IAS/Park City Mathematics Series
Volume 1, 1995

Introduction to Quantum Field Theory for Mathematicians

Jeffrey M. Rabin

Introduction

The aim of these lectures is to introduce mathematicians to the basic physical concepts and mathematical structures of classical mechanics, quantum mechanics, and quantum field theory. From the physicist's perspective the presentation will be very standard, being a highly compressed version (minus the detailed examples and computational techniques) of the standard graduate courses in these subjects. I have deliberately avoided translating the material completely into the language and notation of modern mathematics, preferring to present the standard physics notation and terminology (but with a running mathematical commentary). My goal is to help the student learn to read the physics literature and converse with its authors on their own terms, as well as to introduce the specific subject matter of quantum field theory (QFT). I have used the standard notation of each field of physics I discuss, but this unavoidably leads to some conflicts when the standard notations of different fields happen to coincide. I have tried to point out such ambiguities when they arise. An extensive set of exercises appears at the end of these lecture notes.

I do not assume prior knowledge of physics beyond $F = ma$, although some previous exposure to quantum mechanics would cushion the shock of absorbing it all at once. The mathematical background of a first-year graduate student should be more than adequate, with perhaps some review of the following topics: the elementary methods of the calculus of variations needed for the derivation of the Euler-Lagrange equation; linear algebra, especially the spectral properties of Hermitian operators in finite-dimensional spaces and the faith that unbounded self-adjoint operators in Hilbert spaces behave similarly; what is meant by a representation of a Lie group or Lie algebra; and some idea of how to work with distributions and Fourier transforms.

These lectures may also be viewed as an introduction for mathematicians to the alien culture of the physics community. One important cultural difference to keep

[1]Department of Mathematics University of California at San Diego, La Jolla, CA 92093.
E-mail address: jrabin@ucsd.edu.

in mind is that the notation used in physics is not designed to emphasize the logical relations of the concepts involved, but rather to facilitate explicit calculations. It is often intentionally ambiguous so that a given formula can be reinterpreted in whatever way makes calculation easiest and most natural, and it discourages one from getting bogged down in questions of rigor. Another difference, in fact, is that these questions of rigor are not addressed from the start, at the stage of definitions, but later, after the heuristic calculations have been done on the basis of provisional definitions. The philosophy is to proceed formally until difficulties are encountered, and only then resolve them. The difficulties often arise from an over-idealized mathematical model of a given physical situation, and only after the calculation is completed can one understand their physical origin and base the correct resolution on the true behavior of the physical system under consideration.

It is not easy to recommend references on the subjects of these lectures for a mathematical audience. Standard graduate physics texts are written for readers whose physical intuition has been developed through exposure to many examples and computations in classical and quantum mechanics and electromagnetism, and who are thoroughly steeped in the alien culture described above. Nevertheless the following suggestions should be useful. For classical mechanics, V. Arnold, *Mathematical Methods of Classical Mechanics* [**1**], or the standard graduate physics text, H. Goldstein, *Classical Mechanics* [**2**]. Among the many textbooks on elementary quantum mechanics, A.Z. Capri, *Nonrelativistic Quantum Mechanics* [**3**] is more careful than most concerning unbounded operators, and R. Shankar, *Principles of Quantum Mechanics* [**4**] contains a discussion of path integrals. For the path integral formulation of quantum mechanics direct from the source, see R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* [**5**]. Two books on QFT which mathematicians have found palatable are P. Ramond, *Field Theory: A Modern Primer* [**6**], and L.H. Ryder, *Quantum Field Theory* [**7**]. See also R.J. Rivers, *Path Integral Methods in Quantum Field Theory* [**8**]. A comprehensive treatment of all aspects of renormalization theory can be found in J. Collins, *Renormalization* [**9**]. Finally, I recommend two books on the rigorous foundations of QFT with the warning that such books are somewhat removed from the actual practice of field theory by physicists, and from the geometric and topological subject matter of "physical mathematics". They serve to reassure mathematicians who find themselves unable to suspend their disbelief that QFT is not built on sand. These are J. Glimm and R. Jaffe, *Quantum Physics* [**10**], and N.N. Bogolubov, A.A. Logunov, and R.T. Todorov, *Introduction to Axiomatic Quantum Field Theory* [**11**].

Acknowledgements

I thank the organizers of the Regional Geometry Institute for the opportunity to deliver this course of lectures. I am indebted to Bruce Driver for acting as my assistant during the course, for proofreading these notes, and for contributing the first two groups of homework exercises.

LECTURE 1
Classical Mechanics

We begin with the elementary formulation of classical mechanics for a single particle of mass m moving on a path $\mathbf{x}(t)$ in $\mathbf{R}^3$, as described by Newton's second law of motion,

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}(\mathbf{x}, t), \tag{1.1}$$

where $\mathbf{F}$ is the force acting on the particle. We restrict ourselves to the case of a conservative force, namely one which is the gradient of some (time-independent) potential energy function V on $\mathbf{R}^3$, $\mathbf{F} = -\nabla V(\mathbf{x})$. Then there is an energy function which is conserved along the particle's trajectory,

$$E = \frac{1}{2}m\left|\frac{d\mathbf{x}}{dt}\right|^2 + V(\mathbf{x}), \tag{1.2}$$

$$\frac{dE}{dt} = m\frac{d\mathbf{x}}{dt}\cdot\frac{d^2\mathbf{x}}{dt^2} + \frac{d\mathbf{x}}{dt}\cdot\nabla V(\mathbf{x}) = 0. \tag{1.3}$$

Although Newton's law of motion is in a sense a complete statement of classical mechanics, there are several reasons to desire another formulation. Newton's law is a vector equation, and it is unpleasant to transform components of vectors into exotic coordinate systems. It requires that all forces acting on a mechanical system be known explicitly, whereas in practice one often describes a force implicitly by its effects, e.g. constraining the trajectory to lie on the surface of a sphere. And it describes the motion locally, in terms of the local acceleration caused by the local force. Newton's law gives almost no insight into the mathematical structure of classical mechanics, nor is it a suitable starting point for quantization of a classical system. And it cannot describe classical systems which do not consist of particles, such as the electromagnetic field. The Lagrangian formulation of mechanics to be discussed now gives a global and coordinate-free characterization of the classical trajectory. Then we will derive the Hamiltonian formulation which reveals symplectic geometry as the mathematical structure of mechanics. Both formulations are

general enough to describe fields as well as particles, and both are suitable for the transition to quantum mechanics. The Hamiltonian approach leads to the canonical formulation of quantum mechanics in terms of self-adjoint operators in a Hilbert space, while the Lagrangian approach leads to the path integral formulation.

Let us generalize to a system of many particles having coordinates $q_i(t)$ and velocities $\dot{q}_i(t) = dq_i/dt$. (Custom dictates Newton's notation in which a dot denotes the time derivative.) For N particles moving in $\mathbf{R}^3$, i runs from 1 to $3N$, and the motion is described by a path in the configuration space $\mathbf{R}^{3N}$. However, we do not assume that the q_i are necessarily rectangular coordinates. Any set of $3N$ variables labeling the particles' locations will do. Given the points in configuration space representing the system at initial time t_1 and final time t_2, we wish to characterize the actual path followed by the system, among all others joining the same points, as the one which extremizes some functional S of the path, called the action. So we introduce

$$S = \int_{t_1}^{t_2} L(q_i, \dot{q}_i) dt. \tag{1.4}$$

The function L is called the Lagrangian and will be specified later. It depends on two independent sets of $3N$ variables, denoted q_i and $\dot{q}_i$ for mnemonic reasons, but the latter are not the derivatives of the former. This is an example of intentionally ambiguous notation. The mnemonic reminds us how to evaluate S on a given path: substitute $q_i(t)$ and $\dot{q}_i(t)$ for the arguments of L and integrate the resulting function of t.

Let $q_i(t)$ be a path for which S is stationary. That is, under the change $q_i(t) \to q_i(t) + \delta q_i(t)$, with $\delta q_i(t)$ an arbitrary (smooth) function vanishing at the endpoints t_1 and t_2, the change δS in the action should be zero to linear order in $\delta q_i(t)$. We compute to linear order

$$\delta S = \int_{t_1}^{t_2} \left(\delta q_i \frac{\partial L}{\partial q_i} + \delta \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} \right) dt \tag{1.5}$$

$$= \int_{t_1}^{t_2} \delta q_i \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) dt + \left[\delta q_i \frac{\partial L}{\partial \dot{q}_i} \right]_{t_1}^{t_2}, \tag{1.6}$$

where we have integrated by parts. The Einstein summation convention is in effect throughout these lectures: a sum is implied over any index such as i which is repeated in a given term. Since $\delta q_i = 0$ at the endpoints, the boundary contribution vanishes and we obtain the Euler-Lagrange equations for the extremal path,

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = 0. \tag{1.7}$$

To recover Newton's equations for a particle we can take

$$L(x_i, \dot{x}_i) = \frac{1}{2} m \dot{x}_i \dot{x}_i - V(x_i), \tag{1.8}$$

leading to the Euler-Lagrange equations

$$-\frac{\partial V}{\partial x_i} - \frac{d}{dt}(m\dot{x}_i) = 0, \tag{1.9}$$

which indeed agree with Newton's law. In general the Euler-Lagrange equations are second-order ODEs which can be expected to have a unique solution given the initial positions and velocities of the particles.

A great advantage of the Lagrangian formulation of mechanics, and a running theme of these lectures, is the beautiful connection it reveals between the symmetries of a physical system and its conservation laws. The first example of a conservation law is evident from Eq. (1.7), namely, if the Lagrangian is independent of a particular coordinate q_i then the quantity $p_i \equiv \partial L/\partial \dot{q}_i$ is conserved during the motion. It is called the momentum conjugate to the coordinate q_i. For example, for a free particle with $V = 0$, L is independent of all coordinates and the three conserved quantities $p_i = m\dot{x}_i$ are just the components of the usual momentum vector of elementary mechanics.

More generally, suppose there is a continuous symmetry (a 1-parameter Lie group action on the space of paths) which leaves the action unchanged regardless of the limits of integration t_1, t_2. Infinitesimally this is a variation $q_i(t) \to q_i(t) + \delta q_i(t)$ for which we no longer assume $\delta q_i(t_1) = \delta q_i(t_2) = 0$. Noether's theorem states that such a symmetry implies the existence of a conserved quantity. To see this, we reverse our viewpoint on Eq. (1.6): the boundary term no longer vanishes automatically, but the first term is zero along the extremal path because that path obeys the Euler-Lagrange equations. Then we have

$$0 = \delta S = \left[\delta q_i \frac{\partial L}{\partial \dot{q}_i} \right]_{t_1}^{t_2}. \tag{1.10}$$

Because t_1 and t_2 can be arbitrary times, we see that $p_i \delta q_i$ is conserved along the extremal trajectory. The simplest example is again the free particle. The fact that the Lagrangian is independent of the coordinates can be recast as the invariance of the Lagrangian under the three infinitesimal translations $q_i \to q_i + \epsilon_i$, which are of the above form with $\delta q_i = \epsilon_i$. By Noether's theorem, $p_i \delta q_i = \epsilon_i p_i$ is conserved. Since the ϵ_i are arbitrary, we recover the conservation of the three components of momentum. In general there is a conserved quantity corresponding to each generator of the Lie algebra of a Lie group of symmetries. However, discrete groups of symmetries do not lead to conservation laws: it was crucial to the entire argument that the symmetries could be taken to lie in an infinitesimal neighborhood of the identity.

The Lagrangian approach to mechanics gives a global and coordinate-free description of the path in configuration space, as well as a deep understanding of the origin of conservation laws. As we will see later, the same framework can describe physical systems such as fields which do not consist of moving particles, simply by choosing a different Lagrangian function. Papers in the physics literature always specify the system under consideration by writing down the Lagrangian. From our present viewpoint, perhaps the only complaint about this formulation of mechanics that comes to mind is that the Euler-Lagrange equations are second order ODEs,

whereas it is first order ODEs which have a nice geometric interpretation as flows along vector fields. We motivate the Hamiltonian formulation of mechanics by the desire to interpret the time evolution of a physical system as such a flow.

We define the Hamiltonian of a system as the function of $6N$ suggestively-named variables $H(q_i, p_i) \equiv p_i\dot{q}_i - L(q_i, \dot{q}_i)$. The meaning of this equation is as follows. The right side is a function of q_i and $\dot{q}_i$. We assume that the defining equations for the momenta, $p_i = \partial L/\partial \dot{q}_i$, can be inverted so as to solve for the $\dot{q}_i$ in terms of q_i and p_i, and we use this to eliminate the $\dot{q}_i$ and reexpress the Hamiltonian in terms of the q_i and p_i only. This procedure is called a Legendre transformation. By the implicit function theorem, we can carry it out provided that

$$\frac{\partial p_i}{\partial \dot{q}_j} = \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \tag{1.11}$$

is a nonsingular matrix. This is always true in simple examples, and we can give a physical argument that it is true in all particle mechanics problems. Recall that L normally contains a kinetic energy term which is quadratic in the velocities $\dot{q}_i$. The matrix will be nonsingular provided that each velocity component contributes to the kinetic energy, which is certainly expected. However, as we will see, it is not true in gauge theories, and this is the origin of all the difficulties in quantizing gauge theories.

Computing the differential of H, we find

$$\begin{aligned} dH &= p_i d\dot{q}_i + \dot{q}_i dp_i - \frac{\partial L}{\partial q_i}dq_i - \frac{\partial L}{\partial \dot{q}_i}d\dot{q}_i \\ &= \dot{q}_i dp_i - \frac{\partial L}{\partial q_i}dq_i, \end{aligned} \tag{1.12}$$

where the first and last terms cancel by definition of the momenta. Thus dH is automatically expressed in terms of the differentials of the variables on which H depends. This in fact motivates the Legendre transformation, and it allows us to read off the partial derivatives of H as

$$\frac{\partial H}{\partial p_i} = \dot{q}_i, \qquad \frac{\partial H}{\partial q_i} = -\frac{\partial L}{\partial q_i}. \tag{1.13}$$

All this follows purely from the definition of H. Now, however, suppose that we are on the trajectory obeying the Euler-Lagrange equations,

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = \dot{p}_i. \tag{1.14}$$

Combining this with Eq. (1.13) gives the alternative characterization of the extremal trajectory in terms of Hamilton's equations of motion,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{1.15}$$

which are first order equations as promised. Furthermore, along the extremal path we have $dH = \dot{q}_i dp_i - \dot{p}_i dq_i$, or, dividing by dt, $dH/dt = 0$, and H itself is a conserved quantity.

For an example of all this formalism we return to the particle in a potential $V(x_i)$, for which

$$\begin{aligned} H &= (m\dot{x}_i)\dot{x}_i - \frac{1}{2}m\dot{x}_i\dot{x}_i + V(x_i) \\ &= \frac{1}{2}m(\dot{x}_i)^2 + V(x_i) \\ &= \frac{p_i^2}{2m} + V(x_i). \end{aligned} \tag{1.16}$$

(Note the summation convention even in p_i^2!) Here the Hamiltonian is conserved because it is the total energy. Hamilton's equations of motion become $\dot{x}_i = p_i/m$ and $\dot{p}_i = -\nabla_i V$, equivalent to Newton's laws.

Now let us compute the time derivative, along the extremal path of course, of an arbitrary function $A(q,p)$ of the q's and p's.

$$\begin{aligned} \frac{dA}{dt} &= \frac{\partial A}{\partial q_i}\dot{q}_i + \frac{\partial A}{\partial p_i}\dot{p}_i \\ &= \frac{\partial A}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial q_i} \\ &\equiv [A, H]. \end{aligned} \tag{1.17}$$

The expression $[A, B]$ defined for any two functions of the variables q, p is called their Poisson bracket, and has the usual properties of a Lie bracket:

$$[A, B] = -[B, A], \tag{1.18}$$

$$[aA + bB, C] = a[A, C] + b[B, C], \tag{1.19}$$

where a, b are constants;

$$[AB, C] = A[B, C] + [A, C]B, \tag{1.20}$$

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0, \tag{1.21}$$

as well as the fundamental or canonical Poisson bracket relations

$$[q_i, q_j] = [p_i, p_j] = 0, \quad [q_i, p_j] = \delta_{ij}. \tag{1.22}$$

The equation $dA/dt = [A, H]$ reduces to Hamilton's equations themselves if $A = q_i$ or p_i, and may be considered the basic dynamical equation of Hamiltonian mechanics. It says that the Hamiltonian is the infinitesimal generator of time

evolution via the Poisson bracket, and makes it clear that the conserved quantities are precisely those whose bracket with the Hamiltonian vanishes.

The Poisson bracket allows us to expose an even deeper connection between symmetries and the associated conserved quantities. Consider a system admitting an infinitesimal symmetry $q_i \to q_i + \epsilon f_i(q)$, with ϵ an "infinitesimal parameter" (a physicist's reminder that only linearized variations are important) and the f_i some functions. The corresponding conserved quantity from Noether's theorem is $Q \equiv p_i f_i(q)$. We compute the Poisson bracket,

$$[q_j, Q] = \frac{\partial q_j}{\partial q_i}\frac{\partial Q}{\partial p_i} - \frac{\partial q_j}{\partial p_i}\frac{\partial Q}{\partial q_i} = \delta_{ij} f_i(q) = f_j(q), \tag{1.23}$$

and therefore $\delta q_j = \epsilon[q_j, Q]$, showing that the conserved quantity is the infinitesimal generator of the symmetry via the Poisson bracket. The status of the Hamiltonian as the generator of translations in time is a special case of this general phenomenon. [Actually, before asserting that Q generates the symmetry we should verify that $\delta p_j = \epsilon[p_j, Q]$ as well. It would take us somewhat afield to discuss the effect of symmetries on the momenta p_j; this would lead us into the geometry of canonical transformations, or symplectic diffeomorphisms. We would quickly discover that if there is a Lie group G of symmetries, of dimension k, then the Poisson bracket algebra of the k conserved quantities Q_n coincides (up to a possible central extension) with the Lie algebra of G.]

The Geometry of Hamiltonian Mechanics

Although we will not use it in these lectures, I want to give a translation dictionary between the concepts of Hamiltonian mechanics developed above and the coordinate-free language of symplectic geometry. This should help mathematicians to better understand the physicists' coordinate-dependent notation, as well as connecting the treatment given here to that in Robert Bryant's lectures in this volume.

The variables denoted q_i were global coordinates on the configuration space $\mathbf{R}^{3N}$; more generally they will be local coordinates on a configuration space M which may be any manifold whose points correspond to the possible configurations of the system under discussion. The variables $\dot{q}_i$ should be viewed as fiber coordinates for the tangent bundle $T(M)$ inasmuch as they describe all possible tangent vectors (velocities) to all possible paths in M. The Lagrangian $L(q, \dot{q})$ is a globally defined function on $T(M)$. Given a smooth path in M, one can lift it to a path in $T(M)$ by lifting it along the fiber $T_p(M)$ at each point p to its tangent vector there. Then one can restrict the Lagrangian to the lifted path and integrate this restriction to obtain the action S.

The momenta p_i are fiber coordinates for the cotangent bundle $T^*(M)$ which is known as the phase space; it follows from the definition $p_i = \partial L/\partial \dot{q}_i$ that under a change of coordinates p_i and $\dot{q}_i$ transform by inverse matrices, corresponding to dual bundles. Then the change of variables from $(q, \dot{q})$ to (q, p) is supposed to be the expression in coordinates of a diffeomorphism from $T(M)$ to $T^*(M)$ which is the identity on M. In the usual case in which L is quadratic in the $\dot{q}_i$ it is a linear isomorphism on the fibers, but in general it may not be linear or even surjective. Recall that there is not even a canonical identification of the fibers $T_p(M)$ and $T_p^*(M)$ unless some additional data such as a metric is given; in the usual case

the kinetic energy term in L is a quadratic form on $T_p(M)$ which serves as this metric. The Hamiltonian $H(q,p)$ is a function on $T^*(M)$. The cotangent bundle of any manifold always has a canonical symplectic structure given by the closed, nondegenerate 2-form $\omega = dq_i dp_i = -d(p_i dq_i)$. This form is globally defined despite being specified in local coordinates: the fact that p_i are fiber coordinates for the cotangent bundle means precisely that $p_i dq_i$ is an invariantly defined 1-form. To any function $A(q,p)$ on $T^*(M)$ we associate a vector field $\tilde{A}$ which is its symplectic gradient; namely the vector field which is dual to the 1-form dA in the sense that $\omega(\tilde{A}, \cdot) = dA$. In coordinates we have

$$\tilde{A} = \left(\frac{\partial A}{\partial p_i}, -\frac{\partial A}{\partial q_i}\right) = \frac{\partial A}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial}{\partial p_i}. \tag{1.24}$$

Then Hamiltonian dynamics is simply the flow along the vector field $\tilde{H}$, and the Poisson bracket is nothing but $[A, B] = \omega(\tilde{A}, \tilde{B})$. It has the properties of a Lie bracket because the correspondence between functions and vector fields associates the Poisson bracket $[A, B]$ to the Lie bracket $[\tilde{B}, \tilde{A}]$. Finally, a conserved quantity Q gives rise to a vector field $\tilde{Q}$ whose flow commutes with that of $\tilde{H}$ and represents the action of the symmetry on the points of phase space; the Lie bracket algebra of the vector fields generating a Lie group of symmetries should reproduce the Lie algebra of that group.

LECTURE 2
Classical Field Theory

We have stressed that the Lagrangian and Hamiltonian formulations of mechanics are applicable to systems such as the electromagnetic and gravitational fields, which (classically) do not consist of particles.[2] These systems are described by PDEs such as Maxwell's equations rather than the ODEs which describe the motion of particles. We consider the mathematically simplest example of a real-valued field $\phi(x)$ on $\mathbf{R}^4$ which satisfies the Klein-Gordon equation,

$$(\Delta + m^2)\phi(x) = 0. \tag{2.1}$$

Here $x = x^\mu = (x^0, x^1, x^2, x^3) = (t, x, y, z)$ is a point in $\mathbf{R}^4$ specifying a location $\mathbf{x} = x^i$ in space and a time t at which the field is measured. Note the convention that Greek indices run from 0 to 3 while Roman indices run from 1 to 3, so that x^i are the spatial components of x^μ. $\Delta = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$ is the wave operator, the four-dimensional Laplacian or D'Alembertian corresponding to the Minkowski or Lorentz metric

$$\eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.2}$$

[Actually the four-dimensional Laplacian is more commonly denoted by a square (four sides) than by Δ (three), but this character wasn't available to my typesetting program!] We abbreviate $\partial_\mu \equiv \partial/\partial x^\mu$, so that $\Delta = \eta^{\mu\nu}\partial_\mu\partial_\nu$. We also make the

[2] Mathematicians sometimes bemoan the absence of a definition of the term "field" in the physics literature. But "field", like "particle", is a physical concept, not a mathematical one. An appropriate answer to the question, "What is a field?" would be, "The electromagnetic field is a field." Fields really exist and can be observed in the laboratory. Of course, within a given theory we represent fields and particles by specific mathematical structures. One theory might say that a particle is a path in $\mathbf{R}^3$ while another says it is an irreducible representation of the Lorentz group. For present purposes a field is a function from R^4 to some algebra, usually the real numbers.

convention that the indices on any tensor can be raised and lowered using the metric, e.g.

$$\partial^\mu \equiv \eta^{\mu\nu}\partial_\nu = (\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}), \tag{2.3}$$

and we write the Klein-Gordon equation as $(\partial_\mu\partial^\mu + m^2)\phi = 0$, or even as $(\partial^2 + m^2)\phi = 0$. The parameter m will ultimately turn out to be the mass of something when we reach QFT. [The contortions with indices express the fact that the metric allows us to identify cotangent vectors (those with upper indices) with tangent vectors (those with lower indices), but this geometric view is overly sophisticated when one is in Euclidean space. Think of them as merely a notational device to keep track of the minus signs arising from the indefinite metric.]

It is crucial to understand the transition from our earlier variables $q_i(t)$ to the field $\phi(x) = \phi(\mathbf{x}, t)$. The field ϕ, like an electric field, is an observable quantity which can be measured at any location $\mathbf{x}$ in space at any time. Similarly q_i is an observable quantity which can be measured, for each i, at any time. Thus, ϕ corresponds to q, t is always the time, and — most importantly — $\mathbf{x}$ corresponds to the index i. Sums over i in the formulas of particle mechanics can be expected to translate into integrals over $\mathbf{x}$ in field theory. For this reason field theory is often described as the mechanics of systems having infinitely many degrees of freedom.

We can obtain the Klein-Gordon equation as the extremal condition for either of the two actions

$$S = -\frac{1}{2}\int_V d^4x\, \phi(\partial_\mu\partial^\mu + m^2)\phi, \tag{2.4}$$

or

$$S = \frac{1}{2}\int_V d^4x\, (\partial_\mu\phi\partial^\mu\phi - m^2\phi^2). \tag{2.5}$$

These differ by a boundary term resulting from integration by parts, but this will not change the extremal condition. The integrand for S is called the Lagrangian density $\mathcal{L}$; technically its integral over space (sum over i) would be the Lagrangian L, although commonly $\mathcal{L}$ itself is called the Lagrangian.

Let us obtain the condition for a general action to be stationary to first order under variations $\delta\phi$ which vanish at the boundary of the volume V. From

$$S = \int_V d^4x\, \mathcal{L}(\phi, \partial_\mu\phi) \tag{2.6}$$

we obtain

$$\delta S = \int_V d^4x \left(\delta\phi\frac{\partial\mathcal{L}}{\partial\phi} + \partial_\mu\delta\phi\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\right) \tag{2.7}$$

$$= \int_V d^4x\, \partial_\mu\left(\delta\phi\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\right) + \int_V d^4x\, \delta\phi\left(\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\right). \tag{2.8}$$

The derivatives appearing here are formal derivatives of $\mathcal{L}$ thought of as a function of the independent variables ϕ and $\partial_\mu\phi$. The first integral is zero by the divergence theorem for $\delta\phi$ vanishing on ∂V, so the Euler-Lagrange equations are

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} = 0. \tag{2.9}$$

For the Lagrangian of Eq. (2.5) they are easily seen to give the Klein-Gordon equation.

