M.Phys Option in Theoretical Physics: C6

John Chalker and Andre Lukas
Physics Department, Oxford University

2010-2011

Chapter 1

Path Integrals in Quantum Mechanics

1.1 Mathematical Background

In this first part of the chapter, we will introduce *functionals* which are one of the main tools in modern theoretical physics and explain how to differentiate and integrate them. In the second part, these techniques will be applied to re-formulate quantum mechanics in terms of functional integrals (path integrals).

1.1.1 Functionals

What is a *functional*? Let us first consider a real function $f : [a, b] \to \mathbb{R}$, mapping all elements $x \in [a, b]$ in the domain $[a, b]$ to real numbers $f(x)$. A functional is similar to a function in that it maps all elements in a certain domain to real numbers, however, the nature of its domain is very different. Instead of acting on all points of an interval or some other subset of the real numbers, the domain of functionals consists of (suitably chosen) classes of functions itself. In other words, given some class $\{f\}$ of functions, a functional $F : \{f\} \to \mathbb{R}$ assigns to each such f a real number $F[f]$.

Differentiation and integration are the two main operations on which the calculus of (real or complex) functions is based. We would like to understand how to perform analogous operations on functionals, that is, we would like to make sense of the expressions

$$\frac{\delta F[f]}{\delta f} , \qquad \int \mathcal{D}f \, F[f] \tag{1.1}$$

for functional differentiation and integration. Here, we denote a functional derivative by $\delta/\delta f$, replacing the ordinary d/dx or $\partial/\partial x$ for functions, and the measure of a functional integral by $\mathcal{D}f$, replacing dx in the case of integrals over functions. A common application of ordinary differentiation is to find the extrema of a function f by solving the equation $f'(x) = 0$. Of course, one can also ask about the extrema of a functional F, that is the functions f which minimise (or maximise) the value $F[f]$. One important application of functional differentiation which we will discuss is how these extrema can be obtained as solutions to the equation

$$\frac{\delta F[f]}{\delta f} = 0 . \tag{1.2}$$

To make our discussion more concrete we will now introduce two (well-known) examples of functionals.

Example: distance between two points

A very simple functional F consists of the map which assigns to all paths between two fixed points the length of the path. To write this functional explicitly, let us consider a simple two-dimensional situation in the (x, y) plane and choose two points (x_1, y_1) and (x_2, y_2). We consider all functions $\{f\}$ defining paths between those two points, that is functions on the interval $[x_1, x_2]$ satisfying $f(x_1) = y_1$ and $f(x_2) = y_2$. The length of a path is then given by the well-known expression

$$F[f] = \int_{x_1}^{x_2} dx \, \sqrt{1 + f'(x)^2} . \tag{1.3}$$

What is the minimum of this functional? Of course, we know that the correct answer is "a line", but how do we prove this formally? Let us approach this problem in a somewhat pedestrian way first before re-formulating it in

terms of functional derivatives along the lines of Eq. (1.2). Consider a small but otherwises abritrary perturbation ϵ of a path f which vanishes at the endpoints, that is, which satisfies $\epsilon(x_1) = \epsilon(x_2) = 0$. We can then define the function

$$l(\lambda) \equiv F[f + \lambda\epsilon] = \int_{x_1}^{x_2} dx \, \sqrt{1 + (f'(x) + \lambda\epsilon'(x))^2} \, , \tag{1.4}$$

where λ is a real parameter. A necessary condition for f to be a minimum of the functional F is then

$$\frac{dl}{d\lambda}(0) = 0 \tag{1.5}$$

Hence, this simple trick has reduced our functional minimisation problem to an ordinary one for the function l. The derivative of l at $\lambda = 0$ can be easily worked out by differentiating "under the integral" in Eq. (1.4). One finds

$$\frac{dl}{d\lambda}(0) = \int_{x_1}^{x_2} dx \, \frac{f'(x)\epsilon'(x)}{\sqrt{1 + f'(x)^2}} \, . \tag{1.6}$$

The derivative on ϵ in this expression can be removed by partial integration, keeping in mind that the boundary term vanishes due to ϵ being zero at x_1 and x_2. This results in

$$\frac{dl}{d\lambda}(0) = -\int_{x_1}^{x_2} dx \, \epsilon(x) \frac{d}{dx} \left[\frac{f'(x)}{\sqrt{1 + f'(x)^2}} \right] \, . \tag{1.7}$$

From Eq. (1.5) this integral needs to vanish and given that ϵ is an arbitrary function this is only the case if the integrand is zero pointwise in x. This leads to the differential equation

$$\frac{d}{dx} \left[\frac{f'(x)}{\sqrt{1 + f'(x)^2}} \right] = 0 \, , \tag{1.8}$$

for the function f. The desired extrema of the length functional F must be among the solutions to this differential equation which are given by the lines

$$f'(x) = \text{const} \, . \tag{1.9}$$

Note that the differential equation (1.8) is second order and consequently its general solution has two integration constant. They are precisely "used up" by implementing the boundary conditions $f(x_1) = y_1$ and $f(x_2) = y_2$, so that we remain with a unique solution, the line between our two chosen points.

Example: action in classical mechanics

In physics, a very important class of functionals are *action functionals*. Let us recall their definition in the context of classical mechanics. Start with n generalised coordinates $\mathbf{q}(t) = (q_1(t), \ldots, q_n(t))$ and a Lagrangian $L = L(\mathbf{q}, \dot{\mathbf{q}})$. Then, the action functional $S[\mathbf{q}]$ is defined by

$$S[\mathbf{q}] = \int_{t_1}^{t_2} dt \, L(\mathbf{q}(t), \dot{\mathbf{q}}(t)) \, . \tag{1.10}$$

It depends on classical paths $\mathbf{q}(t)$ between times t_1 and t_2 satisfying the boundary conditions $\mathbf{q}(t_1) = \mathbf{q}_1$ and $\mathbf{q}(t_2) = \mathbf{q}_2$. *Hamilton's principle* states that the classical path of the system minimises the action functional S. To work out the implications of Hamilton's principle we follow the same steps as in the previous example. We define the function

$$l(\lambda) \equiv S[\mathbf{q} + \lambda\boldsymbol{\epsilon}] = \int_{t_1}^{t_2} dt \, L(\mathbf{q} + \lambda\boldsymbol{\epsilon}, \dot{\mathbf{q}} + \lambda\dot{\boldsymbol{\epsilon}}) \, . \tag{1.11}$$

where we have introduced small but arbitrary variations $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ of the coordinates, satisfying $\boldsymbol{\epsilon}(t_1) = \boldsymbol{\epsilon}(t_2) = 0$. As before, we work out the derivative of l at $\lambda = 0$

$$\frac{dl}{d\lambda}(0) = \int_{t_1}^{t_2} dt \, \left[\frac{\partial L}{\partial q_i}(\mathbf{q}, \dot{\mathbf{q}}) \, \epsilon_i + \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}, \dot{\mathbf{q}}) \, \dot{\epsilon}_i \right] \tag{1.12}$$

$$= \int_{t_1}^{t_2} dt \, \epsilon_i \left[\frac{\partial L}{\partial q_i}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}, \dot{\mathbf{q}}) \right] \, , \tag{1.13}$$

where the last step follows from partial integration (and the boundary term vanishes due to $\boldsymbol{\epsilon}(t_1) = \boldsymbol{\epsilon}(t_2) = 0$ as previously). The ϵ_i vary independently and, hence, for the above integral to vanish the bracketed expression in (1.13) must be zero for each index value i. This gives rise to the *Euler-Lagrange equations*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{\partial L}{\partial q_i}(\mathbf{q}, \dot{\mathbf{q}}) = 0 \; . \tag{1.14}$$

The solutions to these equations represent the classical paths of the system which minimise the action functional. They contain $2n$ integration constants which are fixed by the boundary conditions $\mathbf{q}(t_1) = \mathbf{q}_1$ and $\mathbf{q}(t_2) = \mathbf{q}_2$.

1.1.2 Functional differentiation

We have seen that, for the above examples, the problem of varying (or differentiating) functionals can be reduced to ordinary differentiation "under the integral". While this approach is transparent it is not particularly convenient for calculations. We would, therefore, like to introduce a short-hand notation for the variation procedure based on the functional derivative $\delta/\delta f(x)$ so that the minimum of a functional F can be obtained from the equation $\delta F[f]/\delta f(x) = 0$. The crucial step in our previous variation calculations was the ordinary differentiation with respect to the parameter λ. In order to reproduce the effect of this differentiation our functional derivative should certainly satisfy all the general properties of a derivative, that is, it should be linear, it should satisfy the product and chain rules of differentiation and we should be able to commute it with the integral. Let us see how far we can get with those assumptions, starting with our first example (1.3).

$$\frac{\delta F[f]}{\delta f(x)} = \frac{\delta}{\delta f(x)} \int_{x_1}^{x_2} d\tilde{x}\, \sqrt{1 + f'(\tilde{x})^2} = \int_{x_1}^{x_2} d\tilde{x}\, \frac{f'(\tilde{x})}{\sqrt{1 + f'(\tilde{x})^2}} \frac{\delta f'(\tilde{x})}{\delta f(x)} \tag{1.15}$$

Our next step has been partial integration and in order to be able to carry this out we need to assume that ordinary and functional derivative commute. Then we find

$$\frac{\delta F[f]}{\delta f(x)} = -\int_{x_1}^{x_2} d\tilde{x}\, \frac{d}{d\tilde{x}} \left[\frac{f'(\tilde{x})}{\sqrt{1 + f'(\tilde{x})^2}} \right] \frac{\delta f(\tilde{x})}{\delta f(x)} \tag{1.16}$$

In the bracket we recognise the desired left-hand-side of the differential equation (1.8). Our last step consisted of removing the integral due to the presence of an arbitrary variation ϵ in the integrand. Here, we can formally reproduce this step by demanding the relation

$$\frac{\delta f(\tilde{x})}{\delta f(x)} = \delta(\tilde{x} - x) \; , \tag{1.17}$$

which can be viewed as a continuous version of the equation $\partial q_i/\partial q_j = \delta_{ij}$ for a set of n coordinates q_i. Using the relation (1.17) in Eq. (1.16) we finally obtain the desired result

$$\frac{\delta F[f]}{\delta f(x)} = -\frac{d}{dx} \left[\frac{f'(x)}{\sqrt{1 + f'(x)^2}} \right] \; . \tag{1.18}$$

To summarise, functional derivation $\delta/\delta f(x)$ can be understood as a linear operation which satisfies the product and chain rules of ordinary differentiation, commutes with ordinary integrals and derivatives and is subject to the "normalisation" (1.17).

Armed with this "definition", let us quickly re-consider our second example, Hamilton's principle of classical mechanics. We find

$$\begin{aligned}
\frac{\delta S[\mathbf{q}]}{\delta q_i(t)} &= \frac{\delta}{\delta q_i(t)} \int_{t_1}^{t_2} d\tilde{t}\, L(\mathbf{q}(\tilde{t}), \dot{\mathbf{q}}(\tilde{t})) && (1.19)\\
&= \int_{t_1}^{t_2} d\tilde{t} \left[\frac{\partial L}{\partial q_j}(\mathbf{q}, \dot{\mathbf{q}}) \frac{\delta q_j(\tilde{t})}{\delta q_i(t)} + \frac{\partial L}{\partial \dot{q}_j}(\mathbf{q}, \dot{\mathbf{q}}) \frac{\delta \dot{q}_j(\tilde{t})}{\delta q_i(t)} \right] && (1.20)\\
&= \int_{t_1}^{t_2} d\tilde{t} \left[\frac{\partial L}{\partial q_j}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{d\tilde{t}} \frac{\partial L}{\partial \dot{q}_j}(\mathbf{q}, \dot{\mathbf{q}}) \right] \frac{\delta q_j(\tilde{t})}{\delta q_i(t)} && (1.21)\\
&= \int_{t_1}^{t_2} d\tilde{t} \left[\frac{\partial L}{\partial q_j}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{d\tilde{t}} \frac{\partial L}{\partial \dot{q}_j}(\mathbf{q}, \dot{\mathbf{q}}) \right] \delta_{ij} \delta(\tilde{t} - t) = \frac{\partial L}{\partial q_i}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}, \dot{\mathbf{q}}) && (1.22)
\end{aligned}$$

where, in the second last step, we have used the obvious generalisation

$$\frac{\delta q_j(\tilde{t})}{\delta q_i(t)} = \delta_{ij}\delta(\tilde{t}-t) \tag{1.23}$$

of Eq. (1.17).

1.1.3 Functional integration

As before, we consider a class of functions $\{f\}$, $f = f(x)$, defined on a specific domain $\{x\}$ and a functional $F[f]$. Here, we slightly generalise our previous set-up and allow x to represent a set of coordinates in various numbers of dimensions. For example, in the context of (quantum) mechanics, x is one-dimensional and represents time, so that we are dealing with functions f of one real variable. In (quantum) field theory, on the other hand, x stands for a four-vector x^μ. As we will see, most of the discussion in this subsection is independent of the space-time dimension, so we do not need to be more specific until later.

We would now like to introduce a functional integral (1.1) over F. While ordinary integrals have an integration range which may consist of an interval in the real numbers, the integration range for functional integrals is a whole class of functions. Such integrals are not easy to define and, in these notes, we will make no attempt at mathematical rigour. Instead, we will concentrate on the features of functional integrals relevant to our applications in quantum mechanics and quantum field theory.

To "define" functional integrals we can take a lead from the definition of ordinary integrals. There, one introduces a grid (or lattice) of points $\{x_i\}$ with separation Δx in space-time and approximates the function $f = f(x)$ by the collection of its values $\mathbf{y} = (y_1, \ldots, y_n)$, where $y_i = f(x_i)$, at the lattice points. The conventional integral over f (in the one-dimensional case) can then be defined as the $\Delta x \to 0$ limit of $\Delta x \sum_i y_i$. For functional integrals we have to integrate over all functions f and, hence, in the discrete case, over all possible vectors $\mathbf{y}$. This suggest, we should "approximate" the functional integral measure $\int \mathcal{D}f$ by multiple ordinary integrals $\int \Pi_i dy_i = \int d^n\mathbf{y}$. We also need a discrete version $\tilde{F}$ of the functional such that $\tilde{F}(\mathbf{y}) \sim F[f]$ for $\mathbf{y} \sim f$. Then, we can formally introduce functional integrals by setting

$$\int \mathcal{D}f\, F[f] \sim \lim_{\Delta x\to 0} \int d^n\mathbf{y}\, \tilde{F}(\mathbf{y})\ . \tag{1.24}$$

In practise, this relation means we can often represent functional integrals by multiple ordinary integrals or, at least, deduce properties of functional integrals from those of multiple integrals.

In general, functional integrals are prohibitively difficult to evaluate and often a lattice calculation (that is, the numerical evaluation of the RHS of Eq. (1.24) on a computer) is the only option. However, there is one class of integrals, namely *Gaussian integrals* and integrals closely related to Gaussian ones, which is accessible to analytical computations. It turns out that such Gaussian functional integrals are of central importance in physics and we will focus on them in the remainder of this section. In keeping with the spirit of Eq. (1.24), the discussion will center on ordinary multiple Gaussian integrals from which we deduce properties of Gaussian functional integrals.

Gaussian integral over a single variable

As a reminder, we start with a simple one-dimensional Gaussian integral over a single variable y. It is given by

$$I(a) \equiv \int_{-\infty}^{\infty} dy\, \exp(-\frac{1}{2}ay^2) = \sqrt{\frac{2\pi}{a}}\ , \tag{1.25}$$

where a is real and positive. The standard proof of this relation involves writing $I(a)^2$ as a two-dimensional integral over y_1 and y_2 and then introducing two-dimensional polar coordinates $r = \sqrt{y_1^2+y_2^2}$ and φ. Explicitly,

$$I(a)^2 = \int_{-\infty}^{\infty} dy_1\, \exp(-\frac{1}{2}ay_1^2) \int_{-\infty}^{\infty} dy_2\, \exp(-\frac{1}{2}ay_2^2) = \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2\, \exp(-\frac{1}{2}a(y_1^2+y_2^2)) \tag{1.26}$$

$$= \int_0^{2\pi} d\varphi \int_0^{\infty} dr\, r \exp(-\frac{1}{2}ar^2) = \frac{2\pi}{a}\ . \tag{1.27}$$

Multidimensional Gaussian integrals

Next we consider n-dimensional Gaussian integrals

$$W_0(\mathbf{A}) \equiv \int d^n\mathbf{y}\, \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y}\right) , \tag{1.28}$$

over variables $\mathbf{y} = (y_1, \ldots, y_n)$, where $\mathbf{A}$ is a symmetric, positive definite matrix. This integral can be easily reduced to a product of one-dimensional Gaussian integrals by diagonalising the matrix $\mathbf{A}$. Consider an orthogonal rotation $\mathbf{O}$ such that $\mathbf{A} = \mathbf{O}\mathbf{D}\mathbf{O}^T$ with a diagonal matrix $\mathbf{D} = \mathrm{diag}(a_1, \ldots, a_n)$. The Eigenvalues a_i are strictly positive since we have assumed that $\mathbf{A}$ is positive definite. Introducing new coordinates $\tilde{\mathbf{y}} = \mathbf{O}\mathbf{y}$ we can write

$$\mathbf{y}^T\mathbf{A}\mathbf{y} = \tilde{\mathbf{y}}^T\mathbf{D}\tilde{\mathbf{y}} = \sum_{i=1}^{n} a_i\tilde{y}_i^2 , \tag{1.29}$$

where the property $\mathbf{O}^T\mathbf{O} = \mathbf{1}$ of orthogonal matrices has been used. Note further that the Jacobian of the coordinate change $\mathbf{y} \to \tilde{\mathbf{y}}$ is one, since $|\det(\mathbf{O})| = 1$. Hence, using Eqs. (1.25) and (1.29) we find for the integral (1.28)

$$W_0(\mathbf{A}) = \prod_{i=1}^{n}\int d\tilde{y}_i \exp(-\frac{1}{2}a_i\tilde{y}_i^2) = (2\pi)^{n/2}(a_1a_2\ldots a_n)^{-1/2} = (2\pi)^{n/2}(\det\mathbf{A})^{-1/2} . \tag{1.30}$$

To summarise, we have found for the multidimensional Gaussian integral (1.28) that

$$W_0(\mathbf{A}) = (2\pi)^{n/2}(\det\mathbf{A})^{-1/2} , \tag{1.31}$$

a result which will be of some importance in the following.

One obvious generalisation of the integral (1.28) involves adding a term linear in $\mathbf{y}$ in the exponent, that is

$$W_0(\mathbf{A}, \mathbf{J}) \equiv \int d^n\mathbf{y}\, \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} + \mathbf{J}^T\mathbf{y}\right) . \tag{1.32}$$

Here $\mathbf{J} = (J_1, \ldots, J_n)$ is an n-dimensional vector which can be viewed as an external source for $\mathbf{y}$. With the inverse matrix $\mathbf{\Delta} = \mathbf{A}^{-1}$ and a change of variables $\mathbf{y} \to \tilde{\mathbf{y}}$, where

$$\mathbf{y} = \mathbf{\Delta}\mathbf{J} + \tilde{\mathbf{y}} \tag{1.33}$$

this integral can be written as

$$W_0(\mathbf{A}, \mathbf{J}) = \exp\left(\frac{1}{2}\mathbf{J}^T\mathbf{\Delta}\mathbf{J}\right)\int d^n\tilde{\mathbf{y}}\, \exp\left(-\frac{1}{2}\tilde{\mathbf{y}}^T\mathbf{A}\tilde{\mathbf{y}}\right) . \tag{1.34}$$

The remaining integral is Gaussian without a linear term, so can be easily carried out using the above results. Hence, one finds

$$W_0(\mathbf{A}, \mathbf{J}) = W_0(\mathbf{A})\exp\left(\frac{1}{2}\mathbf{J}^T\mathbf{\Delta}\mathbf{J}\right) = (2\pi)^{n/2}(\det\mathbf{A})^{-1/2}\exp\left(\frac{1}{2}\mathbf{J}^T\mathbf{\Delta}\mathbf{J}\right) . \tag{1.35}$$

The exponential in Gaussian integrals can be considered as a probability distribution and, hence, we can introduce expectation values for functions $O(\mathbf{y})$ by

$$\langle O\rangle_0 \equiv \mathcal{N}\int d^n\mathbf{y}\, O(\mathbf{y})\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y}\right) . \tag{1.36}$$

The normalisation $\mathcal{N}$ is chosen such that $\langle 1\rangle = 1$ which implies

$$\mathcal{N} = W_0(\mathbf{A})^{-1} = (2\pi)^{-n/2}(\det\mathbf{A})^{1/2} . \tag{1.37}$$

The sub-script 0 refers to the fact that the expectation value is computed with respect to a Gaussian distribution. As we will see later, such expectation values play a central role when it comes to extracting physical information from a quantum field theory. Of particular importance are the *moments* of the Gaussian probability distribution (or

l-point functions) which are given by the expectation values of monomials, $\langle y_{i_1} y_{i_2} \dots y_{i_l} \rangle_0$. From Eqs. (1.32) and (1.35) they can be written as

$$\langle y_{i_1} y_{i_2} \dots y_{i_l} \rangle_0 = \mathcal{N} \left[\frac{\partial}{\partial J_{i_1}} \dots \frac{\partial}{\partial J_{i_l}} W_0(\mathbf{A}, \mathbf{J}) \right]_{\mathbf{J}=0} = \left[\frac{\partial}{\partial J_{i_1}} \dots \frac{\partial}{\partial J_{i_l}} \exp\left(\frac{1}{2} \mathbf{J}^T \mathbf{\Delta} \mathbf{J} \right) \right]_{\mathbf{J}=0} . \tag{1.38}$$

The first equality above suggests an interpretation of the Gaussian integral $W_0(\mathbf{A}, \mathbf{J})$ with source $\mathbf{J}$ as a *generating function* for the l-point functions. The second equality provides us with a convenient way of calculating l-point functions explicitly. It is easy to see, for example, that the two-point and four-point functions are given by

$$\langle y_i y_j \rangle_0 = \Delta_{ij} , \qquad \langle y_i y_j y_k y_l \rangle_0 = \Delta_{ij}\Delta_{kl} + \Delta_{ik}\Delta_{jl} + \Delta_{il}\Delta_{jk} . \tag{1.39}$$

Every differentiation with respect to J_i in Eq. (1.38) leads to a factor of $\Delta_{ij} J_j$ in front of the exponential. Unless J_j is removed by another differentiation the contribution vanishes after setting $\mathbf{J}$ to zero. In particular, this means that all odd l-point functions vanish. For the even l-point functions one deduces that the Δ_{ij} terms which appear are in one-to-one correspondence with all ways of pairing up the indices $i_1, \dots, i_l$. Hence, for even l we have

$$\langle y_{i_1} \dots y_{i_l} \rangle_0 = \sum_{\text{pairings } p} \Delta_{i_{p_1} i_{p_2}} \dots \Delta_{i_{p_{l-1}} i_{p_l}} \tag{1.40}$$

where the sum runs over all pairings $p = \{(p_1, p_2), \dots, (p_{l-1}, p_l)\}$ of the numbers $\{1, \dots, l\}$. This statement is also known as *Wick's theorem*. For $l = 2$ and $l = 4$ this formula reproduces the results (1.39) for the two- and four-point functions.

We would now like to use some of the above formulae for multidimensional Gaussian integrals to infer analogous formulae for Gaussian functional integrals, following the corrrespondence (1.24). We start with the analogue of $W_0(A, J)$ in Eq. (1.32), the *generating functional*

$$W_0[A, J] = \int \mathcal{D}f \exp\left(-\frac{1}{2} \int dx \int d\tilde{x}\, f(\tilde{x}) A(\tilde{x}, x) f(x) + \int dx\, J(x) f(x) \right) . \tag{1.41}$$

Note, that the summations in the expondent of Eq. (1.32) have been replaced by integrations. The "kernel" $A(\tilde{x}, x)$ is the generalisation of the matrix $\mathbf{A}$ and, in our applications, will be usually a differential operator. It is straightforward to write down the analogue of the result (1.35) for the generating functional

$$W_0[A, J] = \exp\left(-\frac{1}{2} \mathrm{tr} \ln A \right) \exp\left(\frac{1}{2} \int d\tilde{x} \int dx J(\tilde{x}) \Delta(\tilde{x}, x) J(x) \right) . \tag{1.42}$$

where $\Delta = A^{-1}$ is the inverse of the operator A and we have used the well-known matrix identity $\det A = \exp \mathrm{tr} \ln A$ to re-write the determinant of A in terms of its trace. (We have absorbed the powers of 2π in front of Eq. (1.35) into the definition of the path integral measure $\mathcal{D}f$.) It is not clear, at this stage, how to compute the inverse and the trace of the operator A, but we will see that, in many cases of interest, this can be accomplished by looking at the Fourier transform of A. For now, we proceed formally and define the l-point functions

$$\langle f(x_1) \dots f(x_l) \rangle_0 = \mathcal{N} \int \mathcal{D}f\, f(x_1) \dots f(x_l) \exp\left(-\frac{1}{2} \int dx \int d\tilde{x}\, f(\tilde{x}) A(\tilde{x}, x) f(x) \right) , \tag{1.43}$$

where the normalisation $\mathcal{N}$ is given by $\mathcal{N} = \exp(\frac{1}{2} \mathrm{tr} \ln A)$. In analogy with Eq. (1.38) we have

$$\langle f(x_1) \dots f(x_l) \rangle_0 = \left[\frac{\delta}{\delta J(x_1)} \dots \frac{\delta}{\delta J(x_l)} \exp\left(\frac{1}{2} \int d\tilde{x} \int dx\, J(\tilde{x}) \Delta(\tilde{x}, x) J(x) \right) \right]_{J=0} . \tag{1.44}$$

This expression can be used to compute the l-point functions in terms of Δ, where Eq. (1.17) should be taken into account. Since functional derivatives have the same general properties than ordinary derivatives Wick's theorem applies and the l-point functions take the form

$$\langle f(x_1) \dots f(x_l) \rangle_0 = \sum_{\text{pairings } p} \Delta(x_{p_1}, x_{p_2}) \dots \Delta(x_{p_{l-1}}, x_{p_l}) \tag{1.45}$$

Steepest descent approximation

Although functional integrals in the context of quantum mechanics or quantum field theory are, in general, non-Gaussian they can be approximated by Gaussian integrals in many situations of physical interest. The steepest descent method is one such approximation scheme which is closely related to the transition between quantum and classical physics (as will be explained later).

