alpha| + \int d\mathbf{k}\,|\psi_{\mathbf{k}}\rangle\langle\psi_{\mathbf{k}}| = 1$$
where the first term includes bound states, which are orthogonal to all scattering states.

10.5 The S-Matrix

We introduce the S-matrix using the simple example of one-dimensional potential scattering.

- With an incoming right-moving wave, we may write the scattered wave as

$$\psi_R(x) \sim \begin{cases} e^{ikx} + re^{-ikx} & x \to -\infty, \\ te^{ikx} & x \to +\infty. \end{cases}$$

  Then $R = |r|^2$ and $T = |t|^2$ give the probability of reflection and transmission, as can be seen by computing the probability fluxes. Conservation of probability requires $R + T = 1$.

- Similarly, we can use left-moving waves, and define

$$\psi_L(x) \sim \begin{cases} t'e^{-ikx} & x \to -\infty, \\ e^{-ikx} + r'e^{ikx} & x \to +\infty. \end{cases}$$

- Since the potential is real, if ψ is a solution, then ψ^* is as well. This gives the identities

$$t' = t, \quad r' = -\frac{r^* t}{t^*}$$

  so that $|r| = |r'|$. These results also appear in classical scattering as a result of time-reversal symmetry. The same symmetry is acting here, as time reversal is complex conjugation.

- As an explicit example, the finite well potential $V(x) = -V_0\theta(x - a/2)\theta(x + a/2)$ has

$$r = \frac{(k^2 - q^2)\sin(qa)e^{-ika}}{(q^2 + k^2)\sin(qa) + 2iqk\cos(qa)}, \quad t = \frac{2iqke^{-ika}}{(q^2 + k^2)\sin(qa) + 2iqk\cos(qa)}, \quad q^2 = \frac{2mV_0}{\hbar^2} + k^2.$$

  We note that there is perfect reflection for low k, no reflection for high k, and also perfect transmission for k so that $\sin(qa) = 0$, i.e. resonant transmission. We also note that $r = r'$. This follows from parity symmetry, as we'll see below.

- We summarize our data in terms of the S-matrix,

$$\begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} I_R \\ I_L \end{pmatrix} + S \begin{pmatrix} O_R \\ O_L \end{pmatrix}, \quad S = \begin{pmatrix} t & r \\ r' & t' \end{pmatrix}$$

  where I_R is an incoming right-moving wave, O_L is an outgoing left-moving wave, and so on. Applying our identities above shows that S is unitary.

Next, we consider a general parity-symmetric potential $V(x) = V(-x)$.

- It is useful to switch to a parity basis,

$$I_+(x) = e^{-ik|x|}, \quad I_-(x) = \mathrm{sign}(x)e^{-ik|x|}, \quad O_+(x) = e^{ik|x|}, \quad O_-(x) = -\mathrm{sign}(x)e^{ik|x|}$$

  which is related by the change of basis

$$\begin{pmatrix} I_+ \\ I_- \end{pmatrix} = M \begin{pmatrix} I_R \\ I_L \end{pmatrix}, \quad \begin{pmatrix} O_+ \\ O_- \end{pmatrix} = M \begin{pmatrix} O_R \\ O_L \end{pmatrix}, \quad M = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

  Applying this transformation, the S-matrix in the parity basis is $S^P = MSM^{-1}$.

- For a parity-symmetric potential, $r = r'$ because $\psi_R(x) = \psi_L(-x)$. Then S^P simplifies to

$$S^P = \begin{pmatrix} S_{++} & \\ & S_{--} \end{pmatrix}, \quad S_{++} = t + r, \quad S_{--} = t - r.$$

  The off-diagonal elements are zero because parity is conserved.

- Combining our identities shows that S_{++} and S_{--} are phases,

$$S_{++} = e^{2i\delta_+(k)}, \quad S_{--} = e^{2i\delta_-(k)}.$$

  This is analogous to how we distilled three-dimensional central force scattering into a set of phases in the partial wave decomposition.

- The S-matrix can also detect bound states. Since the algebra used to derive $r(k)$ and $t(k)$ never assumed that k was real, the same expressions hold for general complex k. Consider a pure imaginary wavenumber $k = i\lambda$ with even parity,

$$\lim_{|x|\to\infty} \psi_+(x) = I_+(x) + S_{++}O_+(x), \quad I_+(x) = e^{\lambda|x|}, \quad O_+(x) = e^{-\lambda|x|}.$$

  It looks like there can't be a bound state solution here, since the I_+ component diverges at infinity. The trick is to rewrite this as

$$\lim_{|x|\to\infty} \psi_+(x) = S_{++}^{-1} I_+(x) + O_+(x)$$

  which gives a valid bound state as long as $S_{++}^{-1} = 0$, which corresponds to a pole in S_{++}. That is, we can identify bound states from poles in S-matrix elements! (The same reasoning works in the original left/right basis, though there are more terms.)

- Some careful algebra shows that

$$S_{++}(k) = -e^{-ika} \frac{q\tan(qa/2) - ik}{q\tan(qa/2) + ik}$$

  which shows that bound states of even parity occur when $\lambda = q\tan(qa/2)$, a familiar result. We can recover the bound state energy from $E = -\hbar^2\lambda^2/2m$.