In field theory there is a much stronger, local version of Noether's theorem. Again we suppose that $\delta\phi$ is an infinitesimal symmetry, so that $\delta S = 0$ for any region V even though $\delta\phi$ no longer vanishes on the boundary. Along the path satisfying the Euler-Lagrange equations the second integral in Eq. (2.8) is now zero, whereas the vanishing of the first integral for all regions V implies that

$$\partial_\mu j^\mu = 0, \quad j^\mu = \delta\phi\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}. \tag{2.10}$$

Separating the time and space components of the current vector $j^\mu = (j^0, \mathbf{j})$ gives the differential conservation law $\partial_0 j^0 = -\nabla\cdot\mathbf{j}$. If we define the total charge at time t by

$$Q(t) = \int d^3x\, j^0(t, \mathbf{x}), \tag{2.11}$$

and assume that $\mathbf{j}$ vanishes fast enough at infinity, then the divergence theorem implies the conservation of charge $dQ/dt = 0$. However, we can also define the charge inside a specified region,

$$Q_V(t) = \int_V d^3x\, j^0(t, \mathbf{x}), \tag{2.12}$$

and deduce that any change in this charge must be accounted for by a flux of the current vector $\mathbf{j}$ through the boundary of V. This local conservation law is much stronger than the overall conservation of Q throughout space, which would allow charge to simply vanish from one region and simultaneously reappear elsewhere without traversing the space between. These conserved currents play a fundamental role in QFT and particle physics. [The terminology stems from electromagnetism, where the Q corresponding to the $U(1)$ symmetry group is indeed the electric charge, and $\mathbf{j}$ is the electric current density. Physicists often refer to any conserved quantity as a charge or current.]

The above discussion gives only the simplest version of Noether's theorem, applicable to an "internal symmetry". For a symmetry such as a spatial rotation, which acts on the coordinates x^μ as well as the field ϕ, the derivation requires more care because the integration region V is different in the new coordinate system. Goldstein [**2**] derives the conserved current

$$j^\mu = (\delta\phi - \delta x^\nu\partial_\nu\phi)\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} + \mathcal{L}\,\delta x^\mu \tag{2.13}$$

in this case.

I will give only a brief discussion of the Hamiltonian formulation of field theory, mainly to fix some notation and introduce the notion of functional derivative which will be useful later. Each field ϕ appearing in the Lagrangian has a conjugate momentum field defined by $\pi(x) \equiv \partial\mathcal{L}/\partial\partial_0\phi$, and the Hamiltonian is defined by

$$H = \int d^3x[\pi(x)\partial_0\phi(x) - \mathcal{L}]. \tag{2.14}$$

H is constructed from the fields at a fixed time t, but as usual it will be conserved and thus independent of t. For example, for the Klein-Gordon field we find $\pi(x) = \partial^0\phi(x)$ and

$$H = \frac{1}{2}\int d^3x\left(\pi^2 + |\nabla\phi|^2 + m^2\phi^2\right), \tag{2.15}$$

a positive definite functional of $\phi(x)$ which can be identified as the energy in the field. The Hamilton equations of motion can be written in nearly the same form they had in particle mechanics by introducing the functional derivative notation,

$$\dot{\phi}(\mathbf{x},t) = \frac{\delta H}{\delta\phi(\mathbf{x})}, \quad \dot{\pi}(\mathbf{x},t) = -\frac{\delta H}{\delta\pi(\mathbf{x})}, \tag{2.16}$$

or alternatively in the Poisson bracket form $\dot{A} = [A, H]$ with

$$[A,B] \equiv \int d^3z\left[\frac{\delta A}{\delta\phi(\mathbf{z})}\frac{\delta B}{\delta\pi(\mathbf{z})} - \frac{\delta A}{\delta\pi(\mathbf{z})}\frac{\delta B}{\delta\phi(\mathbf{z})}\right]. \tag{2.17}$$

The functional derivative appearing in these equations is merely the physicists' way of expressing the (Fréchet) derivative of a functional on an infinite-dimensional function space. To define the derivative of a functional F on a function space at a particular point ϕ, a mathematician would linearize the functional about that point, writing $F[\phi + \delta\phi] = F[\phi] + F_\phi[\delta\phi] + O(\delta\phi^2)$, and define the linear map F_ϕ to be the derivative. Physicists, however, insist on expressing such linear maps in terms of their integral kernels. For example, the simple linear functional $\phi \to \phi(0)$ is thought of in terms of the Dirac delta function, $\phi(0) = \int dx\,\phi(x)\delta(x)$. Similarly, the functional derivative is the integral kernel of F_ϕ:

$$F_\phi[\delta\phi] = \int d^3x\,\delta\phi(\mathbf{x})\frac{\delta F}{\delta\phi(\mathbf{x})}. \tag{2.18}$$

The integral d^3x here is appropriate for Hamiltonian mechanics, in which the fields are thought of as functions on $\mathbf{R}^3$ at each fixed time; later we will encounter functional derivatives with respect to functions on $\mathbf{R}$ or $\mathbf{R}^4$. An alternative definition,

$$\frac{\delta F}{\delta\phi(\mathbf{y})} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\{F[\phi(\mathbf{x}) + \epsilon\delta^3(\mathbf{x}-\mathbf{y})] - F[\phi(\mathbf{x})]\}, \tag{2.19}$$

where $\delta^3(\mathbf{x}-\mathbf{y})$ is the Dirac delta distribution in $\mathbf{x}$ with support at $\mathbf{y}$, shows that the functional derivative measures the response of the functional F to a change in the function ϕ supported at a point $\mathbf{y}$; it does depend on the point $\mathbf{y}$ in general. Some basic examples (in four dimensions) which will be useful later are

$$\frac{\delta}{\delta\phi(y)}\int d^4x\,\phi^n(x) = n\phi^{n-1}(y), \tag{2.20}$$

$$\frac{\delta}{\delta\phi(y)}\exp\int d^4x\,J(x)\phi(x) = J(y)\exp\int d^4x\,J(x)\phi(x). \tag{2.21}$$

Returning to mechanics we also have $\delta\phi(\mathbf{x})/\delta\phi(\mathbf{y}) = \delta^3(\mathbf{x}-\mathbf{y})$, which leads to the canonical brackets

$$[\phi(\mathbf{x}), \pi(\mathbf{y})] = \delta^3(\mathbf{x}-\mathbf{y}). \tag{2.22}$$

The fact that the canonical Poisson bracket is a distribution rather than a smooth function reflects our choice of coordinates for field theory. The configuration space M is the space of smooth functions $\phi(\mathbf{x})$ on $\mathbf{R}^3$, perhaps with some boundary conditions at infinity. Compare the configuration space $\mathbf{R}^3$ for a single particle, which can be viewed as the space of functions on the index set $i \in \{1,2,3\}$, recalling that $\mathbf{x}$ is analogous to i. The coordinate function q_i on $\mathbf{R}^3$ assigns to any function on the index set its value on the index i. Similarly, $\phi(\mathbf{x})$ above is the coordinate function on M which assigns to any function on $\mathbf{R}^3$ its value at $\mathbf{x}$. $\pi(\mathbf{x})$ is a similar coordinate function on the fibers of $T^*(M)$. But we know that the values of functions at points are never the best coordinates on function spaces; much better coordinates are things like Fourier coefficients which are the integrals of the functions against various kernels. For each test function f on $\mathbf{R}^3$ we can introduce a better coordinate function $\phi[f] = \int d^3x f(\mathbf{x})\phi(\mathbf{x})$ and similarly for $\pi[f]$. Then the Poisson bracket becomes nonsingular; $[\phi[f], \pi[g]] = \int d^3x f(\mathbf{x})g(\mathbf{x})$. These ideas will reappear in QFT.

As an example of a physically important field theory, in fact a gauge theory, let us now discuss electromagnetism. The classical electromagnetic field in a vacuum is described by a pair of time-dependent vector-valued functions on $\mathbf{R}^3$, the electric field $\mathbf{E}(x)$ and the magnetic field $\mathbf{B}(x)$. In some system of electromagnetic units, in which the speed of light c is equal to unity, they obey Maxwell's equations,

$$\nabla\cdot\mathbf{B} = 0, \quad \nabla\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t} = 0, \tag{2.23}$$

$$\nabla\cdot\mathbf{E} = 0, \quad \nabla\times\mathbf{B} - \frac{\partial\mathbf{E}}{\partial t} = 0. \tag{2.24}$$

These equations assume a more covariant appearance in terms of the antisymmetric electromagnetic field strength tensor,

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{bmatrix}, \tag{2.25}$$

or the corresponding 2-form $F = \frac{1}{2}F^{\mu\nu}dx_\mu dx_\nu$. The Hodge dual ${}^*F$ corresponds to the dual tensor $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\kappa}F_{\lambda\kappa}$, where $\epsilon^{\mu\nu\lambda\kappa}$ is totally antisymmetric in its indices and $\epsilon^{0123} = +1$. Then Maxwell's equations (2.23) read $\partial_\mu \tilde{F}^{\mu\nu} = 0$ or $dF = 0$, while Eqns. (2.24) become $\partial_\mu F^{\mu\nu} = 0$ or $d^*F = 0$. Since we are in Euclidean space, $dF = 0$ is solved by $F = dA$ or $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ for some 1-form called the potential. Of course A is not unique; we have the freedom to change it by any gauge transformation $A \to A + d\Lambda$ or $A^\mu \to A^\mu + \partial^\mu \Lambda$ with an arbitrary function Λ.

If we assume the representation $F = dA$, then the remaining Maxwell equations $\partial_\mu F^{\mu\nu} = 0$ can be obtained by varying A^μ in the gauge-invariant action

$$S = -\frac{1}{4}\int d^4x\, F_{\mu\nu}F^{\mu\nu} = -\frac{1}{2}\int F \wedge {}^*F. \tag{2.26}$$

Alternatively, both these equations and the relation between A and F can be obtained from

$$S = -\frac{1}{2}\int d^4x\, F_{\mu\nu}(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{4}\int d^4x\, F_{\mu\nu}F^{\mu\nu} \tag{2.27}$$

by varying A and F independently.

We see that the equations of this gauge theory coincide with the equations for a connection on a $U(1)$ principal fiber bundle over $\mathbf{R}^4$. A is the connection viewed as a $U(1)$ Lie algebra ($= \mathbf{R}$) valued 1-form on the base space, and $F = dA + \frac{1}{2}[A, A]$ is the curvature, with $[A, A] = 0$ for this Abelian algebra. $A \to A + d\Lambda$ is the change in A under a change of bundle trivialization. To see this, write an element of the gauge group as $g = e^{i\Lambda}$, so that $g^{-1}dg = id\Lambda$ and $A \to A + d\Lambda$ can be rewritten as $A \to g^{-1}Ag - ig^{-1}dg$, which looks more familiar as $iA \to g^{-1}iAg + g^{-1}dg$. So the mathematician's connection is iA; this relative factor i between mathematicians' and physicists' conventions occurs frequently. Its origin is that physicists want A to be real, or more generally Hermitian, whereas mathematicians prefer skew Hermitian.

From the action of Eq. (2.26) we compute the momenta $\pi^\mu = \partial\mathcal{L}/\partial\partial_0 A_\mu$ conjugate to A^μ as $\pi^0 = 0, \pi^i = F^{i0} = E^i = \partial^i A^0 - \partial^0 A^i$. As advertised, in this gauge theory the velocity $\partial_0 A_0$ contributes nothing to the kinetic energy, consequently $\pi^0 = 0$ and we cannot solve for the velocities in terms of the momenta to perform the Legendre transform. It is easy to understand why this phenomenon is linked with gauge invariance. If it were possible to pass to the Hamiltonian formulation, we would obtain a unique solution to the initial value problem for the equations of motion: given the initial point in phase space, simply flow along the Hamiltonian vector field. However, the initial value problem cannot have a unique solution

in a gauge theory: a gauge transformation with a time-dependent function Λ can produce a different solution with the same initial data.

One way to evade this difficulty is by gauge-fixing: seeking solutions to the equations of motion which obey some additional condition which picks out one A from each gauge orbit. In general there are topological obstructions to such a choice of gauge [**12**], but evidently it is enough to pick out a subset of A's whose stabilizer contains only time-independent gauge transformations. In electrodynamics, either the temporal gauge $A^0 = 0$ or the Coulomb gauge $\nabla \cdot \mathbf{A} = 0$ plus $A^0 = 0$ is a satisfactory choice, but in quantized non-Abelian gauge theories a straightforward gauge-fixing is computationally intractable. The extensive machinery of BRST quantization and Faddeev-Popov ghost fields has been developed to handle such systems. For an introduction, see [**13, 14, 15**].

LECTURE 3
The Lorentz Group and Spinors

The particles which are produced and studied in accelerators have velocities comparable to the velocity of light c (which is 1 in our units!), and consequently exhibit clearly the effects predicted by the special theory of relativity. The mathematical content of special relativity is that the Lorentz group, in fact the Poincaré group, (or at least the connected component of the identity in these groups) must act on the space of solutions to the equations of motion of particles and fields.

In relativistic physics we are concerned with inertial observers, namely observers for whom free particles appear to move with constant velocity. Intuitively such observers are themselves unaccelerated (e.g. not observing the world from a roller coaster) but may be in uniform motion relative to one another. Each such observer sets up a three dimensional coordinate system as well as a clock to measure time, and assigns to each event of interest a set of coordinates in the resulting coordinate system on $\mathbf{R}^4$. The coordinate transformations relating the coordinate systems of different observers comprise the Poincaré group; the subgroup fixing the origin of $\mathbf{R}^4$ is the Lorentz group (by convention the origins of the spatial coordinate systems of the corresponding observers coincide at time $t = 0$). The Poincaré group must act on the space of solutions of the equations of motion of any physical system, so that each orbit can be interpreted as a single physical solution viewed in the coordinates of all possible inertial observers. The simplest way to ensure this in the Lagrangian formulation of mechanics is to require the action to be Poincaré invariant. By Noether's theorem, any relativistic theory will then contain conserved quantities whose Poisson bracket algebra is that of the Poincaré group; these quantities are the energy, momentum, and angular momentum of the system.

The basic result of special relativity is the identification of the Lorentz group as the group $O(1,3)$ of linear transformations of $\mathbf{R}^4$ preserving the quadratic form $\eta_{\mu\nu}x^\mu x^\nu = t^2 - x^2 - y^2 - z^2$. The null vectors of this form satisfy $x^2 + y^2 + z^2 = t^2$, which is the equation of a sphere in $\mathbf{R}^3$ of radius t. If a pulse of light is emitted at the origin at $t = 0$ it will reach this sphere at time t, so the constancy of the speed of light for all observers implies that the Lorentz group must preserve this sphere. Some more physical input, such as the homogeneity and isotropy of space, is needed to conclude that the quadratic form and not merely its nullspace is preserved. The Poincaré group is the semidirect product of the Lorentz group with the translation group of $\mathbf{R}^4$.

Let the 4×4 matrix Λ belong to the Lorentz group, so that $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ is a Lorentz transformation (humor me regarding the index locations). The condition of preserving the quadratic form implies $\Lambda^T \eta \Lambda = \eta$. Taking the determinant of this equation, we learn that $\det \Lambda = \pm 1$. Taking its 00 component gives $(\Lambda^0{}_0)^2 = 1 + \Lambda^i{}_0 \Lambda^i{}_0 \geq 1$. Therefore, the Lorentz group consists of at least — in fact, exactly — four connected components, characterized by the signs of $\det \Lambda$ and $\Lambda^0{}_0$. The identity component, for which both signs are $+$, is the proper Lorentz group. The other components contain various reflections in the space and time axes, which in fact are known not to be exact symmetries of the laws of particle physics. (The violations of parity and time-reversal invariance were among the revolutionary experimental discoveries of the 1950s and '60s in particle physics.)

To expose the important relation between the Lorentz group and $SL(2,C)$ we associate to any vector x^μ in $\mathbf{R}^4$ the Hermitian matrix

$$X = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix} = x^\mu \sigma_\mu. \tag{3.1}$$

Here the Pauli matrices σ_μ form a convenient basis for the space of 2×2 Hermitian matrices,

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.2}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.3}$$

They obey $\sigma_i \sigma_j = i\epsilon_{ijk}\sigma_k$ and are orthogonal with respect to the inner product $\mathrm{Tr}(\sigma_\mu \sigma_\nu) = 2\delta_{\mu\nu}$. This allows us to invert relations like Eq. (3.1) by dotting each side with σ_ν to obtain $x^\mu = \frac{1}{2}\mathrm{Tr}(X\sigma_\mu)$.

We note that under the correspondence (3.1) the Lorentz norm corresponds to the determinant: $x^\mu x_\mu = \det X$. For any matrix A belonging to $SL(2,C)$, the complex 2×2 matrices of determinant 1, the transformation $X' = AXA^\dagger$ preserves the Hermiticity and the determinant of X and consequently induces a Lorentz transformation on x^μ, proper in fact because $SL(2,C)$ is connected ($A^\dagger$ is the standard physics notation for the adjoint or Hermitian conjugate of a matrix, the complex conjugate transposed). We can find this Lorentz transformation as follows:

$$x'^\mu = \frac{1}{2}\mathrm{Tr}(X'\sigma_\mu) = \frac{1}{2}\mathrm{Tr}(AXA^\dagger\sigma_\mu) = \frac{1}{2}x^\nu \mathrm{Tr}(A\sigma_\nu A^\dagger \sigma_\mu). \tag{3.4}$$

Thus $\Lambda^\mu{}_\nu = \frac{1}{2}\mathrm{Tr}(\sigma_\mu A \sigma_\nu A^\dagger)$. It also follows that

$$\Lambda^\mu{}_\nu \sigma_\nu = A^\dagger \sigma_\mu A, \tag{3.5}$$

a formula which we will need shortly. It is not difficult to show (exercise) that we have defined a homomorphism of $SL(2,\mathbf{C})$ onto the proper Lorentz group, with kernel exactly $\{\pm 1\}$. Therefore the proper Lorentz group is doubly connected, and $SL(2,\mathbf{C})$ is its universal cover.

Suppose that two observers O and O' each measure some physical quantities, such as the electromagnetic field components, at a point P to which they assign coordinates x^μ and $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ respectively, obtaining the results $F^{\mu\nu}(x)$ and $F'^{\mu\nu}(x')$, respectively. The simplest assumption one can make is that their results are related by the matrix associated to Λ in some representation of the Lorentz group which is characteristic of the particular quantities measured.

Example 3.1. $\phi(x)$ appearing in the Klein-Gordon equation is called a scalar field because it is associated to the trivial representation: $\phi'(x') = \phi(x)$. Indeed, the Klein-Gordon equation is invariant under this representation in the sense that if $(\Delta + m^2)\phi(x) = 0$ then automatically $(\Delta' + m^2)\phi'(x') = 0$, because m is a scalar and Δ is a Lorentz-invariant differential operator.

Example 3.2. Maxwell's equations are invariant if the potential A transforms according to the fundamental vector representation of the Lorentz group, $A'^\mu(x') = \Lambda^\mu{}_\nu A^\nu(x)$, while the field strength F transforms by the tensor product of two vector representations, $F'^{\mu\nu}(x') = \Lambda^\mu{}_\sigma \Lambda^\nu{}_\kappa F^{\sigma\kappa}(x)$. We see the motivation for the fanatical index conventions: any product of tensors which is summed over a pair of upper and lower indices will transform correctly as a tensor having the remaining indices. Contracting all pairs of indices so that none remain gives a Lorentz scalar, suitable for use as a Lagrangian — the physicist's solution to the problem of invariant theory.

Example 3.3. A representation of $SL(2,\mathbf{C})$ assigns to the elements $\pm A$ either the same matrix, or a pair of matrices $\pm M$. In the former case it induces a true representation of the Lorentz group, while in the latter case we say we have a "double-valued representation" of the Lorentz group. There are in fact two inequivalent complex two-dimensional representations of $SL(2,\mathbf{C})$: left-handed spinors which obey $\psi'_L(x') = A\psi_L(x)$, and right-handed spinors obeying $\psi'_R(x') = (A^\dagger)^{-1}\psi_R(x)$. Each ψ is a two-component complex column vector. The direct sum of these two "chiral representations" is called the Dirac spinor representation, and it acts on four-component complex column vectors by $\psi'(x') = S(\Lambda)\psi(x)$, where

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}, \quad S = \begin{pmatrix} A & 0 \\ 0 & A^{\dagger -1} \end{pmatrix}. \tag{3.6}$$

Before constructing a Lorentz-invariant wave equation for $\psi(x)$, let me clarify the physical interpretation of double-valued representations. The notation $S(\Lambda)$ is misleading because the Lorentz transformation Λ only determines S up to sign. Therefore, the relation between the values of ψ measured by observers O and O' is also ambiguous by a sign. Of course I have not yet given a physical interpretation of ψ or a procedure for measuring it, and in fact no satisfactory interpretation exists outside the context of QFT. However, part of the interpretation must be that the overall sign of ψ is an arbitrary convention and not determined by the physical measurements. Observable quantities such as the energy or momentum associated to the field ψ will turn out to be quadratic in ψ and unaffected by this sign ambiguity.

For convenience in manipulating Dirac spinors we define 4×4 matrices analogous to σ_μ, the Dirac γ matrices which in terms of 2×2 subblocks are

$$\gamma^0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}, \qquad \gamma^j = \begin{pmatrix} 0 & -\sigma_j \\ \sigma_j & 0 \end{pmatrix}. \tag{3.7}$$

They satisfy the Clifford algebra $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}I$, where I is the 4×4 identity matrix which is often omitted from this and similar formulas. This is a hint of the well-known relationship between Clifford algebras and the (pseudo-)orthogonal Lie groups. Note that γ^0 is Hermitian but γ^j is skew Hermitian. The projection operators onto the two-dimensional subspaces of the left and right handed spinors are $(1 \pm \gamma^5)/2$, where

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.8}$$

(The story behind the notation γ^5 is this. In the early days of relativity some authors used four-dimensional coordinates x^μ with $\mu = 0, 1, 2, 3$ while others preferred $\mu = 1, 2, 3, 4$. Consequently one author's γ^4 was likely to be another's γ^0, and to avoid confusion this new matrix had to be called γ^5.) Two crucial properties of the γ matrices are $S^{-1} = \gamma^0 S^\dagger \gamma^0$, which can be verified directly, and $S^{-1}\gamma^\mu S = \Lambda^\mu{}_\nu \gamma^\nu$, which is a consequence of Eq. (3.5). They imply that in a certain sense the index on γ^μ behaves like a Lorentz vector index and can be used as such to form invariants.

For example, consider the four complex functions (actually 1×1 matrices) $j^\mu = \bar{\psi}\gamma^\mu\psi$, where $\bar{\psi} \equiv \psi^\dagger\gamma^0$ is the "Dirac conjugate" of ψ. I claim that under Lorentz transformations they behave as the components of a vector field. Proof:

$$\begin{aligned} j'^\mu(x') &= \bar{\psi}'(x')\gamma^\mu\psi'(x') = \psi^\dagger(x)S^\dagger\gamma^0\gamma^\mu S\psi(x) = \bar{\psi}(x)\gamma^0 S^\dagger\gamma^0\gamma^\mu S\psi(x) \\ &= \bar{\psi}(x)S^{-1}\gamma^\mu S\psi(x) = \Lambda^\mu{}_\nu\bar{\psi}(x)\gamma^\nu\psi(x) = \Lambda^\mu{}_\nu j^\nu(x). \end{aligned} \tag{3.9}$$

A similar computation shows that the following is a Lorentz invariant action:

$$\int d^4x\, \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x), \tag{3.10}$$

whose variation with respect to the components of ψ yields the Dirac equation

$$(i\gamma^\mu\partial_\mu - m)\psi = 0. \tag{3.11}$$

The parenthesized expression is a 4×4 matrix differential operator; the constant m (which really means mI) will ultimately be the mass of some particle, such as an electron, which is somehow related to ψ. By applying the operator $i\gamma^\nu\partial_\nu + m$ to both sides of the Dirac equation, one can verify that each component of ψ individually obeys the Klein-Gordon equation. In terms of the right and left handed spinors the Dirac equation reads

$$\begin{bmatrix} i(\sigma_0\partial_0 - \sigma_j\partial_j)\psi_R \\ i(\sigma_0\partial_0 + \sigma_j\partial_j)\psi_L \end{bmatrix} = \begin{bmatrix} m\psi_L \\ m\psi_R \end{bmatrix}. \tag{3.12}$$

Physicists call the self-adjoint operator $i\gamma^\mu\partial_\mu - m$ the Dirac operator, whereas mathematicians would use this term for one operator of the adjoint pair $i(\sigma_0\partial_0 \pm \sigma_j\partial_j)$ which map the chiral subspaces to each other. We note that for $m = 0$ we can write an invariant equation for each chiral spinor separately, but for $m \neq 0$ Lorentz invariance necessarily couples them. This is the motivation for introducing the reducible Dirac representation. If, as some experiments now suggest, the electron's neutrino has a nonzero mass, we would have to abandon the traditional view of the "left-handed neutrino" described by the chiral spinor ψ_L alone.

The Dirac equation is invariant under the $U(1)$ action $\psi \to e^{i\theta}\psi$, and the corresponding conserved current is precisely j^μ of Eq. (3.9). It represents the electric current carried by the moving particles (e.g. electrons) which are (somehow) described by the Dirac equation. As promised earlier it is quadratic in the components of ψ and therefore insensitive to the sign ambiguity in those components. As a last note on the Dirac equation, note that we are free to change basis in the four dimensional Dirac representation space by $\psi \to K\psi$ and $\gamma^\mu \to K\gamma^\mu K^{-1}$ with K a constant matrix. This leads to different representations of the γ matrices which are more convenient in various situations when it is not important to separate the chiral components of ψ.