As usual, we start with simple one-dimensional integrals and work our way towards path integrals. Consider the integral

$$I(\xi) = \int dy \, \exp(-s(y)/\xi) \tag{1.46}$$

with a real function $s = s(y)$ and a real positive parameter ξ. (In our physical examples the role of ξ will be played by $\hbar$.) We would like to study the value of this integral for small ξ. For such values, the main contributions to the integral come from regions in y close to the minima of s, that is points $y^{(c)}$ satisfying $s'(y^{(c)}) = 0$. Around such a minimum we can expand

$$s(y) = s(y^{(c)}) + \frac{1}{2}\xi a\epsilon^2 + \dots \, , \tag{1.47}$$

where $\epsilon = (y - y^{(c)})/\sqrt{\xi}$, $a = s''(y^{(c)})$ and the dots stand for terms of order higher than ξ. Then, we can approximate

$$I(\xi) \simeq \sqrt{\xi}\exp(-s(y^{(c)})/\xi) \int d\epsilon \, \exp\left(-\frac{1}{2}a\epsilon^2\right) \, . \tag{1.48}$$

The remaining integral is Gaussian and has already been computed in Eq. (1.25). Inserting this result we finally find

$$I(\xi) = \sqrt{\frac{2\pi\xi}{a}} \, \exp(-s(y^{(c)})/\xi) \, . \tag{1.49}$$

To leading order the integral is simply given by the exponential evaluated at the minimum $y^{(c)}$ and the square root pre-factor can be interpreted as the first order correction.

Let us generalise this to the multi-dimensional integrals

$$W(\xi) = \int d^n\mathbf{y} \, \exp(-S(\mathbf{y})/\xi) \, , \tag{1.50}$$

where $\mathbf{y} = (y_1, \dots, y_n)$. Around a solution $\mathbf{y}^{(c)}$ of $\frac{\partial S}{\partial y_i}(\mathbf{y}) = 0$ we expand

$$S(\mathbf{y}) = S(\mathbf{y}^{(c)}) + \frac{1}{2}\xi\boldsymbol{\epsilon}^T\mathbf{A}\boldsymbol{\epsilon} + \dots \, , \tag{1.51}$$

where $\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{y}^{(c)})/\sqrt{\xi}$ and $A_{ij} = \frac{\partial^2 S}{\partial y_i \partial y_j}(\mathbf{y}^{(c)})$. Using this expansion and the result (1.31) for the multi-dimensional Gaussian integral we find

$$W(\xi) = (2\pi\xi)^{n/2} \, (\det\mathbf{A})^{-1/2} \exp(-S(\mathbf{y}^{(c)})/\xi) \, . \tag{1.52}$$

We would now like to apply the steepest descent method to the generating functional for l-point functions. In analogy with our procedure for Gaussian integrals, we can introduce this generating function by

$$W(\mathbf{J}, \xi) = \int d^n\mathbf{y} \exp\left(-\frac{1}{\xi}(S(\mathbf{y}) - \mathbf{J}^T\mathbf{y})\right) \tag{1.53}$$

so that the l-point functions are given by

$$\langle y_{i_1} \dots y_{i_l}\rangle \equiv \mathcal{N} \int d^n\mathbf{y} \, y_{i_1} \dots y_{i_l} \exp(-S(\mathbf{y})/\xi) = \xi^l \mathcal{N} \left[\frac{\partial}{\partial J_{i_1}} \dots \frac{\partial}{\partial J_{i_l}} W(\mathbf{J}, \xi)\right]_{\mathbf{J}=0} \, . \tag{1.54}$$

Minima $\mathbf{y}^{(c)}$ of the exponent in Eq. (1.53) are obtained from the equations

$$J_i = \frac{\partial S}{\partial y_i} \, . \tag{1.55}$$

Applying the result (1.52) to the generating function (1.53) one finds

$$W(\mathbf{J}, \xi) \sim (\det\mathbf{A})^{-1/2} \exp\left(-\frac{1}{\xi}(S(\mathbf{y}^{(c)}) - \mathbf{J}^T\mathbf{y}^{(c)})\right) \tag{1.56}$$

From Eq. (1.55), we should think of $\mathbf{y}^{(c)}$ as a function of the source $\mathbf{J}$ which has to be inserted into the above result for $W(\mathbf{J},\xi)$. The l-point functions in the steepest descent approximation can then be computed by inserting the so-obtained $W(\mathbf{J},\xi)$ into Eq. (1.54).

Finally, we apply the steepest descent method to functional integrals of the form

$$W_\xi = \int \mathcal{D}f \, \exp(-S[f]/\xi) \, , \tag{1.57}$$

where $S[f]$ is a functional. In our examples S will be the action of a physical system. The steepest descent approximation now amounts to an expansion around the "classical solution" $f^{(c)}$ characterised by

$$\frac{\delta S}{\delta f}[f^{(c)}] = 0 \, . \tag{1.58}$$

With

$$A(\tilde{x},x) = \frac{\delta^2 S}{\delta f(\tilde{x})\delta f(x)}[f^{(c)}] \tag{1.59}$$

we find in analogy with the multi-dimensional steepest descent approximation (1.52)

$$W_\xi \sim \exp\left(-\frac{1}{2}\mathrm{tr}\ln A\right)\exp\left(-S[f^{(c)}]/\xi\right) \, . \tag{1.60}$$

Note that the leading contribution to the functional integral in this approximation comes from the classical path $f^{(c)}$. We can now introduce the generating functional

$$W_\xi[J] = \int \mathcal{D}f \, \exp\left[-\frac{1}{\xi}\left(S[f] - \int dx\, J(x)f(x)\right)\right] \, , \tag{1.61}$$

which generates the l-point functions

$$\langle f(x_{i_1})\ldots f(x_{i_l})\rangle \equiv \mathcal{N}\int \mathcal{D}f \, f(x_{i_1})\ldots f(x_{i_l})\exp(-S[f]/\xi) = \xi^l \mathcal{N}\left[\frac{\delta}{\delta J(x_{i_1})}\ldots\frac{\delta}{\delta J(x_{i_1})}W_\xi[J]\right]_{\mathbf{J}=0} \, . \tag{1.62}$$

The steepest descent approximation can then be applied around the solution $f^{(c)} = f^{(c)}(J)$ of

$$J(x) = \frac{\delta S}{\delta f(x)} \tag{1.63}$$

and leads to the result

$$W_\xi[J] \sim \exp\left(-\frac{1}{2}\mathrm{tr}\ln A\right)\exp\left[-\frac{1}{\xi}\left(S[f^{(c)}] - \int dx\, J(x)f^{(c)}(x)\right)\right] \tag{1.64}$$

1.1.4 Perturbation theory for non-Gaussian functional integrals

Perturbation theory is an approximation scheme which can be applied to integrals which differ only slightly from Gaussian ones. Let us start with the multi-dimensional case and consider the non-Gaussian integral (1.50) with the specific exponent $S = \frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} + \lambda V(\mathbf{y})$, that is,

$$W_\lambda(\mathbf{A}) = \int d^n\mathbf{y}\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} - \lambda V(\mathbf{y})\right) \, . \tag{1.65}$$

(We can set ξ to one for the purpose of this section.) Here, we take the "interaction term" $V(\mathbf{y})$ to be a polynomial in $\mathbf{y}$ and λ is the "coupling constant". Expanding the part of the exponential which involves the "perturbation" λV, we can write

$$W_\lambda(\mathbf{A}) = W_0(\mathbf{A})\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\langle V(\mathbf{y})^k\rangle_0 \, , \tag{1.66}$$

where the expectation value $\langle\ldots\rangle_0$ is defined with respect to the Gaussian measure, as in Eq. (1.36). Given that $V(\mathbf{y})$ is polynomial each of these expectation values can be calculated using Wick's theorem (1.40). Of course, the resulting expressions will be progressively more complicated for increasing powers in λ but, provided λ is

sufficiently small, we can hope that the integral is well-approximated by cutting the infinite series in (1.66) off at some finite value of k. Let us see how this works in practise by focusing on the example of a quartic interaction term

$$V(\mathbf{y}) = \frac{1}{4!}\sum_{i=1}^{n} y_i^4 \,. \tag{1.67}$$

Neglecting terms of order λ^3 and higher we have

$$\frac{W_\lambda(\mathbf{A})}{W_0(\mathbf{A})} = 1 - \frac{\lambda}{4!}\sum_i \langle y_i^4\rangle_0 + \frac{\lambda^2}{2(4!)^2}\sum_{i,j}\langle y_i^4 y_j^4\rangle_0 + \mathcal{O}(\lambda^3) \tag{1.68}$$

$$= 1 - \frac{\lambda}{8}\sum_i \Delta_{ii}^2 + \frac{\lambda^2}{128}\sum_i \Delta_{ii}^2 \sum_j \Delta_{jj}^2 + \frac{\lambda^2}{48}\sum_{i,j}(3\Delta_{ii}\Delta_{jj}\Delta_{ij}^2 + \Delta_{ij}^4) + \mathcal{O}(\lambda^3) \tag{1.69}$$

where Wick's theorem (1.40) has been used in the second step. There is a very useful representation of expressions of the form $\Delta_{i_1 i_2}\ldots\Delta_{i_{l-1} i_l}$ in terms of diagrams. First, recall that the indices $i, j, \ldots$ label points $x_i, x_j, \ldots$ of our discrete space-time. This suggests we should represent each such index by a point (or dot). Each Δ_{ij} is then represented by a line connecting dot i and dot j. One may think of the quantity Δ_{ij} as "moving the system" from x_i to x_j and it is, therefore also called *propagator*. Given these rules, the diagrams for the terms at order λ and order λ^2 in Eq. (1.69) are shown in Figs. 1.1 and 1.2, respectively. Obviously, these diagrams are constructed by

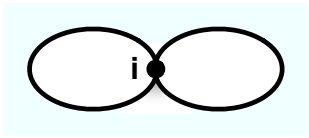

Figure 1.1: Feynman diagram contributing to $W_\lambda(\mathbf{A})$ at first order in perturbation theory for a quartic interaction.

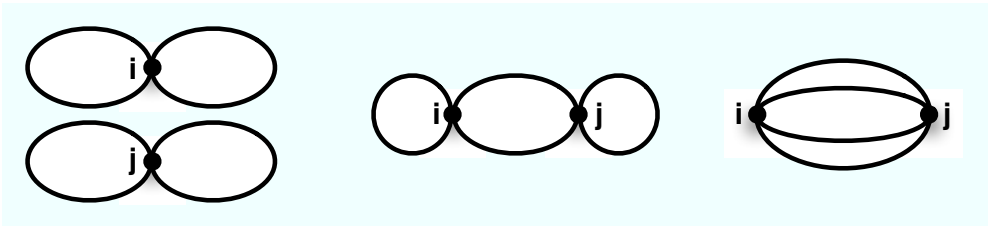

Figure 1.2: Feynman diagrams contributing to $W_\lambda(\mathbf{A})$ at second order in perturbation theory for a quartic interaction.

joining up four-leg vertices, one for each power of the coupling λ. The reason we are dealing with four-leg vertices is directly related to the quartic nature of the interaction term, as an inspection of Wick's theorem (1.40) shows. The diagrams associated to the order λ^k terms in the perturbation expansion can then be obtained by connecting the legs of k four-vertices in all possible ways.

We are also interested in computing the l-point functions

$$\langle y_{i_1}\ldots y_{i_l}\rangle_\lambda = W_\lambda(\mathbf{A})^{-1}\mathcal{G}^{(l)}_{i_1\ldots i_l}\,, \qquad \mathcal{G}^{(l)}_{i_1\ldots i_l} \equiv \int d^n\mathbf{y}\, y_{i_1}\ldots y_{i_l}\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} - \lambda V(\mathbf{y})\right) \tag{1.70}$$

in perturbation theory. Here, we have defined the *Green functions* $\mathcal{G}^{(l)}_{i_1\ldots i_l}$. Since we have already discussed $W_\lambda(\mathbf{A})$ the Green functions are what we still need to compute, in order to fully determine the l-point functions. One way to proceed is to simply expand the Green function in powers of λ and write the individual terms as Gaussian expectation values (1.36). This leads to

$$\mathcal{G}^{(l)}_{i_1\ldots i_l} = W_0(\mathbf{A})\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\langle y_{i_1}\ldots y_{i_l}V(\mathbf{y})^k\rangle_0\,. \tag{1.71}$$

Each Gaussian expectation value in this expansion can be calculated using Wick's theorem. Alternatively, in analogy with the Gaussian case, we may also introduce the generating function

$$W_\lambda(\mathbf{A},\mathbf{J}) = \int d^n\mathbf{y}\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} - \lambda V(\mathbf{y}) + \mathbf{J}^T\mathbf{y}\right) \tag{1.72}$$

so that the Green functions can be written as

$$\mathcal{G}^{(l)}_{i_1\ldots i_l} = \left[\frac{\partial}{\partial J_{i_1}}\ldots\frac{\partial}{\partial J_{i_l}} W_\lambda(\mathbf{A},\mathbf{J})\right]_{\mathbf{J}=0} . \tag{1.73}$$

We can now expand the generating function into a perturbation series

$$W_\lambda(\mathbf{A},\mathbf{J}) = \sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\int d^n\mathbf{y}\, V(\mathbf{y})^k \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y}+\mathbf{J}^T\mathbf{y}\right) = \sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!} V\left(\frac{\partial}{\partial\mathbf{J}}\right)^k W_0(\mathbf{A},\mathbf{J}) , \tag{1.74}$$

around the Gaussian generating function W_0. The partial differential operator $V(\partial/\partial\mathbf{J})$ is obtained by formally replacing y_i with $\partial/\partial J_i$ in the function $V = V(\mathbf{y})$. Combining the last two equations we find

$$\begin{aligned}\mathcal{G}^{(l)}_{i_1\ldots i_l} &= \sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\left[\frac{\partial}{\partial J_{i_1}}\ldots\frac{\partial}{\partial J_{i_l}} V\left(\frac{\partial}{\partial\mathbf{J}}\right)^k W_0(\mathbf{A},\mathbf{J})\right]_{\mathbf{J}=0} &(1.75)\\ &= W_0(\mathbf{A})\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\left[\frac{\partial}{\partial J_{i_1}}\ldots\frac{\partial}{\partial J_{i_l}} V\left(\frac{\partial}{\partial\mathbf{J}}\right)^k \exp\left(\frac{1}{2}\mathbf{J}^T\mathbf{\Delta}\mathbf{J}\right)\right]_{\mathbf{J}=0} , &(1.76)\end{aligned}$$

where we have used Eq. (1.35) in the second step. From our result (1.38) for Gaussian l-point functions this expression for the Green function is of course equivalent to the earlier one (1.71). Either way, we end up having to compute each term in the perturbation series applying Wick's theorem. Let us see how this works explicitly by computing the two-point function for the quartic perturbation (1.67) up to first order in λ. We find for the Green function

$$\begin{aligned}\mathcal{G}^{(2)}_{ij}/W_0(\mathbf{A}) &= \langle y_i y_j\rangle_0 - \lambda\langle y_i y_j V(y_m)\rangle_0 + \mathcal{O}(\lambda^2) &(1.77)\\ &= \Delta_{ij} - \frac{\lambda}{24}\Delta_{ij}\sum_m <y_m^4>_0 - \frac{\lambda}{2}\sum_m \Delta_{mi}\Delta_{mm}\Delta_{mj} + \mathcal{O}(\lambda^2) . &(1.78)\end{aligned}$$

The three terms in this expression can be represented by the Feynman diagrams shown in Fig. 1.3. At zeroth order

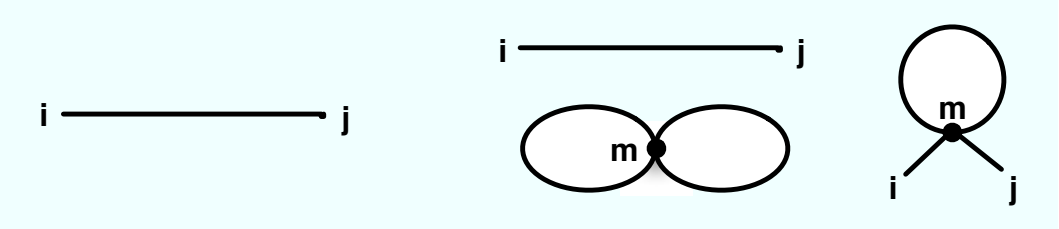

Figure 1.3: Feynman diagram contributing to two-leg Green function at zeroth and first order in perturbation theory for a quartic interaction.

in λ the two-point function is simply given by the propagator and at first order we have two diagrams. The first of these, however, is disconnected and consists of a propagator and a "vacuum bubble" which we have earlier (see Fig. 1.1) recognised as one of the contributions to $W_\lambda(\mathbf{A})$. If we compute the two-point function from Eq. (1.70) by inserting the Green function (1.78) and our earlier result (1.69) for $W_\lambda(\mathbf{A})$ (dropping terms of order λ^2) we find that this disconnected term cancels, so that

$$\langle y_i y_j\rangle_\lambda = \Delta_{ij} - \frac{\lambda}{2}\sum_m \Delta_{mi}\Delta_{mm}\Delta_{mj} + \mathcal{O}(\lambda^2) . \tag{1.79}$$

Hence, the two-point function up to first order in λ consists of the two connected diagrams in Fig. 1.3 only. The cancellation of disconnected diagrams in the two-point function is an attractive feature since it reduces the number of diagrams and removes "trivial" contributations. However, in general, l-point functions still do contain disconnected diagrams. For example, the four-point function at order λ^2 has a disconnected contribution which consists of a propagator plus the third diagram in Fig 1.3. It is, therefore, useful to introduce a new generating function $Z_\lambda(\mathbf{A},\mathbf{J})$ whose associated Green functions

$$G^{(l)}_{i_1\ldots i_l} = \left[\frac{\partial}{\partial J_{i_1}}\ldots\frac{\partial}{\partial J_{i_l}} Z_\lambda(\mathbf{A},\mathbf{J})\right]_{\mathbf{J}=0} , \tag{1.80}$$

the so-called *connected Green functions*, only represent connected diagrams. It turns out that the correct definition of this new generating function is

$$Z_\lambda(\mathbf{A},\mathbf{J}) = \ln W_\lambda(\mathbf{A},\mathbf{J}) \; . \tag{1.81}$$

One indication that this is the right choice comes from the result (1.69) for the vacuum contribution $W_\lambda(\mathbf{A})$ which can also be written as

$$\frac{W_\lambda(\mathbf{A})}{W_0(\mathbf{A})} = \exp\left[-\frac{\lambda}{8}\sum_i \Delta_{ii}^2 + \frac{\lambda^2}{48}\sum_{i,j}(3\Delta_{ii}\Delta_{jj}\Delta_{ij}^2 + \Delta_{ij}^4)\right] + \mathcal{O}(\lambda^3) \; . \tag{1.82}$$

Note that the third term in Eq. (1.69) which corresponds to the disconnected diagram in Fig. 1.2 has indeed been "obsorbed" by the exponential. It turns out that this is a general feature. As a result, connected Green functions can be evaluated in terms of Feynman diagrams in the same way ordinary Green function can, except that only connected Feynman diagrams are taken into account.

As before, the above results for multiple integrals can be formally re-written for functional integrals. We can simply carry out the replacements $y_{i_p} \to f(x_p)$, $\mathcal{G}^{(l)}_{i_1\ldots i_l} \to \mathcal{G}^{(l)}(x_1,\ldots,x_l)$, $\Delta_{i_p i_q} \to \Delta(x_p,x_q)$, $\partial/\partial J_{i_p} \to \delta/\delta J(x_p)$ and $\sum_{i_p} \to \int dx_p$. This leads to the formalism for perturbative quantum (field) theory based on functional integrals. Perturbation theory is arguably the most important method to extract physical information from quantum field theories and we will come back to it more explicitly later in the course.

1.2 Quantum Mechanics

1.2.1 Background

From Young's slits to the path integral

As inspiration, recall the two-slit interference experiment (see Fig.1.4). One of the key features of quantum mechanics is that the total amplitude $A_{\rm total}$ at a point on the screen is related to the contributions A_1 and A_2 from the two possible paths, 1 and 2, by $A_{\rm total} = A_1 + A_2$. The path integral formulation of quantum mechanics provides a way of thinking in these terms, even in a situation when, in contrast to the two-slits experiment, there are no such obviously defined classical paths.

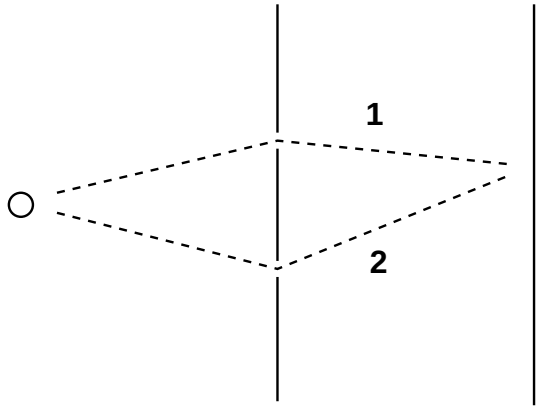

Figure 1.4: A two-slit interference experiment.

The time evolution operator

The standard approach to quantum mechanics in a first course is to go quickly from the time-dependent Schrödinger equation to the time-independent one, and to discuss eigenfunctions of the Hamiltonian. The path integral approach is different. Here the focus is on time-evolution, and eigenfunctions of the Hamiltonian remain in the background. We start from the time-dependent Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial\psi(t)}{\partial t} = H\psi(t) \, .$$

If the Hamiltonian is time-independent, the solution can be written as

$$\psi(t) = \exp(-\mathrm{i}Ht/\hbar)\psi(0)$$

Here, $\mathrm{e}^{-\mathrm{i}Ht/\hbar}$ is the time-evolution operator. It is a function (the exponential) of another operator (the Hamiltonian), and we need to remember that functions of operators can be defined via the Taylor series for the function involved.

1.2.2 Path integral for the evolution operator

Derivation

We are going to think of the time evolution operator as the product of a large number N of contributions from individual time steps t/N. We will do this for a particle moving in one dimension with a time-independent Hamiltonian $H = T + V$ consisting as usual of a potential energy V and a kinetic energy T, written in terms of the momentum operator p and particle mass m as $T = p^2/2m$. We can introduce time steps simply by writing

$$\mathrm{e}^{-\mathrm{i}Ht/\hbar} = \left(\mathrm{e}^{-\mathrm{i}Ht/N\hbar}\right)^N . \tag{1.83}$$

Next, we'd like to think of the particle's path as passing through points x_n at the successive time-steps nt/N. We can do this by introducing position eigenstates $|x\rangle$ and inserting the identity in the form

$$\mathbf{1} = \int_{-\infty}^{\infty} \mathrm{d}x |x\rangle\langle x| \tag{1.84}$$

between each factor on the right of Eq. (1.83). In this way we get

$$\langle x_0|\mathrm{e}^{-\mathrm{i}Ht/\hbar}|x_N\rangle = \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_{N-1} \langle x_0|\mathrm{e}^{-\mathrm{i}Ht/N\hbar}|x_1\rangle\langle x_1|\mathrm{e}^{-\mathrm{i}Ht/N\hbar}|x_2\rangle \ldots \langle x_{N-1}|\mathrm{e}^{-\mathrm{i}Ht/N\hbar}|x_N\rangle .$$

The next job is to evaluate matrix elements of the form

$$\langle x_n|\mathrm{e}^{-\mathrm{i}Ht/N\hbar}|x_{n+1}\rangle .$$

The way to do this is by separating the contributions from T and V, using the fact that, for $N \to \infty$

$$\left(\mathrm{e}^{-\mathrm{i}Ht/N\hbar}\right)^N = \left(\mathrm{e}^{-\mathrm{i}Tt/N\hbar}\mathrm{e}^{-\mathrm{i}Vt/N\hbar}\right)^N .$$

Now we use two versions of the resolution of the identity: the one in Eq. (1.84) based on position eigenstates, which diagonalise V, and a second one based on momentum eigenstates $|p\rangle$, which diagonalise T. We have

$$\langle x_n|\mathrm{e}^{-\mathrm{i}Tt/N\hbar}\mathrm{e}^{-\mathrm{i}Vt/N\hbar}|x_{n+1}\rangle = \int \mathrm{d}p_n \langle x_n|\mathrm{e}^{-\mathrm{i}Tt/N\hbar}|p_n\rangle\langle p_n|\mathrm{e}^{-\mathrm{i}Vt/N\hbar}|x_{n+1}\rangle .$$

The action of the operator V on the state $|x_{n+1}\rangle$ is simple: $V|x_{n+1}\rangle = V(x_{n+1})|x_{n+1}\rangle$, where $V(x)$ denotes the c-number function. Similarly for p on $|p_n\rangle$: $p|p_n\rangle = p_n|p_n\rangle$, where p is the operator and p_n is a number. Then

$$\int \mathrm{d}p_n \langle x_n|\mathrm{e}^{-\mathrm{i}Tt/N\hbar}|p_n\rangle\langle p_n|\mathrm{e}^{-\mathrm{i}Vt/N\hbar}|x_{n+1}\rangle = \int \mathrm{d}p_n \mathrm{e}^{-\mathrm{i}p_n^2 t/2mN\hbar}\langle x_n|p_n\rangle\langle p_n|x_{n+1}\rangle\mathrm{e}^{-\mathrm{i}V(x_{n+1})t/N\hbar} .$$

Next, the matrix element between states from our two bases is $\langle x|p\rangle = (2\pi\hbar)^{-1/2}\mathrm{e}^{\mathrm{i}px/\hbar}$ (check that you can justify this), and so

$$\int \mathrm{d}p_n \mathrm{e}^{-\mathrm{i}p_n^2 t/2mN\hbar}\langle x_n|p_n\rangle\langle p_n|x_{n+1}\rangle = (2\pi\hbar)^{-1}\int \mathrm{d}p_n \mathrm{e}^{-\mathrm{i}p_n^2 t/2mN\hbar - \mathrm{i}p_n(x_{n+1}-x_n)/\hbar} = \left(\frac{mN}{2\pi\mathrm{i}\hbar t}\right)^{1/2} \mathrm{e}^{\mathrm{i}mN(x_{n+1}-x_n)^2/2\hbar t} .$$

Putting everything together, and writing the interval between time-slices as $\epsilon = t/N$, we get

$$\langle x_0|\mathrm{e}^{-\mathrm{i}Ht/\hbar}|x_N\rangle = \lim_{N\to\infty}\left(\frac{m}{2\pi\mathrm{i}\epsilon\hbar}\right)^{N/2}\int \mathrm{d}x_1 \ldots \int \mathrm{d}x_{N-1} \exp\left(\frac{i\epsilon}{\hbar}\sum_{n=0}^{N-1}\left[\frac{m}{2}\left(\frac{x_{n+1}-x_n}{\epsilon}\right)^2 - V(x_{n+1})\right]\right) . \tag{1.85}$$

Finally, we interpret the sum in the argument of the exponential as a time integral, writing

$$\frac{\mathrm{i}\epsilon}{\hbar}\sum_{n=0}^{N-1}\left[\frac{m}{2}\left(\frac{x_{n+1}-x_n}{\epsilon}\right)^2 - V(x_{n+1})\right] = \frac{\mathrm{i}}{\hbar}\int_0^t \mathcal{L}\,\mathrm{d}t' \equiv \frac{\mathrm{i}}{\hbar}S[x(t')]$$

where $\mathcal{L}$ is the Lagrangian

$$\mathcal{L} = \frac{m}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}t'}\right)^2 - V(x)$$

and $S[x(t')]$ is the action for the system on the path $x(t')$. Our final result is

$$\langle x_\mathrm{i}|\mathrm{e}^{-\mathrm{i}Ht/\hbar}|x_\mathrm{f}\rangle = \mathcal{N}\int \mathcal{D}[x(t')]\mathrm{e}^{\mathrm{i}S[x(t']/\hbar}\,. \tag{1.86}$$

The notation here is as follows: $\int \mathcal{D}[x(t')]$ indicates a functional integral over paths $x(t')$, in this case starting from x_i at $t'=0$ and ending at x_f at $t'=t$, and $\mathcal{N}$ is a normalisation factor, which is independent of what the start and end points are.