Finally, a comment on the geometric interpretation of fields. In these lectures, we always take $\mathbf{R}^4$ as our model of spacetime. We have of course exploited the simple geometry of $\mathbf{R}^4$ to simplify our formulas, but this necessarily obscures the general situation. On a general spacetime manifold M we would have a principal fiber bundle P whose structure group is the full symmetry group of our physical theory, $G = \text{Lorentz} \times U(1) \times \cdots$, together with vector bundles associated to various representations of G. Some fields, such as the electromagnetic potential A, are viewed as connections on P, while others such as ψ are sections of the associated bundles. The action is some invariant integral over M which depends on the fields but not on arbitrary choices such as local trivializations. The Lorentz group factor of G acts on the frame bundle of M, which is one factor of P, rather than on M itself. In our case we have implicitly used the simple geometry of $M = \mathbf{R}^4$ in several ways. Since $\mathbf{R}^4$ is contractible, all bundles are trivial. Hence we can pick global trivializations and view connections and sections as functions on the base space. Under a change of global trivialization the structure group acts on these functions according to various representations, as we have described. Furthermore, $\mathbf{R}^4$ can be identified with its tangent space at any point. Hence we can set the frame bundle connection (gravitational field) to zero, and restrict ourselves to changes of trivialization which are uniform over M and hence preserve this condition. Then the Lorentz group can be viewed as acting on M rather than on the tangent spaces. A physicist would say it has become a global symmetry rather than a local one.

LECTURE 4
Quantum Mechanics

Very small objects, such as molecules, atoms, and subatomic particles, do not obey the laws of classical mechanics. Quantum mechanics was developed during the 1920s in order to account for their behavior. The Schrödinger equation is to quantum mechanics what Newton's second law is to classical mechanics: a simple and, in principle, complete statement of the basic physics which however is not the most powerful method for solving practical problems, and which moreover obscures the true mathematical structures of the theory. Nevertheless we will begin with this equation because of its familiarity to some readers.

A particle of mass m is described in quantum mechanics by a complex-valued "wave function" $\psi(\mathbf{x}, t)$ (not to be confused with the four-component Dirac spinor despite the coincidence of the standard notations) which obeys the Schrödinger wave equation,

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{x})\psi, \tag{4.1}$$

where $V(\mathbf{x})$ is the same potential energy function that appeared in classical mechanics, and $\hbar = h/2\pi = 1.055 \times 10^{-27}$ erg-sec is a fundamental constant called Planck's constant (Actually h was originally Planck's constant, but since $h/2\pi$ appears in most formulas a special notation was introduced for it). We will shortly adopt units in which $\hbar = 1$ as we did for the speed of light. The standard way to solve such a PDE is of course by separation of variables, seeking solutions of the form $\psi(\mathbf{x}, t) = \psi(\mathbf{x})\exp(-iEt/\hbar)$, where E is a constant having the dimensions of energy. This leads to the eigenvalue problem

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi, \tag{4.2}$$

known as the time-independent Schrödinger equation. The physical interpretation of the wave function, also known as the probability amplitude,[3] is that $|\psi(\mathbf{x}, t)|^2 d^3x$

[3] The term amplitude refers to anything which must be squared in order to produce the quantity of physical significance. It derives from wave theory, where the energy of a wave is proportional to its amplitude squared.

represents the probability of finding the particle located in an infinitesimal volume element d^3x at the point $\mathbf{x}$ at time t. ψ should belong to $L^2(\mathbf{R}^3)$ so that the total probability can be normalized to unity, $\int d^3x\,|\psi|^2 = 1$, and this provides the boundary condition for the eigenvalue problem. The fundamental tenet of quantum mechanics is that no amount of knowledge concerning the initial conditions or the physical environment of the particle will enable one to predict more about its future motion than the probabilities given by ψ. For those accustomed to classical mechanics, this failure of determinism can be difficult to accept. It has been said that if you believe quantum mechanics when you first learn it, then you didn't understand it!

We will now give the Hamiltonian formulation of quantum mechanics which reveals its true mathematical context as the spectral theory of operators in Hilbert space. The setting is completely different from classical mechanics, but the central fact that time evolution is a flow generated by the Hamiltonian is preserved. The description of any particular physical system starts with a complex vector space $\mathcal{H}$ associated to that system and called "The Hilbert Space" whether it is one or not (generally it fails to be one only in that the scalar product may not be positive definite). Usually $\mathcal{H}$ is infinite-dimensional, often the space of complex L^2 functions on the classical configuration space M, sometimes tensored with another finite-dimensional vector space. The spectral properties of the Hermitian operators on $\mathcal{H}$ play a central role.

This creates a problem because the operators of interest are generally unbounded and often have no eigenfunctions in $\mathcal{H}$. For example, in $L^2(\mathbf{R})$ we will be interested in the operator $-id/dx$, whose eigenfunctions are $\exp ipx$ with eigenvalue p. Even in the best case, when p is real, these do not belong to $\mathcal{H}$. Even worse, the operator x (that is, multiplication by x) has no eigenfunctions at all unless we allow distributions like $\delta(x-a)$, an "eigenfunction" with eigenvalue a. Of the many ways to deal with this situation, I will describe only the setup known as a rigged[4] Hilbert space. This is a nested sequence $\Omega \subset \mathcal{H} \cong \mathcal{H}^* \subset \Omega^*$, where Ω is a "nuclear subspace" dense in $\mathcal{H}$, and Ω^* is its dual space. The definition of a nuclear subspace is complicated, but when $\mathcal{H}$ is a function space Ω can be the functions decreasing faster than any power of $|x|$ at infinity. Then the "eigenfunctions" encountered above can be viewed as distributions belonging to Ω^*, e.g. $\exp ipx$ is reinterpreted as the linear functional $f(x) \to \int dx f(x) \exp ipx$, namely the Fourier transform. Further, the inverse transform $f(x) = \int (dp/2\pi)\hat{f}(p) \exp -ipx$ can be viewed as the expansion of the vector $f(x)$ in $\mathcal{H}$ in terms of the "basis" of eigenfunctions of the Hermitian operator $-id/dx$. In this sense, the spectral theory needed to formulate quantum mechanics works fine even though the relevant eigenfunctions are not in $\mathcal{H}$.

Returning to quantum mechanics, each state of the physical system is supposed to be represented by a normalized (unit length) vector in $\mathcal{H}$. The state of a system is the totality of information about the system at a given time needed to solve the initial value problem for its evolution starting at that time. In classical mechanics the states were represented by the points of phase space, the values of all coordinates and momenta at a given time. Next, the observables of the system — the measurable quantities such as positions or momenta of particles, their energies or

[4] "Rigged" not in the sense of prearranged for dishonest purposes, as a rigged election, but in the sense of well-equipped for a voyage, as a rigged schooner.

angular momenta and so forth, which were functions on phase space in classical mechanics — are represented by Hermitian (self-adjoint) linear operators (usually unbounded) in $\mathcal{H}$. Notation: vectors in $\mathcal{H}$ are denoted by $|\psi>$, where ψ is any convenient mnemonic label; e.g. an eigenvector of some operator A with eigenvalue a might be denoted by $|a>$. The scalar product $(\phi, A\psi)$ is written $<\phi|A|\psi>$ and is called a matrix element of A; if there is an orthonormal basis of $\mathcal{H}$ containing the vectors $|\phi>$ and $|\psi>$ then this is an entry in the matrix of A relative to this basis. The symbol $<\phi|$ appearing by itself denotes the vector in V^* which is dual to $|\phi>$ via the scalar product. The notation $<\phi|A|\psi>$ is intentionally ambiguous, readable as a scalar product or as the evaluation of a dual vector on a vector.

Suppose then that the system is in a state $|\psi>$ and we measure some observable represented by the Hermitian operator A — what result will we obtain? The answer depends on the spectrum of A. It is conventional to discuss separately the cases of discrete and continuous spectrum, although a unified discussion could be given in terms of the spectral projection operators associated to A. We also assume for simplicity that the eigenvalues are all simple (nondegenerate). Thus, let A have eigenvectors $|a>$ obeying $A|a> = a|a>$ and normalized so that $<a'|a> = \delta_{aa'}$ for discrete spectrum, or $<a'|a> = \delta(a-a')$ for continuous spectrum.[5] Then the result of the measurement will be one of the eigenvalues a of A, with probability $|<a|\psi>|^2$ for the discrete spectrum. For continuous spectrum $|<a|\psi>|^2 da$ is the probability that the result lies between a and $a+da$. The total probability is 1 because of the normalization of $|\psi>$, and the eigenvalues are real because A is Hermitian.

We will frequently use the identity $\sum_a |a><a| = 1$ for the orthonormal eigenvectors of any Hermitian operator, called "inserting a complete set of states". Applying both sides to any vector gives $\sum_a |a><a|\psi> = |\psi>$, which is just the expansion of $|\psi>$ in an orthonormal basis. (It is conventional to write such formulas as if the spectrum were discrete, with the understanding that the sum means an integral over a spectral measure in general.) As an example of its use, let us calculate the average value obtained over many measurements of A in state $|\psi>$. This is the sum of the possible values weighted by their probabilities,

(4.3)
$$\sum_a a|<a|\psi>|^2 = \sum_a a\,<\psi|a><a|\psi> = \sum_a <\psi|A|a><a|\psi> = <\psi|A|\psi>,$$

giving a direct physical interpretation to the diagonal matrix elements of A.

Among the observables are the coordinate functions q_i and p_i on phase space themselves, which classically obeyed the Poisson bracket relations (1.22). We assume that the corresponding operators in quantum mechanics obey

(4.4)
$$[q_i, q_j] = [p_i, p_j] = 0, \qquad [q_j, p_k] = i\delta_{jk},$$

where the bracket is now the commutator, $[A, B] = AB - BA$. The factor i (which means the multiplication operator by the constant i) is necessary because the commutator of two Hermitian operators is not Hermitian but skew Hermitian. As a

[5] In the rigged Hilbert space framework, this means that for any two vectors in Ω, $|\psi> = \int da' f(a')|a'>$ and $|\phi> = \int da\, g(a)|a>$, we should have $<\psi|\phi> = \int da f^*(a)g(a)$.

first approximation, any classical function on phase space is assumed to go over to the same function of the operators p and q upon quantization. This prescription is ambiguous because, for example, $pq^2 = q^2p = qpq$ classically but not quantum-mechanically, but in simple physical examples the correct ordering of operators can be fixed.

To complete the postulates of quantum mechanics we must give the rule for time evolution. Among the observables of a system is the total energy; the corresponding Hermitian operator is the Hamiltonian H. We assume that if a system is initially in the state $|\psi(0)>$, then its state at time t will be

$$|\psi(t)> = e^{-itH}|\psi(0)>, \tag{4.5}$$

so that time evolution is a one parameter group of unitary transformations generated by H. The Schrödinger equation follows immediately in the form

$$i\frac{\partial}{\partial t}|\psi(t)> = H|\psi(t)> . \tag{4.6}$$

[This discussion assumes that H has no explicit dependence on t. If it does, the Schrödinger equation remains correct, but the resulting unitary time evolution operator $U(t)$ is not simply given by $\exp -itH$.] Any state can be expanded in the basis of eigenstates of H, which satisfy $H|E> = E|E>$ and change only by a phase under time evolution: $|E(t)> = \exp -iEt|E(0)>$. Therefore, diagonalizing H solves the time evolution problem for all states. These eigenstates of H are often called stationary states, because all probabilities $|<\psi|E(t)>|^2$ with fixed $|\psi>$ are constant in time.

Let us complete this circle of ideas by deriving the Schrödinger equation (4.2) from this general framework. The Hilbert space for a particle in $\mathbf{R}^3$ is $L^2(\mathbf{R}^3)$, so a state $|\psi>$ is a square-integrable function $\psi(\mathbf{x})$. The fundamental commutation relations for coordinates and momenta can be satisfied by choosing the operators x_i to be multiplication by x_i, and p_j to be $-id/dx_j$. Furthermore, a rough statement of the von Neumann uniqueness theorem is that this is the *only* realization of the canonical commutation relations up to unitary equivalence. The Hamiltonian operator should then be

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x}) = -\frac{1}{2m}\nabla^2 + V(\mathbf{x}), \tag{4.7}$$

and the Schrödinger equation is nothing but the eigenvalue problem for H. Recalling that $|\mathbf{y}> = \delta^3(\mathbf{x}-\mathbf{y})$ is an eigenfunction of the position operators x_i with eigenvalues y_i, we have

$$<\mathbf{y}|\psi> = \int d^3x\, \delta^3(\mathbf{x}-\mathbf{y})\psi(\mathbf{x}) = \psi(\mathbf{y}), \tag{4.8}$$

so that the general postulates of quantum mechanics give the interpretation of $\psi(\mathbf{y})$ as the probability amplitude for finding the particle located near $\mathbf{y}$.

It will be helpful later to be familiar with the eigenfunctions of H for a free particle, the case $V = 0$. Then $H = \mathbf{p}^2/2m$, and the eigenfunctions $e^{i\mathbf{p}\cdot\mathbf{x}}$ of $\mathbf{p} = -id/d\mathbf{x}$ are also the eigenfunctions of H, with eigenvalues $\mathbf{p}^2/2m$. Note that $\mathbf{p}$ here refers to either the operator or its eigenvalues depending upon the context. We see that H has continuous spectrum, and its eigenfunctions do not belong to $L^2(\mathbf{R}^3)$ but must be viewed as elements of Ω^* in the rigged Hilbert space setup. The correctly normalized eigenstates satisfying $<\mathbf{p}|\mathbf{p}'> = \delta^3(\mathbf{p}-\mathbf{p}')$ are $|\mathbf{p}> = (2\pi)^{-3/2}e^{i\mathbf{p}\cdot\mathbf{x}}$.

Symmetries in Quantum Mechanics

When a physical system is invariant under some Lie symmetry group G, we assume that a unitary representation of G acts in the Hilbert space and commutes with the time evolution operator $\exp -iHt$. For each one-parameter subgroup of G we will have an Abelian group of unitary linear operators $U(\theta)$ which act on a state of the system $|\psi>$ to give the state of the system after the symmetry transformation, e.g. after rotation by angle θ about some axis. The unitarity guarantees that scalar products are preserved; since the observable predictions of quantum mechanics are given by various scalar products this ensures that the symmetry transformation does not change the results of experiments. Stone's theorem guarantees that each one-parameter subgroup has a Hermitian generator A such that $U(\theta) = \exp -i\theta A$, and it follows that the observable A commutes with the Hamiltonian, $[A, H] = 0$. Therefore A is conserved, in the sense that any matrix element

$$<\phi(t)|A|\psi(t)> = <\phi(0)|e^{iHt}Ae^{-iHt}|\psi(0)> = <\phi(0)|A|\psi(0)> \tag{4.9}$$

is independent of time. Thus, in quantum mechanics we have a very direct proof that conserved quantities are the generators of symmetries.

Symmetries are of great practical use in diagonalizing H in concrete problems. Because $[A, H] = 0$, the eigenstates of H can be chosen to be simultaneously eigenstates of A, whose spectrum may already be known. In fact, for any element g of the symmetry group G, and any eigenstate $|E>$ of H, it is easy to see that $g|E>$ is another eigenstate having the same eigenvalue. Therefore, each eigenspace of H carries a unitary representation of G! Knowledge of the unitary representations of common symmetry groups such as $SO(3)$ is thus extremely valuable in understanding the spectrum of an invariant Hamiltonian, which is why it is taught in quantum mechanics courses under titles such as "theory of angular momentum".

Heisenberg Picture

Our entire discussion of quantum mechanics so far has been in the so-called Schrödinger picture which leads to the Schrödinger equation as the description of dynamics. In this picture the vectors representing the states of the system change with time, but the operators representing physical observables do not. This situation is reversed in the Heisenberg picture, which is more useful in QFT.

The Heisenberg picture is obtained by a time-dependent unitary map of $\mathcal{H}$ onto itself. Each state $|\psi>$ at time t is mapped to $e^{iHt}|\psi>$, and each operator A is mapped to $e^{iHt}Ae^{-iHt}$. Because the map is unitary, it preserves all scalar products and matrix elements of operators, hence all physically measurable quantities.[6]

[6] Recall that I also spoke of symmetries as preserving all measurable quantities, even though symmetries act unitarily on the states and leave the operators unchanged, hence do not preserve matrix elements. This is actually as it should be: a rotation of a system in space should not preserve the matrix elements of p_x, the momentum component along the x axis, unless the coordinate

Because it inverts the time evolution operator on the states, the state of a system in the Heisenberg picture never changes. Instead, the operator representing a given observable will change with time according to $A(t) = e^{iHt}Ae^{-iHt}$, or, infinitesimally,

$$\frac{dA}{dt} = -i[A, H]. \tag{4.10}$$

This replaces the Schrödinger equation as the description of dynamics. Note that it is identical to the classical Hamilton equation of motion (1.17) with the Poisson bracket replaced by $-i$ times the commutator. This makes it especially clear that the conserved quantities are those which commute with the Hamiltonian. Finally note that the canonical commutation relations $[q_j, p_k] = i\delta_{jk}$ hold in the Heisenberg picture only if the operators involved are evaluated at equal times. The commutator $[q_j(t), p_k(t')]$ is more complicated and depends on the specific form of H.

axes are rotated along with the system. That second rotation, not the first, would produce a new operator having the same matrix elements as the original p_x.

LECTURE 5
The Simple Harmonic Oscillator

The simple harmonic oscillator is the most important exactly solvable system in quantum mechanics. It has been called the only exactly solvable system in the sense that virtually all others can be reduced to it by a change of variables. It is the basis for QFT as well as a first approximation to the description of any oscillating system, from atoms in a solid to gravitational radiation detectors.

At the classical level we have a particle of mass m moving along the x axis and attached to the origin by a spring of force constant k, so that the force acting on the particle is $F = -kx$. Newton's equation of motion $m\ddot{x} + kx = 0$ has the solution $x(t) = A\sin(\omega t + \phi)$, a sinusoidal oscillation with frequency $\omega^2 = k/m$. The Hamiltonian for the system is

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2. \tag{5.1}$$

The Schrödinger equation for this Hamiltonian can be solved directly in terms of special functions (Hermite polynomials), but much more insight results from diagonalizing H using methods of Lie algebra representation theory.

We introduce two new operators in the Hilbert space $L^2(\mathbf{R})$,

$$a = x\sqrt{\frac{m\omega}{2}} + ip\sqrt{\frac{1}{2m\omega}}, \tag{5.2}$$

$$a^\dagger = x\sqrt{\frac{m\omega}{2}} - ip\sqrt{\frac{1}{2m\omega}}. \tag{5.3}$$

They are indeed adjoint operators, and they obey $[a, a^\dagger] = 1$ as a consequence of $[x, p] = i$. Further, the Hamiltonian can be written as

$$H = \frac{1}{2}\omega(a^\dagger a + aa^\dagger) = \omega(a^\dagger a + \frac{1}{2}) \equiv \omega(N + \frac{1}{2}). \tag{5.4}$$

Diagonalizing H is then equivalent to diagonalizing the Hermitian operator N, which is easily found to obey $[N, a^\dagger] = a^\dagger$ and $[N, a] = -a$. Together with $[a, a^\dagger] =$

1, these relations define a Lie algebra called the Heisenberg algebra, and by studying its representations we will determine the spectrum of N.

Let $|n>$ be a normalized eigenstate of N, so that $N|n>=n|n>$. We have

$$Na^\dagger|n> = (a^\dagger N + a^\dagger)|n> = (n+1)a^\dagger|n>, \tag{5.5}$$

$$Na|n> = (aN - a)|n> = (n-1)a|n> . \tag{5.6}$$

This says that when $a^\dagger$ (resp., a) acts on an eigenstate of N, it gives a new eigenstate with eigenvalue raised (resp., lowered) by one unit. In particular it would seem that by repeatedly applying a we could obtain eigenstates of N or H having negative eigenvalues. However, H is a sum of squares of Hermitian operators and as such cannot have negative eigenvalues. The only escape from this paradox is that repeated application of a must eventually terminate with a state obeying $a|\psi>=0$. Because a is in reality a first-order differential operator, this state is unique up to normalization.[7] By applying $a^\dagger$ we see that $N|\psi>=0$. We should therefore denote this state, called the ground state or vacuum state, as $|0>$ (not to be confused with 0, the zero vector in the Hilbert space). By repeated application of $a^\dagger$ to $|0>$ we obtain eigenstates of N with all nonnegative integers as eigenvalues. This is the complete eigenvalue spectrum of N, since if there were a state with nonintegral eigenvalue then repeated application of a would bypass $|0>$ and contradict the positivity of H.

If the state $|n>$ is normalized, we can compute the norm squared of $a^\dagger|n>$ as

$$<n|aa^\dagger|n> = <n|a^\dagger a + 1|n> = n+1. \tag{5.7}$$

It follows inductively that $|n> \equiv (n!)^{-1/2}(a^\dagger)^n|0>$ is a normalized eigenstate of N with eigenvalue n for each positive integer n. The action of the raising and lowering operators on these normalized states is

$$a^\dagger|n> = \sqrt{n+1}|n+1>, \quad a|n> = \sqrt{n}|n-1> . \tag{5.8}$$

The uniqueness of the ground state was shown above. This is a question about the irreducibility of this particular representation of the Heisenberg algebra in $L^2(\mathbf{R})$ which, as we saw, is not determined by the algebra alone but by the specific differential operator representing a. We can now show inductively that in fact each eigenspace of N is one-dimensional. Suppose n is the smallest eigenvalue for which there exists a normalized eigenstate $|n'>$ linearly independent of $|n>$. Then

$$n|n'> = a^\dagger a|n'> = \sqrt{n}a^\dagger|n-1> = n|n>, \tag{5.9}$$

a contradiction. This completes the determination of the spectrum of N as usually given in physics texts. However, since the spectrum of an unbounded operator may not be exhausted by its eigenvalues (equivalently, there may be more eigenfunctions in Ω^*), we should check that the states $|n>$ are already complete in $\mathcal{H}$. This is done in Glimm and Jaffe [**10**], for example.

[7] By solving the differential equation $a|\psi>=0$ we can obtain the corresponding wave function as $\psi(x) = C\exp(-m\omega x^2/2)$.

We thus have a ladder of simple eigenvectors of H, with eigenvalues $(n+\frac{1}{2})\omega$. A measurement of the energy of an oscillator in the state $|n>$ is certain to yield this result. Matrix elements $<m|A|n>$ of any operator $A(x,p)$ of physical interest are easily computed by reexpressing A in terms of a and $a^\dagger$ and using Eqs. (5.8). It is often stated that the oscillator can emit or absorb energy only in integer multiples of ω, by making a "quantum jump" between pairs of states $|n>$. To explain this, suppose that the oscillator interacts with some external system, such as an electric field, for some time interval $0 < t < T$. During this interval, the Hamiltonian of the oscillator is changed by the addition of a potential energy term describing the interaction. Before and after this interval the states $|n>$ are eigenstates of H, but during the interval they are not. Under time evolution, an initial state $|n>$ at $t=0$ may evolve to a linear combination $\sum c_m|m>$ at $t=T$. If the interaction term in H is small, the coefficients c_m can be calculated by using perturbation theory to approximately diagonalize the new Hamiltonian. A measurement of the energy in the final state of the oscillator will yield the result $(m+\frac{1}{2})\omega$ with probability $|c_m|^2$ (and, according to the controversial "reduction of the state vector" postulate, the oscillator will be left in the corresponding state $|m>$, thus completing the jump). The difference between this energy and the initial energy, which is a multiple of ω, is absorbed from or emitted to the external system.

LECTURE 6
The Path Integral Formulation of Quantum Mechanics

The Hilbert space formalism for quantum mechanics corresponds to the Hamiltonian version of classical mechanics in that the Hamiltonian is the infinitesimal generator for the time evolution. We will now derive an expression for quantum mechanical time evolution in terms of the action of the various paths the system might follow classically. This Lagrangian formulation of quantum mechanics leads us to path integrals. Our treatment will be very standard; an instructive alternative approach can be found in [**16**].

We consider for simplicity a particle moving in one dimension, with Hamiltonian $H = p^2/2m + V(x)$. Suppose that at time t the particle is located at x, so that its state is the eigenstate $|x>$ of the operator x. We want to compute the probability amplitude that the particle will be found at position x' at a later time t', denoted $<x', t'|x, t>$. This can be computed as

$$<x', t'|x, t> = <x'|e^{-iH(t'-t)}|x> . \tag{6.1}$$

In the Schrödinger picture the interpretation is that the initial state is $|x>$, we apply the time evolution operator to obtain the state $e^{-iH(t'-t)}|x>$ at time t', and we project this onto the eigenstate $|x'>$. In the Heisenberg picture, the state is initially and always the eigenstate $|x, t> = e^{iHt}|x>$ of the position operator at time t, $x(t) = e^{iHt}x(0)e^{-iHt}$, where $x(0)$ coincides with the standard position operator of the Schrödinger picture, and we take the component of this along the eigenstate $|x', t'> = e^{iHt'}|x'>$ of $x(t')$. In either case, we should not be surprised to encounter singular distributions in calculating this quantity in view of the singular nature of the eigenstates of position. From a mathematical point of view it should be regarded as the integral kernel of the time evolution operator. Namely, given any two states $|\phi>$ and $|\psi>$ we can insert two complete sets of position eigenstates to obtain

$$\begin{aligned} <\phi|e^{-iH(t'-t)}|\psi> &= \int dx\, dx' <\phi|x'><x'|e^{-iH(t'-t)}|x><x|\psi> \\ &= \int dx\, dx' \phi^*(x')K(x, t; x', t')\psi(x). \end{aligned} \tag{6.2}$$

The plan will be to compute $<x',t'|x,t>$ first for infinitesimal $\Delta t = t' - t$, and then to somehow integrate up to finite time intervals. We compute

$$\begin{aligned} <x'|e^{-iH\Delta t}|x> &= <x'|e^{-i\Delta tV(x)}e^{-i\Delta tp^2/2m}[1+O(\Delta t^2)]|x> \\ &= e^{-i\Delta tV(x)} <x'|e^{-i\Delta tp^2/2m}|x> \\ &= e^{-i\Delta tV(x)} \int dpdp' <x'|p'><p'|e^{-i\Delta tp^2/2m}|p><p|x> \\ &= e^{-i\Delta tV(x)} \int \frac{dpdp'}{2\pi} e^{-i\Delta tp^2/2m}\delta(p'-p)e^{ip'x'}e^{-ipx} \\ &= e^{-i\Delta tV(x)} \int \frac{dp}{2\pi} e^{ip(x'-x)}e^{-i\Delta tp^2/2m}, \end{aligned} \tag{6.3}$$

where we have inserted complete sets of eigenstates of p in the third line. (By now the reader should distinguish effortlessly between the operators x and p and their eigenvalues x and p!) The resulting Gaussian integral is only convergent if we assume that Δt has a small negative imaginary part. We will discuss the meaning of this assumption extensively below, but for now we simply look up the integral and obtain

$$<x',t'|x,t> = \left(\frac{m}{2\pi i\Delta t}\right)^{1/2} e^{im(x'-x)^2/2\Delta t}e^{-i\Delta tV(x)}. \tag{6.4}$$

This formula has a remarkable interpretation. Imagine that the particle travels at constant velocity from point x at time t to point x' at time t', so that its path in an xt plane is a straight line segment. The constant velocity must be $(x'-x)/\Delta t$, so that the kinetic energy is $m(x'-x)^2/2\Delta t^2$. When $<x',t'|x,t>$ is used as an integral kernel, the rapidly oscillating phase ensures that the dominant contributions for small Δt occur for x' near x. Then the potential energy is approximately $V(x)$ and the exponent in our formula is just the action for this straight-line path! That is,

(6.5)

$$<x',t'|x,t> = \left(\frac{m}{2\pi i\Delta t}\right)^{1/2} \exp i\int_t^{t'} [\frac{1}{2}m\dot{x}^2 - V(x)]dt = \left(\frac{m}{2\pi i\Delta t}\right)^{1/2} \exp iS,$$

for infinitesimal Δt.