The significance of Eq (1.86) is that it expresses the quantum amplitude — for a particle to propagate from a starting point x_i to a finishing point x_f in time t — in terms of an integral over possible paths, with the contribution from each path weighted by a phase factor $\mathrm{e}^{\mathrm{i}S/\hbar}$ that is expressed in terms of a classical quantity, the action S for the path. In this sense, we have succeeded in extending the treatment of interference in the context of Young's slits to a general quantum system.

Correspondence principle

One of the simplest and most attractive features of the path integral formulation of quantum mechanics is that it provides a framework in which we can see classical mechanics arising as a limiting case. We expect classical mechanics to emerge in the limit $\hbar \to 0$ (or more precisely, when $\hbar$ is small compared to the characteristic scale in S). Within the path integral, this limit is just the one in which we can apply the steepest descents approximation, since then the argument of the exponential in Eq. (1.86) is multiplied by the large quantity $\hbar^{-1}$. Using that approximation, the dominant contribution to the path integral will come from the vicinity of the path for which the action S is stationary. But we know from Hamilton's principle that this path is precisely the classical one, and so we have reached the desired conclusion.

1.2.3 Path integral in statistical mechanics

With some simple changes, we can express the Boltzmann factor, and hence other quantities in statistical mechanics, as a path integral. Comparing the time evolution operator $\mathrm{e}^{-\mathrm{i}Ht/\hbar}$ with the Boltzmann factor $\mathrm{e}^{-\beta H}$, we see that we require the substitution $\mathrm{i}t = \beta\hbar$. As a consequence, for the time t', which parameterises points on a path and has limits $0 \le t' \le t$, we substitute $\tau = \mathrm{i}t'$, which has limits $0 \le \tau \le \beta\hbar$. Now we can write a matrix element of the Boltzmann factor as

$$\langle x_\mathrm{i}|\mathrm{e}^{-\beta H}|x_\mathrm{f}\rangle = \mathcal{N}\int \mathcal{D}[x(\tau)]\mathrm{e}^{-S[x(\tau)]/\hbar} \tag{1.87}$$

with

$$S[x(\tau)] \;=\; \int_0^{\beta\hbar} d\tau\left[\frac{m}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 + V(x)\right]\,. \tag{1.88}$$

The crucial point to note about Eq. (1.88) is that because of the substitution $\tau = \mathrm{i}t'$, there is a change in the relative sign of the terms in the action coming from kinetic and potential energies. With this sign change, $S[x(\tau)]$ is often referred to as the Euclidean action, by analogy with the change from a Lorentz metric to a Euclidean one, produced by a similar factor of i. Our result, Eq. (1.87), gives matrix elements of the Boltzmann factor for a quantum system as a functional integral over classical paths, this time with paths weighted by a *real* factor, $\mathrm{e}^{-S[x(\tau)]/\hbar}$.

Quantities we want to calculate in the context of statistical mechanics are the partition function $\mathcal{Z}$ and thermal averages, for example the average $\langle f(x)\rangle$ of a function $f(x)$ of the coordinate x of the particle. These are related in a simple way to the matrix elements of the Boltzmann factor that we have discussed. Consider first the partition function, and recall its definition for a system with energy levels E_i: $\mathcal{Z} = \sum_i \mathrm{e}^{-\beta E_i}$. Clearly, this is simply

$\mathcal{Z} = \mathrm{Tr} e^{-\beta\mathcal{H}}$, written in the eigenbasis of $\mathcal{H}$. Since the trace of a matrix is invariant under change of basis, we can use this second form, which leads us to

$$\mathcal{Z} = \int \mathrm{d}x \, \langle x|e^{-\beta H}|x\rangle \equiv \mathcal{N} \int \mathcal{D}[x(\tau)] e^{-S[x(\tau)]/\hbar} \, ,$$

where the path integral is now over periodic paths, satisfying $x(\beta\hbar) = x(0)$. In a similar fashion, we can evaluate thermal averages as

$$\langle f(x)\rangle = \mathcal{Z}^{-1} \int \mathcal{D}[x(\tau)] f(x(0)) e^{-S[x(\tau)]/\hbar} \, .$$

1.2.4 Two examples

While the path integral is very important as a way of thinking about quantum mechanics and for treating more advanced problems, it turns out not to be a particularly convenient way to handle elementary ones. Nevertheless, we should see how it works for a couple of simple examples.

Evaluation for a free particle

As a first illustration, consider a free particle moving in one dimension. We want to start from Eq (1.85), set $V(x) = 0$ and evaluate the multiple integrals explicitly. Within the argument of the exponential, each coordinate x_n appears inside two quadratic terms. The integrals that we need to evaluate are therefore of the form

$$\left(\frac{\alpha\beta}{\pi^2}\right)^{1/2} \int_{-\infty}^{\infty} \mathrm{d}x \exp\left(-\alpha[u-x]^2 - \beta[x-v]^2\right) = \left(\frac{\alpha\beta}{\pi(\alpha+\beta)}\right)^{1/2} \exp\left(-\frac{\alpha\beta}{\alpha+\beta}[u-v]^2\right) . \qquad (1.89)$$

[Prove this by completing the square in the argument of the exponential.] We use this result to integrate out x_{N-1}, then x_{N-2} and so on. After completing k of the integrals we obtain

$$\left(\frac{m}{2\pi \mathrm{i}\epsilon\hbar}\right)^{N/2} \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_{N-1} \exp\left(-\frac{m}{2\mathrm{i}\epsilon\hbar} \sum_{n=0}^{N-1} (x_{n+1}-x_n)^2\right) = \left(\frac{m}{2\pi\mathrm{i}\epsilon\hbar}\right)^{(N-k-1)/2} \cdot \left(\frac{m}{2\pi\mathrm{i}[k+1]\epsilon\hbar}\right)^{1/2} \times$$
$$\times \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_{N-k-1} \exp\left(-\frac{m}{2\mathrm{i}\epsilon\hbar} \sum_{n=0}^{N-k-2} (x_{n+1}-x_n)^2 - \frac{m}{2\mathrm{i}[k+1]\epsilon\hbar} (x_N - x_{N-k-1})^2\right)$$

For $k = N-1$ this gives

$$\langle x_0|e^{\mathrm{i}Ht/\hbar}|x_N\rangle = \left(\frac{m}{2\pi\mathrm{i}\hbar t}\right)^{1/2} \exp\left(-\frac{m}{2\mathrm{i}\hbar t}(x_N - x_0)^2\right) . \qquad (1.90)$$

To check whether this is correct, we should evaluate the same quantity using a different method. For a free particle, this can be done by inserting into the expression for the matrix element of the time evolution operator a single resolution of the identity in terms of momentum eigenstates. Repeating manipulations very similar to the ones we used to derive the path integral, we have

$$\begin{aligned}
\langle x_0|e^{-\mathrm{i}Ht/\hbar}|x_N\rangle &= (2\pi\hbar)^{-1} \int \mathrm{d}p \, \langle x_0|p\rangle\langle p|e^{-\mathrm{i}Ht/\hbar}|x_N\rangle \\
&= (2\pi\hbar)^{-1} \int \mathrm{d}p \exp\left(-\frac{\mathrm{i}p^2 t}{2m\hbar} + \mathrm{i}p(x_0 - x_N)/\hbar\right) \\
&= \left(\frac{m}{2\pi\mathrm{i}\hbar t}\right)^{1/2} \exp\left(-\frac{m}{2\mathrm{i}\hbar t}(x_N - x_0)^2\right)
\end{aligned}$$

— confirming our earlier result, Eq. (1.90).

Evaluation for a harmonic oscillator

As a second example, we will evaluate the thermal average $\langle x^2\rangle$ for a one-dimensional harmonic oscillator, with the potential energy

$$V(x) = \frac{\kappa}{2} x^2 \, .$$

In terms of the matrix element of the Boltzmann factor for which we have a path integral representation, this thermal average is

$$\langle x^2 \rangle = \frac{\int \mathrm{d}x\, x^2 \langle x|\mathrm{e}^{-\beta H}|x\rangle}{\int \mathrm{d}x\, \langle x|\mathrm{e}^{-\beta H}|x\rangle}\,. \tag{1.91}$$

For the harmonic oscillator $S[(\tau)]$ is quadratic in $x(\tau)$, and so the path integral is Gaussian. To evaluate it, we should diagonalise the Euclidean action, which is

$$S[x(\tau)] \;=\; \int_0^{\beta\hbar} d\tau \left[\frac{m}{2}\left(\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau}\right)^2 + \frac{\kappa}{2}x(\tau)^2\right].$$

To do that, we simply expand $x(\tau)$ in Fourier components, writing

$$x(\tau) = \frac{1}{\sqrt{\beta\hbar}} \sum_{n=-\infty}^{\infty} \phi_n \mathrm{e}^{2\pi \mathrm{i} n\tau/\beta\hbar} \qquad \text{and} \qquad \phi_n = \frac{1}{\sqrt{\beta\hbar}} \int_0^{\beta\hbar} \mathrm{d}\tau x(\tau) \mathrm{e}^{-2\pi \mathrm{i} n\tau/\beta\hbar}\,.$$

Since $x(\tau)$ is real, $\phi_{-n} = (\phi_n)^*$, and so we can take as independent variables the real and imaginary parts of ϕ_n for $n > 0$, together with ϕ_0 which is real. The Jacobian for this change of variables is unity, because with our choice of normalisation the Fourier transform is an expansion of $x(\tau)$ in an orthonormal basis. In this way we have

$$\int \mathcal{D}[x(\tau)] \equiv \int \mathrm{d}\phi_0 \prod_{n>0} \int \mathrm{d}\Re\phi_n \int \mathrm{d}\Im\phi_n \qquad \text{and} \qquad S[x(\tau)] = \sum_{n>0} \left[m\left(\frac{2\pi n}{\hbar\beta}\right)^2 + \kappa\right] |\phi_n|^2 + \frac{1}{2}\kappa\phi_0^2\,.$$

Averaging the amplitudes ϕ_n with the weight $\mathrm{e}^{-S[x(\tau)]/\hbar}$, and introducing the oscillator frequency $\omega = \sqrt{\kappa/m}$, we have

$$\langle \phi_m \phi_{-n} \rangle = \delta_{m,n} \frac{\hbar/\kappa}{1 + (2\pi n/\beta\hbar\omega)^2}\,.$$

We obtain finally (using a contour integral to evaluate a sum on n)

$$\langle x^2(0) \rangle = \frac{1}{\beta\hbar} \left\langle \left(\sum_n \phi_n\right)^2 \right\rangle = \frac{1}{\beta\kappa} \sum_n \frac{1}{1 + (2\pi n/\beta\hbar\omega)^2} = \frac{\hbar\omega/\kappa}{2\tanh(\beta\hbar\omega/2)} = \frac{\hbar\omega}{\kappa}\left[\frac{1}{\mathrm{e}^{\beta\hbar\omega} - 1} + \frac{1}{2}\right]. \tag{1.92}$$

The rightmost form of the result is recognisable as the Planck formula for the average energy in the oscillator multiplied by κ^{-1}, which is to be expected since this energy has equal kinetic and potential contributions, and the potential energy is $\kappa\langle x^2\rangle/2$.

To summarise, we've succeeded in writing the Boltzmann factor for the harmonic oscillator as a path integral, and in evaluating this path integral, and the result we've obtained matches what is familiar from more elementary treatments. Of course, the approach seems rather laborious for an elementary problem like this one, but it brings real advantages for more advanced ones.

1.3 Further reading

- Functional derivatives (or calculus of variations) are covered by Boas and by Riley, Bence and Hobson, in their textbooks on *Mathematical Methods*.
- Multidimensional Gaussian integrals are discussed by Zinn-Justin in the introductory chapters of *Quantum Field Theory and Critical Phenomena* and of *Path integrals in Quantum Mechanics*.
- Extended treatments of path integrals in quantum mechanics are given by Zinn-Justin in both of the above books and by Schulman in *Techniques and Applications of Path Integration*.
- The reference to Feynman's original paper is: Rev. Mod. Phys. **20**, 367 (1948).
- Path integrals in quantum mechanics are introduced in the first two chapters of Bailin and Love, *Introduction to Gauge Theory*.

Be warned that, except for the first one, these references were not written with undergraduate readers in mind.

Chapter 2

Stochastic Processes and Path Integrals

2.1 Introduction

In this chapter we turn to a class of problems from classical statistical physics known as stochastic processes. We will discuss stochastic processes from several points of view, including the path integral formulation.

Recalling our discussion of path integrals in quantum mechanics, it is worth stressing a distinction between the path integral for the time evolution operator and the one for the Boltzmann factor. For the time evolution operator, the weight for a path $x(t)$ is the phase factor $\exp(\mathrm{i}S[x(t)]/\hbar)$, while for the Boltzmann factor, the weight for a path $x(\tau)$ is the real, positive quantity $\exp(-S[x(\tau)]/\hbar)$. In this second case, we have the possibility of viewing the weights as probabilities attached to the paths. For stochastic processes, one literally has probabilities associated with various paths that a system may follow as a function of time. A path integral treatment of stochastic processes therefore gives us a context in which functional integrals take on a very clear and concrete meaning.

An example of a stochastic process that we will discuss in some detail is Brownian motion: a small particle, suspended in a fluid and observed under a microscope, moves in a random way because it is buffeted by collisions with water molecules. The theory of Brownian motion, developed by Einstein in 1905, gave important confirmation of ideas of statistical mechanics and a way to determine from observations the value of Avogadro's number.

2.2 Random variables

We start by listing some definitions and introducing some useful ideas. There is some overlap between the material in this section and our earlier discussion of Gaussian integrals, but also some differences in emphasis: here we are more restrictive, in the sense that we concentrate on functions of one variable, rather than many variables; but we are also more general, in the sense that we consider distributions which are not necessarily Gaussian.

A random variable X is specified by giving the set of possible values $\{x\}$ that it can take (which may be discrete or continuous), and by a normalised probability distribution $P_X(x)$.

2.2.1 Moments of X

$$\langle X^n \rangle = \int x^n P_X(x)\,\mathrm{d}x$$

mean	$\langle X \rangle$
variance	$\langle X^2 \rangle - \langle X \rangle^2 \equiv \sigma^2$
standard deviation	σ

2.2.2 Characteristic function

The characteristic function is the Fourier transform of the probability distribution.

$$\phi_X(k) = \langle \mathrm{e}^{\mathrm{i}kX} \rangle \equiv \int P_X(x)\mathrm{e}^{\mathrm{i}kx}\mathrm{d}x = \sum_{n=0}^{\infty} \frac{(\mathrm{i}k)^n}{n!}\langle X^n \rangle$$

It is also called the *moment generating function*, since

$$\langle X^n \rangle = \frac{1}{\mathrm{i}^n} \frac{\mathrm{d}^n}{\mathrm{d}k^n} \phi_X(k) \Big|_{k=0} .$$

2.2.3 Cumulants

Write as a definition of the cumulants $C_n(X)$

$$\phi_X(k) = \exp\left(\sum_n C_n(X) \frac{(\mathrm{i}k)^n}{n!} \right) .$$

Then

$$C_n(X) = \frac{1}{\mathrm{i}^n} \left. \frac{\mathrm{d}^n}{\mathrm{d}k^n} \right|_{k=0} \ln(\phi_X(k)) .$$

We have

$$\begin{aligned} C_1(X) &= \langle X \rangle \\ C_2(X) &= \langle X^2 \rangle - \langle X \rangle^2 \qquad \text{etc} \end{aligned}$$

In a rough sense, the size of the nth cumulant tells you how much the nth moment deviates from the value you would have expected for it, given only knowledge of the lower order moments.

2.2.4 Gaussian distribution

Defined by

$$P_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2} \right)$$

which gives

$$\phi_X(k) = \exp(\mathrm{i}kx_0 - \frac{1}{2}\sigma^2 k^2) .$$

From this, we see that the cumulants are: $C_1(X) = x_0$, $C_2(X) = \sigma^2$ and $C_n(X) = 0$ for $n \geq 3$. Hence, in the language of cumulants, a Gaussian distribution is completely specified by the values of its first two cumulants.

2.2.5 Random Walks

Consider a set of N random variables, for simplicity taken to be independently and identically distributed:

$$Y_N = X_1 + X_2 + \ldots + X_N .$$

We can think of N as labelling discrete time, and of Y_N as a random walk in one dimension. The mean of Y_N is

$$\langle Y_N \rangle = N \langle X \rangle$$

and the variance Σ_N^2 is

$$\langle Y_N^2 \rangle - \langle Y_n \rangle^2 = \sum_{i,j=1}^{N} [\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle] = N\sigma^2$$

so that $\Sigma_N = \sqrt{N}\sigma$.

2.2.6 The central limit theorem

Consider $S_N \equiv Y_N/N$. Then, for large N: *the distribution of S_N converges towards a Gaussian (with mean $\langle X \rangle$ and variance $\sigma/\sqrt{N}$) irrespective of the distribution of X, provided only that $\langle X \rangle$ and $\langle X^2 \rangle$ are not infinite.*

Proof via the characteristic function.

We have

$$\phi_{S_N}(k) = \langle \mathrm{e}^{\mathrm{i}kS_N} \rangle = \langle \mathrm{e}^{\mathrm{i}\frac{k}{N}\sum_{j=1}^{N} X_j} \rangle = \langle \mathrm{e}^{\mathrm{i}\frac{k}{N}X} \rangle^N = [\phi_X(k/N)]^N \,.$$

Writing this in terms of cumulants, we have

$$\exp\left(\sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!} C_m(S_N)\right) = \exp\left(N \sum_{m=0}^{\infty} \frac{(\mathrm{i}k/N)^m}{m!} C_m(X)\right) \,,$$

and so

$$C_m(S_N) = C_m(X) \cdot N^{1-m} \,.$$

This shows that for large N, only $C_1(S_N)$ and $C_2(S_N)$ are non-vanishing. Such a feature of the cumulants is unique to a Gaussian distribution, and so S_N must be Gaussian distributed.

To demonstrate explicitly that this is the case, we can consider the calculation of the probability distribution for S_N, which we write as $P_{S_N}(s)$. It can be obtained as the inverse Fourier transform of the characteristic function for S_N, and so we have

$$P_{S_N}(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}ks} \phi_{S_N}(k) \mathrm{d}k \,. \tag{2.1}$$

Now, taking account only of the first two cumulants, we have

$$|\phi_{S_N}(k)| \approx \exp\left(-\frac{C_2(X)k^2}{2N}\right) \,.$$

This makes it clear that only the range $|k| \lesssim \sqrt{N}$ gives a significant contribution to the integral. But in this range, the corrections involving higher cumulants are small: they involve

$$N \cdot \left(\frac{k}{N}\right)^m \lesssim N^{1-m/2} \,,$$

which goes to zero as $N \to \infty$ for $m \geq 3$. Retaining only $C_1(S_N)$ and $C_2(S_N)$ in $\phi_{S_N}(k)$, we can evaluate the Fourier transform to obtain

$$P_{S_N}(s) \approx \left(\frac{N}{2\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{N(s-\langle X \rangle)^2}{2\sigma^2}\right) \,.$$

Note that what we have done is a steepest descents evaluation of Eq. (2.1).

2.3 Stochastic Processes

A random variable or variables evolving as a function of time (or some other coordinate): $Y(t)$.

For example, the position and velocity of a particle that is undergoing Brownian motion.

A stochastic process is characterised by an (infinite) sequence of probability densities

$$\begin{aligned} P_1(y_1, t_1) &\qquad \text{the probability that } Y = y_1 \text{ at } t = t_1 \\ &\vdots \\ P_n(y_1, t_1; y_2, t_2 \ldots y_n, t_n) &\quad \text{joint probability that } Y = y_1 \text{ at } t = t_1 \text{ and } Y = y_2 \text{ at } t = t_2 \text{ etc.} \end{aligned}$$

Some general properties of these densities are

$$\begin{aligned} P_n(y_1, t_1 \ldots) \geq 0 &\qquad \text{probabilities are non-negative} \\ \int P_n(y_1, t_1 \ldots y_n, t_n) \mathrm{d}y_1 \ldots \mathrm{d}y_n = 1 &\qquad \text{probabilities are normalised} \\ \int P_n(y_1, t_1 \ldots y_n, t_n) \mathrm{d}y_n = P_{n-1}(y_1, t_1 \ldots y_{n-1}, t_{n-1}) &\qquad \text{the sequence is } \textit{reducible} \end{aligned}$$

2.3.1 Correlation functions

One can define an infinite sequence of correlation functions:

$$\begin{aligned} \langle Y(t_1)\rangle &= \int \mathrm{d}y_1 P_1(y_1,t_t) y_1 \\ \langle Y(t_1)Y(t_2)\rangle &= \int\int \mathrm{d}y_1 \mathrm{d}y_2 P_2(y_1,t_1;y_2,t_2) y_1 y_2 \\ &\vdots \end{aligned}$$

2.3.2 Stationary Processes

Stationary processes are ones that are invariant under a change in the origin for time, so that

$$P_n(y_1,t_1\ldots y_n,t_n) = P_n(y_1,t_1+\tau\ldots y_n,t_n+\tau)$$

for all n and for any choice of τ. For a stationary process $\langle Y(t_1)\rangle$ is independent of t_1, $\langle Y(t_1)Y(t_2)\rangle$ depends only on the difference $t_1 - t_2$, and so on.

2.3.3 Conditional probabilities

We write the conditional probability that $Y = y_2$ at t_2, *given that* $Y = y_1$ at t_1, as $P_{1|1}(y_2,t_2|y_1,t_1)$. This obeys

$$\int \mathrm{d}y_2 P_{1|1}(y_2,t_2|y_1,t_1) = 1$$

since Y must take *some* value at t_2. Conditional and unconditional probabilities are related by

$$P_2(y_1,t_1;y_2,t_2) = P_{1|1}(y_2,t_2|y_1,t_1)\cdot P_1(y_1,t_1)\,.$$

The idea of conditional probabilities generalises in the obvious way: the quantity

$$P_{1|n-1}(y_n,t_n|y_1,t_1;y_2,t_2\ldots y_{n-1},t_{n-1})$$

is the probability to find $Y = y_n$ at t_n, *given that* $Y = y_1$ at t_1, $Y = y_2$ at t_2 etc.

2.3.4 Markov Processes

A process without memory, except through the values of the random variables.

This is the most important special class of stochastic processes, because processes of this kind are common and are much easier to study than more general ones.

Consider $t_1 < t_2 < \ldots < t_n$: a Markov process is one for which

$$P_{1|n-1}(y_n,t_n|y_1,t_1;y_2,t_2\ldots y_{n-1},t_{n-1}) = P_{1|1}(y_n,t_n|y_{n-1},t_{n-1})$$

for all n. In such a case, at time t_{n-1} one can predict the future (y_n at t_n) from present information (y_{n-1}) without additional knowledge about the past (the values of $y_{n-2}\ldots y_1$). For example, for a Brownian particle, one might expect to predict the future velocity of a particle given only the current value of the velocity, although the attempt would fail if information about the current states of the fluid molecules is important for the prediction.

A Markov process is fully determined by the two functions: $P_1(y_1,t_1)$ and $P_{1|1}(y_2,t_2;y_1,t_1)$. The absence of memory makes Markov processes tractable.

Any $P_1(y,t)$ and $P_{1|1}(y_2,t_2|y_1,t_1)$ define a Markov process, provided they satisfy certain conditions. The main ones are the *Chapman-Kolmogorov equation* (also known as the *Smoluchowski* equation), which is the condition that for any $t_3 > t_2 > t_1$

$$P_{1|1}(y_3,t_3|y_1,t_1) = \int \mathrm{d}y_2 P_{1|1}(y_3,t_3|y_2,t_2)\, P_{1|1}(y_2,t_2|y_1,t_1),$$

and the *evolution equation*, which is the condition that, for any $t_2 > t_1$

$$P_1(y_2,t_2) = \int \mathrm{d}y_1 P_{1|1}(y_2,t_2|y_1,t_1) P_1(y_1,t_1)\,.$$

2.3.5 A stationary Markov process

A Markov process which is invariant under translation in time.

That is to say, $P_1(y_1, t_1)$ does not depend on t_1, and $P_{1|1}(y_2, t_2; y_1, t_1)$ depends only on $t_2 - t_1$, y_1 and y_2.

2.4 Markov Chains

The simplest version of a Markov process has only a finite number N of discrete values possible for the random variable Y, and involves discrete (integer) time-steps. Such processes are called Markov chains.

For a Markov chain $P_1(y, t)$ is an N-component vector and $P_{1|1}(y_2, t+1; y_1, t) \equiv \mathbf{T}$ is an $N \times N$ matrix.

2.4.1 An example

Consider a two-state process, in which $Y = 1$ or $Y = 2$. Evolution from time t to time $t + 1$ can be specified as follows:

Evolution: $Y = 1 \to Y = 1$ has probability $1 - q$

Evolution: $Y = 1 \to Y = 2$ has probability q

Evolution: $Y = 2 \to Y = 2$ has probability $1 - r$

Evolution: $Y = 2 \to Y = 1$ has probability r

Then $P_1(y, t)$ can be represented by the vector

$$\mathbf{a}(t) = \begin{pmatrix} a_1(t) \\ a_2(t) \end{pmatrix}$$

with $a_1(t) + a_2(t) = 1$. The evolution equation is

$$\mathbf{a}(t+1) = \mathbf{T} \cdot \mathbf{a}(t)$$

with

$$\mathbf{T} = \begin{pmatrix} 1-q & r \\ q & 1-r \end{pmatrix} .$$

Note that, in general, $\mathbf{T}$ is a square matrix with non-negative entries, but that it is not necessarily symmetric.