To compute $<x',t'|x,t>$ for a finite time interval, we subdivide the interval by choosing intermediate times $t' > t_1 > t_2 > \cdots > t_n > t$ and insert complete sets of eigenstates of the Heisenberg operators $x(t_i)$ to obtain

(6.6)

$$<x',t'|x,t> = \int dx_1 \cdots dx_n <x',t'|x_1,t_1><x_1,t_1|x_2,t_2> \cdots <x_n,t_n|x,t> .$$

As $n \to \infty$, by our previous computation, each of these matrix elements is proportional to $\exp iS$, where S is the action for a straight-line path joining the appropriate points. Multiplying them together, we obtain $\exp iS$ with S the action for a zigzag path composed of straight line segments joining the intermediate points (x_i,t_i) (Figure 6.1). The integral over the x_i is an integral over the space of all

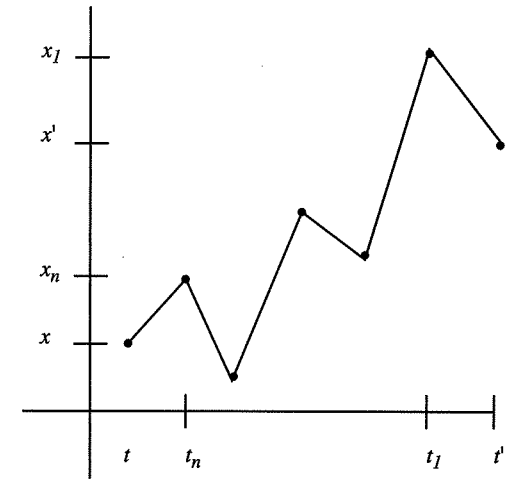

Figure 6.1. Piecewise linear path from x at t to x' at t'. The integration over the locations x_i of the corners approximates the Feynman path integral.

piecewise linear paths with corners at the intermediate times t_i. In the limit this should approach an integral over the space of all continuous paths from (x,t) to (x',t') with each path weighted according to its action, formally,

$$<x',t'|x,t> = N\int [Dx] \exp i\int_t^{t'} [\frac{1}{2}m\dot{x}^2 - V(x)]dt, \tag{6.7}$$

with $[Dx]$ a formal measure and N a formal normalization constant. This path integral is the basic formula of this approach to quantum mechanics. Clearly our calculations have been heuristic, and much remains to be done to rigorously establish such a formula. One must prove that the limit $n \to \infty$ exists, and equals the matrix element we wanted to compute. This will involve a careful estimate of the errors made in our computation for "infinitesimal" Δt. Then one should prove that the limit of the measures on the spaces of piecewise linear paths exists and gives a measure on continuous paths. In fact, no one has ever succeeded in doing this, and the reason is closely related to our need to assume a small imaginary part for Δt.

The time evolution operator $\exp -iHt$ whose matrix elements we have been computing makes sense for certain complex, as well as real, values of t. Recall that the spectrum of H in general is bounded below but not above. Therefore $\exp -iHz$ is a very nice operator, analytic in z, for z in the lower half of the complex z plane. However, in the upper half plane it gives exponentially increasing weight to the higher eigenvalues of H and is too unbounded to possess an integral kernel $<x',t'|x,t>$. The real axis, containing the values of physical interest, is the boundary between these regions, and as such formal computations here can be dangerous. (By the spectral theorem, $\exp -iHz$ is strongly continuous for Im $z \leq 0$ and we are interested in the continuous boundary values of an analytic function.) The small negative imaginary part assumed for Δt above simply amounts to computing the kernel by analytic continuation from slightly below the real axis. This turns out to be such a good idea that we will carry it to the extreme and compute the kernel by

analytic continuation from the negative imaginary axis. We parametrize the imaginary axis by $z = -i\tau$ with τ real, and compute the kernel of $\exp -iHz = \exp -H\tau$. Retracing our derivation of the path integral, or just formally substituting $t = -i\tau$ in the result, gives the Feynman-Kac formula

(6.8)

$$<x',\tau'|x,\tau> \equiv <x'|e^{-H(\tau'-\tau)}|x> = N\int [Dx] \exp -\int_{\tau}^{\tau'} d\tau[\frac{1}{2}m\dot{x}^2 + V(x)].$$

The integral is over all paths from (x,τ) to (x',τ'). This formula can indeed be rigorously proven by the means sketched previously. The important estimates are contained in the Trotter product formula,

(6.9)

$$e^{-(A+B)} = \lim_{n\to\infty} (e^{-A/n}e^{-B/n})^n$$

for suitable operators A and B. The resulting measure on the space of continuous paths in the case $V(x) = 0$ is (conditional) Wiener measure. Neither the constant N, the formal Lebesgue measure $[Dx]$, nor the exponential of the kinetic part of the action (because Wiener measure is concentrated on nondifferentiable paths) make sense separately, but only their product which is Wiener measure. The $\pi/2$ rotation of our attention from the real axis to the imaginary is called "Wick rotation to Euclidean space", named for the physicist G.C. Wick. We will see shortly how to view it as a literal rotation. Although at the moment we are not concerned with Lorentz invariance, in QFT we will be, and the term "Euclidean space" refers to the effect of the rotation in changing the indefinite Minkowski metric $t^2 - x^2 - y^2 - z^2$ to the definite Euclidean metric $-\tau^2 - x^2 - y^2 - z^2$. The quantity $\frac{1}{2}m\dot{x}^2 + V(x)$ is called the Euclidean Lagrangian (and sometimes coincides with the Hamiltonian). Our earlier "Minkowski" path integral formula is viewed as a mnemonic for the Feynman-Kac formula plus analytic continuation in $\Delta z = -i(\tau' - \tau)$ back to the real axis.

The path integral would not be the versatile tool that it is if it could only compute the kernel of the time evolution operator. It can also be used to compute matrix elements of the position operators in the Heisenberg picture. Consider the computation of

(6.10)

$$<x',t'|x(t_1)\cdots x(t_n)|x,t>,$$

where $t' > t_1 > \cdots > t_n > t$. Inserting complete sets of eigenstates of $x(t_i)$ at the intermediate times allows us to rewrite this as

(6.11)

$$\int dx_1 \cdots dx_n \, x_1 \cdots x_n <x',t'|x_1,t_1> \cdots <x_n,t_n|x,t>,$$

and the path integral representation for the remaining matrix elements gives

(6.12)

$$<x',t'|x(t_1)\cdots x(t_n)|x,t> = N\int [Dx] x(t_1)\cdots x(t_n) \exp iS,$$

where $x(t_i)$ in the integral denotes the location of the path $x(t)$ at time t_i. Note that the factors $x(t_i)$ on the right side of this equation are just numbers (actually functionals of the paths) and can be rearranged freely, while those on the left are noncommuting operators and must be arranged in decreasing order of their time arguments. That is, the path integral on the right automatically computes the matrix element of a product of $x(t)$ operators in a particular order. We define the time ordering symbol T so that $T[x(t_1)\cdots x(t_n)]$ is the product of the operators in decreasing time order, regardless of their order as written. Then we have the path integral formula,

(6.13)

$$<x',t'|T[x(t_1)\cdots x(t_n)]|x,t> = N\int [Dx] x(t_1)\cdots x(t_n) \exp iS,$$

with no restriction on the order of the intermediate times t_i.

Even more cleverly, we can write a generating functional for all such matrix elements at once. We add a so-called source term $x(t)J(t)$ to the Lagrangian, where $J(t)$ is a given function with support in the interval $[t,t']$. The path integral with this new Lagrangian is a functional of $J(t)$, and we can take its functional derivatives. We obtain

(6.14)

$$<x',t'|T[x(t_1)\cdots x(t_n)]|x,t> = (-i)^n \frac{\delta^n}{\delta J(t_1)\cdots\delta J(t_n)} N\int [Dx] \exp i\int_t^{t'} dt[\frac{1}{2}m\dot{x}^2 - V(x) + xJ]\Big|_{J=0},$$

because each functional derivative on the right brings down a factor $x(t_i)$ into the integrand, reproducing our previous formula.

This may strike you as the ultimate in cleverness, but there is one more step to take. Even more useful than matrix elements between states like $|x,t>$ are matrix elements between the eigenstates of the Hamiltonian, especially matrix elements in the ground state $|0>$. The following trick allows us to compute these as well by means of path integrals. Let the source $J(t)$ now have support in a subinterval $[a,b]$ of $[t,t']$. Adding the source term to the Lagrangian will change the Hamiltonian as well, into, say, H^J. The corresponding unitary time evolution operator $U^J(t_1,t_2)$ between times t_1 and t_2 will not coincide with $\exp -iH(t_2 - t_1)$ when the source is nonzero. Its matrix elements are

(6.15)

$$\begin{aligned} <x',t'|x,t>^J &= <x'|U^J(t,t')|x> \\ &= <x'|e^{-iH(t'-b)}U^J(a,b)e^{-iH(a-t)}|x> . \end{aligned}$$

We now insert complete sets of eigenstates $|n>$ of H, with eigenvalues E_n (which are *not* generally eigenstates of H^J) to obtain

(6.16)

$$\begin{aligned} <x',t'|x,t>^J &= \sum_{m,n} <x'|e^{-iH(t'-b)}|m><m|U^J(a,b)|n><n|e^{-iH(a-t)}|x> \\ &= \sum_{m,n} <x'|m><n|x> e^{-iE_m(t'-b)}e^{-iE_n(a-t)} <m|U^J(a,b)|n> . \end{aligned}$$

Now we want to take the limits $t \to -\infty$, $t' \to +\infty$. This makes sense when we recall that $<x',t'|x,t>^J$ is the analytic continuation of the Euclidean matrix element $<x',\tau'|x,\tau>$, which has a representation as above with $e^{-iE_m\cdots}$ replaced by $e^{-E_m\cdots}$, so that we can let $\tau \to -\infty$, $\tau' \to +\infty$. Equivalently, we let $t \to -\infty$, $t' \to +\infty$ in the complex plane, along a line with a small negative slope rather than the real axis. We also assume, as is usual in QFT for just this reason, that a constant has been added to H to make the ground state energy equal to zero. Then in the limit only the ground state contributes to $<x',t'|x,t>^J$, which becomes

$$<x',t'|x,t>^J = <x'|0><0|x>\, Z(J), \tag{6.17}$$

where, by definition, the vacuum functional or partition function $Z(J)$ is

$$Z(J) = <0|U^J(a,b)|0> = <0|U^J(t,t')|0>\,. \tag{6.18}$$

We have the normalization condition $Z(0) = 1$, but in general $Z(J) \neq 1$ because $|0>$ is not an eigenstate of H^J when $J \neq 0$. Rewriting (6.17) as

$$Z(J) = \frac{<x',t'|x,t>^J}{<x'|0><0|x>}, \tag{6.19}$$

we obtain the path integral formula for $Z(J)$,

$$Z(J) = N\int [Dx] \exp iS^J, \tag{6.20}$$

with N chosen so that $Z(0) = 1$. The integral is over all paths having fixed but arbitrary endpoints x', x at the times $\pm\infty$. Furthermore, repeating these arguments for the matrix elements (6.14) gives

$$<0|T[x(t_1)\cdots x(t_n)]|0> = (-i)^n \frac{\delta^n}{\delta J(t_1)\cdots\delta J(t_n)} Z(J)\bigg|_{J=0}. \tag{6.21}$$

$Z(J)$ is thus the generating functional for a whole series of ground state matrix elements which turn out to have direct physical meaning in QFT. They are called Green's functions, correlation functions, or n-point functions, and in a real sense they contain all the physical predictions of the theory.

Before we learn how to actually compute path integrals, let me add two comments about their importance in physics. First, the path integral formalism is the most elegant way to understand the connection between quantum and classical mechanics. With physical units restored, the path integral weight $\exp iS$ is actually $\exp iS/\hbar$. In the limit $\hbar \to 0$ it is reasonable to try to compute the integral by the method of stationary phase. The integral over paths should be dominated by the neighborhood of the path which extremizes S, which is precisely our Lagrangian formulation of classical mechanics. The asymptotic expansion of the path integral for small $\hbar$ is known as the semiclassical or WKB (Wentzel, Kramers, and Brillouin)

approximation in physics. Second, the Euclidean path integral (6.8) has an interpretation in terms of statistical mechanics which leads to the modern convergence between the subjects of statistical mechanics and QFT. In a small nutshell, statistical mechanics is a probability theory in which a physical system at temperature T has a probability proportional to $\exp -E/kT$ of being found in a state having energy E, with k being Boltzmann's constant. The Euclidean path integrals for a particle can be reinterpreted as moments of this probability measure for some continuous one-dimensional system, e.g. $x(\tau)$ is the pressure or density in a one-dimensional gas at location τ and the Euclidean action is the energy of the gas. Note that τ has become a spatial coordinate while x is the value of a field at that location. The term "partition function" itself comes from statistical mechanics, where $Z(J)$ is related to the number of ways of partitioning the system's total energy among its constituent particles.

LECTURE 7
The Harmonic Oscillator via Path Integrals

We will now see how path integrals are actually computed, by calculating $Z(J)$ for the simple harmonic oscillator. Just as the oscillator is the only exactly solvable problem in physics, so this type of "Gaussian" path integral is the only one that can honestly be evaluated. As such our computation is fundamental for all applications of path integrals in QFT.

We want to compute

$$Z(J) = N \int [Dx] \exp i \int dt\,(L + xJ), \qquad Z(0) = 1, \tag{7.1}$$

where

$$L = -\frac{1}{2} mx(\frac{d^2}{dt^2} + \omega^2)x. \tag{7.2}$$

This form of the Lagrangian differs from that of Eq. (1.8) by an integration by parts, like the forms in Eqs. (2.4,2.5). It exposes the central role of the self-adjoint operator $\frac{d^2}{dt^2} + \omega^2$. To see how to proceed, compare the finite-dimensional integral over column vectors x in $\mathbf{R}^n$

$$\int d^n x \exp(-x^T A x + J^T x), \tag{7.3}$$

with J a fixed vector and A a nonsingular symmetric matrix. (In fact, the finite-dimensional integrals over piecewise-linear paths which approximate the path integral have this form.) We complete the square

$$-x^T A x + J^T x = -(x - \frac{1}{2} A^{-1} J)^T A (x - \frac{1}{2} A^{-1} J) + \frac{1}{4} J^T A^{-1} J$$

and use the translation-invariance of Lebesgue measure to substitute $y = x - \frac{1}{2}A^{-1}J$, obtaining

$$e^{J^T A^{-1} J/4} \int d^n y\, e^{-y^T A y}. \tag{7.4}$$

We can normalize the result to unity when $J = 0$ by simply dividing by the integral, so that the result is obtained from the translation-invariance of the measure without ever evaluating an integral.

Since the formal Lebesgue measure $[Dx]$ is also translation-invariant,[8] we can apply the same procedure to $Z(J)$ as soon as we know "the" inverse for the operator $A = \frac{d^2}{dt^2} + \omega^2$. That is, we need to solve the differential equation

$$(\frac{d^2}{dt^2} + \omega^2) f(t) = g(t). \tag{7.5}$$

As we have previously mentioned, physicists like to express linear operators in terms of their integral kernels, so that the solution is written in terms of a Green's function $G(t, s)$ as

$$f(t) = \int ds\, G(t, s) g(s), \tag{7.6}$$

and the statement that this inverts A becomes

$$(\frac{d^2}{dt^2} + \omega^2) G(t, s) = \delta(t - s). \tag{7.7}$$

Not all solutions of this equation are of the form $G(t, s) = G(t - s)$, because of the freedom to add arbitrary solutions of the homogeneous equation, but I assert with the benefit of hindsight that the one we are seeking is. We compute it via its Fourier transform,

$$\hat{G}(k) = \int dt\, G(t) e^{-ikt}, \tag{7.8}$$

$$G(t) = \int \frac{dk}{2\pi} \hat{G}(k) e^{ikt}, \tag{7.9}$$

which satisfies

$$(-k^2 + \omega^2)\hat{G}(k) = 1. \tag{7.10}$$

The solution to this equation in the sense of distributions is a bit tricky,

$$\hat{G}(k) = P\frac{-1}{k^2 - \omega^2} + h(k)\delta(k^2 - \omega^2). \tag{7.11}$$

[8] That is, Wiener measure has the properties that would be expected if it were.

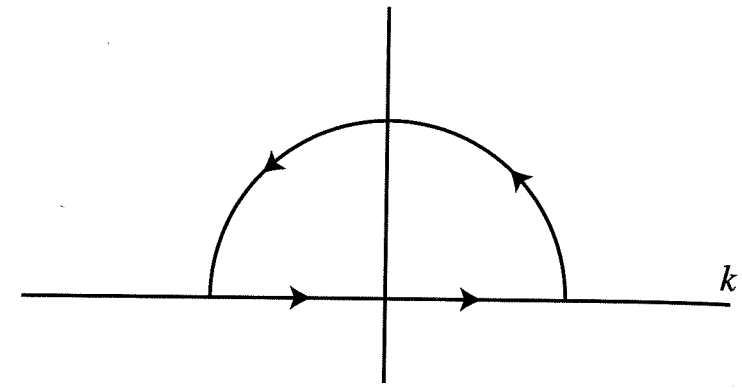

Figure 7.1. Integration contour in the k plane for computing $G(t)$ from its Fourier transform, $t > 0$.

The point is that the operator A has a kernel ("zero modes" to a physicist), so is not uniquely invertible until boundary conditions are specified for Eq. (7.5). $G(t)$ can be changed by addition of functions in the kernel, which are combinations of $e^{\pm i\omega t}$. Its Fourier transform is thus only determined up to distributions concentrated at $k = \pm\omega$. Yet another way to say this is to note that the inverse Fourier transform

$$G(t) = \int \frac{dk}{2\pi} \frac{-1}{k^2 - \omega^2} e^{ikt} \tag{7.12}$$

is ambiguous because the integration contour (the real axis) passes through two poles at $k = \pm\omega$. A prescription for handling these poles amounts to a choice of boundary conditions and selects one of the possible Green's functions.

The appropriate boundary conditions in quantum mechanics are specified implicitly by the condition that the correlation functions obtained from $Z(J)$ should be related by analytic continuation to those given by a Euclidean path integral. The Euclidean path integral for the harmonic oscillator is calculated by the same technique of completing the square, but the relevant self-adjoint operator becomes

$$-\frac{d^2}{d\tau^2} + \omega^2. \tag{7.13}$$

The corresponding Euclidean Green's function is, unambiguously,

$$G_E(k) = \frac{1}{k^2 + \omega^2}, \tag{7.14}$$

because this function has no poles on the real axis.[9] The Minkowski space Green's function we seek is the one related by analytic continuation to the Fourier transform of $G_E(k)$. To determine it, return to Eq. (7.12) and close the contour with a semicircle at infinity, in the upper half plane if $t > 0$ (Figure 7.1). This function

[9] One may wonder why the Euclidean Green's function is unambiguously determined when the operator (7.13) still has a kernel, spanned by the functions $\exp \pm\omega t$. The point is that these functions are not in L^2 or even in the set of distributions Ω^*.

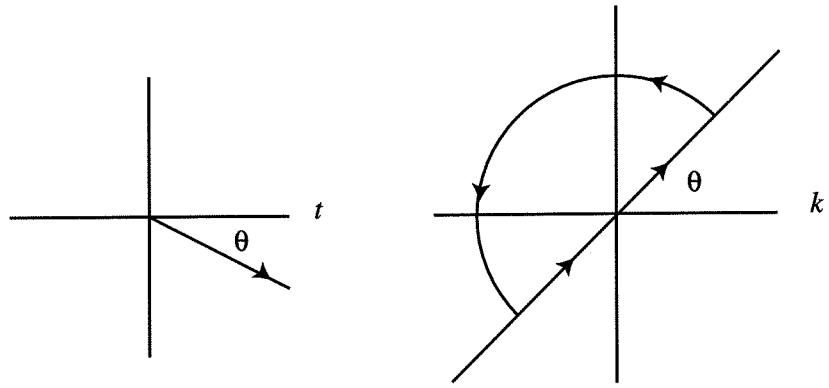

Figure 7.2. Wick rotation of the contour for computing $G(t)$, with t on the indicated line in the fourth quadrant.

$G(t)$ is analytic in some region of the complex t plane. For t in the fourth quadrant, say on the line indicated in Figure 7.2, the semicircle does contribute to the integral, but we can fix that by rotating the contour in the k plane, providing it does not cross any poles of the integrand. After a rotation by $\pi/2$ (the Wick rotation), t is on the negative imaginary axis and $G(t)$ becomes the Euclidean Green's function. Therefore, $G(t)$ is the analytic continuation of $G_E(\tau)$ provided only that the integration contour did not cross poles. This specifies the correct treatment of the poles on the real axis: the contour must pass below the one at $k = -\omega$ and above the one at $k = +\omega$, or equivalently the poles must be moved infinitesimally off the axis to $k = \pm\sqrt{\omega^2 - i\epsilon}$. The resulting Green's function,

$$G(t) = \int \frac{dk}{2\pi} \frac{-1}{k^2 - \omega^2 + i\epsilon} e^{ikt}, \tag{7.15}$$

often called the Feynman Green's function or Feynman propagator, is easily evaluated by residues as

$$G(t) = \frac{i}{2\omega}[\theta(t)e^{-i\omega t} + \theta(-t)e^{i\omega t}], \tag{7.16}$$

where $\theta(t) = 0$ for $t < 0$, $\theta(t) = 1$ for $t > 0$ is the Heaviside step function rather than the Riemann theta function. Of course, we could easily have determined the correct $G(t)$ by explicitly evaluating $G_E(\tau)$ and analytically continuing, but the method used works in QFT also, and motivates the term "Wick rotation" as well. Comparing the explicit form of $G(t)$ with the time evolution operator e^{-iHt}, we can see why the Feynman propagator is often described as propagating positive energy states (ω) into the future ($t > 0$) and negative energy states ($-\omega$) into the past ($t < 0$). In QFT this is related to the behavior of antiparticles and can be used to give a physical derivation of the correct Green's function.

Armed with the correct Green's function, we easily complete the square and finish the computation of $Z(J)$:

$$Z(J) = \exp\frac{i}{2m}\int dt\, ds\, J(t)G(t-s)J(s). \tag{7.17}$$

By taking two functional derivatives we obtain the two-point correlation function

$$\begin{aligned} <0|T[x(t)x(t')]|0> &= (-i)^2\frac{\delta^2}{\delta J(t)\delta J(t')}Z(J)|_{J=0} \\ &= -\frac{i}{m}\frac{\delta}{\delta J(t')}[\int ds J(t)G(t-s)]Z(J)|_{J=0} \\ &= -\frac{i}{m}G(t-t') \\ &= \frac{1}{2m\omega}[\theta(t-t')e^{-i\omega(t-t')} + \theta(t'-t)e^{i\omega(t-t')}]. \end{aligned} \tag{7.18}$$

Note that the second functional derivative has to act on the $J(t)$ brought down by the first one in order to get a nonzero result upon setting $J = 0$. Up to a factor, the two-point function coincides with the Green's function. This is why correlation functions in general are called Green's functions.

There is a simple rule for computing the n-point correlation functions, called Wick's theorem, which is really a general fact about the moments of a Gaussian measure. When we compute the functional derivative

$$\frac{\delta^n}{\delta J(t_1)\cdots\delta J(t_n)}Z(J)|_{J=0}, \tag{7.19}$$

the first derivative brings down a $J(t_1)$, among other things, from the exponential in $Z(J)$. Unless some other derivative acts on this $J(t_1)$, the result vanishes for $J = 0$. The nonvanishing terms arise from all possible ways of pairing the functional derivatives so that one acts on the $J(t)$ brought down by the other. The n-point function is a sum over all pairings of the product of factors $-\frac{i}{m}G(t_i - t_j)$ for each pairing. For example, the four-point function is

$$\begin{aligned} <0|T[x(t_1)x(t_2)x(t_3)x(t_4)]|0> = -\frac{1}{m^2}[&G(t_1-t_2)G(t_3-t_4) \\ &+G(t_1-t_3)G(t_2-t_4) + G(t_1-t_4)G(t_2-t_3)]. \end{aligned} \tag{7.20}$$

The n-point functions for n odd vanish since a complete pairing is not possible.