2.4.2 Mathematical digression

On properties of eigenvectors for square matrices that are not Hermitian, but that are diagonalisable by a similarity transformation

In general there are distinct left and right eigenvectors associated with each eigenvalue. Using Dirac notation, we have

$$\mathbf{M}|R_\alpha\rangle = \lambda_\alpha |R_\alpha\rangle \quad \text{and} \quad \langle L_\alpha|\mathbf{M} = \langle L_\alpha|\lambda_\alpha$$

with

$$\{\langle L_\alpha|\}^\dagger \equiv |L_\alpha\rangle \neq |R_\alpha\rangle \, ,$$

where $\mathbf{M}$ is an $N \times N$ matrix, $|R\rangle$ is an N-component column vector and $\langle L|$ is an N-component row vector.

The two sets of eigenvectors are *biorthogonal*, meaning that, with an appropriate choice of normalisation, one has

$$\langle L_\alpha|R_\beta\rangle = \delta_{\alpha\beta} \, ,$$

but $\langle L_\alpha|L_\beta\rangle$ and $\langle R_\alpha|R_\beta\rangle$ have no special properties.

To prove this, start from

$$\mathbf{M}|R_\alpha\rangle = \lambda_\alpha|R_\alpha\rangle \quad (*) \qquad \text{and} \qquad \langle L_\beta|\mathbf{M} = \langle L_\beta|\lambda_\beta \quad (**)$$

Take $\langle L_\beta| \cdot (*) - (**) \cdot |R_\alpha\rangle$ to obtain

$$0 = (\lambda_\alpha - \lambda_\beta)\langle L_\beta|R_\alpha\rangle \,.$$

Hence $\langle L_\beta|R_\alpha\rangle = 0$ for $\lambda_\beta \neq \lambda_\alpha$, while by choice of normalisation (and of basis in the case of degeneracy) we can set $\langle L_\beta|R_\alpha\rangle = \delta_{\alpha\beta}$, so including the case $\lambda_\beta = \lambda_\alpha$.

A useful idea is that of *spectral decomposition*, meaning that we can write $\mathbf{M}$ in terms of its eigenvectors and eigenvalues, as

$$\mathbf{M} = \sum_\alpha |R_\alpha\rangle\lambda_\alpha\langle L_\alpha| \,.$$

2.4.3 Now apply these ideas to Markov chains

One eigenvalue of the evolution operator and the associated *left* eigenvector are easy to find. Consider (in component form)

$$a_i(t+1) = \sum_j T_{ij} a_j(t) \,.$$

From conservation of probability, we must have

$$\sum_i a_i(t+1) = \sum_i a_i(t)$$

for *any* $\mathbf{a}(t)$, which can only be true if $\sum_i T_{ij} = 1$ for all j. From this, we can conclude that one eigenvalue of $\mathbf{T}$ is $\lambda_1 = 1$, and that the associated left eigenvector is $\langle L_1| = (1, 1, \ldots 1)$. The corresponding right eigenvector is not so trivial. It has physical significance for the long-time limit of the Markov process. To see this, consider

$$\mathbf{a}(n) = (\mathbf{T})^n \mathbf{a}(0).$$

Now. from the spectral decomposition and biorthogonality, we see that

$$(\mathbf{T})^n = \sum_\alpha |R_\alpha\rangle\lambda_\alpha^n\langle L_\alpha| \,.$$

To avoid divergences, one requires $|\lambda_\alpha| \leq 1$ for all α. in addition, for a large class of problems λ_1 is the unique eigenvalue with largest magnitude, so that $|\lambda_\alpha| < 1$ for $\alpha \neq 1$. Then in the limit $n \to \infty$ we have $\mathbf{T}^n \to |R_1\rangle\langle L_1|$. From this we can conclude that the limiting distribution is given by $|R_1\rangle$.

2.4.4 Returning to the two-state process of Sec. 2.4.1

By solving

$$\begin{pmatrix} 1-q & r \\ q & 1-r \end{pmatrix} \begin{pmatrix} a_1 \\ 1-a_1 \end{pmatrix} = \begin{pmatrix} a_1 \\ 1-a_1 \end{pmatrix}$$

we find $a_1 = r/(r+q)$, which gives us the limiting distribution.

2.5 Brownian motion

Now we will switch from these generalities to the specific example of Brownian motion, which we will consider from three points of view: first, using a stochastic differential equation (that is, a differential equation which includes a random term) called the *Langevin equation*; second, using a partial differential equation for the time evolution of probability, known as the *Fokker-Planck equation*; and third, using a path integral.

2.5.1 Langevin Equation

Considering Brownian motion as an example, we can take a microscopic approach to the evolution of the velocity, $v(t)$. For simplicity, we take $v(t)$ to be a scalar – either one component of the velocity for a particle moving in three dimensions, or simply the velocity of a particle constrained to move in one dimension. The Langevin equation is

$$\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\gamma v(t) + \eta(t) \,, \tag{2.2}$$

where the term involving the constant γ represents viscous drag exerted by the fluid on the particle, which tends to reduce its velocity, and the function $\eta(t)$ represents a random force on the particle, due to collisions with molecules of the fluid. Since $\eta(t)$ is random, it can by characterised by its correlation functions. In view of the central limit theorem, it is often sufficient to know just the first two moments. We take this force to be zero on average, and take its values at times separated by more than a microscopic interval to be uncorrelated, so that

$$\langle \eta(t) \rangle = 0 \qquad \text{and} \qquad \langle \eta(t)\eta(t') \rangle = \Gamma \delta(t-t') \, .$$

In this way, the parameter Γ characterises the strength of the noise. Note that averages over realisations of the noise are given by a Gaussian functional integral:

$$\langle \ldots \rangle \equiv \int \mathcal{D}[\eta(t)] \ldots \mathrm{e}^{-\frac{1}{2\Gamma}\int \mathrm{d}t\, \eta^2(t)} \, .$$

Now, we can solve Eq. (2.2) explicitly for any noise history $\eta(t)$ in the standard way for a first-order ordinary differential equation, using an integrating factor. We find

$$v(t) = v(0)\mathrm{e}^{-\gamma t} + \int_0^t \mathrm{d}t' \mathrm{e}^{-\gamma(t-t')}\eta(t') \tag{2.3}$$

which can be checked by substituting back. From this solution, we can calculate averages. We find

$$\langle v(t) \rangle = v(0)\mathrm{e}^{-\gamma t}$$

and

$$\begin{aligned} \langle [v(t)]^2 \rangle &= [v(0)]^2 \mathrm{e}^{-2\gamma t} + \mathrm{e}^{-2\gamma t} \int_0^t \int_0^t \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{e}^{\gamma(t_1+t_2)} \langle \eta(t_1)\eta(t_2) \rangle \\ &= [v(0)]^2 \mathrm{e}^{-2\gamma t} + \frac{\Gamma}{2\gamma}[1 - \mathrm{e}^{-2\gamma t}] \, , \end{aligned} \tag{2.4}$$

and also

$$\begin{aligned} \lim_{t\to\infty} \langle v(t-\tau/2)v(t+\tau/2) \rangle &= \lim_{t\to\infty} e^{-2\gamma t} \int_0^{t-\tau/2} \int_0^{t+\tau/2} \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{e}^{\gamma(t_1+t_2)} \langle \eta(t_1)\eta(t_2) \rangle \\ &= \frac{\Gamma}{2\gamma} \mathrm{e}^{-\gamma|\tau|} \, . \end{aligned} \tag{2.5}$$

2.5.2 Fluctuation-dissipation relation

So far, it has seemed that the two constants specifying the Langevin equation, γ (viscous drag) and Γ (noise strength), are independent. In fact, of course, they both have their microscopic origins in collisions of the Brownian particle with fluid molecules, and for this physical reason there is an important relation connecting them. We get it by looking at behaviour in the long-time limit, and using information that we know from statistical mechanics – specifically, the equipartition theorem.

From Eq. (2.4), for $t \to \infty$ we have

$$\langle v^2(t) \rangle = \frac{\Gamma}{2\gamma} \, .$$

But from equipartion, the kinetic energy of the particle (which we have taken to move in one dimension only) is $k_\mathrm{B}T/2$, and so $\langle v^2(t) \rangle = k_\mathrm{B}T/m$, with m the mass of the Brownian particle. To analyse an experiment, one can go further: if the particle has a simple, known shape (for example, it is spherical), then γ can be calculated in terms of the particle dimensions and the fluid viscosity, by solving the Navier-Stokes equations. In this way, the strength of molecular noise is fixed in absolute terms.

2.5.3 Displacement of a Brownian particle

Experimentally, it is hard to observe the instantaneous velocity of a Brownian particle. Instead, one observes its position at a sequence of times which are not closely enough spaced for its velocity to remain constant between each observation. To interpret experiments, it is therefore important to calculate moments of position $x(t)$, and in

particular $\langle x^2(t)\rangle$. One obvious approach to this calculation is to obtain $x(t)$ in terms of $\eta(t)$ by integrating the expression for $v(t)$ given in Eq. (2.3), and to average moments of $x(t)$ over realisations of $\eta(t)$. An alternative is to derive and solve equations for the time evolution of the moments, as follows.

Consider

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x^2(t)\rangle = 2\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle\,.$$

Differentiating again, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle = \langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}v(t)\rangle + \langle [v(t)]^2\rangle\,.$$

Now, we can simplify the first term on the right side of this equation by using the Langevin equation to substitute for $v(t)$, and we know the second term from equipartition. In this way we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle = -\gamma\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle + \frac{k_\mathrm{B}T}{m} + \langle x(t)\eta(t)\rangle\,. \tag{2.6}$$

Since $\langle x(t)\eta(t)\rangle = 0$, we can integrate Eq. (2.6), obtaining

$$\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle = C\mathrm{e}^{-\gamma t} + \frac{k_\mathrm{B}T}{\gamma m}\,,$$

where C is an integration constant. Taking as our initial condition $x(0) = 0$, we find

$$\langle x(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)\rangle = \frac{k_\mathrm{B}T}{\gamma m}[1 - \mathrm{e}^{-\gamma t}]\,.$$

Integrating again, we have

$$\langle x^2(t)\rangle = \frac{2k_\mathrm{B}T}{\gamma m}\int_0^t (1-\mathrm{e}^{-\gamma t'})\mathrm{d}t' = \frac{2k_\mathrm{B}T}{\gamma m}[t - \frac{1}{\gamma}(1-\mathrm{e}^{-\gamma t})]\,.$$

To appreciate this result, it is useful to think about limiting cases. At short times ($\gamma t \ll 1$)

$$\langle x^2(t)\rangle \approx \frac{2k_\mathrm{B}T}{\gamma m}[t - \frac{1}{\gamma}(\gamma t - \frac{(\gamma t)^2}{2}\ldots)] = \frac{k_\mathrm{B}T}{m}t^2\,,$$

which, reasonably enough, is ballistic motion at the mean square speed expected from equipartition. At long times ($\gamma t \gg 1$)

$$\langle x^2(t)\rangle = \frac{2k_\mathrm{B}T}{\gamma m}t\,.$$

In this case, the mean square displacement grows only linearly with time, which we recognise as diffusive motion. The diffusion coefficient, $2k_\mathrm{B}T/\gamma m$ involves system-specific constants γ and m: as indicated above, γ can be calculated in terms of the size of particles and the viscosity of the fluid, and m can likewise be determined independently. A measurement of this diffusion constant therefore constitutes a measurement of Boltzmann's constant k_B, which is related to the gas constant R via Avogadro's number. Since R is known from elementary measurements on nearly ideal gases, we have a determination of Avogadro's number. The relevant theory was published by Einstein in 1905, and early experimental results were provided by Perrin in 1910.

2.5.4 Fokker-Planck Equation

As we have seen, the treatment of Brownian motion using the Langevin equation involves first analysing behaviour with a given realisation of the random force $\eta(t)$, then averaging over all realisations. An alternative approach is to consider the probability distribution of the velocity. The Fokker-Planck equation determines the time evolution of this probability. It can be derived starting from the Langevin equation, as follows.

Consider time evolution from t to $t + \Delta t$ with $\Delta t > 0$. By analogy with Eq. (2.3) we have

$$v(t+\Delta t) = v(t)\mathrm{e}^{-\gamma\Delta t} + \int_t^{t+\Delta t} \mathrm{d}t'\mathrm{e}^{-\gamma(t-t')}\eta(t') \tag{2.7}$$

so that the velocity change $\Delta v \equiv v(t+\Delta t) - v(t)$ is Gaussian ditributed if $\eta(t')$ is. Now, following the general approach to Markov processes, as introduced above, the probability distribution $P_1(v,t)$ for the velocity v of a Brownian particle at time t satisfies the (integral) evolution equation

$$P_1(v, t+\Delta t) = \int \mathrm{d}u P_{1|1}(v, t+\Delta t|u,t) P_1(u,t)\,. \tag{2.8}$$

We would like to get from this to a differential equation, by taking the limit of small Δt, in which we expect that $|v-u|$ will also be small. Some care is necessary in the derivation, because we should think of $P_{1|1}(v, t+\Delta t|u,t)$ as giving the probability distribution of v, for a fixed u, while in Eq. (2.8) the quantity v is fixed and the integral is on u. To deal with this, we change variables from u to $\Delta v = v - u$, and then Taylor expand in Δv, obtaining

$$\begin{aligned} P_1(v, t+\Delta t) &= \int \mathrm{d}(\Delta v)\, P_{1|1}(v, t+\Delta t|v-\Delta v, t) P_1(v-\Delta v, t) \\ &= \int \mathrm{d}(\Delta v)\, P_{1|1}(v-\Delta v+\Delta v, t+\Delta t|v-\Delta v, t) P_1(v-\Delta v, t) \\ &= \int \mathrm{d}(\Delta v) \sum_{n=0}^{\infty} \frac{(-\Delta v)^n}{n!} \frac{\partial^n}{\partial v^n} P_{1|1}(v+\Delta v, t+\Delta t|v, t) P_1(v,t)\,. \end{aligned}$$

Note in the middle line of this equation the substitution $v = v - \Delta\ + \Delta v$, used so that the combination $v - \Delta v$ appears uniformly in the integrand, facilitating the Taylor expansion. The moments of Δv which appear here,

$$\int \mathrm{d}(\Delta v) (\Delta v)^n P_{1|1}(v+\Delta v, t+\Delta t|v,t) \quad \equiv \quad \langle (\Delta v)^n \rangle\,,$$

take for small Δt the values

$$\langle \Delta v \rangle = -\gamma v \cdot \Delta t + \mathcal{O}(\Delta t^2)\,, \qquad \langle (\Delta v)^2 \rangle = \Gamma \cdot \Delta t + \mathcal{O}(\Delta t^2) \qquad \text{and} \qquad \langle (\Delta v)^n \rangle \lesssim \mathcal{O}(\Delta t^2) \quad \text{for} \quad n \geq 3\,.$$

Hence

$$P_1(v, t+\Delta t) = P_1(v,t) + \gamma \Delta t \frac{\partial}{\partial v}\, v P_1(v,t) + \frac{\Gamma\, \Delta t}{2} \frac{\partial^2}{\partial v^2} P_1(v,t) + \mathcal{O}(\Delta t^2)$$

and hence

$$\partial_t P_1(v,t) = \gamma \partial_v\, v P_1(v,t) + \frac{\Gamma}{2} \partial_v^2 P_1(v,t)\,. \tag{2.9}$$

which is the Fokker-Planck equation. Its solution is simplest to find in the limit $t \to \infty$, because then $P_1(v,t)$ is independent of initial conditions and of t. It is simple to show that

$$P_1(v, \infty) = \left(\frac{\gamma}{\pi\Gamma}\right)^{1/2} \exp\left(-\gamma v^2/\Gamma\right)\,,$$

which is the Maxwell distribution familiar from kinetic theory (recall from the fluctuation-dissipation relation that $\gamma/\Gamma = m/(2k_{\mathrm{B}}T)$).

2.5.5 Diffusion equation

Now we switch our attention from evolution of the velocity to evolution of the position. Note that the time-dependence of position by itself is *not* a Markov process, since future values of position depend not only on the current value of position, but also on the current velocity. It is possible to treat the combined evolution of $x(t)$ and $v(t)$, but for simplicity we will avoid this. We do so by noting that velocity relaxes on the timescale γ^{-1}: provided we treat evolution of $x(t)$ only on much longer timescales, $v(t)$ is simply a random function, like the force $\eta(t)$ that appears in the Langevin equation. We have

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = v(t)\,. \tag{2.10}$$

From Eq. (2.5) the moments of the velocity, for $t \gg \gamma^{-1}$, are

$$\langle v(t) \rangle = 0 \qquad \text{and} \qquad \langle v(t) v(t') \rangle = \frac{\Gamma}{2\gamma} \mathrm{e}^{-\gamma|t-t'|} \sim \frac{\Gamma}{\gamma^2} \delta(t-t')$$

where the replacement of $(\gamma/2)\mathrm{e}^{-\gamma|t-t'|}$ by $\delta(t-t')$ is appropriate if other quantities that depend on $t-t'$ vary much more slowly. Clearly, Eq. (2.10) is like the Langevin equation, Eq. (2.2), but with $x(t)$ replacing $v(t)$ and with $\gamma = 0$. We can therefore adapt many of the results we have already derived. In particular, by comparison with the Fokker-Planck equation, Eq. (2.9), the evolution equation for $P_1(x,t)$, the probability distribution for position, is

$$\partial_t P_1(x,t) = \frac{\Gamma}{2\gamma^2}\partial_x^2 P_1(x,t)\,, \tag{2.11}$$

which we recognise as the diffusion equation, with diffusion coefficient $D = \Gamma/2\gamma^2$. Its solution, with an initial condition that the particle is at position x_0 at time t_0, is

$$P_1(x,t) \equiv P_2(x,t;x_0,t_0) = (2\pi D|t-t_0|)^{-1/2}\exp\left(-\frac{[x-x_0]^2}{4D|t-t_0|}\right)\,. \tag{2.12}$$

2.5.6 Path integral for diffusion

We can recast the description we have built up for the motion of a Brownian particle on timescales much longer than γ^{-1} as a path integral. The most streamlined way to do so is to note that, as $v(t)$ is a Gaussian random variable with $\langle v(t)\rangle = 0$ and $\langle v(t)v(t')\rangle = (\Gamma/\gamma^2)\delta(t-t')$, its distribution is given by the Gaussian functional

$$\exp\left(-\frac{\gamma^2}{2\Gamma}\int \mathrm{d}t[v(t)]^2\right)\,.$$

This is equivalent to the Euclidean path integral for a free particle, on making the identification

$$\frac{\gamma^2}{\Gamma} \equiv \frac{1}{2D} = \frac{m}{\hbar}\,.$$

Alternatively, we can consider evolution through a sequence of discrete intervals, dividing a total time t into N equal timeslices, each of duration $\epsilon = t/N$. Extending Eq. (2.12), we have

$$P_n(x_0,0;x_1,\epsilon;\ldots:x_n,n\epsilon;\ldots x_N,t) = (4\pi D\epsilon)^{-N/2}\exp\left(-\frac{\epsilon}{4D}\sum_{n=0}^{N-1}\left(\frac{x_{n+1}-x_n}{\epsilon}\right)^2\right)\,, \tag{2.13}$$

which is the equivalent of Eq. (1.85).

A general feature of typical paths that contribute to a path integral can be read off from Eq. (2.13), by considering the typical distance a Brownian particle moves in a short time. In a time ϵ the variance of the displacement is $\langle(x_{n+1}-x_n)^2\rangle = 2D\epsilon$. The characteristic velocity averaged over this time interval is

$$\frac{\langle(x_{n+1}-x_n)^2\rangle^{1/2}}{\epsilon} = \left(\frac{2D}{\epsilon}\right)^{1/2}\,,$$

which diverges as $\epsilon \to 0$. In other words, paths are continuous (since $\langle(x_{n+1}-x_n)^2\rangle \to 0$ as $\epsilon \to 0$), but not differentiable. They are in fact fractal objects, with dimension two, in contrast to one-dimensional smooth curves.

2.6 Further Reading

- L. E. Reichl, *A Modern Course in Statistical Physics* (Edward Arnold). Chapters 5 and 6 give a basic introduction to probability theory, the central limit theorem and the Master equation.
- F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill). Chapter 15 covers Master equations, Langevin equations and Fokker-Planck equations.
- N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland). A reasonably complete and quite advanced reference book.
- L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley). Chapter 9 gives a careful account of the link between Brownian motion and the so-called Wiener integral.

Chapter 3

Statistical mechanics in one dimension

In this chapter we treat some problems from statistical mechanics that are selected partly because the mathematical methods used to study them have close links with the ideas we have met in the context of the path integral formulation of quantum mechanics and in the theory of stochastic processes. One common feature of the systems we discuss, which makes their behaviour interesting, is that they are built from particles (or other microscopic degrees of freedom) which have *interactions* between them. In this crucial sense, they differ from the simplest problems of statistical mechanics, which involve only non-interacting particles: the ideal classical gas of kinetic theory, and the ideal Fermi and Bose gases of elementary quantum statistical mechanics. Interparticle interactions can lead to behaviour quite different from that of non-interacting systems, with the possibility of spontaneous order and symmetry breaking at low temperature, and phase transitions as temperature is varied. In general, it is impossible or very difficult to treat the statistical mechanics of interacting systems exactly. It can be done, however, for a variety of one-dimensional models, and the necessary techniques are introduced in this chapter: the criterion of tractability is the second reason for the selection of problems we make here. As we will see, it is a general feature of one-dimensional systems that they do not show phase transitions or symmetry-breaking at non-zero temperature. Nevertheless, they serve to show how interactions can have a controlling influence. We will return to the topic of phase transitions in interacting systems in higher dimensions, using approximate methods, later in the course. The one-dimensional models we define in this chapter all have obvious extensions to higher dimensions.

3.1 Lattice models

As often in physics, it is useful to make models that are intended to include only the essential features of the problem we are concerned with. Details that are believed to be irrelevant are omitted, so that these models are simple but not completely realistic. For our present purposes, we need models with two ingredients: some microscopic degrees of freedom, and a form for the energy. We will refer to the latter as the Hamiltonian for the model, even in cases where there are no equations of motion, so that there is no link to classical mechanics. We start with models in which the microscopic degrees of freedom are defined at the sites of a lattice. This lattice may represent the atomic lattice of a crystalline solid, or it may simply be a convenient mathematical construct.

3.1.1 Ising model

Several interesting and much-studied models are inspired by the phenomenon of magnetism. In these models, the microscopic degrees of freedom are intended to represent atomic magnetic moments: we will refer to them as *spins*, although we take them to be classical variables. In the simplest case, the Ising spin S_i at site i of a lattice is a scalar with two states: $S_i = \pm 1$. We will consider a one-dimensional lattice, and take ferromagnetic exchange interactions of strength J (a positive quantity) between neighbouring spins, so that configurations in which neighbouring spins have the same orientation are lower in energy. We also include the Zeeman energy of spins in an external magnetic field of strength h (in scaled units). The Hamiltonian is

$$\mathcal{H} = -J \sum_i S_i S_{i+1} - h \sum_i S_i \,. \tag{3.1}$$

It is sometimes useful to specify periodic boundary conditions for a system of N sites, by setting $S_{N+1} \equiv S_1$ and taking the sums on i in Eq. (3.1) to run from 1 to N.

In a ground state all spins have the same orientation, so that the first term of $\mathcal{H}$ is minimised. With $h = 0$ there are two such states: (i) $S_i = +1$ for all i, and (ii) $S_i = -1$ for all i. Hence, at zero temperature, when the system is in a ground state, there is long-range order. Also, at $h = 0$, the model has symmetry under global spin inversion (the energy is unchanged if the directions of all spins are reversed via $S_i \to -S_i$): this symmetry is broken in each of the ground states. With $h \neq 0$ there is a unique ground state: $S_i = \text{sign}(h)$.

3.1.2 Lattice gas

A lattice gas model inludes some of the features present in a real, non-ideal gas, but within a simplifying framework in which atoms occupy sites of a lattice. Hard core repulsion between atoms is built in by allowing the occupation number n_i of site i to take only two values: $n_i = 0$ or $n_i = 1$. In addition, there is an attractive energy between atoms on neighbouring sites of magnitude V, and the concentration of atoms is controlled by a chemical potential μ. The Hamiltonian is

$$\mathcal{H} = -V \sum_i n_i n_{i+1} - \mu \sum_i n_i \,. \tag{3.2}$$

Both the Ising model and the lattice gas have microscopic degrees of freedom with two possible states, and one can be mapped onto the other: for a lattice in which each site has z neighbours (z=2 in one dimension) the replacements $S_i = 2n_i - 1$, $4J = V$, and $2h - 2zJ = \mu$ give $\mathcal{H}_{\rm Ising} = \mathcal{H}_{\rm lattice\ gas} +$ constant.

3.1.3 Classical XY and Heisenberg models

Magnetic systems can also be represented using classical spins that are two or three component unit vectors, in place of Ising spins, giving the XY and Heisenberg models. In each case, writing the spin at site i as $\vec{S}_i$, with $|\vec{S}_i| = 1$ and an external field as $\vec{h}$, the Hamiltonian is

$$\mathcal{H} = -J \sum_i \vec{S}_i \cdot \vec{S}_{i+1} - \vec{h} \cdot \sum_i \vec{S}_i \,. \tag{3.3}$$

For the XY model we can alternatively use an angle θ_i to specify the orientation of spin $\vec{S}_i$, writing

$$\mathcal{H} = -J \sum_i \cos(\theta_{i+1} - \theta_i) - h \sum \cos(\theta_i) \,. \tag{3.4}$$

At $\vec{h} = \vec{0}$ both these models are symmetric under global rotations of the spins; they both also have continuous sets of ground states, in which all spins are aligned but the common direction is arbitrary.

3.1.4 Potts model

In some circumstances a symmetry other than rotational symmetry is appropriate. In the q-state Potts model, microscopic variables σ_i take an integer number q possible states, $\sigma_i = 1 \ldots q$, and the Hamiltonian is symmetric under global permutations of these states, with

$$\mathcal{H} = -J \sum_i \delta(\sigma_i, \sigma_{i+1}) \,, \tag{3.5}$$

where we use the notation $\delta(\sigma_i, \sigma_{i+1})$ for a Kronecker delta. With ferromagnetic interactions ($J > 0$), all σ_i are the same in a ground state and there are q such states. For $q = 2$ the Potts model is equivalent to the Ising model.

3.2 Continuum models

As an alternative to these lattice models, it is also useful to decribe such systems in terms of fields (the microscopic degrees of freedom) that are defined as a function of continuous spatial coordinates $\mathbf{r}$ (or x in one dimension). If the underlying physical problem involves a lattice (as in a crystalline solid), then a continuum description may arise if we take a coarse-grained view of the system, in which information at the scale of the lattice spacing is averaged out. In such an approach it is natural to replace discrete-valued lattice variables, such as Ising spins, by fields that take values from a continuous range.

3.2.1 Scalar φ^4 theory

A continuum model equivalent to the Ising model should be based on a scalar field $\varphi(x)$, since Ising spins are scalars. The Hamiltonian at $h = 0$ should be invariant under global inversion $\varphi(x) \to -\varphi(x)$ and should have two ground states, with $\varphi(x) = \pm\varphi_0$. In addition, it should favour energetically states in which $\varphi(x)$ is constant, since for the lattice model neighbouring spins have higher energy if they are oppositely orientated. These considerations lead us to

$$\mathcal{H} = \int \mathrm{d}x \left[\frac{J}{2} \left(\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x} \right)^2 + V(\varphi(x)) - h\varphi(x) \right] \tag{3.6}$$

where the potential $V(\varphi(x))$ is chosen to have minima at $\varphi(x) = \pm\varphi_0$, and in its simplest form is

$$V(\varphi(x)) = \frac{t}{2}\varphi^2(x) + \frac{u}{4}\varphi^4(x)\,. \tag{3.7}$$

We require $u > 0$ so that the energy has a lower bound; if in addition $t > 0$, then $V(\varphi(x))$ has a single minimum at $\varphi(x) = 0$, whereas for $t < 0$ it has two minima at $\varphi(x) = \pm(-t/u)^{1/2}$.

3.2.2 Continuum limit of the XY model

We can write down a continuum version of the XY model without making the step that took us from Ising spins to the field $\varphi(x)$, since the states available to vector spins form a continuous set. Starting from Eq. (3.4), it is natural to introduce a field $\theta(x)$ and write

$$\mathcal{H} = \int \mathrm{d}x \left[\frac{J}{2} \left(\frac{\mathrm{d}\theta(x)}{\mathrm{d}x} \right)^2 - h\cos(\theta(x)) \right]\,. \tag{3.8}$$

3.3 Statistical mechanics

For completeness, we recall some of the main results of statistical mechanics. Consider a configuration (or microstate) of one of the models introduced above. Its energy is given by the Hamiltonian $\mathcal{H}$. When the system is in equilibrium with a heat bath at temperature T, then writing $\beta = 1/k_\mathrm{B}T$, where k_B is Boltzmann's constant, the probability for it to adopt a particular microstate is proportional to the Boltzmann factor $\exp(-\beta\mathcal{H})$. The normalisation constant for these probabilities is the partition function

$$\mathcal{Z} = \sum_{\text{states}} \mathrm{e}^{-\beta\mathcal{H}}\,. \tag{3.9}$$

Here, the sum on states indicates literally a sum on discrete states for the Ising and Potts models, and multiple integrals on spin orientations at each site for the lattice XY and Heisenberg models, while for continuum models it denotes a functional integral over field configurations. We will be concerned with thermal averages of observables: averages over configurations, weighted with the Boltzmann factor. We use the notation

$$\langle \ldots \rangle = \mathcal{Z}^{-1} \sum_{\text{states}} \ldots \mathrm{e}^{-\beta\mathcal{H}} \tag{3.10}$$

where $\ldots$ stands for the observable. One example is the internal energy E, the average energy of the system, which can be calculated from the partition function via

$$E \equiv \langle \mathcal{H} \rangle = -\frac{\partial}{\partial\beta} \ln(\mathcal{Z})\,. \tag{3.11}$$

Other thermodynamic quantities can also be obtained from the partition function. In particular, the (Helmholtz) free energy is

$$F = -k_\mathrm{B}T \ln(\mathcal{Z}) \tag{3.12}$$

and the entropy S is

$$S = \frac{E - F}{T}\,. \tag{3.13}$$

When we come to develop intuition about the behaviour of models in statistical mechanics, it is useful to remember the expression for the entropy of a system in the microcannonical ensemble, with W accessible states: $S = k_{\rm B} \ln(W)$. In addition, it is helpful to recall that free energy F is minimised for a system in equilibrium.

To characterise the behaviour of interacting systems, we will be particularly concerned with thermal averages of products of microscopic variables. For example, for the Ising model we are interested in the magnetisation $\langle S_i \rangle$ and the two-point correlation function $\langle S_i S_j \rangle$. The partition function, viewed as a function of external field h, is the generating function for these correlation functions, provided we allow the external field to take independent values h_i at each site. In particular, for the Ising model we have

$$\langle S_i \rangle = \frac{1}{\beta} \frac{\partial}{\partial h_i} \ln(\mathcal{Z}) \qquad \text{and} \qquad \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle = \frac{1}{\beta^2} \frac{\partial^2}{\partial h_i \partial h_j} \ln(\mathcal{Z}) \,. \tag{3.14}$$

Moreover, the magnetisation m and the magnetic susceptibility χ can be obtained from derivatives with respect to the strength of an external field that is uniform: for a system of N sites

$$m \equiv \frac{1}{N} \sum_{i=1}^{N} \langle S_i \rangle = \frac{1}{\beta N} \frac{\partial}{\partial h} \ln(\mathcal{Z}) \qquad \text{and} \qquad \chi \equiv \frac{\partial m}{\partial h} = \frac{\beta}{N} \sum_{i,j=1}^{N} \left[\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle \right] . \tag{3.15}$$

In a similar way, for the continuum theories we have introduced, we allow the external field to be a function of position, $h(x)$, so that the partition function is a functional, and then functional derivatives give correlation functions, as in Chapter 1.

The dependence of the two-point correlation function on separation provides a measure of the influence of interactions on the behaviour of the system. In the high temperature limit ($\beta \to 0$) all configurations are weighted equally in a thermal average, spins fluctuate independently at each site, and the correlation function is short range, meaning that it falls to zero at large separation (in fact, in this limit $\langle S_i S_j \rangle$=0 unless i=j). In the opposite limit of zero temperature ($\beta \to \infty$), the system is in its ground state: as we have seen, all spins then adopt the same configuration and the correlation function is infinite-ranged. We would like to understand in detail how behaviour interpolates between these two limiting cases as temperature varies.

3.4 Transfer matrix method

The transfer matrix method provides a general formalism for solving one-dimensional models with short range interactions in classical statistical mechanics. It can also be formulated for systems in higher dimensions, but is then tractable only in special cases. We will describe it for a general system, but using the notation of the one-dimensional Ising model, Eq. (3.1).

A first step is to divide the one-dimensional system into a series of slices, analogous to the time steps used in the path integral formulation of quantum mechanics. The slices must be chosen long enough that interactions couple only degrees of freedom in neighbouring slices. We denote the degrees of freedom in the i-th slice by S_i and write the Hamiltonian as a sum of terms, each involving only the degrees of freedom in adjacent slices:

$$\mathcal{H} = \sum_i \mathcal{H}(S_i, S_{i+1}) \,.$$

The Boltzmann factor, being the exponential of this sum, is a product of terms:

$$\mathrm{e}^{-\beta \mathcal{H}} = \prod_i T(S_i, S_{i+1}) \qquad \text{with} \qquad T(S_i, S_{i+1}) = \mathrm{e}^{-\beta \mathcal{H}(S_i, S_{i+1})} \,. \tag{3.16}$$

Because interactions couple only adjacent slices, only two terms in this product depend on a given S_i: $T(S_{i-1}, S_i)$ and $T(S_i, S_{i+1})$. Moverover, summation on S_i acts just like matrix multiplication. This leads us to define the transfer matrix $\mathbf{T}$: for the case in which S_i takes M values (M=2 for the one-dimensional Ising model with only nearest-neighbour interactions) it is an $M \times M$ matrix with rows and columns labelled by the possible values of S_i and S_{i+1}, and matrix elements $T(S_i, S_{i+1})$ as defined in Eq. (3.16). For a system of N slices and periodic boundary conditions (so that i runs from 1 to N), the partition function is simply the matrix trace

$$\mathcal{Z} = \mathrm{Tr}\, \mathbf{T}^N \,.$$

Alternatively, with fixed configurations for S_1 and S_N, and without periodic boundary conditions, the partition function is the matrix element

$$\mathcal{Z} = \mathbf{T}^{N-1}(S_1, S_N) \,.$$

The advantage of the transfer matrix approach is that in this way calculations are reduced to study of an $M \times M$ matrix, independent of system size N. We will assume for simplicity that $\mathcal{H}(S_i, S_{i+1})$=$\mathcal{H}(S_{i+1}, S_i)$ so that $\mathbf{T}$ is symmetric and its eigenvectors $|\alpha\rangle$ can be chosen to form a complete, orthonormal set. We order the associated eigenvalues by magnitude (the first is in fact necessarily positive and non-degenerate): $\lambda_0 > |\lambda_1| \geq \ldots |\lambda_{M-1}|$, with

$$\mathbf{T}|\alpha\rangle = \lambda_\alpha |\alpha\rangle\,, \qquad \langle\alpha|\beta\rangle = \delta_{\alpha\beta} \qquad \text{and} \qquad \mathbf{T} = \sum_\alpha |\alpha\rangle\lambda_\alpha\langle\alpha|\,.$$

Using this notation, it is simple to write down a power of the transfer matrix: we have

$$\mathbf{T}^N = \sum_\alpha |\alpha\rangle\lambda_\alpha^N\langle\alpha|\,.$$

The free energy per slice with periodic boundary conditions is therefore

$$\frac{F}{N} = -\frac{k_\mathrm{B}T}{N}\ln(\mathcal{Z}) = -k_\mathrm{B}T\ln(\lambda_0) - \frac{k_\mathrm{B}T}{N}\ln\left(1 + \sum_{\alpha=1}^{M-1}[\lambda_\alpha/\lambda_0]^N\right)\,.$$

In the thermodynamic limit ($N \to \infty$), $[\lambda_\alpha/\lambda_0]^N \to 0$ for $\alpha \geq 1$, and so the free energy density $f = \lim_{N\to\infty} F/N$ is simply

$$f = -k_\mathrm{B}T\ln(\lambda_0)\,. \tag{3.17}$$

Clearly, we can obtain other thermodynamic quantities, including the energy density E/N, the entropy density S/N, the magnetisation m and magnetic susceptibility χ by this route.

To determine correlation functions, one might imagine we should first evaluate a generating function dependent on field values h_i at each site. Within the transfer matrix apporach, however, this is not convenient quantity to consider, because if field values vary at different sites, the transfer matrices are different for each slice and transfer matrix products no longer have simple expressions in terms of powers of the eigenvalues. Instead, we extend the transfer matrix formalism by defining diagonal matrices $\mathbf{C}$ with diagonal elements $C(S_i, S_i)$ that are functions of the degrees of freedom within a slice, chosen to reproduce the required correlation function. For example, for the one-dimensional Ising model with N sites and periodic boundary conditions, taking $C(S_i, S_i) = S_i$ we have

$$\langle S_i\rangle = \frac{\mathrm{Tr}\,\left[\mathbf{T}^i\mathbf{C}\mathbf{T}^{N-i}\right]}{\mathrm{Tr}\,\left[\mathbf{T}^N\right]} \qquad \text{and} \qquad \langle S_i S_{i+x}\rangle = \frac{\mathrm{Tr}\,\left[\mathbf{T}^i\mathbf{C}\mathbf{T}^x\mathbf{C}\mathbf{T}^{N-i-x}\right]}{\mathrm{Tr}\,\left[\mathbf{T}^N\right]}\,.$$

As happened for the free energy density, these expressions simplify greatly in the thermodynamic limit when written in terms of the eigenvalues and eigenvectors of the transfer matrix: they reduce to

$$\lim_{N\to\infty}\frac{\mathrm{Tr}\,\left[\mathbf{T}^i\mathbf{C}\mathbf{T}^{N-i}\right]}{\mathrm{Tr}\,\left[\mathbf{T}^N\right]} = \langle 0|\mathbf{C}|0\rangle \tag{3.18}$$

and

$$\lim_{N\to\infty}\frac{\mathrm{Tr}\,\left[\mathbf{T}^i\mathbf{C}\mathbf{T}^x\mathbf{C}\mathbf{T}^{N-i-x}\right]}{\mathrm{Tr}\,\left[\mathbf{T}^N\right]} = \sum_\alpha \langle 0|\mathbf{C}|\alpha\rangle\langle\alpha|\mathbf{C}|0\rangle\left(\frac{\lambda_\alpha}{\lambda_0}\right)^x\,. \tag{3.19}$$

In summary, diagonalisation of the transfer matrix provides a path to calculating all quantities of physical interest.

3.5 Transfer Matrix solution of the Ising model in one dimension

Let's illustrate these ideas by using the transfer matrix approach to solve the one-dimensional Ising model, Eq. (3.1). We take the Hamiltonian for a single slice to be

$$\mathcal{H}(S_i, S_{i+1}) = -JS_iS_{i+1} - \frac{h}{2}(S_i + S_{i+1})\,.$$

Note that there was an element of choice here: we have used a symmetric form for the Zeeman energy, which will in turn ensure that the transfer matrix is symmetric. The transfer matrix is

$$\mathbf{T} = \begin{pmatrix} \exp(\beta[J+h]) & \exp(-\beta J) \\ \exp(-\beta J) & \exp(\beta[J-h]) \end{pmatrix}\,. \tag{3.20}$$

For simplicity, we will set h=0. Then the eigenvalues are $\lambda_0 = 2\cosh(\beta J)$ and $\lambda_1 = 2\sinh(\beta J)$, and the eigenvectors are $|0\rangle = 2^{-1/2}(1,1)^{\rm T}$ and $|1\rangle = 2^{-1/2}(1,-1)^{\rm T}$. The matrix representing the spin operator is

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and its matrix elements in the basis of eigenstates are $\langle 0|\mathbf{C}|0\rangle = \langle 1|\mathbf{C}|1\rangle = 0$ and $\langle 0|\mathbf{C}|1\rangle = \langle 1|\mathbf{C}|0\rangle = 1$.

We are now in a position to write down some results. From Eq. (3.17), the free energy density is

$$f = -k_{\rm B}T\ln(2\cosh(\beta J)).$$

As a check, we should examine its behaviour in the high and low temperature limits. At high temperature ($\beta \to 0$), $f \sim -k_{\rm B}T\ln(2)$. This is as expected from the relation between free energy, energy and entropy, Eq. (3.13): in the high-temperature limit neighbouring spins are equally likely to be parallel or antiparallel, and so the entropy per spin (from the general formula $S = k_{\rm B}\ln(W)$, with W the number of accessible states) is $S = k_{\rm B}\ln(2)$, while the average energy is $\langle\mathcal{H}\rangle$=0. Conversely, at low temperature ($\beta \to \infty$), $f = -J$, which arises because in this limit the system is in a ground state, with neighboring spins parallel, so that $\langle\mathcal{H}\rangle = -NJ$ and $S = 0$. Beyond these checks, the most interesting and important feature of our result for the free energy density is that it is analytic in temperature for all $T > 0$. As we will discuss later in the course, phase transitions are associated with singularities in the free energy density as a function of temperature, and analyticity of f in the one-dimensional Ising model reflects the absence of a finite-temperature phase transition. The model in fact has a critical point at $T = 0$, and f is non-analytic there (compare the limits $T \to 0^+$ and $T \to 0^-$).

What happens in the model at low temperatures is most clearly revealed by the form of the correlation functions, although for this we have to go to the two-point function. The one-point function, or magnetisation is trivial:

$$\langle S_i\rangle = \langle 0|\mathbf{C}|0\rangle = 0,$$

which is a consequence of symmetry at $h = 0$ under global spin reversal. The two-point correlation function between spins separated by a distance x is

$$\langle S_i S_{i+x}\rangle = \left(\frac{\lambda_1}{\lambda_0}\right)^{|x|} = \exp(-|x|/\xi) \qquad \text{with} \qquad \frac{1}{\xi} = \ln[\coth(\beta J)].$$

We see that correlations decay exponentially with separation, with a lengthscale ξ. This lengthscale, termed the correlation length, increases monotonically with decreasing temperature and diverges as $T \to 0$. Its asymptotic form at low temperature ($\beta J \gg 1$) is

$$\xi = \frac{1}{\ln[\coth(\beta J)]} \sim \frac{1}{\ln[1+2{\rm e}^{-2\beta J}]} \sim \frac{1}{2}{\rm e}^{2\beta J}. \tag{3.21}$$

Useful physical insight into this result comes from a simple picture of typical states at low temperature. It is clear that they consist of long sequences of parallel spins, with occasional reversals in orientation, as in Fig. 3.1. In these circumstances it is natural to focus on the reversals, called domain walls or kinks, as the elementary

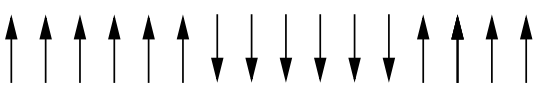

Figure 3.1: A typical low-temperature configuration of the one-dimensional Ising model.

excitations in the system. Their average spacing sets the correlation length. Moreover, each domain wall costs an energy $2J$, the difference between the two possible values of JS_iS_{i+1}. It also has an entropy associated with it, since it can be placed between any one of $\mathcal{O}(1/\xi)$ neighbouring pairs of spins, without disrupting the overall arrangement of irregularly spaced kinks. An estimate of the free energy for a chain of L sites is therefore

$$F \sim 2J\frac{L}{\xi} - k_{\rm B}T\frac{L}{\xi}\ln(\xi). \tag{3.22}$$

The *actual* value of ξ at a given temperature can be estimated as the one that minimises Eq. (3.22), yielding $\xi \sim \exp(2\beta J)$ for $\beta J \gg 1$, in reasonable agreement with our earlier, detailed calcuation.

3.6 Statistical mechanics in one dimension and quantum mechanics

There is a general relationship between the statistical mechanics of a classical system in $d+1$ dimensions at finite temperature, and the (Euclidean-time) quantum theory of a many-body system in d dimensions at zero temperature. Under this mapping, thermal fluctuations in the classical system become zero-point fluctuations in the quantum system. We will examine this relationship as it applies to the one-dimensional, classical statistical-mechanical systems we have met in this chapter. For these examples, since $d{+}1{=}1$, we have $d{=}0$, meaning that the quantum theory is particularly simple: rather than being a theory for a many-body quantum system in a finite number of dimensions, it involves just a single particle. In this way, for $d{=}0$ we map the statistical-mechanical problem to a problem in quantum mechanics, while for $d{>}0$ we would arrive at a problem in quantum field theory. In the following, we establish three different variants of this connection. All are based on viewing the spatial dimension of the statistical-mechanical system as the imaginary time direction for a corresponding quantum system.

3.6.1 Ising model and spin in a transverse field

Consider the transfer matrix $\mathbf{T}$ for the one-dimensional Ising model. We want to view this as the imaginary time evolution operator $\exp(-\tau\mathcal{H}_{\rm Q}/\hbar)$ for a quantum system with Hamiltonian $\mathcal{H}_{\rm Q}$, where τ is the duration in imaginary time equivalent to the distance in the Ising model between neighbouring sites. Since $\mathbf{T}$ is a 2×2 matrix, so must $\mathcal{H}_{\rm Q}$ be. That suggests we should regard $\mathcal{H}_{\rm Q}$ as the Hamiltonian for a single spin of magnitude $S=1/2$, for which the Pauli matrices $\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_z$ provide a complete set of operators. Anticipating the final result, note that with

$$\boldsymbol{\sigma_x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and α a constant, one has $\exp(\alpha\boldsymbol{\sigma}_x) = \cosh(\alpha)\mathbf{1} + \sinh(\alpha)\boldsymbol{\sigma}_x$. Matching this against Eq. (3.20) for $h=0$, we find that

$$\mathbf{T} = \mathrm{e}^{\beta J}\mathbf{1} + \mathrm{e}^{-\beta J}\boldsymbol{\sigma_x} \equiv A\mathrm{e}^{\alpha\boldsymbol{\sigma_x}}$$

with $\tanh(\alpha) = e^{-2\beta J}$ and $A=\sqrt{2\sinh(2\beta J)}$. In this way we can read off the Hamiltonian for the equivalent quantum system: $\tau\mathcal{H}_Q/\hbar = -\alpha\boldsymbol{\sigma}_x + \ln(A)$. In addition, we see that the lowest eigenvalue ϵ_0 of $\mathcal{H}_{\rm Q}$ is related to the largest eigenvalue of the transfer matrix via $\ln(\lambda_0) = -\tau\epsilon_0/\hbar$. Also, the inverse correlation length (in units of the lattice spacing) is related to the splitting between the ground state and first excited state eigenvalues of $\mathcal{H}_{\rm Q}$: we have $\xi^{-1} = \tau(\epsilon_1-\epsilon_0)/\hbar = 2\alpha$.

3.6.2 Scalar φ^4 and a particle in double-well potential

For the one-dimensional Ising model, the mapping we have just set out between classical statistical mechanics and quantum mechanics takes us to a quantum problem with a finite dimensional Hilbert space, since the number of possible states for an Ising spin is finite. By contrast, one-dimensional statistical mechanical problems with continuous degrees of freedom map onto quantum problems with infinite-dimensional Hilbert spaces. In particular, one-dimensional statistical mechanics problems with n continuous degrees of freedom at each point in space are equivalent to quantum problems involving a single particle moving in n dimensions.

Let's examine how this works for one-dimensional φ^4 theory, starting from Eq. (3.6). The partition function $\mathcal{Z}$ for a system of length L with fixed values $\varphi(0)$ and $\varphi(L)$ for the field at $x{=}0$ and $x{=}L$ is given by the functional integral

$$\mathcal{Z} = \int \mathcal{D}[\varphi(x)]\mathrm{e}^{-\beta\mathcal{H}}$$

over functions $\varphi(x)$ satisfying the boundary conditions. Referring back to Eqns. (1.87) and (1.88), we see that $\beta\mathcal{H}$ in the classical statistical mechanics problem plays the same role as the Euclidean action $S/\hbar$ in a path integral expression for the Boltzmann factor arising in a quantum problem: we use $\mathcal{H}_{\rm Q}$ to denote the Hamiltonian of the quantum problem. The translation dictionary is as follows.

Classical statistical mechanics	**Quantum mechanics**
position	imaginary time
system length L	imaginary time interval $\beta\hbar$
field $\varphi(x)$	particle coordinate
thermal energy $k_{\rm B}T$	Planck's constant $\hbar$
exchange stiffness J	particle mass m

The action is that for a particle of mass J moving in a potential $V(x)$: reversing the steps of section 1.2.3, we can read off the quantum Hamiltonian as

$$\mathcal{H}_{\rm Q} = -\frac{1}{2\beta^2 J}\frac{{\rm d}^2}{{\rm d}\varphi^2} + V(\varphi)\,. \tag{3.23}$$

Knowledge of the eigenvalues ϵ_α and eigenfunctions $|\alpha\rangle$ of $\mathcal{H}_{\rm Q}$, which satisfy $\mathcal{H}_{\rm Q}|\alpha\rangle = \epsilon_\alpha|\alpha\rangle$, gives access to thermodynamic quantities and correlation functions for the classical system. In particular, for a classical system extending over $-\infty < x < \infty$, the arguments leading to Eq. (3.19) also give

$$\langle\varphi(x_1)\varphi(x_2)\rangle = \sum_\alpha \langle 0|\varphi|\alpha\rangle\langle\alpha|\varphi|0\rangle\, {\rm e}^{-\beta(\epsilon_\alpha-\epsilon_0)|x_1-x_2|}$$

Now, although we cannot find the eigenfunctions of $\mathcal{H}_{\rm Q}$ exactly, we know quite a lot about them for the case of interest, in which $V(\varphi)$ is a quartic double-well potential. In particular, since $V(-\varphi) = V(\varphi)$, all eigenfunctions have definite parity. The ground state wavefunction, $\langle\varphi|0\rangle$ has even partiy and is nodeless, while the first excited state wavefunction, $\langle\varphi|1\rangle$, is odd, having a single node at $\varphi = 0$. From this it follows that $\langle 0|\varphi|0\rangle = 0$ and $\langle 0|\varphi|1\rangle \neq 0$. The inverse correlation length, governing the correlation function at large separation $|x_1 - x_2|$ is therefore $\xi^{-1} = \beta(\epsilon_1 - \epsilon_0)$. At low temperatures, the form of the lowest and first excited eigenstates is as sketched in Fig. 3.2. In the language of quantum mechanics, the splitting between eigenvalues $\epsilon_1 - \epsilon_0$ arises because of tunneling through the barrier between the two minima of $V(\varphi)$ and is exponentially small for small T, which leads to a correlation length exponentially large in β, as for the Ising model, Eq. (3.21).

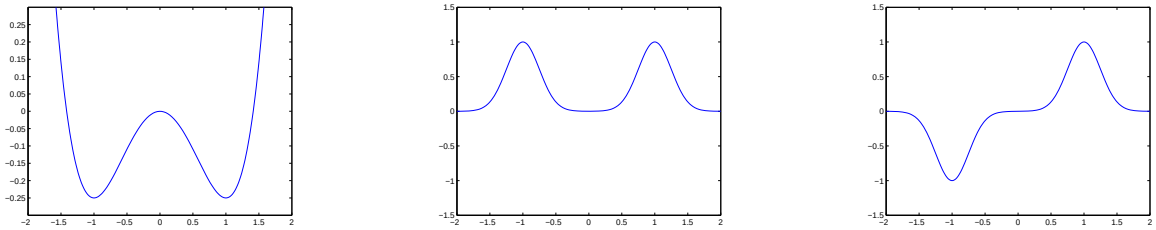

Figure 3.2: Left: the potential $V(\varphi)$ appearing in $\mathcal{H}_{\rm Q}$, Eq. (3.23). Centre and right: form of the lowest two eigenfunctions of $\mathcal{H}_{\rm Q}$, $\langle\varphi|0\rangle$ and $\langle\varphi|1\rangle$, for small T

3.6.3 One-dimensional XY model and the quantum rotor

An exactly parallel treatment can be applied to the XY chain, defined in Eq. (3.8). For this case we find

$$\mathcal{H}_{\rm Q} = -\frac{1}{2\beta^2 J}\frac{{\rm d}^2}{{\rm d}\theta^2}$$

subject to the condition that eigenfunctions are periodic in θ with period 2π. The eigenfunctions and eigenvalues are of course

$$\langle\theta|n\rangle = (2\pi)^{-1/2}\exp({\rm i}n\theta) \qquad \text{and} \qquad \epsilon_n = n^2/(2\beta^2 J) \qquad \text{with} \qquad n = 0, \pm 1, \ldots\,.$$

The correlation functions we use to characterise behaviour of the system should be constructed in a way that respects the periodic nature of the coordinate θ. An obvious candidate is $\langle {\rm e}^{i[\theta(x_1)-\theta(x_2)]}\rangle$. As in our previous examples, we can express this in terms of the eigenfunctions and eigenvalues of $\mathcal{H}_{\rm Q}$. We find

$$\langle {\rm e}^{i[\theta(x_1)-\theta(x_2)]}\rangle = \sum_n \langle 0|{\rm e}^{-i\theta}|n\rangle\langle n|{\rm e}^{i\theta}|0\rangle {\rm e}^{-\beta(\epsilon_n-\epsilon_0)|x_1-x_2|}\,.$$

The correlation length for this model is therefore $\xi = [\beta(\epsilon_1 - \epsilon_0)]^{-1} = 2J/(k_{\rm B}T)$. As for the Ising model, ξ diverges in the limit $T \to 0$, which is expected since for $T = 0$ the system adopts a ground state with $\theta(x)$ indepndent of x. The divergence of ξ, howover, is much less rapid in the XY model than in the Ising model. The reason for this is that, whereas excitations in the Ising model cost a minimum energy, the kink energy $2J$, long wavelength twists of $\theta(x)$ in the XY model can have arbitrarily low energy. As a consequence, thermal excitations are more effective in disordering the XY model at low temperature than for the Ising model, leading to a shorter correlation length in the XY model.