LECTURE 8
Quantum Field Theory: Free Fields

Quantum field theory, to a first approximation, is the relativistic quantum mechanics of systems having infinitely many degrees of freedom, such as fields. We want to construct a quantum mechanics of the electromagnetic field, or the Klein-Gordon field, just as we constructed the quantum mechanics of a particle moving in a potential. We will in fact discuss the Klein-Gordon equation so as to avoid the complications associated with nontrivial representations of the Lorentz group and with gauge invariance.

We recall the action

$$S = \frac{1}{2}\int d^4x(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2), \tag{8.1}$$

which leads to the Klein-Gordon equation $(\partial_\mu\partial^\mu + m^2)\phi = 0$, and the corresponding Hamiltonian

$$H = \frac{1}{2}\int d^3x(\pi^2 + |\nabla\phi|^2 + m^2\phi^2), \tag{8.2}$$

where $\pi(x) = \partial^0\phi(x)$. At a fixed time, the infinitely many coordinate and momentum variables $\phi(\mathbf{x},t), \pi(\mathbf{x},t)$ should be reinterpreted as Hermitian operators acting in a Hilbert space $\mathcal{H}$ and satisfying commutation relations modeled on their classical Poisson brackets,

$$[\phi(\mathbf{x},t), \pi(\mathbf{y},t)] = i\delta^3(\mathbf{x}-\mathbf{y}), \tag{8.3}$$

$$[\phi(\mathbf{x},t), \phi(\mathbf{y},t)] = [\pi(\mathbf{x},t), \pi(\mathbf{y},t)] = 0. \tag{8.4}$$

We are adopting the Heisenberg picture in which the operators are time-dependent and the canonical commutators hold only at equal times. Of course, we have the option of going to the Schrödinger picture, transferring the time dependence to the states. However, this obscures the Lorentz invariance of the theory since the Lorentz group mixes the time and space components of x, and therefore the

Heisenberg picture is generally more convenient in QFT. One can check that the Heisenberg equations of motion

$$\dot{\pi}(x) = -i[\pi(x), H], \qquad \dot{\phi}(x) = -i[\phi(x), H] \tag{8.5}$$

are equivalent to the Klein-Gordon equation.

Unfortunately, the singular δ distribution in the canonical commutators shows that $\phi(x)$ and $\pi(x)$ cannot be simply operator-valued functions on $\mathbf{R}^4$, since then their commutator would also be such a function. We encountered this problem in the Hamiltonian formulation of classical field theory, and resolved it by choosing better coordinate and momentum functions on phase space. Similarly here we postulate that for each test function $f(\mathbf{x})$ on $\mathbf{R}^3$ there are time-dependent operators $\phi[f,t], \pi[f,t]$ formally given by

$$\phi[f,t] = \int d^3x f(\mathbf{x})\phi(\mathbf{x},t), \qquad \pi[f,t] = \int d^3x f(\mathbf{x})\pi(\mathbf{x},t). \tag{8.6}$$

Then $\phi(x), \pi(x)$ should be viewed as operator-valued distributions, and the canonical commutators actually mean, for example,

$$[\phi[f,t], \pi[g,t]] = i\int d^3x\, f(\mathbf{x})g(\mathbf{x}). \tag{8.7}$$

It happens that this interpretation of the fields is correct in noninteracting, or free, field theories such as the Klein-Gordon theory, but that spatial smearing alone is not sufficient to produce well-defined operators in interacting field theories. In those cases we assume the existence of fields $\phi[f] = \int d^4x f(x)\phi(x)$, and similarly for π, for each test function f on $\mathbf{R}^4$ rather than $\mathbf{R}^3$. In general one cannot make sense of equal time commutators in these theories.

We will now solve the Klein-Gordon, or free Hermitian scalar field, theory. This means that we will construct a Hilbert space $\mathcal{H}$ containing operator-valued distributions $\phi(x), \pi(x)$ satisfying the Klein-Gordon equation and the canonical commutators. Rather than producing these objects from thin air, I will give the heuristic derivation which is standard in the physics literature, and only then connect it with the precise mathematical definitions. The mathematician should try to appreciate how some familiar mathematical objects are described in the language and notation of physics.

If we view the Klein-Gordon equation as simply a PDE of a standard type for a real-valued function $\phi(x)$ on $\mathbf{R}^4$, the solution can be obtained easily by means of a Fourier transform:

$$\phi(x) = \int (dk)[a(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-E_\mathbf{k}t)} + a^\dagger(\mathbf{k})e^{-i(\mathbf{k}\cdot\mathbf{x}-E_\mathbf{k}t)}], \tag{8.8}$$

$$\pi(x) = \int (dk)(-iE_\mathbf{k})[a(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-E_\mathbf{k}t)} - a^\dagger(\mathbf{k})e^{-i(\mathbf{k}\cdot\mathbf{x}-E_\mathbf{k}t)}]. \tag{8.9}$$

Several comments are necessary concerning the notation here. The suggestively-named functions $a(\mathbf{k}), a^\dagger(\mathbf{k})$ on $\mathbf{R}^3$ are complex conjugates of each other, so that $\phi(x)$ is real. The Klein-Gordon equation is satisfied provided only that $E_\mathbf{k} = \sqrt{|\mathbf{k}|^2 + m^2}$; we define a momentum 4-vector $k^\mu = (E_\mathbf{k}, \mathbf{k})$ which lies on the hyperboloid $k^2 = m^2$ in $\mathbf{R}^4$ (often referred to as the "mass shell"), so that $k \cdot x = k^\mu x_\mu = E_\mathbf{k}t - \mathbf{k}\cdot\mathbf{x}$. Finally, the Fourier integrals over $\mathbf{R}^3$ use a Lorentz-invariant measure for convenience,

$$\int (dk) = \int \frac{d^3k}{(2\pi)^3 2E_\mathbf{k}} = \int \frac{d^4k}{(2\pi)^4} 2\pi\delta(k^2 - m^2)\theta(k^0), \tag{8.10}$$

where the last form makes the Lorentz invariance obvious.

In QFT we use the above solution to motivate a change of variables: at some arbitrary but fixed time $t = x^0$ we replace the variables $\phi(x), \pi(x)$ by the $a(\mathbf{k}), a^\dagger(\mathbf{k})$ introduced above. These are now adjoint operators for each $\mathbf{k}$ (actually distributions again), given explicitly by

$$a^\dagger(\mathbf{k}) = \int d^3x\, e^{-ik\cdot x}[E_\mathbf{k}\phi(x) - i\pi(x)], \tag{8.11}$$

$$a(\mathbf{k}) = \int d^3x\, e^{ik\cdot x}[E_\mathbf{k}\phi(x) + i\pi(x)]. \tag{8.12}$$

Any resemblance to the change of variables we made in solving the simple harmonic oscillator is totally intentional. These new creation and annihilation operators are found to satisfy

$$[a(\mathbf{k}), a^\dagger(\mathbf{p})] = (2\pi)^3 2E_\mathbf{k}\delta^3(\mathbf{k} - \mathbf{p}), \tag{8.13}$$

$$[a(\mathbf{k}), a(\mathbf{p})] = [a^\dagger(\mathbf{k}), a^\dagger(\mathbf{p})] = 0. \tag{8.14}$$

In terms of these operators the Hamiltonian reads

$$H = \frac{1}{2}\int (dk)E_\mathbf{k}[a^\dagger(\mathbf{k})a(\mathbf{k}) + a(\mathbf{k})a^\dagger(\mathbf{k})] \tag{8.15}$$

$$= \int (dk)[E_\mathbf{k}a^\dagger(\mathbf{k})a(\mathbf{k}) + 8\pi^3 E_\mathbf{k}^2\delta^3(\mathbf{0})], \tag{8.16}$$

where the second line results from the formal application of the commutator (8.13).

In the meaningless expression $\int (dk)8\pi^3 E_\mathbf{k}^2\delta^3(\mathbf{0})$ we have encountered the first of the famous infinities of QFT, and the easiest to deal with. The Hamiltonian H should be viewed as describing uncountably many simple harmonic oscillators labeled by the vectors $\mathbf{k}$, and we expect that there will be a vacuum state $|0>$ satisfying $a(\mathbf{k})|0> = 0$ for all $\mathbf{k}$. The constant term involving $\delta^3(\mathbf{0})$ is infinite because it is the sum of the ground state energies $\omega/2$ of infinitely many oscillators [compare Eq. (5.4)]. We propose to subtract an infinite constant from H so that the energy of the vacuum state will be zero instead, so that we use simply

$$H = \int (dk)E_\mathbf{k}a^\dagger(\mathbf{k})a(\mathbf{k}). \tag{8.17}$$

Two common justifications for this procedure are the following.

(1) Even in classical mechanics $V(\mathbf{x})$ and hence H are arbitrary by an additive constant; only differences in energy are measurable and any state can be defined to have zero energy.

(2) The only function of a Hamiltonian is to generate the correct Heisenberg equations of motion. Since one can check that both forms (8.15) and (8.17) of H have the same commutators with $a(\mathbf{k})$ and $a^\dagger(\mathbf{k})$, they are equally acceptable.

In fact, our original expression (8.2) for H was only formal. It contained expressions such as $\pi(x)^2$ which, as products of distributions, are not generally meaningful. We are now defining what those products meant, with the interpretation of H in terms of oscillators as a guide, and by definition they do not include the infinite constant. A convenient way to formulate the definition is in terms of normal-ordered products. We say that a polynomial in creation and annihilation operators $P(a^\dagger, a)$ has been normal ordered, denoted $:P(a^\dagger, a):$, when each term has been reordered so that all $a^\dagger$'s stand to the left of all a's. This reordering is done by fiat, as if all operators commuted, and not by using the canonical commutation relations to interchange them. For example,

$$\frac{1}{2}:a^\dagger(\mathbf{k})a(\mathbf{k}) + a(\mathbf{k})a^\dagger(\mathbf{k}):= a^\dagger(\mathbf{k})a(\mathbf{k}), \tag{8.18}$$

so that our modification of the Hamiltonian simply amounts to saying that all products in it should be interpreted as normal ordered. It is often the case that normal ordered products of distributions make sense when general products do not. It is a useful fact that the vacuum expectation value of a normal ordered product is always zero, which in our case means that the energy $<0|H|0>$ of the vacuum vanishes, as we wished.

The Hamiltonian H commutes with several other operators, which are therefore conserved quantities. As examples we have the number operator,

$$N = \int (dk) a^\dagger(\mathbf{k})a(\mathbf{k}), \tag{8.19}$$

and the momentum operator,

$$\mathbf{P} = \int (dk)\mathbf{k}a^\dagger(\mathbf{k})a(\mathbf{k}), \tag{8.20}$$

which, like H, annihilate the vacuum. We can combine H and $\mathbf{P}$ into a 4-momentum operator

$$P^\mu = (H, \mathbf{P}) = \int (dk) k^\mu a^\dagger(\mathbf{k})a(\mathbf{k}), \tag{8.21}$$

which is indeed the generator of the translation group in $\mathbf{R}^4$,

$$[\phi(x), P^\mu] = i\partial^\mu \phi(x). \tag{8.22}$$

As in the harmonic oscillator, we expect that the eigenstates of H can be obtained by acting on the vacuum $|0>$ with finite products of the creation operators $a^\dagger(\mathbf{k})$

for various $\mathbf{k}$. By using the commutation relations as we did then, we can verify that acting with $a^\dagger(\mathbf{k})$ on a simultaneous eigenstate of N and P^μ gives a new eigenstate with eigenvalues increased by unity for N and by k^μ for P^μ. Recall that $k^2 = m^2$, or $E_{\mathbf{k}}^2 = |\mathbf{k}|^2 + m^2$. This is nothing but the relation between the energy and momentum of a free particle of mass m in relativity. Although we did not have time to derive it earlier, we can make it plausible now. With physical units restored it reads $E_{\mathbf{k}} = \sqrt{|\mathbf{k}|^2c^2 + m^2c^4}$, and in the nonrelativistic limit $c \to \infty$ this becomes $E_{\mathbf{k}} = mc^2 + |\mathbf{k}|^2/2m$, showing the energy mc^2 corresponding to the (rest) mass as well as the Newtonian kinetic energy $|\mathbf{k}|^2/2m$.

We have been lead to the following remarkable conclusion. The eigenstates of H are of the form $a^\dagger(\mathbf{k}_1)\cdots a^\dagger(\mathbf{k}_n)|0>$. The total energy and momentum of this state are exactly those of a collection of n free particles of momenta $\mathbf{k}_i$, the operator N counts the particles, and the operators $a^\dagger(\mathbf{k}_i)$ and $a(\mathbf{k}_i)$ create and annihilate them. These particles are the quanta of the field, and this is the very deep connection between fields and particles in QFT: the stationary states of a free field theory are multiparticle states. The quantization of the electromagnetic field is technically harder but conceptually the same, and the resulting particles are photons. The Klein-Gordon equation describes spinless particles such as pi mesons. (The spin refers essentially to the highest weight of the representation of the Lorentz group by which the field transforms; it is zero for the trivial representation.)

Let us now translate the physicist's description of the Hilbert space into rigorous mathematics. The only thing wrong with the description above, really, is that the creation and annihilation "operators" are still operator-valued distributions and must be smeared in $\mathbf{k}$ to get actual operators. Physically, they create and annihilate free particles of definite momenta, and we recall that even in quantum mechanics the free particle states lie outside the Hilbert space $L^2(\mathbf{R}^3)$, in a rigged extension. So, rigorously: the Hilbert space $\mathcal{H}$ will be the space of all sequences of complex functions of the form

$$F = \{F_0, F_1(\mathbf{k}), F_2(\mathbf{k}_1, \mathbf{k}_2), \dots\}, \tag{8.23}$$

where F_n is a symmetric function of n vectors from $\mathbf{R}^3$, such that the norm to be defined below is finite. The sequence $\{1, 0, 0, \dots\}$ is the vacuum state previously denoted $|0>$. The sequence $\{0, 0, \dots, F_n(\mathbf{k}_1, \dots, \mathbf{k}_n), 0, \dots\}$ is a smeared n-particle state which we would have previously written as

$$\frac{1}{n!}\int (dk_1)\cdots(dk_n)F_n(\mathbf{k}_1, \dots, \mathbf{k}_n)a^\dagger(\mathbf{k}_1)\cdots a^\dagger(\mathbf{k}_n)|0>, \tag{8.24}$$

where any antisymmetric part of F_n would not contribute to this equation because the $a^\dagger$s all commute. The scalar product is

$$<F, G> = F_0^* G_0 + \sum_{n=1}^{\infty} \frac{1}{n!}\int (dk_1)\cdots(dk_n)F_n^*(\mathbf{k}_1, \cdots \mathbf{k}_n)G_n(\mathbf{k}_1, \cdots \mathbf{k}_n). \tag{8.25}$$

The nuclear subspace Ω of the rigged Hilbert space can be taken as the sequences with only finitely many nonzero entries, and each F_n decreasing very fast as any

$|\mathbf{k}_i| \to \infty$. For each test function $f(\mathbf{k})$ there is an operator

$$a^\dagger[f] = \int (dk) f(\mathbf{k}) a^\dagger(\mathbf{k}) \tag{8.26}$$

which acts on a sequence F by

$$[a^\dagger[f]F]_n(\mathbf{k}_1, \dots, \mathbf{k}_n) = \sum_{j=1}^{n} f(\mathbf{k}_j) F_{n-1}(\mathbf{k}_1, \dots, \hat{\mathbf{k}}_j, \dots, \mathbf{k}_n), \tag{8.27}$$

where $\hat{\mathbf{k}}_j$ means that this argument is omitted. Similarly,

$$[a[f]F]_n(\mathbf{k}_1, \dots, \mathbf{k}_n) = \int (dk) f(\mathbf{k}) F_{n+1}(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_n). \tag{8.28}$$

All these formulas follow from (8.24) and the canonical commutators (8.13).

Returning to the original Klein-Gordon fields $\phi(x)$, we construct for each test function $g(\mathbf{x})$ an operator $\phi[g,t] = \int d^3x\, g(\mathbf{x})\phi(\mathbf{x},t)$ by using Eq. (8.8). We see that $\phi[g,t]$ is expressed in terms of $a[\hat{g}]$ and $a^\dagger[\hat{g}]$, where $\hat{g}$ is the Fourier transform of g. The resulting operators do satisfy the Klein-Gordon equation in the sense of distributions. The Hamiltonian of Eq. (8.17) exists as an operator in $\mathcal{H}$ and generates time evolution according to the Heisenberg equations.

The operator-valued distributions $\phi(x), a(\mathbf{k}), a^\dagger(\mathbf{k})$ themselves make sense as operators from Ω into Ω^* in the rigged Hilbert space. For example,

$$a^\dagger(\mathbf{p})|0> = |1_\mathbf{p}> = \{0, (2\pi)^3 2E_\mathbf{p}\delta^3(\mathbf{k}-\mathbf{p}), 0, \dots\}, \tag{8.29}$$

which is a sequence of distributions rather than functions. Normal-ordered products make sense in the same way, e.g.,

$$a^\dagger(\mathbf{p})a(\mathbf{p})\{0, F_1(\mathbf{k}), 0, 0, \dots\} = \{0, (2\pi)^3 2E_\mathbf{p} F_1(\mathbf{p})\delta^3(\mathbf{k}-\mathbf{p}), 0, 0, \dots\}, \tag{8.30}$$

and of course $a^\dagger(\mathbf{p})a(\mathbf{p})|0> = 0$. However, other products do not exist in even this weak sense, e.g. formally $a(\mathbf{p})a^\dagger(\mathbf{p})|0> = (2\pi)^3 2E_\mathbf{p}\delta^3(\mathbf{0})|0>$, which was the difficulty that motivated us to introduce normal ordering in the first place.

A Hilbert space of this type, which arises in free field theory and rigorously consists of sequences of functions as above, is called a Fock space (after V.A. Fock). In the physics literature Fock spaces are defined by giving scarcely more than the commutators of the relevant creation and annihilation operators, and it is up to the reader to supply the sequence-space construction if desired. However, a little experience doing calculations in these spaces should convince mathematicians of the convenience of the physicist's way of describing them.

The path integral machinery for computing Green's functions applies in free field theories with barely a change in notation, since the relevant integrals are still

Gaussian. For the Klein-Gordon theory we define

$$Z(J) = N\int [D\phi] \exp iS, \tag{8.31}$$

$$S = \int d^4x [-\frac{1}{2}\phi(\partial_\mu\partial^\mu + m^2)\phi + J(x)\phi(x)], \tag{8.32}$$

where intuitively we integrate over all continuous functions $\phi(x)$ having specified but arbitrary limits at $t \to \pm\infty$. The calculation is again simply completing the square; all we need is the Green's function satisfying

$$(\partial_\mu\partial^\mu + m^2)G(x-y) = \delta^4(x-y), \tag{8.33}$$

which is

$$G(x-y) = \int \frac{d^4k}{(2\pi)^4} \frac{-1}{k^2 - m^2 + i\epsilon} e^{ik\cdot(x-y)}, \tag{8.34}$$

where the $i\epsilon$ prescription is fixed by the Euclidean Green's function $(k^2+m^2)^{-1}$ and a Wick rotation of the k^0 axis. This Green's function can be evaluated in terms of Bessel functions, but its Fourier transform will be all we need. The partition function is

$$Z(J) = \exp \frac{i}{2} \int d^4x\, d^4y\, J(x)G(x-y)J(y), \tag{8.35}$$

and as usual the correlation functions are given by

$$<0|T[\phi(x_1)\cdots\phi(x_n)]|0> = (-i)^n \frac{\delta^n}{\delta J(x_1)\cdots\delta J(x_n)} Z(J)|_{J=0}. \tag{8.36}$$

Axioms

I will now present a set of axioms for QFT which reflect our (limited) experience with the Klein-Gordon field. Mathematicians find it reassuring to see that there actually are rules of the game which cannot be revised at will. In the 1960s, before renormalization was deeply understood, the mathematical foundation of QFT was felt to be insecure due to the cavalier treatment of formally infinite quantities. Its applicability beyond the subject of electrodynamics, for example to the strong interactions of particle physics, was also in doubt. The goal of axiomatic quantum field theory was to rigorously prove the folk theorems of the subject, and to learn which difficulties of the theory were inherent in its structure and which reflected unjustified approximations or calculational methods. Today it is somewhat disconnected from field theory as practiced by (most) physicists, but it is reassuring to see how much can be rigorously proven. Note that these axioms are intended to describe realistic quantum field theories in $\mathbf{R}^4$, so they refer to the Lorentz group and metric. They must be modified for topological and conformal field theories where spacetime may be compact and no metric is given.

Axiom 1. (*Quantum Mechanics*) States and observables of a quantum field are represented as usual by vectors and operators in a rigged Hilbert space $\mathcal{H}$, as per the axioms of quantum mechanics.

Axiom 2. (*Relativistic Invariance*) A continuous unitary representation of the (double cover of the) Poincaré group acts in the Hilbert space, an operator $U(\Lambda, a)$ representing the transformation $x' = \Lambda x + a$. We define the 4-momentum operator P^μ by $U(1, a) = \exp iP^\mu a_\mu$. Its eigenvalues k^μ lie in or on the forward light cone, $k^2 \geq 0$. There is a unique (up to a phase) Poincaré invariant vector $|0>$ called the vacuum.

Axiom 3. (*Fields*) For each test function $f(x)$ on $\mathbf{R}^4$ there is a set of operators $\phi_i[f] = \int d^4x f(x)\phi_i(x)$. All these operators and their adjoints are defined on a common domain Ω, which is dense in $\mathcal{H}$, closed under their action and that of $U(\Lambda, a)$, and is a linear set containing $|0>$. All the matrix elements $<\Psi|\phi_i[f]|\Psi'>$ are distributions in f. Under the Poincaré group the fields transform by

$$U(\Lambda, a)\phi_i[f(x)]U^{-1}(\Lambda, a) = \sum_j M_{ij}(\Lambda^{-1})\phi_j[f(\Lambda^{-1}(x-a))], \tag{8.37}$$

where M_{ij} is some matrix representation of $SL(2, \mathbf{C})$ and this equation holds when applied to any vector in Ω.

Axiom 4. (*Microscopic Causality*) If the supports of f and g are spacelike separated, i.e. $(x-y)^2 < 0$ for every x and y in the support of f and g respectively, then

$$[\phi_i[f], \phi_j[g]]_\pm \equiv \phi_i[f]\phi_j[g] \pm \phi_j[g]\phi_i[f] = 0, \tag{8.38}$$

where the $-$ sign must occur if the operators $\phi_i[f]$ represent observables.

Axiom 5. (*Cyclicity of the Vacuum*) The set of vectors obtained by applying polynomials in the $\phi_i[f]$ and their adjoints to the vacuum is dense in Ω, hence in $\mathcal{H}$.

We did not verify microscopic causality for the scalar field, but it is easy to compute $[\phi(x), \phi(y)]$ from our solution and to verify that it is a Lorentz invariant function of $x - y$. Since spacelike separated points can be Lorentz transformed so that $x^0 = y^0$ (exercise), microscopic causality follows from the equal time commutators. This axiom has many important physical consequences. The uncertainty principle (exercise) places no restriction on the accuracy with which commuting operators can be measured. The fact that measurements at spacelike separated points do not interfere with each other reflects the fact that no signal can propagate between them (faster than light). This axiom also ensures that time-ordered correlation functions are Lorentz invariant, which is not obvious a priori. The instruction to arrange operators in order of their time arguments is not Lorentz invariant, because the time order of spacelike separated events is not invariant (exercise). But precisely in this case the operators commute, so their order is irrelevant. Finally, the spin-statistics theorem says that the $+$ sign occurs in axiom (4) for fields associated to double-valued representations of the Lorentz group. As a consequence, each of

the four components of the quantized Dirac field obeys $\psi_i[f, t]^2 = 0$ (recall that in free field theories spatial smearing is enough). Therefore, in contrast to the scalar field, there is no vector in the Dirac Hilbert space describing two particles both in the same state f. This is the famous Pauli exclusion principle which is responsible for the shell structure of atoms and ultimately the stability of all matter! Particles obeying the exclusion principle are called fermions, while the others are bosons.

The cyclicity axiom (5) is a substitute for the properties of the creation and annihilation operators arising from canonical equal time commutation relations, which do not make sense when smearing over time as well as space is necessary. Since any vector can be approximated by fields acting on the vacuum, the diagonal vacuum matrix elements of operators determine all others. This partially accounts for the central importance of the n-point functions in field theory. The rest of the explanation will emerge in the next section.