3.7 Further reading

- K. Huang *Introduction to Statistical Physics* (CRC Press). A good review of ideas from statistical mechanics that form the background for this chapter.
- J. M. Yeomans *Statistical Mechanics of Phase Transitions* (OUP). Chapters 2 and 5 give an introduction to lattice models and transfer matrix methods.
- J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, *The Theory of Critical Phenomena* (OUP). Chapter 3 gives a self-contained introduction to models and to transfer matrix methods.

Chapter 4

Classical Field Theory

In this chapter, we will develop the Lagrangian approach to the classical theory of fields, focusing on field theories with scalar and vector fields. Our discussion will culminate in the discussion of scalar electrodymanics, a theory which couples Maxwell's theory of electrodymanics to scalar fields. As we will see, symmetries play an important role in constructing all these theories. Traditionally, symmetries of physical theories have often been identified only after a proper mathematical formulation of the theory. For example, the fact that Maxwell's theory of electrodynamics is covariant under Lorentz transformations was only discovered significantly after its the first formulation. In modern theoretical physics, this traditional relation between theories and their symmetries is frequently reversed. One starts by identifying the symmetries of the given physical situation and then writes down the (most general) theory compatible with these symmetries. This approach has been immensely fruitful and has played a major role, for example, in constructing the standard model of particle physics. It is, therefore, crucial to understand the relevant symmetries (or groups in Mathematical language) and the objects they act on (representations, in Mathematical language) first.

4.1 Symmetries

4.1.1 Definition of groups and some examples

The word "symmetry" in Physics usually (although not always) refers to the Mathematical structure of a *group*, so we begin with the following definition.

Definition A group G is a set with a map $\cdot : G \times G \to G$ ("multiplication") satisfying the three conditions
1) $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$ for all $g_1, g_2, g_3 \in G$ (associativity)
2) There exists an $e \in G$ such that $g \cdot e = g$ for all $g \in G$ (neutral element)
3) For each $g \in G$ there exist a $g^{-1} \in G$ such that $g \cdot g^{-1} = e$ (inverse element)

It is easy to prove from the above axioms that the neutral element e is unique, that it is also the neutral element when acting from the left, that is $e \cdot g = g$ for all $g \in G$, that the right-inverse g^{-1} is uniquely defined for each $g \in G$ and that it is also the left-inverse, that is $g^{-1} \cdot g = e$. If, in addition to the three axions above, $g_1 \cdot g_2 = g_2 \cdot g_1$ is satisfied for all $g_1, g_2 \in G$ the group is called *Abelian*.

Well-known groups are the integers with respect to addition and the real and complex numbers with respect to addition and multiplication. All these groups are Abelian. Here are some more interesting groups which will play a role in our field theory constructions. Consider first the group $\mathbb{Z}_N = \{0, 1, \cdots, N-1\}$ with "multiplication" defined by $n_1 \cdot n_2 = (n_1 + n_2) \bmod N$. This group is obviously finite (that is, it has a finite number of elements) and Abelian. Another Abelian example is given by the complex numbers of unit length, $U(1) = \{z \in \mathbb{C} | \, |z| = 1\}$, with group multiplication the ordinary multiplication of complex numbers. Not only is this group infinite but, as it corresponds to the unit circle in the complex plane, it is also "continuous" and one-dimensional. Examples for non-Abelian groups are provided by the unitary groups $\mathrm{SU}(n)$ which consist of all complex $n \times n$ matrices U satisfying $U^\dagger U = \mathbf{1}$ and $\det(U) = 1$, with ordinary matrix multiplication as the group multiplication and the unit matrix as the neutral element. Matrix multiplication does in general not "commute" which causes the non-Abelian character of the unitary groups. The simplest non-trivial example of a unitary group on which we will focus later is $\mathrm{SU}(2)$. Solving the unitary conditions $U^\dagger U = \mathbf{1}$ and $\det(U) = 1$ by inserting an arbitrary 2×2 matrix with

complex entries it is easy to show that $\mathrm{SU}(2)$ can be written as

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\beta^\star & \alpha^\star \end{pmatrix} \mid \alpha, \beta \in \mathbb{C} \text{ and } |\alpha|^2 + |\beta|^2 = 1 \right\} . \tag{4.1}$$

This shows that we can think of $\mathrm{SU}(2)$ as the unit sphere in four-dimensional Euklidean space and, hence, that it is a three-dimensional continuous group. Such continuous groups are also called *Lie groups* in Mathematical parlance and we will discuss some of their elementary properties in due course. We can solve the constraint on α and β in Eq. (4.1) by setting $\alpha = \sqrt{1-|\beta|^2}e^{i\sigma}$ and $\beta = -\beta_2 + i\beta_1$ which leads to the explicit parameterization

$$U = \begin{pmatrix} \sqrt{1-|\beta|^2}e^{i\sigma} & -\beta_2 + i\beta_1 \\ \beta_2 + i\beta_1 & \sqrt{1-|\beta|^2}e^{-i\sigma} \end{pmatrix} , \tag{4.2}$$

of $\mathrm{SU}(2)$ in terms of β_1, β_2 and σ.

4.1.2 Representations of groups

Let us denote by $\mathrm{Gl}(n)$ the group of invertible $n \times n$ matrices with real (or complex) entries. We can think of these matrices as linear transformations acting on an n-dimensional real (or complex) vector space $V \sim \mathbb{R}^n$ (or $V \sim \mathbb{C}^n$).

Definition A representation R of a group G is a map $R : G \to \mathrm{GL}(n)$ which satisfies $R(g_1 \cdot g_2) = R(g_1)R(g_2)$.

In other words, a representation assigns to each element of a group a matrix such that these matrices multiply "in the same way" as the associated group elements. In this way, the group is realised or "represented" as a set of matrices. Given a representation by $n \times n$ matrices we can also think of the group as acting on the n-dimensional vector space V, via the representation matrices $R(g) \in \mathrm{Gl}(n)$. The dimension n of this vector space is also referred to as the *dimension of the representation*. In a physics context, the elements of this vector space can be thought of as the fields (more precisely, the fields at each fixed point in space-time) and, hence, group representations provide the appropriate mathematical structure to describe symmetries acting on fields. The mathematical problem of finding all representations of a given group then translates into the physics problem of finding all fields on which the symmetry can act and, hence, amounts to a classification of all possible objects from which a field theory which respects the symmetry can be "built up". We will discuss explicit examples of this later on. For now, let us present a few simple examples of representations. The trivial representation which exists for all groups is given by $R(g) = \mathbf{1}$, for all $g \in G$, so each group element is represented by the unit matrix in a given dimension. For the group $\mathbb{Z}_N$ and each integer q we can write down the representation

$$R_q(n) = \begin{pmatrix} \cos(2\pi qn/N) & \sin(2\pi qn/N) \\ -\sin(2\pi qn/N) & \cos(2\pi qn/N) \end{pmatrix} , \tag{4.3}$$

by real two-dimensional rotation matrices over the vector space $V = \mathbb{R}^2$, where $n = 0, \ldots, N-1$. We can restrict the value of q to the range $0, \ldots, N-1$ (as two values of q which differ by N lead to the same set of matrices) and this provides, in fact, a complete list of representations for $\mathbb{Z}_N$. Equivalently, we can write down the same representations over a one-dimensional complex vector space $V = \mathbb{C}$ where they take the form $R_q(n) = \exp(2\pi iqn/N)$. If, for a given representation R_q, we denote elements of the vector space $V = \mathbb{C}$ by Φ then the group acts on them as $\Phi \to R_q(n)\Phi = \exp(2\pi iqn/N)\Phi$. In this case, Φ is said to have "charge" $-q$ in physics language.

Representations R_q for $U(1)$ (on $V = \mathbb{C}$) are just as easily obtained by writing

$$R_q(e^{i\alpha}) = e^{iq\alpha} , \tag{4.4}$$

where $\alpha \in [0, 2\pi]$. For R_q to be continuous when going around the circle, q must be an integer, however, unlike in the $\mathbb{Z}_N$ case it is not otherwise restricted. The above representations R_q for q an arbitrary integer, in fact, provide all (continuous) representations of $U(1)$. As before, a (complex) field transforming as $\Phi \to R_q(e^{i\alpha})\Phi$ is said to have charge $-q$. Also note that charge $q = 0$ corresponds to the trivial representation. The groups $\mathrm{SU}(n)$ are already given by matrices, so we can think about them as representing themselves. This representation is n-complex dimensional and is also called the *fundamental representation*. On a complex vector $\Phi = (\phi_1, \ldots, \phi_n)$ is acts as $\Phi \to U\Phi$, where $U \in \mathrm{SU}(n)$. However, this is by no means the only representation of $\mathrm{SU}(n)$, in fact, there is an infinite number of them, as we will see.

There are a number of general ways of constructing new representations from old ones which should be mentioned. For a representation $R : G \to \mathrm{Gl}(n)$ of a group G there is a *complex conjugate representation* $R^\star$ defined by $R^\star(g) = R(g)^\star$, that is, each group element is now represented by the complex conjugate of the original representation matrix. Applying this to $\mathrm{SU}(n)$ leads to the *complex conjugate of the fundamental representation* $U \to U^\star$. For two representations R_1 and R_2 of a group G with dimensions n_1 and n_2 one can consider the *direct sum representation* $R_1 \oplus R_2$ with dimension $n_1 + n_2$ defined by the block-diagonal matrices

$$(R_1 \oplus R_2)(g) = \begin{pmatrix} R_1(g) & 0 \\ 0 & R_2(g) \end{pmatrix} . \tag{4.5}$$

A representation such as this is called *reducible* and, conversely, a representation which cannot be split into smaller blocks as in (4.5) is called *irreducible*. For example, the direct sum representation

$$R(e^{i\alpha}) = \mathrm{diag}(e^{i\alpha}, e^{-i\alpha}) \tag{4.6}$$

of $U(1)$ consisting of a charge $+1$ and -1 representation realises an explicit embedding of $U(1)$ into $\mathrm{SU}(2)$. Another, less trivial way of combining the two representations R_1 and R_2 to a new one is the *tensor representation* $R_1 \otimes R_2$ with dimension $n_1 n_2$ defined by $(R_1 \otimes R_2)(g) = R_1(g) \times R_2(g)$ [1]. In general, a tensor representation $R_1 \times R_2$ is not irreducible and decomposes into a sum of irreducible representations $R^{(i)}$, so one can write

$$R_1 \otimes R_2 = \bigoplus_i R^{(i)} \tag{4.7}$$

This is also referred to as *Clebsch-Gordon decomposition*.

4.1.3 Lie groups and Lie algebras

To understand representations of Lie groups we should look at their structure more closely. The matrices M of a Lie group form a continuous (differentiable) family $M = M(\mathbf{t})$ where $\mathbf{t} = (t^1, \ldots, t^m)$ are m real parameters and we adopt the convention that $M(\mathbf{0}) = \mathbf{1}$. An example for such a parametrisation for the case of $\mathrm{SU}(2)$ has been given in Eq. (4.2), where the three parameters are $(t^1, t^2, t^3) = (\beta_1, \beta_2, \sigma)$. Let us now look in more detail at the neighbourhood of the identity element, corresponding to small values of the parameters $\mathbf{t}$, where we can expand

$$M(\mathbf{t}) = \mathbf{1} + \sum_i t^i T_i + \mathcal{O}(\mathbf{t}^2) , \quad \text{with} \quad T_i = \frac{\partial M}{\partial t^i}(\mathbf{0}) . \tag{4.8}$$

The matrices T_i are called the *generators* of the Lie group and the vector space $\mathcal{L}(G) = \{t^i T_i\}$ spanned by these matrices is referred to as *Lie algebra*. In the case of $\mathrm{SU}(2)$, the generators are given by (i times) the Pauli matrices, as differentiating Eq. (4.2) shows. In general, there is a theorem which states that the group (or, rather, a neighbourhood of the identity of the group) can be reconstructed from the Lie algebra by the *exponential map*

$$M(\mathbf{t}) = \exp(t^i T_i) . \tag{4.9}$$

Now consider two matrices $M(\mathbf{t})$ and $M(\mathbf{s})$ and the product

$$M(\mathbf{t})^{-1} M(\mathbf{s})^{-1} M(\mathbf{t}) M(\mathbf{s}) = \mathbf{1} + \sum_{i,j} t^i s^j [T_i, T_j] + \cdots , \tag{4.10}$$

where $[\cdot, \cdot]$ is the ordinary matrix commutator. Since the product on the LHS of Eq. (4.10) is an element of the group, we conclude that the commutators $[T_i, T_j]$ must be elements of the Lie algebra and can, hence, be written as

$$[T_i, T_j] = f_{ij}{}^k T_k . \tag{4.11}$$

The coefficients $f_{ij}{}^k$ are called the *structure constants* of the Lie algebra $\mathcal{L}(G)$. More accurately, the Lie-algebra $\mathcal{L}(G)$ is then the vector space $\mathcal{L}(G) = \{t^i T_i\}$ together with the commutator bracket $[\cdot, \cdot]$. The concept of a representation can now also be defined at the level of the Lie algebra.

Definition A representation r of a Lie algebra $\mathcal{L}$ is a linear map which assigns to elements $T \in \mathcal{L}$ matrices $r(T)$ such that $[r(T), r(S)] = r([T, S])$ for all $T, S \in \mathcal{L}$.

[1] For two matrices M and N the product $M \times N$ can be thought of as the matrix obtained by replacing each entry in M by a block consisting of that entry times N. A useful property of this product is $(M_1 \times N_1)(M_2 \times N_2) = (M_1 M_2) \times (N_1 N_2)$.

Note this is equivalent to saying that the representation matrices $r(T_i)$ commute in the same way as the generators T_i, so $[r(T_i), r(T_j)] = f_{ij}{}^k r(T_k)$, with the same structure constants $f_{ij}{}^k$ as in Eq. (4.11). It is usually easier to find representations of Lie algebras than representations of groups. However, once a Lie-algebra representation has been found the associated group representation can be re-constructed using the exponential map (4.9). Concretely, for a Lie-algebra representation $T_i \to r(T_i)$ the corresponding group representation is $e^{t^i T_i} \to e^{t^i r(T_i)}$. Recall that the dimension of the representation r is defined to be the dimension of the vector space on which the representation matrices $r(T)$ (or the associated group elements obtained after exponentiating) act, that is, it is given by the size of the matrices $r(T)$. This dimension of the representation r is not to be confused with the dimension of the Lie-algebra itself, the latter being the dimension of the Lie algebra $\mathcal{L}(G)$ as a vector space of matrices.

Example $\mathrm{SU}(2)$
Let us see how all this works for our prime example $\mathrm{SU}(2)$. Consider an $\mathrm{SU}(2)$ matrix U close to the identity matrix and write [2] $U = 1 + iT + \ldots$, where T is a Lie algebra element. Then, evaluating the conditions $U^\dagger U = \mathbf{1}$ and $\det(U) = 1$ at linear level in T, one finds the constraints $T^\dagger = T$ and $\mathrm{tr}(T) = 0$. In other words, the Lie algebra $\mathcal{L}(\mathrm{SU}(2))$ of $\mathrm{SU}(2)$ consists of all traceless, hermitian 2×2 matrices. Note that this space is three-dimensional. A convenient basis of generators τ_i for this Lie algebra is obtained from the Pauli matrices σ_i. Recall that they satisfy the useful identities

$$\sigma_i \sigma_j = \delta_{ij} \mathbf{1} + i\epsilon_{ijk}\sigma_k \,, \quad [\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \,, \quad \mathrm{tr}(\sigma_i \sigma_j) = 2\delta_{ij} \,. \tag{4.12}$$

Hence, the Lie algebra of $\mathrm{SU}(2)$ is spanned by the generators

$$\tau_i = \frac{1}{2}\sigma_i \quad \text{with} \quad [\tau_i, \tau_j] = i\epsilon_{ijk}\tau_k \,, \tag{4.13}$$

and the structure constants are simply given by the Levi-Civita tensor. While the dimension of the Lie algebra $\mathcal{L}(\mathrm{SU}(2))$ is 3 (as it is spanned by three Pauli matrices), the dimension of the $\mathrm{SU}(2)$ representation defined by the Pauli matrices is 2 (since they are 2×2 matrices).

Finite $\mathrm{SU}(2)$ matrices are then obtained by exponentiating

$$U = \exp(it^i \tau_i) = \exp(it^i \sigma_i/2) \,. \tag{4.14}$$

Note that the generator τ_3 corresponds to the $U(1)$ subgroup (4.6) of $\mathrm{SU}(2)$. The commutation relations (4.13) of the $\mathrm{SU}(2)$ Lie algebra are identical to the commutation relations of the angular momentum operators in quantum mechanics. Hence, we already know that the finite-dimensional representations of this algebra can be labelled by a "spin" j, that is an integer or half-integer number $j = 0, 1/2, 1, 3/2, \ldots$. For a given j the dimension of the representation is $2j+1$ and the representation space is spanned by states $|jm\rangle$, where $m = -j, -j+1, \ldots, j-1, j$. The two-dimensional representation for $j = 1/2$ of course corresponds to the explicit representation of the algebra in terms of Pauli matrices which we have written down above. The complex conjugate of the fundamental is also a two-dimensional representation and, on purely dimensional grounds, must also be identified with the $j = 1/2$ representation.

We also know from quantum mechanics that the tensor product of two representations characterised by j_1 and j_2 decomposes into the irreducible representations with j in the range $|j_1 - j_2|, |j_1 - j_2| + 1, \ldots, j_1 + j_2$. This is an explicit example of a Clebsch-Gordon decomposition. It is customary to refer to representations by their dimensions, that is, write for example, the $j = 1/2$ representation as $\mathbf{2}$ (or $\bar{\mathbf{2}}$ for the conjugate) and the $j = 1$ representation as $\mathbf{3}$. With this notation, examples of $\mathrm{SU}(2)$ Clebsch-Gordon decompositions are

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3} \,, \qquad \mathbf{2} \otimes \mathbf{3} = \mathbf{2} \oplus \mathbf{4} \,. \tag{4.15}$$

Actions should be invariant under a symmetry group and, hence, it is of particular importance to understand the singlets which occur in a Clebsch-Gordan decomposition. They will tell us about the invariant terms which are allowed in an action. For example, if we have a field Φ which transforms as a doublet under $\mathrm{SU}(2)$, the first of Eqs. (4.15) tells us that we should be able to write a quadratic term in Φ, corresponding to the direction of the singlet on the right-hand side.

Example $\mathrm{SO}(3)$
Another important Lie group is $\mathrm{SO}(3)$, the group of three-dimensional rotations, consisting of real 3×3 matrices

[2] In the physics literature it is conventional to include a factor of i in front of the generators T.

$\det(\Lambda)$	${\Lambda^0}_0$	name	contains	given by
$+1$	≥ 1	$L_+^\uparrow$	$\mathbf{1}_4$	$L_+^\uparrow$
$+1$	≤ -1	$L_+^\downarrow$	PT	$PTL_+^\uparrow$
-1	≥ 1	$L_-^\uparrow$	P	$PL_+^\uparrow$
-1	≤ -1	$L_-^\downarrow$	T	$TL_+^\uparrow$

Table 4.1: The four disconnected components of the Lorentz group. The union $L_+ = L_+^\uparrow \cup L_+^\downarrow$ is also called the proper Lorentz group and $L^\uparrow = L_+^\uparrow \cup L_-^\uparrow$ is called the orthochronos Lorentz group (as it consists of transformations preserving the direction of time). $L_+^\uparrow$ is called the proper orthochronos Lorentz group.

O satisfying $O^T O = \mathbf{1}$ and $\det(O) = 1$. Writing $O = \mathbf{1} + iT$ with (purely imaginary) generators T, the relation $O^T O = \mathbf{1}$ implies $T = T^\dagger$ and, hence, that the Lie-algebra of $\mathrm{SO}(3)$ consists of 3×3 anti-symmetric matrices (multiplied by i). A basis for this Lie algebra is provided by the three matrices T_i defined by

$$(T_i)_{jk} = -i\epsilon_{ijk} \,, \tag{4.16}$$

which satisfy the commutation relations

$$[T_i, T_j] = i\epsilon_{ijk} T_k \,. \tag{4.17}$$

These are the same commutation relations as in Eq. (4.13) and, hence, the T_i form a three-dimensional (irreducible) representation of (the Lie algebra of) $\mathrm{SU}(2)$. This representation must fit into the above classification of $\mathrm{SU}(2)$ representations by an integer or half-integer number j and, simply on dimensional grounds, it has to be identified with the $j = 1$ representation.

4.1.4 The Lorentz group

The Lorentz group is of fundamental importance for the construction of field theories. It is the symmetry associated to four-dimensional Lorentz space-time and should be respected by field theories formulated in Lorentz space-time. Let us begin by formally defining the Lorentz group. With the Lorentz metric $\eta = \mathrm{diag}(1, -1, -1, -1)$ the Lorentz group L consists of real 4×4 matrices Λ satisfying

$$\Lambda^T \eta \Lambda = \eta \,. \tag{4.18}$$

Special Lorentz transformations are the identity $\mathbf{1}_4$, parity $P = \mathrm{diag}(1, -1, -1, -1)$, time inversion $T = \mathrm{diag}(-1, 1, 1, 1)$ and the product $PT = -\mathbf{1}_4$. We note that the four matrices $\{1_4, P, T, PT\}$ form a finite sub-group of the Lorentz group. By taking the determinant of the defining relation (4.18) we immediately learn that $\det(\Lambda) = \pm 1$ for all Lorentz transformations. Further, if we write out Eq. (4.18) with indices

$$\eta_{\mu\nu} {\Lambda^\mu}_\rho {\Lambda^\nu}_\sigma = \eta_{\rho\sigma} \tag{4.19}$$

and focus on the component $\rho = \sigma = 0$ we conclude that $({\Lambda^0}_0)^2 = 1 + \sum_i ({\Lambda^i}_0)^2 \geq 1$, so either ${\Lambda^0}_0 \geq 1$ or ${\Lambda^0}_0 \leq -1$. This sign choice for ${\Lambda^0}_0$ combined with the choice for $\det(\Lambda)$ leads to four classes of Lorentz transformations which are summarised in Table 4.1. Also note that the Lorentz group contains three-dimensional rotations since matrices of the form

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & O \end{pmatrix} \tag{4.20}$$

satisfy the relation (4.18) and are hence special Lorentz transformations as long as O satisfies $O^T O = \mathbf{1}_3$.

To find the Lie algebra of the Lorentz group we write $\Lambda = \mathbf{1}_4 + iT + \ldots$ with purely imaginary 4×4 generators T. The defining relation (4.18) then implies for the generators that $T = -\eta T^T \eta$, so T must be anti-symmetric in the space-space components and symmetric in the space-time components. The space of such matrices is six-dimensional and spanned by

$$J_i = \begin{pmatrix} 0 & 0 \\ 0 & T_i \end{pmatrix} \,,\; K_1 = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \,,\; K_2 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \,,\; K_3 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \,, \tag{4.21}$$

(j_+, j_-)	dimension	name	symbol
$(0,0)$	1	scalar	ϕ
$(1/2,0)$	2	left-handed Weyl spinor	χ_L
$(0,1/2)$	2	right-handed Weyl spinor	χ_R
$(1/2,0)\oplus(0,1/2)$	4	Dirac spinor	ψ
$(1/2,1/2)$	4	vector	A_μ

Table 4.2: Low-dimensional representations of the Lorentz group.

where T_i are the generators (4.16) of the rotation group. Given the embedding (4.20) of the rotation group into the Lorentz group the appearance of the T_i should not come as a surprise. It is straightforward to work out the commutation relations

$$[J_i, J_j] = i\epsilon_{ijk}J_k\,, \quad [K_i, K_j] = -i\epsilon_{ijk}J_k\,, \quad [J_i, K_j] = i\epsilon_{ijk}K_k\,. \tag{4.22}$$

The above matrices can also be written in a four-dimensional covariant form by introducing six 4×4 matrices $\sigma_{\mu\nu}$, labelled by two anti-symmetric four-indices and defined by

$$(\sigma_{\mu\nu})^\rho{}_\sigma = i(\eta^\rho_\mu\eta_{\nu\sigma} - \eta_{\mu\sigma}\eta^\rho_\nu)\,. \tag{4.23}$$

By explicit computation one finds that $J_i = \frac{1}{2}\epsilon_{ijk}\sigma_{jk}$ and $K_i = \sigma_{0i}$. Introducing six independent parameters $\epsilon^{\mu\nu}$, labelled by an anti-symmetric pair of indices, a Lorentz transformation close to the identity can be written as

$$\Lambda^\rho{}_\sigma \simeq \delta^\rho{}_\sigma - \frac{i}{2}\epsilon^{\mu\nu}(\sigma_{\mu\nu})^\rho{}_\sigma = \delta^\rho{}_\sigma + \epsilon^\rho{}_\sigma;\,. \tag{4.24}$$

The commutation relations (4.22) for the Lorentz group are very close to the ones for $\mathrm{SU}(2)$ in Eq. (4.13). This analogy can be made even more explicit by introducing a new basis of generators

$$J_i^\pm = \frac{1}{2}(J_i \pm iK_i)\,. \tag{4.25}$$

In terms of these generators, the algebra (4.22) takes the form

$$[J_i^\pm, J_j^\pm] = i\epsilon_{ijk}J_k^\pm\,, \quad [J_i^+, J_j^-] = 0\,, \tag{4.26}$$

that is, precisely the form of two copies (a direct sum) of two $\mathrm{SU}(2)$ Lie-algebras. Irreducible representations of the Lorentz group can therefore be labelled by a pair (j_+, j_-) of two spins and the dimension of these representations is $(2j_+ +1)(2j_- +1)$. A list of a few low-dimensional Lorentz-group representations is provided in Table 4.2. Field theories in Minkowski space usually require Lorentz invariance and, hence, the Lorentz group is of fundamental importance for such theories. Since it is related to the symmetries of space-time it is often also referred as *external symmetry* of the theory. The classification of Lorentz group representations in Table 4.2 provides us with objects which transform in a definite way under Lorentz transformations and, hence, are the main building blocks of such field theories. In these lectures, we will not consider spinors in any more detail but focus on scalar fields ϕ, transforming as singlets, $\phi \to \phi$ under the Lorentz group, and vector fields A_μ, transforming as vectors, $A_\mu \to \Lambda_\mu{}^\nu A_\nu$.

4.2 General classical field theory

4.2.1 Lagrangians and Hamiltonians in classical field theory

In this subsection, we develop the general Lagrangian and Hamiltonian formalism for classical field theories. This formalism is in many ways analogous to the Lagrangian and Hamiltonian formulation of classical mechanics. In classical mechanics the main objects are the generalised coordinates $q_i = q_i(t)$ which depend on time only. Here, we will instead be dealing with fields, that is functions of four-dimensional coordinates $x = (x^\mu)$ on Minkowski space. Lorentz indices $\mu, \nu, \cdots = 0, 1, 2, 3$ are lowered and raised with the Minkowski metric $(\eta_{\mu\nu}) = \mathrm{diag}(1, -1, -1, -1)$ and its inverse $\eta^{\mu\nu}$. For now we will work with a generic set of fields $\phi_a = \phi_a(x)$ before discussing scalar and vector fields in more detailed in subsequent sections. Recall that the Lagrangian in

classical mechanics is a function of the generalised coordinates and their first (time) derivatives. Analogously, we start with a field theory *Lagrangian density* $\mathcal{L} = \mathcal{L}(\phi_a, \partial_\mu \phi_a)$ which is a function of the fields ϕ_a and their first space-time derivatives $\partial_\mu \phi_a$. The field theory action can then be written as

$$S = \int d^4x \, \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x)) \; , \tag{4.27}$$

where the integration ranges over all of Minkowski space. Our first task is to derive the Euler-Lagrange equations for such a general field theory by applying the variational principle to the above action. One finds

$$\begin{aligned} 0 &= \frac{\delta S}{\delta \phi_a(x)} = \frac{\delta}{\delta \phi_a(x)} \int d^4\tilde{x} \, \mathcal{L}(\phi_b(\tilde{x}), \partial_\mu \phi_b(\tilde{x})) = \int d^4\tilde{x} \, \left[\frac{\partial \mathcal{L}}{\partial \phi_b} \frac{\delta \phi_b(\tilde{x})}{\delta \phi_a(x)} + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_b)} \frac{\delta \partial_\mu \phi_b(\tilde{x})}{\delta \phi_a(x)} \right] & (4.28) \\ &= \int d^4\tilde{x} \, \left[\frac{\partial \mathcal{L}}{\partial \phi_b} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_b)} \right] \frac{\delta \phi_b(\tilde{x})}{\delta \phi_a(x)} = \left(\frac{\partial \mathcal{L}}{\partial \phi_a} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \right)(x) & (4.29) \end{aligned}$$

where we have used the generalisation of Eq. (1.17)

$$\frac{\delta \phi_b(\tilde{x})}{\delta \phi_a(x)} = \delta_b^a \delta^4(x - \tilde{x}) \tag{4.30}$$

in the last step. Further, we have assumed that the boundary term which arises from the partial integration in the second last step vanishes due to a vanishing variation at infinity. Hence, the Euler-Lagrange equations for general field theories take the form

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} - \frac{\partial \mathcal{L}}{\partial \phi_a} = 0 \; . \tag{4.31}$$

With the *conjugate momenta* defined by

$$\pi^a = \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi_a)} \tag{4.32}$$

the *Hamiltonian density* $\mathcal{H}$ and the *Hamiltonian* H can, in analogy with classical mechanics, be written as

$$\mathcal{H} = \pi^a \partial_0 \phi_a - \mathcal{L} \; , \quad H = \int d^3x \, \mathcal{H} \; . \tag{4.33}$$

4.2.2 Noether's theorem

In classical mechanics, *Noether's theorem* relates symmetries and conserved quantities of a theory. We will now derive the field theory version of this theorem. Let us start with a general set of infinitesimal transformations parameterised by small, continuous parameters ϵ^α and acting an the fields by

$$\phi_a(x) \to \phi'_a(x) = \phi_a(x) + \Phi_{a\alpha}(x) \epsilon^\alpha \; , \tag{4.34}$$

where $\Phi_{a\alpha}(x)$ are functions which encode the particular type of symmetry action. We assume that these transformations leave the action (4.27) invariant and, hence, change the Lagrangian density by total derivatives only. This means the transformation of the Lagrangian density is of the form

$$\mathcal{L} \to \mathcal{L} + \epsilon^\alpha \partial_\mu \Lambda^\mu_\alpha(x) \; , \tag{4.35}$$

with certain functions $\Lambda^\mu_\alpha(x)$ which can be computed for each type of symmetry action. Let us now compare this variation of $\mathcal{L}$ with the one induced by transformation (4.34) of the fields. One finds

$$\begin{aligned} \mathcal{L} &\to \mathcal{L} + \frac{\partial \mathcal{L}}{\partial \phi_a} \Phi_{a\alpha} \epsilon^\alpha + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \partial_\mu \Phi_{a\alpha} \epsilon^\alpha & (4.36) \\ &= \mathcal{L} + \epsilon^\alpha \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \Phi_{a\alpha} \right) - \epsilon^\alpha \left(\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} - \frac{\partial \mathcal{L}}{\partial \phi_a} \right) \Phi_{a\alpha} & (4.37) \end{aligned}$$

The last term vanishes thanks to the Euler-Lagrange equations (4.31) and equating the remaining variation of $\mathcal{L}$ with the one in Eq. (4.35) it follows that

$$\partial_\mu j^\mu_\alpha = 0 \quad \text{where} \quad j^\mu_\alpha = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \Phi_{a\alpha} - \Lambda^\mu_\alpha \; . \tag{4.38}$$

Hence, for each symmetry generator ϵ^α we obtain a *conserved current* j^μ_α, that is, a current with vanishing divergence. Each such current can be used to define a *conserved charge* Q_α by

$$Q_\alpha = \int d^3x\, j^0_\alpha \tag{4.39}$$

Using Eq. (4.38) and assuming that fields fall off sufficiently rapidly at infinity it follows that

$$\dot{Q}_\alpha = \int d^3x\, \partial_0 j^0_\alpha = \int d^3x \partial_i j_{\alpha i} = 0\ , \tag{4.40}$$

and, hence, that the charges Q_α are indeed time-independent.

4.2.3 Translation invariance and energy-momentum tensor

Let us apply Noether's theorem to the case of translations in space-time, acting on the fields as

$$\phi_a(x) \to \phi_a(x+a) = \phi_a(x) + a^\nu \partial_\nu \phi_a(x)\ . \tag{4.41}$$

The role of the symmetry parameters ϵ^α is here played by the infinitesimal translations a^ν. Therefore, the index α which appears in the general equations above becomes a space-time index ν. Under a translation the Lagrangian density changes as

$$\mathcal{L} \to \mathcal{L} + a^\mu \partial_\mu \mathcal{L} = \mathcal{L} + a^\nu \partial_\mu \left(\delta^\mu_\nu \mathcal{L}\right)\ . \tag{4.42}$$

Comparing the last two equations with the general formulae (4.34) and (4.35) we learn that $\Phi_{a\nu} = \partial_\nu \phi_a$ and $\Lambda^\mu_\nu = \delta^\mu_\nu \mathcal{L}$. Inserting this into the general result (4.38) leads to four currents $T^\mu{}_\nu = j^\mu_\nu$ given by

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \partial_\nu \phi_a - \delta^\mu{}_\nu \mathcal{L}\ . \tag{4.43}$$

For a translation-invariant theory they are conserved, that is, they satisfy $\partial_\mu T^\mu{}_\nu = 0$. The tensor $T^\mu{}_\nu$ is called the *energy-momentum* tensor and its associated charges

$$P_\nu = \int d^3x\, T^0{}_\nu \tag{4.44}$$

represent the conserved energy and momentum of the system. In particular, the conserved energy P_0 is explicitly given by

$$P_0 = \int d^3x \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \phi_a)} \partial_0 \phi_a - \mathcal{L}\right) = \int d^3x\, \mathcal{H} = H\ , \tag{4.45}$$

that is, by the Hamiltonian (4.33).

4.2.4 How to construct classical field theories

Before we move on to examples, it may be useful to present a "recipe" for how to construct explicit field theories. The standard steps involved are:

- Choose a group which corresponds to the symmetries of the theory. Normally, the symmetries include the *external symmetry*, that is Lorentz symmetry. In addition, there may be *internal symmetries* which do not act on space-time indices but internal indices. (We will see explicit examples of such internal symmetries shortly.)

- Choose a set of representations of the symmetry group. This fixes the field content of the theory and the transformation properties of the fields.

- Write down the most general action invariant under the chosen symmetry (with at most two derivatives in each term) for the fields selected in the previous step. Normally, only polynomial terms in the fields are considered and an upper bound on the degree of the polynomials is imposed (for example by requiring that the theory does not contain (coupling) constants with negative energy dimension).

4.3 Scalar field theory

4.3.1 A single real scalar field

Lagrangian and equations of motion
In a Lorentz invariant field theory, the simplest choice of field content is that of a single real scalar field $\phi = \phi(x)$, which corresponds to a single representation of the Lorentz group with $(j_+, j_-) = (0, 0)$. The Lagrangian density for this theory is given by

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) \ , \tag{4.46}$$

where the first term is referred to as *kinetic energy* and $V = V(\phi)$ is the *scalar potential*. Let us discuss the dimensions of the various objects in this Lagrangian. The standard convention in particle physics is to set $\hbar = c = 1$, so that both time and space are measured in terms of inverse energy units and the action has to be dimension-less. With this convention, the measure d^4x has -4 energy units and hence, for the action to be dimensionless, we need each term in the Lagrangian density $\mathcal{L}$ to be of energy dimension $+4$. Given that the derivatives ∂_μ have dimension one, the scalar field must have dimension one as well, so that the kinetic energy term has overall dimension 4. Then, for a monomial term $\lambda_n\phi^n$ with coupling λ_n in the scalar potential to be of dimension 4 the coupling λ_n must have dimension $4-n$. If we want to avoid couplings with negative energy dimensions (which normally cause problems in the associated quantum theory) we need to restrict $n \leq 4$ and, hence, the scalar potential has the form

$$V = \frac{1}{2}m^2\phi^2 + \frac{1}{3!}\lambda_3\phi^3 + \frac{1}{4!}\lambda\phi^4 \ . \tag{4.47}$$

(A possible linear term in ϕ can be removed through a re-definition of ϕ by a shift.) Note that m and λ_3 have dimension 1 and λ is dimensionless. The quadratic term in V is called a *mass term* with mass m and the other terms represent couplings. Applying the Euler-Lagrange equations (4.31) to the above Lagrangian leads to the equation of motion

$$\Box\phi + V'(\phi) = 0 \tag{4.48}$$

for ϕ, where $\Box = \partial_\mu\partial^\mu$ and the prime denotes the derivative with respect to ϕ. For non-vanishing λ_3 or λ solutions to this equation are not easy to find.

The free equation of motion and its general solution
In the free case, that is for couplings $\lambda_3 = \lambda = 0$, the equation of motion reduces to the so-called *Klein-Gordon equation*

$$(\Box + m^2)\phi = 0 \ , \tag{4.49}$$

for which a general solution can be written down. To do this we insert the Fourier transform

$$\phi(x) = \int d^4k\, e^{-ikx}\tilde{\phi}(k) \tag{4.50}$$

of ϕ into the Klein-Gordon equation, resulting in

$$(\Box + m^2)\phi = -\int d^4x\, e^{-ikx}(k^2 - m^2)\tilde{\phi}(k) = 0 \ . \tag{4.51}$$

Since the Fourier transform can be inverted we conclude that $(k^2 - m^2)\tilde{\phi}(k) = 0$ and, hence, that $\tilde{\phi}$ can be written in the form $\tilde{\phi}(k) = \delta(k^2 - m^2)\tilde{\varphi}(k)$ for some function $\tilde{\varphi}$. Inserting this result for $\tilde{\phi}$ into the Fourier transform (4.50) and using [3]

$$\delta(k^2 - m^2) = \frac{1}{2w_{\mathbf{k}}}\left(\delta(k_0 - w_{\mathbf{k}}) + \delta(k_0 + w_{\mathbf{k}})\right) \ , \quad w_{\mathbf{k}} = \sqrt{\mathbf{k}^2 + m^2} \tag{4.52}$$

one finds, after integrating over k_0, that

$$\phi(x) = \int d^3\tilde{k}\left(a_+(k)e^{-ikx} + a_-^\star(k)e^{ikx}\right) \ . \tag{4.53}$$

[3] This follows from the well-known delta-function identity $\delta(f(x)) = \sum_{x_0:f(x_0)=0}\frac{1}{|f'(x_0)|}\delta(x - x_0)$.

where we have defined the measure

$$d^3\tilde{k} = \frac{d^3k}{(2\pi)^3 2w_{\mathbf{k}}} \tag{4.54}$$

which is Lorentz invariant as a consequence of Eq. (4.52). Further, the coefficients are defined as $a_+(k) = (2\pi)^3\tilde{\varphi}(w_{\mathbf{k}}, \mathbf{k})$ and $a_-(k) = (2\pi)^3\tilde{\varphi}(-w_{\mathbf{k}}, \mathbf{k})^\star$ and the four-vector k in the exponents is now understood as $(k_\mu) = (w_{\mathbf{k}}, \mathbf{k})$. We note from Eq. (4.54) that

$$2w_{\mathbf{k}}\delta^3(\mathbf{k} - \mathbf{q}) \ . \tag{4.55}$$

is Lorentz invariant as well and can be viewed as a covariant version of the three-dimensional delta function. Up to this point we have, effectively, solved the Klein-Gordon equation for a complex scalar field. However, imposing a reality condition on the solutions is easy and leads to $a_+(k) = a_-(k) \equiv a(k)$. The final result for the general solution of the Klein-Gordon equation for a real scalar field then reads

$$\phi(x) = \int d^3\tilde{k}\left(a(k)e^{-ikx} + a^\star(k)e^{ikx}\right) \ . \tag{4.56}$$

Hamiltonian, symmetries and conserved currents

We now return to the development of the general formalism. From Eq. (4.32) the conjugate momentum π is given by

$$\pi = \partial_0\phi \tag{4.57}$$

and, using Eq. (4.33), this implies the Hamiltonian density

$$\mathcal{H} = \frac{1}{2}\pi^2 + \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + V \ . \tag{4.58}$$

For the stress energy tensor we find by inserting into Eq. (4.43)

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}\eta_{\mu\nu}\partial_\rho\phi\partial^\rho\phi + \eta_{\mu\nu}V \ . \tag{4.59}$$

In accordance with the general formula (4.45) we therefore find for the energy

$$P_0 = \int d^3x\, T_{00} = \int d^3x\, \left(\frac{1}{2}\pi^2 + \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + V\right) = \int d^3x\, \mathcal{H} \ . \tag{4.60}$$

Our theory is also invariant under Lorentz transformations and from Noether's theorem we expect associated conserved currents which we will now derive. First, recall from Eq. (4.24) that an infinitesimal Lorentz transformation on x^μ can be written as $x^\mu \to x^\mu + \epsilon^\mu{}_\nu x^\nu$ where $\epsilon_{\mu\nu}$ is anti-symmetric. On the field ϕ this transformation acts as

$$\phi(x) \to \phi(x - \epsilon x) = \phi(x) + \epsilon^{\mu\nu}x_\mu\partial_\nu\phi(x) \tag{4.61}$$

and a similar transformation law holds for the Lagrangian density

$$\mathcal{L} \to \mathcal{L} + \epsilon^{\mu\nu}\partial_\rho\left(\delta^\rho_\nu x_\mu\mathcal{L}\right) \ . \tag{4.62}$$

Comparing with Eqs. (4.34) and (4.35) we learn that the symmetry parameters ϵ^α are here given by $\epsilon^{\mu\nu}$, so we have to replace the index α in our general equations with an anti-symmetric pair of space-time indices. Further, we have $\Phi_{\mu\nu} = 2x_{[\mu}\partial_{\nu]}\phi$ and $\Lambda^\rho_{\mu\nu} = 2x_{[\mu}\delta^\rho_{\nu]}\mathcal{L}$. Inserting this into Eq. (4.38) we find the conserved currents $M^\rho_{\mu\nu}$ given by

$$M^\rho_{\mu\nu} = x_\mu T^\rho_\nu - x_\nu T^\rho_\mu \ , \tag{4.63}$$

with the energy momentum tensor for a scalar field theory defined in Eq. (4.59). From Noether's theorem we know that these currents are divergence-free, $\partial_\rho M^\rho_{\mu\nu} = 0$, and imply the existence of conserved charges

$$M_{\mu\nu} = \int d^3x\, M^0_{\mu\nu} = \int d^3x\, \left(x_\mu T^0_\nu - x_\nu T^0_\mu\right) \ . \tag{4.64}$$

They can be interpreted as the conserved angular momentum of the theory.

The $\mathbb{Z}_2$ symmetric theory and spontaneous symmetry breaking
So far, we have only imposed external symmetries on the scalar field theory. An internal symmetry which may be considered is a $\mathbb{Z}_2$ symmetry which acts as $\phi(x) \to -\phi(x)$. This transformation leaves all terms except the cubic one in the Lagrangian (4.46), (4.47) invariant. Hence, if we impose this symmetry on our theory the cubic term in the scalar potential has to be dropped for the Lagrangian to be invariant and we are left with [4]

$$V = V_0 + \frac{1}{2}m^2\phi^2 + \frac{1}{4!}\lambda\phi^4 \ . \tag{4.65}$$

In the following, we assume invariance under this $\mathbb{Z}_2$ symmetry and work with the scalar potential (4.65).

Let us discuss the simplest type of solutions to the theory, namely solutions for which $\phi(x) = v$ takes on a constant value v, independent of space-time. From Eq. (4.48) such constant fields solve the equation of motion if

$$V'(v) = 0 \ , \tag{4.66}$$

so we are instructed to look at extrema of the scalar potential V. In fact, to minimise the energy (4.60) we should be focusing on minima of the scalar potential V. We will also refer to such a solution of the classical theory as a *vacuum*. If the quartic coupling λ is negative the scalar potential is unbounded from below and the energy of a constant field configuration tends to minus infinity for large field values. To avoid such an unphysical situation we assume that $\lambda > 0$ in the following. Then we should distinguish two cases which are illustrated in Fig. 4.1.

- $m^2 \geq 0$: In this case there is a single minimum at $\phi = v = 0$. This solution is mapped into itself under the action $\phi \to -\phi$ of the $\mathbb{Z}_2$ symmetry and we say that the symmetry is unbroken in this vacuum.
- $m^2 < 0$: In this case, $\phi = 0$ is a maximum of the potential and there are two minima at

$$\phi = v = \pm\sqrt{\frac{-6m^2}{\lambda}} \ . \tag{4.67}$$

  Neither minimum is left invariant under the $\mathbb{Z}_2$ action $\phi \to -\phi$ (in fact the two minima are mapped into each other under $\mathbb{Z}_2$) and we say that the symmetry is *spontaneously broken*. In general, spontaneous breaking of a symmetry refers to a situation where a symmetry of a theory is partially or fully broken by a vacuum solution of the theory. The potential value at the minima is given by

$$V(v) = V_0 + \frac{1}{4}m^2v^2 = V_0 - \frac{1}{24}\lambda v^4 \ . \tag{4.68}$$

  Just as the constant V_0 which we have included earlier, the potential value at the minima does not affect any of the physics discussed so far. However, if we couple our theory to gravity, it turns out that $V(v)$ acts like a cosmological constant Λ in the Einstein equations. Cosmological constraints tell us that Λ cannot be much bigger than $\mathcal{O}(\mathrm{meV}^4)$. On the other hand, there is no obvious constraint on $V(v)$. Unless there is a cancellation of the two terms in Eq. (4.68), one would expect $V(v)$ to be of the order of the symmetry breaking scale v to the fourth power. Electroweak symmetry is broken spontaneously by a mechanism similar to the above (and we will study a model related to this in Section (4.3.4)) at a scale of $v \sim \mathrm{TeV}$. Hence, the "natural" cosmological constant which arises at electroweak symmetry breaking is about 60 orders of magnitude larger than the observational limit. So, we have to assume that the two terms in Eq. (4.68) cancel each other to a precision of 60 digits. This enormous "fine tuning" is one of the manifestations of what is referred to as the *cosmological constant problem*. The question of why the cosmological constant is as small as it is is one of the most important unresolved problems in modern physics.

4.3.2 Complex scalar field with $\mathrm{U}(1)$ symmetry

Lagrangian and equations of motion
The next simplest scalar field theory is one for two real scalar fields ϕ_1 and ϕ_2. In this case, a more interesting symmetry can be imposed on the theory, namely an $\mathrm{SO}(2)$ symmetry under which the doublet (ϕ_1, ϕ_2) transforms with the charge q representation as

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \to \begin{pmatrix} \cos(q\alpha) & \sin(q\alpha) \\ -\sin(q\alpha) & \cos(q\alpha) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \ . \tag{4.69}$$

[4] For our subsequent discussion, we add a constant V_0 to the potential. This constant does not affect the ϕ equations of motion.

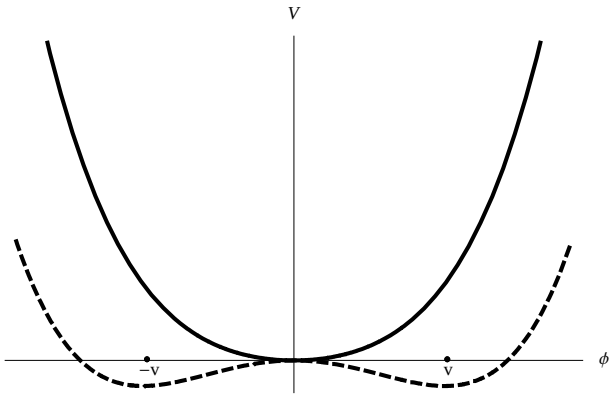

Figure 4.1: Shape of scalar potential (4.65) for $m^2 \geq 0$ (solid line) and $m^2 < 0$ (dashed line). In the latter case the position v of the minima is given by Eq. (4.67).

Such a symmetry, which does not act on space-time indices but on internal indices is also called an *internal symmetry*. For now, we will study the case of *global internal symmetries*, that is symmetries for which the transformation is the same everywhere in space-time. This means the group parameter $\alpha \in [0, 2\pi]$ is independent of the space-time coordinates x^μ. To discuss the two-scalar field theory with this global $\mathrm{SO}(2)$ symmetry and scalar charge q explicitly it proves convenient to arrange the two real scalars into a single complex one

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \; . \tag{4.70}$$

On this complex scalar, the $\mathrm{SO}(2)$ symmetry acts via the charge q representation (4.4) of $\mathrm{U}(1)$, that is

$$\phi \to \exp(-iq\alpha)\phi \; . \tag{4.71}$$

The complex conjugate $\phi^\star$ transforms as $\phi^\star \to \exp(iq\alpha)\phi^\star$ and, hence, corresponds to a representation with charge $-q$. Allowed terms in the Lagrangian density have to be $\mathrm{U}(1)$ invariant which is equivalent to saying that their total charge needs to be zero. For example, the term ϕ^2 has total charge $2q$ and cannot appear while the term $\phi^\star\phi$ has charge zero and is allowed. In general, we can only allow terms with the same number of ϕ and $\phi^\star$, so that the general $\mathrm{U}(1)$ invariant Lagrangian density reads

$$\mathcal{L} = \partial_\mu\phi^\star\partial^\mu\phi - V(\phi, \phi^\star) \; , \quad V = V_0 + m^2\phi^\star\phi + \frac{\lambda}{4}(\phi^\star\phi)^2 \; . \tag{4.72}$$

Note that it is essential for the invariance of the kinetic term that the group parameter α is space-time independent, that is, that the symmetry is global. For the equation of motion for ϕ we find from the Euler-Lagrange equation (4.31)

$$\Box\phi + \frac{\partial V}{\partial\phi^\star} = \Box\phi + m^2\phi + \frac{\lambda}{2}(\phi^\star\phi)\phi = 0 \; . \tag{4.73}$$

For $\lambda = 0$ this is the Klein-Gordan equation for a complex scalar field whose general solution has already been obtained in Eq. (4.53).