The Dirac Field

As a second example of a free field theory satisfying these axioms, and to indicate the origin of the spin-statistics connection, I will briefly sketch the quantization of the Dirac equation. We recall the Lagrangian

$$\mathcal{L} = \bar{\psi}_\beta(i(\gamma^\mu)_{\beta\alpha}\partial_\mu - m\delta_{\beta\alpha})\psi_\alpha, \tag{8.39}$$

where for clarity I have introduced indices $\alpha, \beta = 1, 2, 3, 4$ labeling the components of ψ and the matrix entries of γ^μ. The momentum conjugate to ψ_α is by definition

$$\pi_\alpha = \frac{\partial \mathcal{L}}{\partial \dot{\psi}_\alpha} = i\bar{\psi}_\beta(\gamma^0)_{\beta\alpha} = i\psi^\dagger_\alpha, \tag{8.40}$$

so the canonical commutation relations should be

$$[\psi_\alpha(\mathbf{x}, t), \psi^\dagger_\beta(\mathbf{y}, t)]_\pm = \delta_{\alpha\beta}\delta^3(\mathbf{x} - \mathbf{y}), \tag{8.41}$$

$$[\psi_\alpha(\mathbf{x}, t), \psi_\beta(\mathbf{y}, t)]_\pm = [\psi^\dagger_\alpha(\mathbf{x}, t), \psi^\dagger_\beta(\mathbf{y}, t)]_\pm = 0, \tag{8.42}$$

where I have left the choice of commutators or anticommutators open for now. The Hamiltonian is constructed as

$$H = \int d^3x\, [\pi_\alpha\dot{\psi}_\alpha - \mathcal{L}] = \int d^3x\, \psi^\dagger(-i\gamma^0\gamma^j\partial_j + m\gamma^0)\psi. \tag{8.43}$$

The general solution of the Dirac equation, obtained via Fourier transform, which will motivate our change of variables to creation and annihilation operators, is

$$\psi(x) = \sum_{s=\pm 1/2} \int (dk)[b(\mathbf{k}, s)u(\mathbf{k}, s)e^{-ik\cdot x} + d^\dagger(\mathbf{k}, s)v(\mathbf{k}, s)e^{ik\cdot x}], \tag{8.44}$$

$$\psi^\dagger(x) = \sum_{s=\pm 1/2} \int (dk)[b^\dagger(\mathbf{k}, s)u^\dagger(\mathbf{k}, s)e^{ik\cdot x} + d(\mathbf{k}, s)v^\dagger(\mathbf{k}, s)e^{-ik\cdot x}]. \tag{8.45}$$

For the notation, review our discussion of the Klein-Gordon equation. Unlike the Klein-Gordon field $\phi(x)$ which was real and observable classically, hence assumed Hermitian in quantum mechanics, the Dirac field was complex even classically. Hence we do not take it to be Hermitian in QFT, which is why there are two distinct sets of creation and annihilation operators $b^\dagger, b$ and $d^\dagger, d$. We indeed have a solution to the Dirac equation provided that $k^2 = m^2$ and that the spinors u and v satisfy the Fourier-transformed Dirac equations

$$(\gamma^\mu k_\mu - m)u(\mathbf{k}, s) = (\gamma^\mu k_\mu + m)v(\mathbf{k}, s) = 0. \tag{8.46}$$

That is, the four spinors $u(\mathbf{k}, \pm 1/2), v(\mathbf{k}, \pm 1/2)$ are the four eigenvectors of the matrix $\gamma^\mu k_\mu$, which has eigenvalues $\pm m$ as a consequence of

$$(\gamma^\mu k_\mu)^2 = \gamma^\mu k_\mu \gamma^\nu k_\nu = \frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu)k_\mu k_\nu = \eta^{\mu\nu} k_\mu k_\nu = k^2 = m^2. \tag{8.47}$$

They are conventionally normalized so that

$$\bar{u}(\mathbf{k}, s)u(\mathbf{k}, s') = -\bar{v}(\mathbf{k}, s)v(\mathbf{k}, s') = 2m\delta_{ss'}. \tag{8.48}$$

The pair of spinors $u(\mathbf{k}, \pm 1/2)$ form a basis for a two-dimensional representation (highest weight $1/2$) of the $SO(3)$ subgroup of the Lorentz group which fixes the vector $\mathbf{k}$, and the same is true of the pair $v(\mathbf{k}, \pm 1/2)$. This is just the spatial rotation group in the rest frame of the particle, the inertial coordinate system in which $\mathbf{k} = 0$. This is what is meant by the statement that the Dirac equation describes particles of spin $1/2$.

In terms of the new creation and annihilation operators, the Hamiltonian becomes

$$H = \sum_s \int (dk) E_{\mathbf{k}} [b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) - d(\mathbf{k}, s)d^\dagger(\mathbf{k}, s)]. \tag{8.49}$$

The operators themselves satisfy the (anti)commutators

$$[b(\mathbf{k}, s), b^\dagger(\mathbf{p}, s')]_\pm = [d(\mathbf{k}, s), d^\dagger(\mathbf{p}, s')]_\pm = (2\pi)^3 2E_{\mathbf{k}} \delta_{ss'} \delta^3(\mathbf{k} - \mathbf{p}), \tag{8.50}$$

with other (anti)commutators vanishing. We want to use these relations to redefine the Hamiltonian by changing the $dd^\dagger$ term into $d^\dagger d$ and dropping the resulting infinite constant. Because of the minus sign already present in H, we will get a Hamiltonian involving either the difference or the sum of the b and d number operators according to our choice of commutators or anticommutators for these operators respectively. The choice of commutators means that creating more d particles will lower the energy, so that the energy is unbounded below. In an interacting theory the vacuum would be unstable to the spontaneous creation of more and more such particles. Thus we are forced to quantize the Dirac field using anticommutators, which is the spin-statistics connection leading to the Pauli

exclusion principle: the vanishing anticommutator of an operator with itself means that

$$b^\dagger(\mathbf{k}, s)b^\dagger(\mathbf{k}, s) = d^\dagger(\mathbf{k}, s)d^\dagger(\mathbf{k}, s) = 0, \tag{8.51}$$

and it is impossible to create two particles of either b or d type with identical values of $\mathbf{k}, s$.

The Hamiltonian is now

$$:H: = \sum_s \int (dk) E_{\mathbf{k}} [b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) + d^\dagger(\mathbf{k}, s)d(\mathbf{k}, s)]. \tag{8.52}$$

It commutes with the following conserved quantities: the particle number

$$N = \sum_s \int (dk)[b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) + d^\dagger(\mathbf{k}, s)d(\mathbf{k}, s)], \tag{8.53}$$

the momentum

$$\mathbf{P} = \sum_s \int (dk)\mathbf{k}[b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) + d^\dagger(\mathbf{k}, s)d(\mathbf{k}, s)], \tag{8.54}$$

the charge

$$Q = e \sum_s \int (dk)[b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) - d^\dagger(\mathbf{k}, s)d(\mathbf{k}, s)], \tag{8.55}$$

which is in fact the charge associated with the conserved current $j^\mu = \bar{\psi}\gamma^\mu\psi$ we have seen before, but normalized to have eigenvalue e in the state $b^\dagger(\mathbf{k}, s)|0>$, and the z-component of spin,

$$S_z = \sum_s \int (dk) s[b^\dagger(\mathbf{k}, s)b(\mathbf{k}, s) + d^\dagger(\mathbf{k}, s)d(\mathbf{k}, s)], \tag{8.56}$$

so called because this operator is part of the generator of rotations about the z axis in the $SO(3)$ subgroup mentioned above. Hence we obtain the particle interpretation of the Dirac field: $b^\dagger(\mathbf{k}, s)$ creates particles (electrons in the application to quantum electrodynamics) of mass m, momentum $\mathbf{k}$, spin component s, and electric charge e, while $d^\dagger(\mathbf{k}, s)$ creates their antiparticles (positrons), which have the same mass, momentum, and spin but opposite charge. We see that QFT predicts the existence of antiparticles as well as the spin-statistics connection. The mathematical description of the Dirac Fock space in terms of sequences of functions is left to the reader. The path integral formulation of the Dirac field, which involves fundamentally new ideas because of the anticommutation relations, is unfortunately beyond the scope of these lectures (but see those of Orlando Alvarez in this volume).

LECTURE 9
Interacting Fields, Feynman Diagrams, and Renormalization

Free field theories describe particles which do not interact with each other. The eigenstates of H, which change only by a phase factor under time evolution, contain particles whose momenta are constant for all time. Interacting particles are described by nonlinear field equations, whose Lagrangians contain higher powers of fields than quadratic. Continuing to avoid the mathematical complications of gauge invariance or nontrivial Lorentz representations, we take as our example the Klein-Gordon field with a quartic interaction,

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{4!}\lambda\phi^4. \tag{9.1}$$

We choose a quartic rather than cubic interaction so that the Hamiltonian,

$$H = \int d^3x[\frac{1}{2}\pi^2 + \frac{1}{2}|\nabla\phi|^2 + \frac{1}{2}m^2\phi^2 + \frac{1}{4!}\lambda\phi^4], \tag{9.2}$$

is positive definite. To get a feeling for the effects of the quartic term, suppose we express it in terms of the $a, a^\dagger$ operators using the change of variables (8.8). It becomes a sum of terms having the structure

$$\int (dk_1)\cdots(dk_4)a(\mathbf{k}_1)\cdots a^\dagger(\mathbf{k}_4)\delta^4(k_1 + \cdots - k_4). \tag{9.3}$$

That is, it can create or destroy any combination of four particles provided that the total energy and momentum do not change. By destroying two particles and then recreating them with different momenta it can describe collisions in which particles exchange momentum. It can also change the total particle number, allowing the description of processes in which new particles are created from the kinetic energy of others.

This so-called $\lambda\phi^4$ field theory can be rigorously constructed in two or three space-time dimensions, but not (thus far) in the physical case of four dimensions. What we know about it (assuming it exists) is a combination of general theorems of axiomatic field theory and perturbative series expansions in powers of λ around

the free limit $\lambda = 0$. We will in fact develop a series expansion for the partition function which is formally given by the path integral

$$Z(J) = \int [D\phi] \exp i \int d^4x [\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{4!}\lambda\phi^4 + J\phi], \tag{9.4}$$

normalized to $Z(0) = 1$. (We know that the quadratic terms in this exponential define an honest Gaussian measure on a path space, but what is hard is to interpret the rest of the exponential as an integrable function with respect to this measure.) From experience with the corresponding one variable integral,

$$\int dx \exp -(x^2 + \lambda x^4), \tag{9.5}$$

we should not expect such an expansion to converge, but ideally it would be an asymptotic expansion for the exact result which would give accurate numerical results for λ small. In quantum electrodynamics, where the expansion parameter is $e^2/\hbar c \approx 1/137$ (e is the charge of an electron), such perturbative calculations do agree with experimental measurements to as many as 12 decimal places.

Most experimental tests of QFT involve scattering experiments in which beams of known particles are created at accelerators and allowed to collide. One measures the probabilities that various types and numbers of particles emerge from the collision with various momenta. In our model field theory we ask for the probability amplitude to begin with m particles of 4-momenta k_i at $t = -\infty$ and to end with n particles of momenta p_i at $t = +\infty$, denoted $<\{p_i\}\text{out}|\{k_i\}\text{in}>$. This is viewed as an element of a matrix, the scattering or S-matrix, whose rows and columns are indexed by the stationary states of the free field theory. Intuitively we expect that any collection of particles will separate as $t \to \pm\infty$ into widely separated, hence noninteracting, particles (and perhaps bound clusters of particles, e.g. atoms). In field theory we express this by some statement such as $\phi(x) \sim Z^{1/2}\phi_{\text{out}}(x)$ for $t \to +\infty$, and $\phi(x) \sim Z^{1/2}\phi_{\text{in}}(x)$ for $t \to -\infty$, where ϕ is the interacting field, ϕ_{in} and ϕ_{out} are free fields obeying the Klein-Gordon equation, and Z is not the partition function but merely some normalization constant which can be shown to be the same for both limits. The in and out states appearing in the S-matrix element are created by the $a^\dagger$ operators in the Fourier expansions of $\phi_{\text{in}}, \phi_{\text{out}}$. The precise meaning of these operator limits is quite subtle, e.g. they certainly do not hold as strong operator equations, but only weakly as relations between certain of their matrix elements. Using them, one can derive the remarkable LSZ (Lehmann-Symanzik-Zimmermann) reduction formula expressing the S-matrix elements in terms of the correlation functions:

$$\begin{aligned} <\{p_i\}\text{out}|\{k_i\}\text{in}> &= (iZ^{-1/2})^{m+n} \int d^4x_1 \cdots d^4x_m d^4y_1 \cdots d^4y_n e^{i(p_j\cdot y_j - k_j\cdot x_j)} \\ &\times(\Delta_{x_1} + m^2)\cdots(\Delta_{y_n} + m^2) <0|T[\phi(x_1)\cdots\phi(y_n)]|0> . \end{aligned} \tag{9.6}$$

In view of the fact that $\Delta + m^2$ becomes $-k^2 + m^2$ under Fourier transform, this says that the Fourier transform of the correlation function must have poles at

the mass shell values $p_i^2 = k_i^2 = m^2$, and the residues of these poles are the measurable scattering amplitudes! This accounts for the emphasis on the computation of correlation functions in field theory. As a check, one finds that in free field theory S conserves particle number and momenta.

Let us now derive the asymptotic expansion for $Z(J)$ and hence for the correlation functions. This is cleverly done by using the source $J(x)$ to replace the quartic term by four functional derivatives:

$$\begin{aligned} Z(J) &= \int [D\phi] \exp\left[\int d^4x \frac{-i\lambda}{4!}\phi^4\right] \exp i \int d^4x [\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + J\phi] \\ &= \int [D\phi] \exp\left[\int d^4x \frac{-i\lambda}{4!}\left(-i\frac{\delta}{\delta J(x)}\right)^4\right] \\ &\qquad \times \exp i \int d^4x [\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + J\phi] \\ &= \exp\left[\int d^4x \frac{-i\lambda}{4!}\left(-i\frac{\delta}{\delta J(x)}\right)^4\right] Z_0(J). \end{aligned} \tag{9.7}$$

That is, $Z(J)$ is expressed as a series of successively higher functional derivatives of $Z_0(J)$, the partition function for the free theory evaluated in Eq. (8.35). The correlation functions follow by taking further derivatives,

$$<0|T[\phi(x_1)\cdots\phi(x_n)]|0> = (-i)^n \frac{\delta^n}{\delta J(x_1)\cdots\delta J(x_n)} Z(J)|_{J=0}. \tag{9.8}$$

They are expressed as power series in λ whose coefficients are various integrals of products of the Green's functions $G(x-y)$ appearing in $Z_0(J)$. Writing out these terms is a tedious combinatorial exercise which is greatly facilitated by Feynman diagrams, which are simply graphical representations of the various terms.

Recall from Wick's theorem that the nonvanishing terms in a repeated functional derivative such as (9.8) come from complete pairings of all the derivatives, the pairing of $\delta/\delta J(x)$ with $\delta/\delta J(y)$ leading to a factor $G(x-y)$ in the result. We represent each possible complete pairing by a graph in which there is a vertex labeled with each argument x of a functional derivative $\delta/\delta J(x)$. Vertices corresponding to paired derivatives are joined by a line. It follows that four lines will meet at each internal vertex [a vertex corresponding to an operator $\delta^4/\delta J(x)^4$ from the expansion of the exponential in Eq. (9.7)], while only one line emerges from an external vertex corresponding to an argument x_i in the correlation function being computed. The following "Feynman rules" specify the contribution to the correlation function (9.8) represented by a given graph. Each internal vertex in the graph represents a factor $-i\lambda$, so that the number of internal vertices in a graph is the order of the corresponding term in the power series expansion in λ. A line connecting points x and y represents a factor $-iG(x-y)$. The resulting expression is to be integrated over all internal vertex labels, leaving a function of the external points only. The correlation function to a given order in λ is the sum of the contributions of all graphs with that number of internal vertices or less.

For some graphs, there is an additional numerical "symmetry factor" in addition to the factors specified in these rules. It arises from an incomplete cancellation of

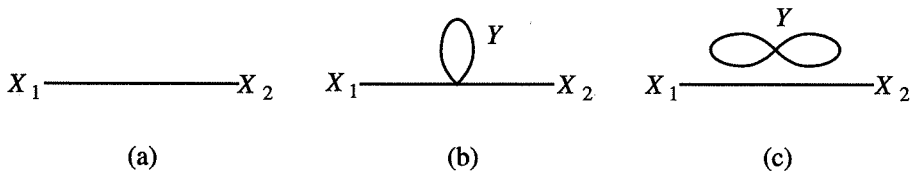

Figure 9.1. Feynman diagrams contributing to the two-point function to order λ.
(a) $-iG(x_1 - x_2)$.
(b) $\frac{\lambda}{2}\int d^4y G(x_1 - y)G(y - x_2)G(0)$.
(c) $\frac{\lambda}{8}G(x_1 - x_2)\int d^4y G^2(0)$.

two numerical factors already present in the series for $Z(J)$: the $1/4!$ in the $\lambda\phi^4$ interaction, and the $1/j!$ in the jth term of the expansion of the exponential in Eq. (9.7). The "normal" situation in a graph is for each internal vertex y to be connected to four distinct neighbor vertices $x_1, \ldots, x_4$. There are 4! distinct ways to choose which of the four derivatives $\delta^4/\delta J(y)^4$ are paired with the $\delta/\delta J(x_i)$, so by letting one graph denote all such choices we eliminate the factor $1/4!$. However, for some graphs this counting is not correct. Whenever an internal vertex is connected to fewer than four distinct neighbors, an extra numerical factor must be attached to the contribution of the graph. For example, in the graph of Figure 4b, there are 4×3 ways to connect x_1, x_2 to y, but then there is only one way to connect y to y. This graph has a factor $4 \times 3/4! = 1/2$. In general, this "symmetry factor" is $1/N$, where N is the order of the permutation group of the lines in the graph, leaving the vertices fixed. Similarly, the $1/j!$ at order j is supposed to be cancelled by the $j!$ ways of labelling the internal vertices with integration variables, so that by drawing only one of the labelings we can omit this factor. Again, there are exceptional situations in which different labelings do not give distinguishable graphs and a symmetry factor is left over, but we will not encounter these.

As an example, let us compute the two point function $<0|T[\phi(x_1)\phi(x_2)]|0>$ to second order in λ. Using Eqs. (9.7,9.8), we have

$$<0|T[\phi(x_1)\phi(x_2)]|0> = (-i)^2\frac{\delta^2}{\delta J(x_1)\delta J(x_2)}\left[1 - \frac{i\lambda}{4!}\int d^4x\,\frac{\delta^4}{\delta J(x)^4}\right.$$
$$\left. - \frac{\lambda^2}{4!4!}\int d^4x\,d^4y\,\frac{\delta^8}{\delta J(x)^4\delta J(y)^4} + \cdots\right] Z_0(J)\,|_{J=0}\,. \tag{9.9}$$

The results of these derivatives to order λ are shown with the corresponding graphs in Figure 4. At zeroth order in λ we have the free field theory Green's function $-iG(x - y)$. At higher order we see at once that some diagrams are topologically connected while others have two (or more) connected components. Each component is a separate multiplicative factor in the corresponding expression; no factors $G(x - y)$ link points in different components. As I will now explain, we should omit the contribution of every disconnected diagram. This is because of the fact, which we have ignored until now, that $Z(J)$ was to be normalized so that $Z(0) = 1$. To accomplish this, all our formulas for correlation functions must be divided by $Z(0)$. In fact, we can explicitly exhibit a factor $Z(0)$ in the two-point function, because $Z(0)$ itself has a Feynman diagram expansion: it is a degenerate case of

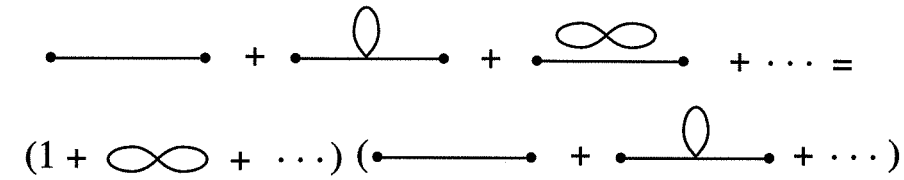

Figure 9.2. Factorization of the two-point function into the zero-point function $Z(0)$ and the series of connected diagrams.

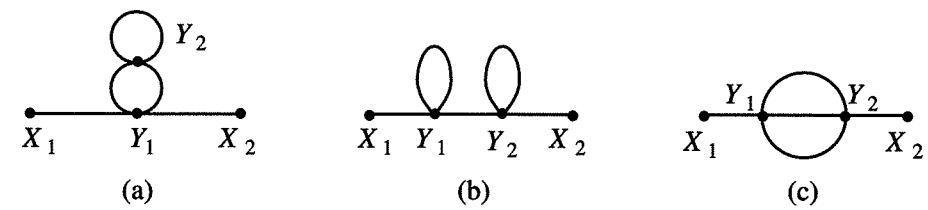

Figure 9.3. Connected graphs for the two-point function at order λ^2.
(a) $\frac{i\lambda^2}{4}\int d^4y_1 d^4y_2 G(x_1 - y_1)G(y_1 - x_2)G^2(y_1 - y_2)G(0)$.
(b) $\frac{i\lambda^2}{4}\int d^4y_1 d^4y_2 G(x_1 - y_1)G(y_1 - y_2)G(y_2 - x_2)G^2(0)$.
(c) $\frac{i\lambda^2}{6}\int d^4y_1 d^4y_2 G(x_1 - y_1)G^3(y_1 - y_2)G(y_2 - x_2)$.

a "zero-point" correlation function and as such corresponds to diagrams having no external vertices at all. Pictorially, the two-point function we computed factors into two series as in Figure 5, where the first series is $Z(0)$ and the second contains only connected diagrams. A similar argument shows that ignoring the disconnected diagrams is the way to correctly normalize every correlation function. Unfortunately, the integral corresponding to the remaining connected diagram contains a factor $G(0)$ which is undefined.

The connected contributions to the two-point function at second order are shown in Figure 6. The second order contribution also involves divergent integrals because of the singularity of the Green's function $G(y_1 - y_2)$ as $y_1 \to y_2$. These are the rest of the notorious infinities of QFT — the serious ones this time. Obtaining sensible results from this series of divergent integrals is the goal of the program of renormalization. The mathematical origin of the divergences is the same as in free field theory: the presence of undefined formal products of distributions in the Lagrangian from which we computed the path integral. Renormalization can in fact be described in mathematical terms as the problem of defining these products of distributions, but we will take a more physical approach.

It will be helpful to study not the correlation functions themselves, but their Fourier transforms. Not only are these are the physically significant quantities according to the LSZ formula, but we only know $G(x - y)$ through its Fourier representation anyway. Mathematically it is standard to study products of distributions in terms of their symbols. Therefore we study the two point function in momentum space,

$$\hat{G}(p_1, p_2) \equiv \int d^4x_1 d^4x_2\, <0|T[\phi(x_1)\phi(x_2)]|0>\, e^{-ip_1\cdot x_1} e^{-ip_2\cdot x_2}. \tag{9.10}$$

For example, computing this transform for the first order contribution of Figure 4b and inserting the integral representation of $G(x-y)$ gives

$$\begin{aligned}&\int d^4x_1\, d^4x_2\, d^4x\, e^{-ip_1\cdot x_1} e^{-ip_2\cdot x_2}\\ &\quad\times\frac{\lambda}{2}\int\frac{d^4k_1\, d^4k_2\, d^4k}{(2\pi)^{12}} e^{ik_1\cdot(x_1-x)} e^{ik_2\cdot(x_2-x)} e^{ik\cdot 0}\\ &\quad\times\frac{-1}{(k_1^2-m^2+i\epsilon)(k_2^2-m^2+i\epsilon)(k^2-m^2+i\epsilon)}\end{aligned}\tag{9.11}$$

Performing the integrals over x_1, x_2, x, this becomes

$$\begin{aligned}&\frac{\lambda}{2}\int d^4k_1\, d^4k_2\, d^4k\, \delta^4(k_1-p_1)\delta^4(k_2-p_2)\delta^4(k_1+k_2)\\ &\quad\times\frac{-1}{(k_1^2-m^2+i\epsilon)(k_2^2-m^2+i\epsilon)(k^2-m^2+i\epsilon)}.\end{aligned}\tag{9.12}$$

Using the delta functions to do two more integrals, we obtain

$$\frac{\lambda}{2}\delta^4(p_1+p_2)\frac{1}{(p_1^2-m^2)(p_2^2-m^2)}\int d^4k\,\frac{-1}{k^2-m^2+i\epsilon},\tag{9.13}$$

which we immediately recognize as quadratically divergent.

By applying the same transform to an arbitrary graph, we obtain a set of Feynman rules for directly writing down the series expansion of the Fourier transforms of the correlation functions. We draw the same diagrams as before, but instead of labeling the vertices we label each line with a 4-momentum variable k. The lines should be directed; with our Fourier transform conventions the momenta on the external lines (those containing an external vertex) should flow into the graph, but the directions on the internal lines are arbitrary. Each line represents a factor $i/(k^2-m^2+i\epsilon)$, and each vertex represents a factor $-i\lambda(2\pi)^4\delta^4(\sum k)$, where the delta function sets to zero the net momentum flowing into the vertex. Multiply by the symmetry factor $1/N$ and integrate the internal momenta with the measure $\prod d^4k/(2\pi)^4$. As in the above example, these rules always produce an overall delta function which expresses the conservation of the total momentum flowing into the graph, and there will always be explicit poles when an external momentum $p^2=m^2$, as required by the LSZ formula. Often we omit these factors for brevity. As another example, Figure 7 shows a second order graph and its contribution.

To analyze and tame the divergences it is helpful to introduce a regularization, or cutoff, for the integrals. The simplest (though not computationally the most useful) way to do this is to fix a large number Λ and restrict each momentum integral $d^4k/(2\pi)^4$ to the region $k^2<\Lambda^2$. Because this region is not compact when k^2 is defined with the Lorentz metric, we should perform the Wick rotation to Euclidean space on all the k^0 integrals first. This amounts to replacing every term k^2 in the Lorentz metric by $-k^2$ in the Euclidean metric. We then compute integrals over balls of radius Λ, and the results, which are finite but depend on Λ, are related by analytic continuation to the integrals of actual interest. The problem is to determine how singular the integrals are in the limit $\Lambda\to\infty$, and

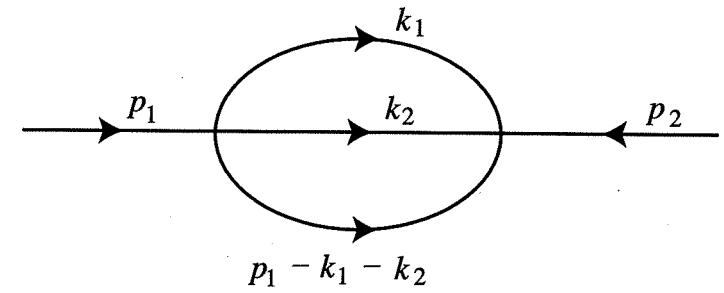

Figure 9.4. The second-order contribution of Figure 6c to the two-point function in the momentum space (Fourier transform) representation:
$(2\pi)^4\delta^4(p_1+p_2)(p_1^2-m^2)^{-1}(p_2^2-m^2)^{-1}\frac{-i\lambda^2}{6}$
$\times\int\frac{d^4k_1 d^4k_2}{(2\pi)^8}(k_1^2-m^2+i\epsilon)^{-1}(k_2^2-m^2+i\epsilon)^{-1}[(p_1-k_1-k_2)^2-m^2+i\epsilon]^{-1}$.

how to modify them to remove this divergence. Of course, the modification cannot be arbitrary, but should be physically motivated so that we know how it affects the physical interpretation of the finite results.

Let me remark that this regularization procedure is precisely what a mathematician might do to study singular products of distributions. For example, to study products of delta functions, we might replace them by delta sequences of smooth functions which become more localized at zero. One way to do this is to impose a cutoff $k^2<\Lambda^2$ on the Fourier integral representation,

$$\delta(x)=\int\frac{dk}{2\pi}e^{ikx}.\tag{9.14}$$

We could then study the singularity in the products of such smooth functions for $\Lambda\to\infty$.

We have estimated the "degree of divergence" of our integrals thus far by "naive power counting": each momentum integration d^4k is counted as Λ^4, each denominator (k^2-m^2) counts as Λ^{-2}, we multiply these factors to obtain Λ^D, and the integral is finite if $D<0$ but divergent if $D\geq 0$. A basic result known as Weinberg's theorem [**17**] (due to Nobel Laureate — but not at the time — Steven Weinberg) justifies this procedure. It shows that the contribution to the integral from a region in which all integration variables are of order Λ does behave as this naive power of Λ (where Λ^0 is understood to mean $\ln\Lambda$). There may be contributions with different behavior from other regions, but they are associated with identifiable subgraphs of the Feynman diagram. This makes possible an inductive study of the divergences.

Let us return to the two-point function $\hat{G}(p_1,p_2)$ to order λ^2 which diverges as Λ^2 by our calculations and Weinberg's theorem. Observe that it is a Lorentz-invariant function of p_1, p_2 supported at $p_1=-p_2$, therefore a function only of $p_1^2=p_2^2\equiv p^2$. By Taylor-expanding the integrands of the Feynman graphs in p, we can write

$$\hat{G}(p^2)=A+Bp^2+C(p^2)^2+\cdots,\tag{9.15}$$

where $A, B, C, \ldots$ are certain integrals over internal momenta k_i obtained by differentiating the original integral and setting $p = 0$. The key point is that each derivative produces more momentum factors in the denominator, thereby lowering the degree of divergence according to Weinberg's theorem. Thus, A diverges as Λ^2, B diverges as $\ln \Lambda$, and higher coefficients are actually finite. The crucial result, which holds for all graphs and is well-known in distribution theory, is that the divergences can always be isolated in the form of a polynomial in the external momenta, in this case $A + Bp^2$. Subtracting this polynomial from $\hat{G}$ leaves a finite result, but we must understand the effect of such subtractions on the interpretation of the theory.

To do this, let us return to the Lagrangian of Eq. (9.1) and ask for the physical interpretation of the parameters m and λ. We know that when $\lambda = 0$ the particles are free and have mass m, which leads us to guess that in general m is their mass and λ is some measure of their interaction strength or charge. To confirm this, we should choose some measure of interaction strength, perhaps the average deflection angle of interacting beams of particles of some specified momentum, and compute it in terms of m and λ using Feynman diagrams. We might also deduce the particle mass in terms of m and λ by comparing deflection angles at various momenta. Given our experience, the result of the computation is likely to be divergent. So let us admit that we do not know the physical meaning of the parameters in the Lagrangian. To indicate this, let us rename them as the "bare" or "unrenormalized" mass m_0 and charge λ_0, in contrast to the "dressed", "physical", or "renormalized" mass m and charge λ which are defined and measured by actual experiments. We also rename the bare field ϕ_0, so that the bare Lagrangian is

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi_0\partial^\mu\phi_0 - \frac{1}{2}m_0^2\phi_0^2 - \frac{1}{4!}\lambda_0\phi_0^4. \tag{9.16}$$

We define the renormalized field by $\phi = Z^{-1/2}\phi_0$, where Z is the constant in the LSZ formula (9.6), so that ϕ is the field which actually becomes free in the limits $t \to \pm\infty$. We assume that Z, m_0, and λ_0 can be related to the physical mass and charge by asymptotic series of the form

$$Z = 1 + a_1\lambda + a_2\lambda^2 + \cdots, \tag{9.17}$$

$$m_0 = m(1 + b_1\lambda + b_2\lambda^2 + \cdots), \tag{9.18}$$

$$\lambda_0 = \lambda(1 + c_1\lambda + c_2\lambda^2 + \cdots), \tag{9.19}$$

where the coefficients may depend on m, Λ, and on the fixed momentum in the experiment defining the physical charge. These series guarantee that in the free theory with $\lambda = 0$ we have $Z = 1$, $\lambda_0 = 0$, and m_0 equal to the physical mass.

Now, perturbation theory has given us an expression for $\hat{G}(p^2)$ in terms of m_0, λ_0, Λ through order λ_0^2. We demand that $Z^{-1}\hat{G}(p^2)$, the correlation function for the fields ϕ rather than ϕ_0, should be finite in the limit $\Lambda \to \infty$ when it is reexpressed in terms of the physical quantities m, λ using the series above to order λ^2. If it is possible to choose the coefficients a_i, b_i, c_i in these series so as to make all the n-point functions finite to all orders in λ in this way, our theory is called renormalizable. In fact, the $\lambda\phi^4$ theory is renormalizable, as is quantum electrodynamics.

One way to see how renormalization works is to substitute the series (9.17)–(9.19) into the bare Lagrangian (9.16), obtaining

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{4!}\lambda\phi^4 \\ & + \frac{1}{2}(a_1\lambda + a_2\lambda^2 + \cdots)\partial_\mu\phi\partial^\mu\phi \\ & - \frac{1}{2}m[(a_1 + b_1)\lambda + (a_2 + b_2 + a_1b_1)\lambda^2 + \cdots]\phi^2 \\ & - \frac{1}{4!}[(c_1 + 2a_1)\lambda + (c_2 + 2a_2 + a_1^2 + 2a_1c_1)\lambda^2 + \cdots]\phi^4.\end{aligned} \tag{9.20}$$

The substitution has produced many new interaction terms in the Lagrangian, called counterterms. When we expand $Z(J)$ in its perturbation series, the new terms will lead to new types of vertices in the Feynman diagrams. The counterterm proportional to $\partial_\mu\phi\partial^\mu\phi$ leads to a vertex where two lines meet representing a factor $p^2(a_1\lambda + \cdots)$ (p is the momentum flowing through the vertex; the p^2 arises from the Fourier transform of the two derivatives). The counterterm proportional to ϕ^2 is another vertex where two lines meet representing the factor $-i(a_1 + b_1)m\lambda + \cdots$, and the ϕ^4 counterterm is a four-point vertex representing $-i(c_1 + 2a_1)\lambda + \cdots$. There will be new diagrams involving the two-point vertices making contributions of the form $A + Bp^2$ to $\hat{G}(p^2)$, and by choosing the coefficients correctly they will precisely cancel the divergences we found earlier. The four-point counterterm is needed to cancel divergences in the four-point function which we have not studied.

In order for this theory to be renormalizable it is clearly necessary that we do not continually encounter new divergences as we compute new Green's functions or as we proceed to higher orders in λ, because we only have three types of counterterms available to cancel them all. To see that this condition is satisfied, we will determine the degree of divergence of an arbitrary diagram having E external lines, V internal vertices, and I internal lines. There are I internal momentum integrals, and $V - 1$ delta functions expressing momentum conservation at each vertex (the -1 is for the overall conservation of external momenta, which will not help us do any integrals), leaving $L = I - V + 1$ integrals to do (this is also the number of independent loops, or generators of the fundamental group, of the graph). There are I denominators containing momenta, so the graph behaves like Λ raised to the power $4L - 2I = 4(I - V + 1) - 2I = 2I - 4V + 4$. Now we need a relation between V and I. Since four lines meet at each vertex, the total number of lines in the graph should be $4V$, but this counts the internal lines twice, so in fact $4V = E + 2I$. Substituting, the graph behaves like Λ^{4-E}. This crucial result shows that the divergences do not get worse as the graphs get larger. In fact, the two-point function is quadratically divergent at all orders, the four-point function is logarithmically divergent (that's what the third counterterm was for!) and the other Green's functions are "superficially convergent" (that is, if they diverge, it is only because they contain two- or four-point subgraphs). From here the proof of renormalizability is a (nontrivial!) combinatorial exercise of checking that the same counterterm that makes a given graph finite will also appear with the correct coefficient to cancel the divergences arising when this graph appears as a subgraph of a larger one.

Let me conclude this introduction to renormalization theory with several comments and hints concerning topics beyond the scope of these lectures.

(1) The key to the renormalizability of $\lambda\phi^4$ theory is our computation of the degree of divergence $4 - E$ for the E-point function. This depends on a delicate balance between the dimension of spacetime, which leads to numerator factors d^4k in Feynman integrals, and the power of ϕ in the interaction term of $\mathcal{L}$, which controls the number of lines meeting at each vertex and indirectly the number of momentum denominators in the integrals. One finds that $\lambda\phi^4$ theory is not renormalizable in more than four dimensions, and is superrenormalizable (diagrams actually become less divergent at higher orders in λ, and beyond a certain order no more counterterms are needed) in less than four dimensions. An easy way to understand this is to apply dimensional analysis to the Lagrangian of Eq. (9.1). In order for the action to be dimensionless, so that $\exp iS$ makes sense, we see from the kinetic term that ϕ must have dimensions of length^{-1}, or of momentum, in four dimensions, and then from the interaction term that λ is dimensionless. Since each Green's function has a fixed dimension, and λ is dimensionless, the momentum integrals at every order in λ have the same dimensions, and therefore the same degree of divergence. In contrast, in more than four dimensions, λ would have dimensions of momentum to some negative power, so that momentum integrals at higher orders must have more momenta in their numerators and hence higher degrees of divergence. In general, a dimensionless expansion parameter is the poor man's test for renormalizability.

(2) The only condition we have imposed on the counterterms a_i, b_i, c_i is that they contain terms with behavior Λ^2 and $\ln\Lambda$ which cancel the divergences of the Feynman integrals for $\Lambda \to \infty$. But we also want to compute the resulting finite Green's functions. Since a_i, b_i, c_i are still arbitrary by constants independent of Λ, the renormalized Green's functions are far from unique. How do we fix or understand this ambiguity? The answer is that we specify that m is the physical mass of the particles, and λ is their charge as measured in a specific experiment. When we compute the results of this experiment via Feynman diagrams and demand that they be equal to m and λ to each order of perturbation theory, this will fix the arbitrary constants in the counterterms to that order. If we choose another definition of the charge, as being the result of some other experiment, the constants will change. In general, there is a standard physical definition of the mass, while the meaning of the charge is more a matter of convention, but even this need not always be the case. For example, in quantum chromodynamics (QCD) we believe that quarks attract each other by such strong forces that they can never be isolated as free particles. In such a case, the mass of a free quark cannot be measured, and several indirect definitions of mass via various experiments are used. What we really have is a two-parameter family of quantum field theories for which we can choose coordinates m and λ in many convenient ways. The freedom to add constants to the counterterms is the invariance of this family under reparametrization, which is one form of what is called the renormalization group.

(3) Suppose we have fixed some experimental definitions of m and λ. In order to get a renormalized theory with these specific numerical values of m and λ, according to Eqs. (9.18,9.19) the bare parameters m_0 and λ_0 must flow in a specific manner with Λ as $\Lambda \to \infty$. (Of course, the asymptotic series may not have any validity for $\Lambda \to \infty$. We are really assuming that they are asymptotic expansions of some expressions which do make sense.) This one-parameter Abelian group action on the two-dimensional space of m_0 and λ_0 parametrized by Λ is another form of the renormalization group. For a more profound discussion of renormalization in terms of this flow, see [**18**].

(4) After renormalization, every term in the perturbation series of every Green's function is finite. However, this says nothing about the convergence of the series itself. As mentioned already, in those two- and three-dimensional field theories which can be constructed rigorously, perturbation theory is only asymptotic to the exact Green's functions, and we hope/pretend this is the case in four dimensions. In some cases Borel summation of perturbation theory gives the exact Green's functions, while in other cases the series is not Borel summable. Nevertheless we know from experiment that the first few terms of the series in quantum electrodynamics give correct results to many decimal places. The mathematician reader should stop to absorb the fact that the renormalizability of electrodynamics is not just a theorem to a physicist: physics students must actually be trained to compute renormalized Green's functions correct down to the signs and factors of 2π!

(5) Recall that a Lagrangian with a continuous symmetry possesses a conserved current $j^\mu, \partial_\mu j^\mu = 0$, which is generally some quadratic expression in the fields. Green's functions involving this current, as given by Feynman integrals before regularization, will formally satisfy

$$\partial_\mu <0|T[j^\mu(x)\phi(x_1)\cdots\phi(x_n)]|0> = 0. \tag{9.21}$$

However, the cutoff Feynman integrals may not satisfy this relation, and there may be no choice of counterterms for which the renormalized Green's functions satisfy it. [Because j^μ is quadratic in the fields, this Green's function is an $(n+2)$-point function with two arguments coinciding. Even after renormalization, Green's functions are normally singular when their arguments coincide. Further counterterms, amounting to a modification of the formal definition of j^μ, are required to make this Green's function finite. This is called the renormalization of composite operators.] In such a case the conservation law $\partial_\mu j^\mu = 0$ was only formal; the rigorously defined operator-valued distribution $j^\mu(x)$ is not conserved, and we say there is an anomaly. Such anomalies typically occur in theories involving Dirac fields and are deeply related to the Atiyah-Singer index theorem for the Dirac operator.

EXERCISES

Self-adjointness and time evolution in quantum mechanics

Let H be a complex separable Hilbert space. An operator A on H is a linear function from a linear subspace $D(A)$ to H; $A : D(A) \to H$ for short. The linear subspace $D(A)$ is called the domain of A. For the most part we are interested only in the case where $D(A)$ is dense in H. Key examples to keep in mind are differential operators on $L^2(\mathbf{R}^n)$, which cannot be defined on all of $L^2(\mathbf{R}^n)$ in any natural way.

Definition. Given two linear operators A and B on H, we say that $A \subset B$ if $D(A) \subset D(B)$ and $A = B|_{D(A)}$.

Definition. Given a densely defined operator A on H we define the adjoint operator $A^\dagger$ on H as follows: an element y in H is in $D(A^\dagger)$ iff there is a z in H such that $< Ax|y > = < x|z >$ for all x in $D(A)$. For such a y we define $A^\dagger y = z$. Remark: if A is a bounded operator on all of H then $D(A^\dagger) = H$ also.

Definition. A densely defined operator A on H is said to be symmetric if $A \subset A^\dagger$. This is equivalent to saying that $< Ax|y > = < x|Ay >$ for all x, y in $D(A)$.

Definition. A densely defined operator A on H is said to be self-adjoint or Hermitian if $A = A^\dagger$, i.e. $A \subset A^\dagger$ and $A^\dagger \subset A$.

1. Show that the following operators are symmetric but not self-adjoint.
   (a) Let $H = L^2([0,1])$ and $D(A) = C^\infty$ functions on $(0,1)$ with compact support, and $Af = f''$ for f in $D(A)$.
   (b) Let H and $D(A)$ be as above, but now take $Af = if'$.
   (c) Let $H = L^2(\Omega)$, where Ω is a bounded open subset of $\mathbf{R}^n$. Take $D(A) = C^\infty$ functions on Ω with compact support, and $Af(x) = \partial_i(a_{ij}(x)\partial_j f)(x)$, where the matrix $a_{ij}(x)$ is Hermitian with C^1 entries.

What is so special about self-adjoint operators? Part of the answer is contained in the following theorem.

Theorem (Stone and von Neumann). *Suppose that for each t in $\mathbf{R}$, $U(t)$ is a unitary operator on H, with $U(0) = 1$. Also assume that $U(t+s) = U(t)U(s)$*

for all s,t and that for each x in H we have $\lim_{t\to 0} U(t)x = x$. (Such a U is called a strongly continuous one-parameter group of unitary transformations, or a continuous symmetry for short.) Define a linear operator A on H as follows: $Ax = -i\lim_{t\to 0}[U(t)x - x]/t$, with $D(A)$ consisting of all x in H for which this limit exists and belongs to H. Then A is a densely defined self-adjoint operator called the infinitesimal generator of U. Conversely if A is a densely defined self-adjoint operator on H, there is a unique continuous symmetry for which A is the infinitesimal generator. (This $U(t)$ is denoted by $\exp itA$.) Furthermore, if $u(t) = e^{itA}u_0$ with u_0 in $D(A)$, then $u(t)$ is in $D(A)$ for all t and is the unique solution to the equation $\dot{u}(t) = iAu(t)$ with $u(0) = u_0$.

Therefore there is a 1-1 correspondence between continuous symmetries and self-adjoint operators A. We will see below that self-adjoint cannot be replaced by symmetric.

2. In each of the problems below a continuous symmetry $U(t)$ of H and a subspace V of H will be given. Let A denote the infinitesimal generator of $U(t)$. In each case show that $V \subset D(A)$ and find the action of A on functions in V.
   (a) $H = L^2(\mathbf{R}^3), U(t)f(x) = f(x + tv)$ for a fixed v in $\mathbf{R}^3$, and V is the C^1 functions with compact support on $\mathbf{R}^3$.
   (b) H and V are as in part (a). Set $U(t)f(x) = f(e^{-tB}x)$ where B is a skew symmetric 3×3 matrix. Also show that $U(t)$ is unitary for each t.
   (c) Let $H = L^2([0, 2\pi])$ and $U(t)f(x) = U_\alpha(t)f(x) = \alpha^{n(x+t)}f(r(x + t))$, where $\alpha \in \mathbf{C}$ with $|\alpha| = 1$ and the functions n and r on $\mathbf{R}$ are defined as follows. For each s in $\mathbf{R}$ let $s = n(s)2\pi + r(s)$ where $n(s)$ is an integer and $0 \le r(s) < 2\pi$. Let $V = V_\alpha = \{f \in C^1([0, 2\pi]) : f(2\pi) = \alpha f(0)\}$. Also show that U_α is a continuous symmetry.

3. Let $H = L^2([0, 2\pi])$ and $Af = -if'$ on $D(A) = C^1$ functions on $(0, 2\pi)$ with compact support. (Note that by part 1b, A is symmetric but not self-adjoint.) For each $\alpha \in \mathbf{C}$ with $|\alpha| = 1$ let A_α denote the infinitesimal generator of U_α defined in problem 2c.
   (a) Show that $A \subset A_\alpha$. So A_α is a self-adjoint extension of A.
   (b) Let $u_0(x)$ be a positive C^∞ function with compact support in $(\pi/2, \pi)$ so that u_0 is in $D(A)$. Show that the unique solution to the partial differential equation $\partial_t u(t, x) = iAu(t, x) = \partial_x u(t, x)$ with $u(0, x) = u_0(x)$ is $u(t, x) = u_0(x + t)$. Notice for $t > 0$ that $u(t, \cdot)$ stays in $D(A)$ as long as $t < \pi/2$. But for t large enough the function $u(t, \cdot)$ is no longer zero at the left end point. This gives an indication of why A being symmetric is not good enough for producing continuous symmetries. For differential operators the choice of domain is essentially a choice of boundary conditions. A choice which makes the differential operator symmetric but not self-adjoint indicates that one has specified boundary conditions which are too restrictive. As a result there is no possible unitary time evolution which will maintain these boundary conditions. This is what happened in this problem. It is an interesting fact that the boundary conditions $f(2\pi) = \alpha f(0)$ with $|\alpha| = 1$ are the only possible boundary conditions which produce a self-adjoint extension of A.

Gaussian measure problems

A basic formula in the study of Gaussian measures is

$$\int_{\mathbf{R}} e^{-ax^2/2}dx = \sqrt{2\pi/a}. \tag{9.22}$$

Definition. The measure μ on $\mathbf{R}^n$ given by

$$d\mu(x) = Z^{-1}e^{-<Ax,x>/2}dx, \tag{9.23}$$

where A is a positive definite matrix and Z is a constant such that $\mu(\mathbf{R}^n) = 1$, is called the mean-zero Gaussian measure on $\mathbf{R}^n$ with covariance $G = A^{-1}$.

4. (a) Find the constant Z.
   (b) Show that $\int_{\mathbf{R}^n} e^{<p,x>}d\mu(x) = e^{<Gp,p>/2}$ for all p in $\mathbf{R}^n$.
   (c) Use analytic continuation to show that $\int_{\mathbf{R}^n} e^{i<p,x>}d\mu(x) = e^{-<Gp,p>/2}$.

   Definition. For any finite measure μ on $\mathbf{R}^n$, the function $\hat{\mu}(p) = \int_{\mathbf{R}^n} e^{i<p,x>}d\mu(x)$ is called the Fourier transform of μ.

   Fact: $\hat{\mu}$ uniquely determines the measure μ.
   (d) For v in $\mathbf{R}^n$ and $f : \mathbf{R}^n \to \mathbf{R}$ "nice", set $\partial_v f(x) = <v, \nabla f(x)>$. Show that

$$\int_{\mathbf{R}^n} \partial_v f(x)d\mu(x) = \int_{\mathbf{R}^n} <Av, x> d\mu(x) \tag{9.24}$$

   for nice f. Use this to show that $\partial_v{}^\dagger = -\partial_v + <Av, \cdot>$.
   (e) Show that $\int_{\mathbf{R}^n} <v, x><w, x> d\mu(x) = <Gv, w>$ for all v, w in $\mathbf{R}^n$.
   (f) (*Wick's theorem.*) Show more generally that

$$\int_{\mathbf{R}^n} <v_1, x> \cdots <v_{2n}, x> d\mu(x) = \sum <Gv_{i_1}, v_{j_1}> \cdots <Gv_{i_n}, v_{j_n}>, \tag{9.25}$$

   where the sum is over all pairings $(i_1, j_1), \ldots, (i_n, j_n)$ of the integers $\{1, 2, \ldots, 2n\}$. Hint: integrate by parts or use part (b).
   (g) Suppose that F and H are bounded measurable functions on $\mathbf{R}^n$ of the form $F(x) = f(\{<v_i, x>\})$ and $H(x) = h(\{<w_j, x>\})$, where the v_i and w_j are a finite collection of vectors in $\mathbf{R}^n$ such that $<Gv_i, w_j> = 0$ for all i, j. Show that

$$\int_{\mathbf{R}^n} F(x)H(x)d\mu(x) = \int_{\mathbf{R}^n} F(x)d\mu(x) \int_{\mathbf{R}^n} H(x)d\mu(x). \tag{9.26}$$

   You may use the fact that F may be approximated in $L^2(d\mu)$ by finite linear combinations of functions of the form $\exp i\sum \alpha_k <v_k, x>$ where the α_k are real numbers.

5. (*Gaussian measures and the heat equation.*) Keep the same notation as in problem 4. Let L be the second order differential operator $L = (1/2)G_{ij}\partial_i\partial_j$. Let e^{tL} be the solution operator to the partial differential equation $u_t = Lu$ (using $L^2(dx)$ functions here). That is, if $u(t,x) = (e^{tL}f)(x)$, where f is a C^2 function on $\mathbf{R}^n$ such that all its derivatives up to second order are in $L^2(dx)$, then $u(t,x)$ is the unique function satisfying $u_t = Lu, u(0,x) = f(x)$, and $u(t,\cdot)$ is in $L^2(dx)$ for each time t. Facts: e^{tL} extends to a bounded operator on $L^2(dx)$ for $t \geq 0$, $e^{tL}(L^2(dx)) \subset C^\infty(\mathbf{R}^n)$ for $t > 0$, and e^{tL} for $t < 0$ is not defined on all of $L^2(dx)$ and is unbounded. Show that

$$(9.27) \qquad e^{tL}f(x) = \int_{\mathbf{R}^n} f(x + t^{1/2}y)d\mu(y) = (p_t * f)(x),$$

where

$$(9.28) \qquad p_t(x) = Z_t^{-1}e^{-<Ax,x>/2t},$$

with Z_t chosen so that $\int_{\mathbf{R}^n} p_t(x)dx = 1$. In particular we conclude that $\int_{\mathbf{R}^n} f(x)d\mu(x) = e^L f(0)$. This formula is useful for computing Gaussian integrals in cases where f is a nice function like a polynomial or an exponential. In these cases $e^{tL}f$ can be computed by using the power series expansion for e^{tL}.

6. (*Normal ordering.*) Normal ordering is a standard operation in QFT. There is a function version of normal ordering and an operator version. In this problem we explore the function version. Fact: loosely speaking, if $t > 0$ and f is a function for which $e^{tL}f$ makes sense, then $e^{tL}f(x)$ is real analytic in x.

Definition. Given a function f on $\mathbf{R}^n$, $:f:(x) = e^{-L}f(x)$ is called the normal ordered version of f.

From the fact above, for this to make sense f must at least be analytic. More precisely f must be of the form $f = e^L g$ for some function g.

(a) Show that $:f:$ is defined if f is a polynomial in x, or if $f(x) = e^{<p,x>}$. Also find $:e^{<p,x>}:$.
(b) Show that $:<v_1,x> \cdots <v_m,x>:= \partial_{\alpha_1} \cdots \partial_{\alpha_m} :e^{\alpha_i <v_i,x>}:$ at $\alpha = 0$ in $\mathbf{R}^m$.
(c) Show that $\partial_v :f:=:\partial_v f:$.
(d) Let M_f denote the multiplication operator $M_f g = fg$. Show that $[e^{tL}, M_{<v,>}] = te^{tL}\partial_{Gv}$ by showing that both sides satisfy the same ordinary differential equation with the same initial condition.
(e) Use parts (c) and (d) to show that $:<v,\cdot> f:=<v,\cdot>:f: -\partial_{Gv} :f:$. A nicer way to write this formula is $:<v,\cdot> f:= \partial^\dagger_{Gv} :f:$. As a corollary of this formula we see that $:\prod_i <v_i,\cdot>:= (\prod_i \partial^\dagger_{Gv})1$.
(f) Let Q_n be the orthogonal projection onto the subspace of $L^2(d\mu)$ orthogonal to the polynomials in x of degree n or less. Show that for $(n+1)$-homogeneous polynomials of the form $f = \prod_{i=1}^{n+1} <v_i,\cdot>$ we have $:f:= f - Q_n f$. This is sometimes how normal ordering is defined. Hint: use the last formula of part (e).

(g) Find a graphical technique for computing integrals of the form $\int_{\mathbf{R}^n} :f_1:\cdots :f_n: d\mu$ where each of the f_i are homogeneous polynomials.
(h) Use part (g) to show that

$$(9.29) \qquad \int_{\mathbf{R}^n} :\prod_{i=1}^k <v_i,\cdot>::\prod_{i=1}^l <w_i,\cdot>: d\mu = \delta_{kl}\sum\prod_{i=1}^k <Gv_i, w_{\sigma(i)}>,$$

where the sum is over all permutations σ of $\{1,2,\ldots,k\}$.

7. In this problem we turn to the operator version of normal ordering.

Notation. Set $a(v) = \partial_{Gv}$ and $a^\dagger(v) = \partial^\dagger_{Gv}$ for each v in $\mathbf{R}^n$. The $a(v)$ are called annihilation operators and the $a^\dagger(v)$ are creation operators. Also set $\phi(v) = a(v) + a^\dagger(v) = M_{<v,>}$, the so-called field operators.

(a) Show that $[a(v), a^\dagger(w)] = <Gv, w> 1$.

Definition. Let O be an operator expressed as a linear combination of products of annihilation and creation operators. The normal ordered version of O, denoted $:O:$, is the operator formed by reordering each term of O so that all creation operators stand to the left of all annihilation operators.

Example. If $O = a(v)a^\dagger(w)$ then $:O:= a^\dagger(w)a(v)$.

Remark. The notation is misleading. The normal ordering operation is not a function on operators but on the formal expression of an operator in terms of creation and annihilation operators. For example by part (a) we can write O of the above example as $O = a^\dagger(w)a(v) + <Gv, w> 1$ which is already in normal ordered form. So unless $<Gv, w> = 0$ we obtain distinct results by normal ordering different expressions of the same operator.

(b) Suppose that A and B are bounded operators on a Hilbert space such that the operator $C = [A,B]$ commutes with both A and B. Show that $e^{A+B} = e^A e^B e^{-[A,B]/2}$. Hint: let $\psi(t) = e^{-tA}e^{t(A+B)}$ and show that $\dot\psi(t) = e^{-tA}Be^{tA}\psi(t)$. Then show that $e^{-tA}Be^{tA} = B - tC$ by differentiation, and solve for ψ.
(c) Suppose that $f(x)$ is an analytic "nice" function on $\mathbf{R}^n$ and that $f(x) = \sum f_\alpha \prod <v_i, x>^{\alpha_i}$, where $v_1,\ldots,v_n$ is a basis for $\mathbf{R}^n$, f_α are complex numbers, and the sum is over all multi-indices $\alpha = (\alpha_1,\ldots,\alpha_n)$ with each α_i a nonnegative integer. Notice that the multiplication operator M_f may be written as $M_f = \sum f_\alpha \prod \phi(v_i)^{\alpha_i}$. Show that $:M_f:= M_{:f:}$ where $:f:$ was defined in the previous problem. Hint: first take $f(x) = e^{<v,x>}$ so that $M_f = e^{\phi(v)}$. Show that $:e^{\phi(v)}:= e^{-<Gv,v>/2}e^{\phi(v)}$ using part (b). Disregard the fact that $a(v)$ and $a^\dagger(v)$ are unbounded operators.

8. (*Relations between $L^2(d\mu)$ and "Fock space" on $\mathbf{R}^n$.*) For notational simplicity we now take $A = I$, so that $G = I$ also. Let $S(\mathbf{R}^n)$ denote the complexified symmetric tensor algebra on $\mathbf{R}^n$. For $v_1,\ldots,v_k$ and $w_1,\ldots,w_l$ in $\mathbf{R}^n$ set $<v_1\cdots v_k, w_1\cdots w_l> = \delta_{kl}\sum\prod <v_i, w_{\sigma(i)}>$, where the sum is over all permutations of $\{1,2,\cdots,k\}$. Extend this form to $S(\mathbf{R}^n)$ by sesquilinearity to get an

inner product. The symmetric Fock space on $\mathbf{R}^n$ is the completion of $S(\mathbf{R}^n)$ with respect to this inner product, and is denoted $F(\mathbf{R}^n)$.

(a) Let $U: F(\mathbf{R}^n) \to L^2(d\mu)$ be given by $U(v_1 \cdots v_k) =: \prod < v_i, \cdot >:$ extended by linearity and continuity to $F(\mathbf{R}^n)$. Show that U is a unitary isomorphism of Hilbert spaces.

(b) Define for v in $\mathbf{R}^n$, $e^v = \sum_{k=0}^{\infty} v^k/k!$ in $F(\mathbf{R}^n)$. Show that $U(e^v) =: e^{<v,\cdot>}:$. This formula could also be used to define U.

(c) Show that $U^{-1}a^\dagger(v_0)U(v_1 \cdots v_k) = v_0 v_1 \cdots v_k$, and that

$$U^{-1}a(v_0)U(w_0 w_1 \cdots w_k) = \sum_{i=0}^{k} < v_0, w_i > w_0 \cdots \hat{w}_i \cdots w_k, \tag{9.30}$$

where the caret means omission of the vector w_i. Compare Equations (8.27,8.28).

9. (*Gaussian measure on infinite dimensional Hilbert space.*) Let H be the Hilbert space of square summable infinite real sequences, $x = (x_1, x_2, \dots)$. Let $< x, y > = \sum_i x_i y_i$ be the inner product on H. Also let V denote the vector space of all real sequences. We say that x in V has finite support if $x_i = 0$ for all sufficiently large i.

Theorem. *There exists a unique measure μ on V such that $\hat{\mu}(p) \equiv \int_V e^{i<p,x>} d\mu(x) = e^{-<p,p>/2}$ for all sequences p of finite support. Informally this is the measure $d\mu(x) = Z^{-1} e^{-<x,x>/2} \prod_{i=1}^{\infty} dx_i$.*

(a) Suppose that $F(x) = f(x_1, \dots, x_n)$ is a function on V depending on only a finite number of variables. Show that $\int_V F(x) d\mu(x) = \int_{\mathbf{R}^n} f(y) d\mu_n(y)$, where μ_n is the Gaussian measure on $\mathbf{R}^n$ with covariance 1.

(b) Let $\alpha \in V$ be a sequence of positive terms. Let $H_\alpha = \{x \in V : \sum \alpha_i^2 x_i^2 < \infty\}$. Prove that $\mu(H_\alpha) = 1$ if $\sum \alpha_i^2 < \infty$ and $\mu(H_\alpha) = 0$ otherwise. Outline:

(i) Notice that $1_{H_\alpha}(x) = \lim_{\epsilon \to 0} \exp -\epsilon^2 \sum_{i=1}^{\infty} \alpha_i^2 x_i^2$, so that

$$\mu(H_\alpha) = \lim_{\epsilon \to 0} \int_V \exp(-\epsilon^2 \sum_{i=1}^{\infty} \alpha_i^2 x_i^2) d\mu(x) \tag{9.31}$$

by the dominated convergence theorem.

(ii) Again by the dominated convergence theorem,

$$\int_V \exp(-\epsilon^2 \sum_{i=1}^{\infty} \alpha_i^2 x_i^2) d\mu(x) = \lim_{n \to \infty} \int_V \exp(-\epsilon^2 \sum_{i=1}^{n} \alpha_i^2 x_i^2) d\mu(x). \tag{9.32}$$

(iii) Compute this last integral explicitly.

Remark. From part (b) above we see that $\mu(H) = 0$! This can also be seen from the strong law of large numbers, which implies that $\lim_{n \to \infty} \sum_{i=1}^{n} x_i^2/n = 1$ for almost every x in V with respect to μ.

(c) For v in H let $v^n = (v_1, v_2, \dots, v_n, 0, 0, \dots)$. Show that $\lim_{n \to \infty} < v^n, \cdot >$ exists in $L^2(d\mu)$. We abuse notation and denote the limit as $< v, \cdot >$ which is an $L^2(d\mu)$ almost everywhere defined function on V.

(d) Show that for v, w in H, $\int_V < v, x >< w, x > d\mu(x) = < v, w >$ and $\int_V e^{i<v,x>} d\mu(x) = e^{-<v,v>/2}$.

Remark. Because of the above result, even though $\mu(H) = 0$ the function $< v, \cdot >$ is still a well-defined element of $L^2(d\mu)$ as an almost everywhere defined function on V. Furthermore the usual Gaussian integral formulas still hold. Because of this most of what was done in Exercises 4 – –8 above is valid for any Hilbert space H provided $< v, \cdot >$ is interpreted properly.

10. In this problem we study Wiener measure which is described informally by

$$d\mu(x) = Z^{-1} \exp\left[-\frac{1}{2}\int_0^1 x'(s)^2 ds\right] \prod_{0 \le s \le 1} dx(s), \tag{9.33}$$

where the measure μ is on continuous functions on $[0, 1]$ such that $x(0) = 0$, and Z is a normalization constant so that μ has total weight unity.

(a) Set $< f, g > = \int_0^1 f(s) g(s) ds$ and $|f|^2 = < f, f >$ for f and g functions on $[0, 1]$. Do an informal computation by completing the square to show that

$$\int e^{<f,x'>} d\mu(x) = e^{-|f|^2/2}. \tag{9.34}$$

(b) Let h be a continuous function and set $f(s) = -\int_s^1 h(u) du$, so that $f' = h$ and $f(1) = 0$. Use this f in part (a) and do an integration by parts to conclude that

$$\int e^{<h,x>} d\mu(x) = e^{<Gh,h>/2}, \tag{9.35}$$

where $Gh(s) \equiv \int_0^1 \min(s, u) h(u) du$.

Remark. The function $\min(s, u)$ is the Green's function for the differential operator $A = -d^2/ds^2$ with a Dirichlet boundary condition at 0 and a Neumann boundary condition at 1. The functions $u_n(s) = C_n \sin(n\pi + \pi/2)s$ for $n \ge 0$ form an orthonormal basis of eigenfunctions of A with eigenvalues $\lambda_n = (n\pi + \pi/2)^2$. From this information it follows from the previous problem that $\mu(L^2(ds)) = 1$ but $\mu(L_1^2(ds)) = 0$, where $L_1^2(ds)$ denotes the functions on $[0, 1]$ with a derivative in L^2. N. Wiener proved the following stronger result.

Theorem. *There is a unique probability measure μ on $V = \{x \in C([0, 1]) : x(0) = 0\}$ such that*

$$\int e^{<h,x>} d\mu(x) = e^{-<Gh,h>/2} \tag{9.36}$$

for all continuous functions h on $[0,1]$.

(The continuous functions $x(s)$ are often called Brownian or Wiener paths.) Since this is the prototypical Euclidean path integral, we will show more directly that the differentiable functions have measure zero. This will also serve to introduce you to stochastic calculus.

(c) Set $h(s) = \sum_{i=1}^{n} h_i \delta(s - s_i)$ with $s_i \in [0,1]$ and h_i real. Plug this h into Eq. (9.36) above to conclude that $[x(s_1), \ldots, x(s_n)]$ is distributed as the Gaussian measure on $\mathbf{R}^n$ with covariance $G_{ij} = \min(s_i, s_j)$.

(d) Let $P = \{0 = s_0 < s_1 < \cdots < s_n = s\}$ be a partition of the interval $[0,s]$. Define the function $V_P(x) = \sum_{i=1}^{n} [x(s_i) - x(s_{i-1})]^2$ for x in V. Show that $V_P(x) \to s$ in $L^2(d\mu)$ as the mesh of P goes to zero. This says that the quadratic variation of a typical path up to s is s. Show that this cannot happen if the path is C^1.

Remark. Part (d) can be written suggestively as $dx(s)^2 = ds$, whereas in ordinary calculus one has $dx(s)^2 = 0$. As a result the calculus formula $df(x(s)) = f'(x(s))dx(s)$ must now be replaced by $df(x(s)) = f'(x(s))dx(s) + (1/2)f''(x(s))ds$ for Brownian paths. This is called Ito's formula and is one of the basic ingredients of stochastic calculus. Of course the formula must be interpreted properly.

Miscellaneous physical problems

11. (*Special Relativity.*)

(a) Show that the linear transformation $x' = \Lambda x$, with

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{9.37}$$

with $|\beta| < 1, \gamma = (1-\beta^2)^{-1/2}$ is a proper Lorentz transformation, and that its inverse is obtained by replacing β by $-\beta$. Show that the spatial (x', y', z') coordinate axes are parallel to the (x, y, z) axes, and that the origin of the primed coordinate system is observed in the unprimed system to move along the positive x axis at velocity β. This Lorentz transformation is called a boost along the positive x axis at velocity β.

(b) (*Time dilation.*) In the primed coordinate system, two events are observed to occur at the same location but separated by the time interval T. Show that in the unprimed system they are separated by the time interval γT.

(c) (*Relativity of simultaneity.*) In the primed coordinate system, two events are observed to occur simultaneously at points on the x' axis a distance L apart. Show that in the unprimed system these events are separated by the time interval $\gamma\beta L$. Which one occurs first? Conversely, suppose that in the unprimed system two events occur at points on the x axis, separated by a distance L and a time interval T with $T^2 - L^2 < 0$. For such "spacelike separated" events, show that β can be chosen to make the events simultaneous in the primed frame.

(d) (*Length contraction.*) A ruler is at rest in the primed system, extending along the x' axis from the origin to $x' = L$. To measure its length, an observer in the unprimed system will note the location of its endpoints at some fixed time (her own t coordinate!) and measure the distance between them. Show that she will obtain L/γ.

12. (*The Lorentz group and $SL(2,\mathbf{C})$.*)
As in Eq. (3.1), we associate to each vector x^μ in $\mathbf{R}^4$ the 2×2 Hermitian matrix

$$X = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}. \tag{9.38}$$

Now define a map from $SL(2,\mathbf{C})$ to the proper Lorentz group by sending $A \in SL(2,\mathbf{C})$ to the Lorentz transformation $X' = AXA^\dagger$. Show that this is in fact a surjective homomorphism with kernel $\{\pm 1\}$.

13. (*Heisenberg's uncertainty principle.*)
Let $|\psi>$ be any vector in $L^2(\mathbf{R})$, and x and p the usual quantum-mechanical operators satisfying $[x,p] = i$. Then the Hermitian operators $\tilde{x} \equiv x - <\psi|x|\psi>$ and $\tilde{p} \equiv p - <\psi|p|\psi>$ measure the deviations of x and p from their mean values in the state $|\psi>$. Note that $[\tilde{x}, \tilde{p}] = i$. Now define $(\Delta x)^2 \equiv <\psi|\tilde{x}^2|\psi>$ and $(\Delta p)^2 \equiv <\psi|\tilde{p}^2|\psi>$, which measure the mean square deviations of x and p from their average values. Use Schwarz' inequality to prove that

$$(\Delta x)^2 (\Delta p)^2 \geq |<\psi|\tilde{x}\tilde{p}|\psi>|^2. \tag{9.39}$$

By writing $\tilde{x}\tilde{p}$ as the sum of Hermitian and skew Hermitian parts, deduce the uncertainty principle,

$$(\Delta x)^2 (\Delta p)^2 \geq 1/4, \tag{9.40}$$

or, with physical units restored,

$$(\Delta x)^2 (\Delta p)^2 \geq \hbar^2/4. \tag{9.41}$$

Show that the ground state of the simple harmonic oscillator is actually a minimum uncertainty state, with equality in this formula.

14. (*Harmonic oscillator Green's function.*) Use our solution of the simple harmonic oscillator in terms of creation and annihilation operators to compute the Green's function

$$G(t - t') = im <0|T[x(t)x(t')]|0>, \tag{9.42}$$

thus verifying that analytic continuation from the Euclidean space path integral does select the correct Green's function. Also show directly from the Heisenberg

equation of motion for $x(t)$ that the Green's function must satisfy

$$(\frac{d^2}{dt^2} + \omega^2)G(t-t') = \delta(t-t'). \tag{9.43}$$

15. (*Angular momentum and $SO(3)$.*) The Lagrangian $L = \frac{1}{2}m|\dot{\mathbf{x}}|^2 - V(|\mathbf{x}|)$ describes a particle moving in $\mathbf{R}^3$ in a spherically symmetric, or central, potential — one which depends only on the particle's distance from the origin. It is invariant under the usual action of the rotation group $SO(3)$ by 3×3 orthogonal matrices multiplying $\mathbf{x}$.
    (a) Show that the conserved quantities associated by Noether's theorem to the generators of rotations about the x, y, and z axes are just the components of the angular momentum vector $\mathbf{L} = \mathbf{x} \times \mathbf{p}$:

$$L_1 = L_x = yp_z - zp_y, \tag{9.44}$$
$$L_2 = L_y = zp_x - xp_z, \tag{9.45}$$
$$L_3 = L_z = xp_y - yp_x, \tag{9.46}$$

    where $\mathbf{p} = m\dot{\mathbf{x}}$.
    (b) In quantum mechanics, where $[x_j, p_k] = i\delta_{jk}$, verify that $[L_j, x_k] = i\epsilon_{jkl}x_l$, $[L_j, p_k] = i\epsilon_{jkl}p_l$, and $[L_j, L_k] = i\epsilon_{jkl}L_l$, where ϵ_{jkl} is ± 1 according as jkl is an even or odd permutation of 123, and is zero if j, k, l are not all distinct. This type of commutation relation with L_j is characteristic of any triple of operators which transform like the components of a vector under rotations. Note that $v_k \to v_k + (i\theta)i\epsilon_{jkl}v_l$ is precisely the change in the kth component of a vector under an infinitesimal rotation by angle θ about the x_j axis.
    (c) Verify that $\mathbf{L}$ is indeed conserved in quantum mechanics by checking that $[H, L_j] = 0$, where the Hamiltonian is $H = |\mathbf{p}|^2/2m + V(|\mathbf{x}|)$. You will find it useful that $[f(\mathbf{x}), p_j] = i\partial_j f(\mathbf{x})$.
    (d) $SO(3)$ also acts on $L^2(\mathbf{R}^3)$ by $Af = f \circ A^{-1}$, where $A \in SO(3), f \in L^2(\mathbf{R}^3)$. Show that the L_j are the generators of this representation by showing that, e.g., $1 - i\theta L_z$ acts as an infinitesimal rotation by angle θ about the z axis.

16. (*The hydrogen atom.*)

    Warning: this problem involves a great deal of tedious calculation! You don't have to do it all!

    The hydrogen atom consists of an electron of mass m and charge e moving in the central electrostatic field of a proton, which we assume to be fixed at the origin. The Hamiltonian for the electron is

$$H = \frac{|\mathbf{p}|^2}{2m} - \frac{e^2}{|\mathbf{x}|}. \tag{9.47}$$

    H acting in $L^2(\mathbf{R}^3)$ has a discrete spectrum of negative eigenvalues giving the atomic energy levels which we will compute. (There is also a continuous spectrum of positive eigenvalues which describe a free electron coming in from infinity with

kinetic energy equal to the eigenvalue, scattering off the proton's electric field, and moving out to infinity again. These will not concern us.)
    (a) H has a large symmetry algebra. In addition to the conserved angular momentum, $[H, \mathbf{L}] = 0$, there is also the *Runge-Lenz vector* $\mathbf{R}$, the Hermitian operator

$$\mathbf{R} = \frac{1}{2m}(\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) - \frac{e^2}{|\mathbf{x}|}\mathbf{x}, \tag{9.48}$$

    for example,

$$R_z = \frac{1}{m}(zp_x^2 + zp_y^2 - xp_xp_z - yp_yp_z + ip_z) - e^2z(x^2+y^2+z^2)^{-1/2}. \tag{9.49}$$

    Show that $[H, \mathbf{R}] = 0$. In classical mechanics, the electron would follow an elliptical orbit with the origin at one focus, and this vector points from the origin to the nearer vertex of the ellipse. Its length is proportional to the eccentricity of the orbit.
    (b) Show further that

$$\mathbf{R} \cdot \mathbf{L} = \mathbf{L} \cdot \mathbf{R} = 0, \tag{9.50}$$

$$[L_j, R_k] = i\epsilon_{jkl}R_l, \tag{9.51}$$

$$|\mathbf{R}|^2 = e^4 + \frac{2}{m}H(|\mathbf{L}|^2 + 1), \tag{9.52}$$

$$[R_j, R_k] = -\frac{2i}{m}H\epsilon_{jkl}L_l. \tag{9.53}$$

    (c) Let $D \subset L^2(\mathbf{R}^3)$ be the span of the eigenvectors of H having negative eigenvalues. We define a Hermitian operator $\mathbf{K}$ with domain D by

$$\mathbf{K} = (-\frac{m}{2H})^{1/2}\mathbf{R}. \tag{9.54}$$

    Further defining $\mathbf{M} = (\mathbf{L} + \mathbf{K})/2$, $\mathbf{N} = (\mathbf{L} - \mathbf{K})/2$, show that

$$[M_j, M_k] = i\epsilon_{jkl}M_l, \tag{9.55}$$

$$[N_j, N_k] = i\epsilon_{jkl}N_l, \tag{9.56}$$

$$[M_j, N_k] = 0, \tag{9.57}$$

$$|\mathbf{M}|^2 = |\mathbf{N}|^2 = (|\mathbf{L}|^2 + |\mathbf{K}|^2)/4, \tag{9.58}$$

$$H = -\frac{me^4}{2(4|\mathbf{M}|^2 + 1)}. \tag{9.59}$$

(d) According to part (c), the operators $\mathbf{M}$ and $\mathbf{N}$ acting in D generate a (unitary) representation of the Lie algebra $su(2) \times su(2)$. Suppose we decompose D as the direct sum of irreducible representations of this algebra. The Casimir operators $|\mathbf{M}|^2$ and $|\mathbf{N}|^2$ commute with the generators and are therefore constant on each irreducible representation by Schur's lemma. Therefore, again by (c), H is also constant on each irrep, so these are the eigenspaces of H. Only those irreps having $|\mathbf{M}|^2 = |\mathbf{N}|^2$ occur, and the value of each of these Casimir operators is $j(j+1)$ in the $2j+1$-dimensional irrep of $su(2)$ having highest weight $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$. Defining the *principal quantum number* $n = 2j + 1 = 1, 2, \ldots$, we deduce that the irrep with highest weights (j, j) is an n^2-dimensional eigenspace of H with eigenvalue $-me^4/2n^2$. In fact, each possible value of n occurs in the decomposition of D, exactly once. To prove this one must construct operators like the $a^\dagger$ and a of the harmonic oscillator which raise and lower the eigenvalue of H. This is even more tedious than what you've already been through. Including these new operators enlarges the $su(2) \times su(2)$ symmetry of the hydrogen atom to $so(4, 2)$. See, if you like, [**19**]. (With physical units restored, the energies of the stationary states of the hydrogen atom are $-me^4/2\hbar^2 n^2 = -13.6/n^2$ electron volts. The ground state energy of -13.6 eV can be measured directly as the negative of the energy required to ionize the atom, and of course the energy differences between states can be measured with extreme precision from the frequencies of photons emitted during a transition between the states. This precision is more than sufficient to measure the corrections to the energies you've derived, arising from such effects as the motion of the proton, the spin of the electron, and the non-instantaneous transmission of the electric force between them, but that's another story.)

17. (*Feynman diagrams.*) Draw the Feynman diagrams contributing to the four point function or vertex function $< 0|T[\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)]|0 >$ in the $\lambda\phi^4$ scalar field theory up to order λ^2. Write down the corresponding terms of the perturbation series as well as their Fourier transforms. How divergent are the integrals?

BIBLIOGRAPHY

1. V. Arnold, *Mathematical Methods of Classical Mechanics*, second edition, Springer-Verlag 1989.
2. H. Goldstein, *Classical Mechanics*, second edition, Addison-Wesley 1980.
3. A.Z. Capri, *Nonrelativistic Quantum Mechanics*, Benjamin-Cummings 1985.
4. R. Shankar, *Principles of Quantum Mechanics*, Plenum Press 1980.
5. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill 1965.
6. P. Ramond, *Field Theory: A Modern Primer*, second edition, Addison-Wesley 1989.
7. L.H. Ryder, *Quantum Field Theory*, Cambridge University Press 1985.
8. R.J. Rivers, *Path Integral Methods in Quantum Field Theory*, Cambridge University Press 1987.
9. J. Collins, *Renormalization*, Cambridge University Press 1984.
10. J. Glimm and R. Jaffe, *Quantum Physics*, second edition, Springer-Verlag 1987.
11. N.N. Bogolubov, A.A. Logunov, and R.T. Todorov, *Introduction to Axiomatic Quantum Field Theory*, Reading: Benjamin 1975.
12. I.M. Singer, "Some remarks on the Gribov ambiguity," Commun. Math. Phys. **60**, 7–12 (1978).
13. M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems*, Princeton University Press 1992.
14. D. Nemeschansky, C. Preitschopf, and M. Weinstein, "A BRST primer," Ann. Phys. **183:2**, 226–268 (1988).
15. B. Kostant and S. Sternberg, "Symplectic reduction, BRS cohomology, and infinite-dimensional Clifford algebras," Ann. Phys. **176:1**, 49–113 (1987).
16. E. Farhi and S. Gutmann, "The functional integral constructed directly from the Hamiltonian", MIT preprint CTP#1943, 1991, submitted to Nucl. Phys. B.
17. W.E. Caswell and A.D. Kennedy, "Asymptotic behavior of Feynman integrals: Convergent integrals", Phys. Rev. **D28**, 3073–3089 (1983).
18. J. Polchinski, "Renormalization and effective Lagrangians", Nucl. Phys. **B231**, 269–295 (1984).
19. B.G. Wybourne, *Classical Groups for Physicists*, Wiley, New York 1974.