Hamiltonian and conserved currents
For the conjugate momenta one finds

$$\pi = \frac{\partial\mathcal{L}}{\partial(\partial_0\phi)} = \partial_0\phi^\star \; , \quad \pi^\star = \frac{\partial\mathcal{L}}{\partial(\partial_0\phi^\star)} = \partial_0\phi \; . \tag{4.74}$$

and, hence, the Hamiltonian density reads

$$\mathcal{H} = \pi\partial_0\phi + \pi^\star\partial_0\phi^\star - \mathcal{L} = \pi^\star\pi + \boldsymbol{\nabla}\phi^\star\cdot\boldsymbol{\nabla}\phi + V(\phi, \phi^\star) \; . \tag{4.75}$$

Being translation and Lorentz invariant the above theory has conserved energy-momentum and angular momentum tensors which can be obtained in complete analogy with the single scalar field case in sub-section 4.3.1. In addition,

the presence of the internal $\mathrm{U}(1)$ symmetry leads to a new type of conserved current which we will now derive. From Eq. (4.71), infinitesimal $\mathrm{U}(1)$ transformations are given by

$$\phi \to \phi - iq\alpha\phi \ , \quad \phi^\star \to \phi^\star + iq\alpha\phi^\star \ . \tag{4.76}$$

Comparing with the general transformation (4.34) we conclude that α plays the role of the (single) symmetry parameter and $\Phi = -iq\phi$, $\Phi^\star = iq\phi^\star$. Since the Lagrangian density is invariant under $\mathrm{U}(1)$ the total derivative terms in Eq. (4.35) vanish and we can set Λ^μ to zero. Inserting this into the general formula (4.38) for the conserved current we find

$$j_\mu = iq(\phi^\star \partial_\mu \phi - \phi \partial_\mu \phi^\star) \ . \tag{4.77}$$

Spontaneous symmetry breaking

As we did before, we would now like to discuss the vacua of the theory, that is solutions to the equation of motion (4.73) with $\phi = v = \mathrm{const}$. For $m^2 \geq 0$ there is a single minimum at $\phi = 0$. This solution is left invariant by the transformations (4.71) and, hence, the $\mathrm{U}(1)$ symmetry is unbroken in this case. For $m^2 < 0$ the shape of the potential is shown in Fig. (4.2). In this case, there is a whole circle of minima

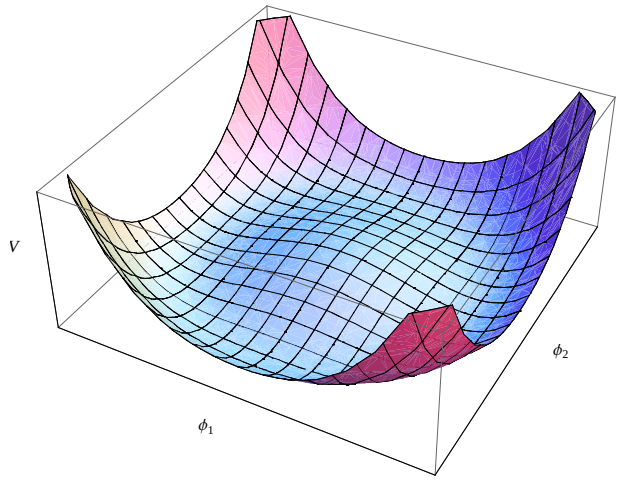

Figure 4.2: Shape of scalar potential (4.72) for $m^2 < 0$.

$$v = \frac{1}{\sqrt{2}} v_0 e^{i\nu} \ , \quad v_0 = \sqrt{-\frac{4m^2}{\lambda}} \ , \tag{4.78}$$

where ν is an arbitrary phase. The existence of this one-dimensional degenerate space of vacua is not an accident but originates from the invariance of the scalar potential under $\mathrm{U}(1)$ transformations

$$V(\phi, \phi^\star) = V(e^{iq\alpha}\phi, e^{-iq\alpha}\phi^\star) \ . \tag{4.79}$$

Indeed, this invariance implies that for every minimum ϕ of V also $e^{iq\alpha}\phi$ is a minimum for arbitrary α. Every particular choice of minimum transforms non-trivially under (4.71) and, hence, the $\mathrm{U}(1)$ symmetry is spontaneously broken. Let us, for convenience, choose the minimum on the ϕ_1 axis (setting the phase $\nu = 0$) so $\phi = v_0/\sqrt{2}$. Around this point we expand the field as

$$\phi = \frac{1}{\sqrt{2}}(v_0 + \varphi_1 + i\varphi_2) \ , \tag{4.80}$$

where φ_1 and φ_2 are small. Inserting this into the potential (4.72) we find

$$V = V_0 + \frac{1}{4} m^2 v_0^2 - 2m^2 \varphi_1^2 + \mathcal{O}(\varphi_1^3, \varphi_2^3) \ . \tag{4.81}$$

This shows that in this vacuum φ_1 is massive with mass $2m$ and φ_2 is massless. This could have been expected as φ_2 corresponds to the direction along the circle of minima, while φ_1 is perpendicular to it. It is, therefore, clear that the appearance of the massless mode φ_2 is directly related to the existence of a circle of minima and, hence, to the spontaneous break-down of the $\mathrm{U}(1)$ symmetry. The appearance of massless scalars for spontaneously broken global symmetries is a general feature known as *Goldstone's theorem* and the corresponding massless scalars are also called *Goldstone bosons*. We will now study this phenomenon in a more general setting.

4.3.3 Spontaneously broken global symmetries and Goldstone's theorem

Let us consider a general scalar field theory with a set of scalar fields $\boldsymbol{\phi} = (\phi^a) = (\phi^1, \ldots, \phi^n)$ and scalar potential $V = V(\boldsymbol{\phi})$. Consider a minimum $\mathbf{v} = (v_1, \ldots, v_n)$ of V, that is a solution of $\frac{\partial V}{\partial \phi^a}(\mathbf{v}) = 0$. Around such a minimum we can expand the potential as

$$V = V(\mathbf{v}) + \frac{1}{2} M_{ab} \boldsymbol{\varphi}^a \boldsymbol{\varphi}^b + \mathcal{O}(\boldsymbol{\varphi}^3) \,, \tag{4.82}$$

where $\boldsymbol{\varphi} = \boldsymbol{\phi} - \mathbf{v}$ and the *mass matrix* M_{ab} is defined by

$$M_{ab} = \frac{\partial^2 V}{\partial \phi^a \partial \phi^b}(\mathbf{v}) \,. \tag{4.83}$$

The eigenvalues of the mass matrix M are the mass squares of the fields around the vacuum $\mathbf{v}$. Now let us assume that our scalar field theory is invariant under a continuous symmetry group G and that the scalar fields transform as $\boldsymbol{\phi} \to R(g)\boldsymbol{\phi}$ under the representation R of G. In particular, this means the scalar potential is invariant, that is

$$V(\boldsymbol{\phi}) = V(R(g)\boldsymbol{\phi}) \tag{4.84}$$

for all $g \in G$. The vacuum $\mathbf{v}$ will in general not respect the full symmetry group G but will spontaneously break it to a sub-group $H \subset G$, so that $R(g)\mathbf{v} = \mathbf{v}$ for $g \in H$ and $R(g)\mathbf{v} \neq \mathbf{v}$ for $g \notin H$. Now introduce infinitesimal transformations $R(g) \simeq \mathbf{1} + it^I T_I$ with generator T_I (in the representation R) and small parameters t^I. We can split these generators into two sets, $\{T_I\} = \{H_i, S_\alpha\}$, where H_i are the generators of the unbroken sub-group H and S_α are the remaining generators corresponding to the broken part of the group. Hence, these two types of generators can be characterised by

$$H_i \mathbf{v} = 0 \,, \quad S_\alpha \mathbf{v} \neq 0 \,. \tag{4.85}$$

Now, write down the infinitesimal version of Eq. (4.84)

$$V(\boldsymbol{\phi}) = V(\boldsymbol{\phi} - it^I T_I \boldsymbol{\phi}) = V(\boldsymbol{\phi}) - it^I \left(\frac{\partial V}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi})\right)^T T_I \boldsymbol{\phi} \,, \tag{4.86}$$

differentiate one more time with respect to $\boldsymbol{\phi}$ and evaluate the result at $\boldsymbol{\phi} = \mathbf{v}$ using that $\frac{\partial V}{\partial \phi^a}(\mathbf{v}) = 0$. This leads to

$$M T_I \mathbf{v} = 0 \,, \tag{4.87}$$

where M is the mass matrix defined above. Every broken generator S_α satisfies $S_\alpha \mathbf{v} \neq 0$ and, hence, leads to an eigenvector of the mass matrix with eigenvalue zero. In other words, every broken generator leads to one massless scalar which is precisely the statement of Goldstone's theorem.

4.3.4 Scalar field theory with symmetry $\mathrm{SU}(2) \times \mathrm{U}(1)$

It may be usefult to illustrate Goldstone's theorem with a less trivial example based on the symmetry $\mathrm{SU}(2) \times \mathrm{U}_Y(1)$. Consider a scalar field theory with an $\mathrm{SU}(2)$ doublet $\boldsymbol{\phi}$ of complex scalar fields which, in addition, carry charge $1/2$ under a $\mathrm{U}(1)$ symmetry. A general $\mathrm{SU}(2) \times \mathrm{U}_Y(1)$ transformation of the scalar field $\boldsymbol{\phi}$ can then be written as

$$\boldsymbol{\phi} \to e^{-i\alpha/2} e^{-it^i \tau_i} \boldsymbol{\phi} \simeq \left(\mathbf{1}_2 - i\alpha Y - it^i \tau_i\right) \boldsymbol{\phi} \tag{4.88}$$

with generators

$$\tau_i = \frac{1}{2}\sigma_i \,, \quad Y = \frac{1}{2}\mathbf{1}_2 \,. \tag{4.89}$$

The general invariant Lagrangian density is

$$\mathcal{L} = \partial_\mu \boldsymbol{\phi}^\dagger \partial^\mu \boldsymbol{\phi} - V(\boldsymbol{\phi}) \,, \quad V = V_0 + m^2 \boldsymbol{\phi}^\dagger \boldsymbol{\phi} + \lambda (\boldsymbol{\phi}^\dagger \boldsymbol{\phi})^2 \,. \tag{4.90}$$

Note that the invariance of this Lagrangian density under $\mathrm{SU}(2)$ is due to the appearance of a singlet in the Clebsch-Gordan decomposition $\bar{\mathbf{2}} \otimes \mathbf{2} = \mathbf{1} + \mathbf{3}$. Provided that $m^2 < 0$, the scalar potential is minimised for

$$\boldsymbol{\phi}^\dagger \boldsymbol{\phi} = v_0^2 = -\frac{m^2}{2\lambda} \tag{4.91}$$

and a particularly simple choice of minimum is provided by

$$\boldsymbol{\phi} = \mathbf{v} = \begin{pmatrix} 0 \\ v_0 \end{pmatrix} . \tag{4.92}$$

Clearly, for this choice is follows that

$$\tau^1\mathbf{v} \neq 0 , \quad \tau^2\mathbf{v} \neq 0 , \quad (\tau^3 - Y)\mathbf{v} \neq 0 , \quad (\tau^3 + Y)\mathbf{v} = 0 . \tag{4.93}$$

Hence, three of the four generators of $\mathrm{SU}(2) \times \mathrm{U}_Y(1)$ are broken, while the generator $\tau^3 + Y$ remains unbroken. This last generator corresponds to a combination of the $\mathrm{U}(1) \subset \mathrm{SU}(2)$ and the additional $\mathrm{U}_Y(1)$ and defines the unbroken $\mathrm{U}(1)$ subgroup. So the induced breaking pattern can be summarised as

$$\mathrm{SU}(2) \times \mathrm{U}_Y(1) \to \mathrm{U}(1) . \tag{4.94}$$

This is precisely the symmetry breaking pattern which arises in the electro-weak sector of the standard model of particle physics. There, $\mathrm{SU}(2) \times U_Y(1)$ is the electro-weak (gauge) symmetry and the unbroken $\mathrm{U}(1)$ corresponds to electromagnetism. In the present case we are working with a global symmetry and Goldstone's theorem tells us that we should have three massless scalars from the three broken generators. In the case of the electro-weak theory, the $\mathrm{SU}(2) \times \mathrm{U}_Y(1)$ symmetry is actually promoted to a *local (or gauge) symmetry* where the symmetry parameters are allowed to depend on space-time. In this case, it turns out that the Goldstone bosons are absorbed by three vector bosons which receive masses from symmetry breaking. This phenomenon is also called the *Higgs effect* and to investigate this in more detail we need to introduce vector fields and gauge symmetries.

4.4 Vector fields, gauge symmetry and scalar electrodynamics

4.4.1 Lagrangian formulation of Maxwell's theory

Covariant electro-magnetism flashback
We know that Maxwell's equations can be formulated in terms of a vector potential A_μ with associated field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu , \tag{4.95}$$

to which the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are related by

$$E_i = F_{0i} , \quad B_i = \frac{1}{2}\epsilon_{ijk}F_{jk} . \tag{4.96}$$

Under Lorentz transformations, A_μ transforms like a vector, that is, $A_\mu \to \Lambda_\mu{}^\nu A_\nu$, and consequently, from Eq. (4.95), $F_{\mu\nu}$ transforms like a tensor, $F_{\mu\nu} \to \Lambda_\mu{}^\rho \Lambda_\nu{}^\sigma F_{\rho\sigma}$. Since it is $F_{\mu\nu}$ which is directly associated to the physical fields it is not surprising that the vector potential A_μ contains some unphysical degrees of freedom. Formally, this is expressed by the fact that a *gauge transformation*

$$A_\mu \to A_\mu + \partial_\mu \Lambda \tag{4.97}$$

on A_μ, parameterized by an arbitrary function [5] $\Lambda = \Lambda(x)$, leaves the field strength tensor $F_{\mu\nu}$ unchanged (as can be easily seen by transforming the RHS of Eq. (4.95)). Let us now write down the most general Lagrangian density for A_μ (up to second order in derivatives) which is Lorentz invariant and invariant under gauge transformations (4.97). Gauge invariance implies that the Lagrangian should depend on A_μ only through the field strength F and, since F contains one derivative the most we should consider is quadratic terms in F. In addition, Lorentz invariance means all indices should be contracted in $\mathcal{L}$. Basically, this leaves only one allowed term [6], namely

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} , \tag{4.98}$$

where we think of $F_{\mu\nu}$ as being given by Eq. (4.95). One finds

$$\frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)} = -\frac{1}{2}F^{\rho\sigma}\frac{\partial(\partial_\rho A_\sigma - \partial_\sigma A_\rho)}{\partial(\partial_\mu A_\nu)} = -\frac{1}{2}F^{\rho\sigma}\left(\delta^\mu_\rho\delta^\nu_\sigma - \delta^\mu_\sigma\delta^\nu_\rho\right) = -F^{\mu\nu} , \quad \frac{\partial\mathcal{L}}{\partial A_\mu} = 0 . \tag{4.99}$$

[5] Not to be confused with a Lorentz transformation!

[6] The term $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$ is also consistent with all stated requirements. However, it can be written as the total derivative $4\partial_\mu(\epsilon^{\mu\nu\rho\sigma}A_\nu\partial_\rho A_\sigma)$ and, hence, does not effect the equations of motion.

Inserting this into the Euler-Lagrange equation (4.31) implies

$$\partial_\mu F^{\mu\nu} = 0 \ , \quad \partial_{[\mu} F_{\nu\rho]} = 0 \ , \tag{4.100}$$

where the second equation is a trivial consequence of the definition (4.95). These are the free Maxwell's equations in covariant form. Splitting indices up into space and time components and inserting Eqs. (4.96) they can be easily shown to be equivalent to the better-known version in terms of the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$. This example illustrates the power of the Lagrangian formulation of field theories. Starting with a simple set of assumptions about the symmetries (Lorentz symmetry and gauge invariance in the present case) and the field content (a single vector field A_μ) one is led to the correct theory by writing down the most general Lagrangian consistent with the symmetries.

Gauge choice and general solution to equations of motion
In terms of A_μ, the free Maxwell theory can be expressed by the single equation

$$\Box A_\mu - \partial_\mu \partial_\nu A^\nu = 0 \ , \tag{4.101}$$

which follows from the first Eq. (4.100) after inserting the definition (4.95) (Note that the second equation (4.100) is automatically satisfied once $F_{\mu\nu}$ is written in terms of A_μ). This equation can be further simplified by exploiting the gauge symmetry (4.97). Gauge invariance allows one, via transformation with an appropriate gauge parameter Λ, to impose a *gauge condition* on A_μ. There are several possibilities for such gauge conditions and here we consider the *Lorentz gauge* defined by

$$\partial_\mu A^\mu = 0 \ . \tag{4.102}$$

This condition has the obvious benefit of being covariant (unlike, for example, the so-called *temporal gauge* which requires $A_0 = 0$) and it simplifies the equation of motion for A_μ to

$$\Box A_\mu = 0 \ . \tag{4.103}$$

Note, however, that the Lorentz gauge does not fix the gauge symmetry completely but leaves a *residual gauge freedom* with gauge parameters Λ satisfying

$$\Box \Lambda = 0 \ . \tag{4.104}$$

Eq. (4.103) is a massless Klein-Gordon equation for a vector field and, hence, can be easily solved using our earlier result (4.56) with an additional μ index attached. This leads to

$$A_\mu(x) = \int d^3\tilde{k} \left(a_\mu(k) e^{-ikx} + a^\star_\mu(k) e^{ikx} \right) \ . \tag{4.105}$$

where $w_{\mathbf{k}} = |\mathbf{k}|$ and $(k_\mu) = (w_{\mathbf{k}}, \mathbf{k})$. In addition, the Lorentz gauge condition demands that the coefficents a_μ satisfy

$$k^\mu a_\mu(k) = 0 \ . \tag{4.106}$$

To exploit this constraint in detail it is useful to introduce a set of polarisation vectors $\epsilon^{(\alpha)}_\mu(k)$, where $\alpha = 0, 1, 2, 3$, with the following properties. The vectors $\epsilon^{(1)}(k)$ and $\epsilon^{(2)}(k)$ are orthogonal to both k and a vector n with $n^2 = 1$ and $n_0 > 0$ in the time direction and they satisfy

$$\epsilon^{(\alpha)}(k) \cdot \epsilon^{(\alpha')}(k) = -\delta^{\alpha\alpha'} \quad \text{for} \quad \alpha, \, \alpha' = 1, 2 \ . \tag{4.107}$$

Further, $\epsilon^{(3)}(k)$ is chosen to be in the (n, k) plane, orthogonal to n and normalised, that is, $n \cdot \epsilon^{(3)}(k) = 0$ and $(\epsilon^{(3)}(k))^2 = -1$. Finally, we set $\epsilon^{(0)} = n$. With these conventions we have an orthogonal set of vectors satisfying

$$\epsilon^{(\alpha)} \cdot \epsilon^{(\alpha')} = \eta^{\alpha\alpha'} \tag{4.108}$$

for all $\alpha, \alpha' = 0, 1, 2, 3$. They can be used to write $a_\mu(k)$ as

$$a_\mu(k) = \sum_{\alpha=0}^{3} a^{(\alpha)}(k) \epsilon^{(\alpha)}_\mu(k) \ , \tag{4.109}$$

where $a^{(\alpha)}(k)$ are four expansion coefficients. The idea of introducing this basis of polarisation vectors is to separate the directions transversal to k, corresponding to $\epsilon^{(1)}(k)$ and $\epsilon^{(2)}(k)$, from the other two directions $\epsilon^{(0)}(k)$

and $\epsilon^{(3)}(k)$. As an example, if we choose a spatial momentum $\mathbf{k}$ pointing in the z-direction, the above vectors are explicitly given by

$$\epsilon^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} , \quad \epsilon^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} , \quad \epsilon^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} , \quad \epsilon^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} . \tag{4.110}$$

Returning to the general case, it is easy to see that the unique choice for $\epsilon^{(3)}(k)$ is

$$\epsilon^{(3)}_\mu(k) = \frac{1}{k_0} k_\mu - n_\mu \quad \text{so that} \quad \epsilon^{(0)}_\mu + \epsilon^{(3)}_\mu = \frac{1}{k_0} k_\mu \ . \tag{4.111}$$

Let us now return to the gauge condition (4.106) and insert the expansion (4.109). The two transverse directions $\epsilon^{(1)}(k)$ and $\epsilon^{(2)}(k)$ drop out of the equation, so that

$$0 = k^\mu a_\mu(k) = (k \cdot \epsilon^{(0)}(k)) a^{(0)}(k) + (k \cdot \epsilon^{(3)}(k)) a^{(3)}(k) = k_0 (a^{(0)}(k) - a^{(3)}(k)) \ . \tag{4.112}$$

Here we have used that

$$k \cdot \epsilon^{(3)}(k) = -k \cdot \epsilon^{(0)}(k) = -k_0 \ , \tag{4.113}$$

in the last step. We conclude that $a^{(0)}(k) = a^{(3)}(k)$ in order to satisfy the gauge condition (4.106) and that the expansion (4.109) can be written as

$$a_\mu(k) = \frac{a^{(0)}(k)}{k_0} k_\mu + \sum_{\alpha=1}^{2} a^{(\alpha)}(k) \epsilon^{(\alpha)}_\mu(k) \ . \tag{4.114}$$

Hence, we are left with the two transversal polarisations and a longitudinal one along the direction of k. Recall that we still have a residual gauge freedom from gauge parameters Λ satisfying the Klein-Gordon equation (4.104). The most general such parameters can be written as

$$\Lambda(x) = \int d^3\tilde{k} \left(\lambda(k) e^{-ikx} + \lambda^\star(k) e^{ikx} \right) \ . \tag{4.115}$$

A gauge transformation (4.97) with Λ of this residual form changes the expansion coefficients $a_\mu(k)$ for the vector field as

$$a_\mu(k) \to a'_\mu(k) = a_\mu(k) - i k_\mu \lambda(k) \ . \tag{4.116}$$

Comparing with Eq. (4.114), it is clear that we can use this residual gauge freedom to remove the longitudinal degree of freedom in $a_\mu(k)$. We are then left with the two transversal polarisations only and we conclude that the number of physical degrees of freedom for a vector field is two. This reduction from four apparent degrees of freedom to two is directly related to the gauge invariance of the theory.

Massive vector field

It is instructive to perform a similar analysis for a massive vector field, that is a vector field A_μ with the additional term $\frac{m^2}{2} A_\mu A^\mu$ added to the Lagrangian density (4.98). Clearly, gauge invariance is explicitly broken for such a massive vector field. The equation of motion reads

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \ . \tag{4.117}$$

Applying ∂_ν to this equation we conclude that $\partial_\nu A^\nu = 0$ and, hence, that Eq. (4.117) can equivalently be written as

$$(\Box + m^2) A_\mu = 0 \ , \quad \partial_\mu A^\mu = 0 \ . \tag{4.118}$$

The first of these equations is a massive Klein-Gordon equation with general solution given by Eq. (4.105) but now with $k^2 = m^2$ instead of $k^2 = 0$. To also satisfy the second equation above we need $k^\mu a_\mu(k) = 0$ which reduces the number of degrees from four to three. Since gauge invariance is not available in this case, no further reduction occurs and we conclude that the number of physical degrees of freedom of a massive vector field is three.

Including a source

Let us return to the massless theory and ask how can we include a source J_μ into this Lagrangian formulation of

Maxwells' theory. This current should appear linearly in the equations of motion (and therefore in the Lagrangian density), so the obvious generalisation of Eq. (4.98) is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - A_\mu J^\mu \ . \tag{4.119}$$

The additional term depends on A_μ explicitly which leads to a non-trivial gauge variation

$$S \to S - \int d^4x\, \partial_\mu \Lambda J^\mu = S + \int d^4x\, \Lambda \partial_\mu J^\mu \ , \tag{4.120}$$

of the action. This apparent breaking of the gauge invariance of the theory can be avoided if we require that

$$\partial_\mu J^\mu = 0 \ , \tag{4.121}$$

that is, if the current J_μ is conserved. With $\frac{\partial \mathcal{L}}{\partial A_\mu} = -J^\mu$ and the first Eq. (4.99) we find from the Euler-Lagrange equations

$$\partial_\mu F^{\mu\nu} = J^\nu \ , \quad \partial_{[\mu}F_{\nu\rho]} = 0 \ , \tag{4.122}$$

which are indeed Maxwell's equations in the presence of a source J_μ. In a fundamental theory, the current J_μ should arise from fields, rather than being put in "by hand" as an external source. We will now study an example for such a theory in which vector fields are coupled to scalars.

4.4.2 Scalar electrodynamics and the Higgs mechanism

Our goal is to write down a theory with a vector field A_μ and a complex scalar $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$. We have already seen individual Lagrangians for these two fields, namely Eq. (4.98) for the vector field and Eq. (4.72) for the scalar field. There is nothing wrong with adding those two Lagrangians, but this leads to a somewhat uninteresting theory where the vector and the scalar are decoupled (that is, there are no terms in this Lagrangian containing both types of fields and consequently their equations of motion are decoupled from one another). Again, the key to constructing a more interesting theory comes from thinking about symmetries. Both the gauge symmetry of electromagnetism and the global $\mathrm{U}(1)$ symmetry of the complex scalar field theory are parameterized by a single parameter. Let us identify those two parameters with one another, that is we set $\alpha = \Lambda$ in Eq. (4.71), and, hence, transform the scalar as

$$\phi(x) \to e^{-iq\Lambda(x)}\phi(x) \ , \tag{4.123}$$

along with the transformation (4.97) of the vector field. The new feature we have introduced in this way is that the formerly global transformation (4.71) of the scalar field has now become local. The scalar potential for ϕ is still invariant under these local $\mathrm{U}(1)$ transformations, however, there is a problem with the kinetic term for ϕ, since the derivative ∂_μ now acts on the local group parameter $\Lambda(x)$. Explicitly, $\partial_\mu\phi$ transforms as

$$\partial_\mu\phi \to e^{-iq\Lambda(x)}(\partial_\mu - iq\partial_\mu\Lambda)\phi \ . \tag{4.124}$$

The additional term proportional to $\partial_\mu\Lambda$ in this transformation is reminiscent of the transformation law (4.97) for the gauge field and A_μ can indeed be used to cancel this term. Define the *covariant derivative*

$$D_\mu = \partial_\mu + iqA_\mu \ . \tag{4.125}$$

Then the covariant derivative of ϕ transforms as

$$D_\mu\phi \to e^{-iq\Lambda(x)}D_\mu\phi \ , \tag{4.126}$$

and the modified kinetic term $(D_\mu\phi)^\star D^\mu\phi$ is gauge invariant. With this modification, we can now combine Eqs. (4.72) and (4.98) to obtain the gauge invariant Lagrangian density

$$\mathcal{L} = (D_\mu\phi)^\star D^\mu\phi - V(\phi,\phi^\star) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \quad V = V_0 + m^2\phi^\star\phi + \frac{\lambda}{4}(\phi^\star\phi)^2 \tag{4.127}$$

This is the Lagrangian for *scalar electrodynamics*. It shows the gauge field A_μ in a new role, facilitating the invariance of the scalar field theory under local symmetry transformations. In fact, had we started with just the globally symmetric scalar field theory (4.72) with the task of finding a locally $\mathrm{U}(1)$ invariant version we would

have been led to introducing a gauge field A_μ. We also note that the covariant derivative in (4.127) has introduced a non-trivial coupling between the scalar field and the vector field.

We will now use scalar electrodynamics as a toy model to study spontaneous breaking of a local $\mathrm{U}(1)$ symmetry. The scalar potential is unchanged from the globally symmetric model and, hence, we can apply the results of Sec. 4.3.2 . For $m^2 \geq 0$ the minimum is at $\phi = 0$ and the symmetry is unbroken. Therefore, we focus on the case $m^2 < 0$ where the potential has a "Mexican hat" shape as in Fig. 4.2. In this case, there is a circle of minima given by Eq. (4.78). Instead of choosing the parameterisation (4.80) for ϕ around the minimum on the ϕ_1 axis it proves more useful in the present context [7] to use

$$\phi = \frac{1}{\sqrt{2}}(v_0 + H)e^{i\chi} \tag{4.128}$$

where $\chi = \chi(x)$ is the Goldstone mode and $H = H(x)$ is the massive mode. Next we perform a gauge transformation with parameter $\Lambda = \chi/q$, that is,

$$\phi \to \phi' = e^{-i\chi}\phi = \frac{1}{\sqrt{2}}(v_0 + H) \; , \quad A_\mu \to A'_\mu = A_\mu + \frac{1}{q}\partial_\mu \chi \; . \tag{4.129}$$

We can now write the Lagrangian density (4.127) in terms of the transformed fields (this leaves the action unchanged) and then insert the explicit expression for ϕ' from the previous equation. One finds for the covariant derivative

$$D_\mu \phi' = \frac{1}{\sqrt{2}}(\partial_\mu H + iq(v_0 + H)A'_\mu) \; . \tag{4.130}$$

and inserting into Eq. (4.127) then leads to

$$\mathcal{L} = \frac{1}{2}\partial_\mu H \partial^\mu H - V(H) - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}q^2 A'_\mu A'^\mu (v_0^2 + 2v_0 H + H^2) \tag{4.131}$$

$$V(H) = V_0 + \frac{1}{4}m^2 v_0^2 - m^2 H^2 + \frac{\lambda}{16}(4v_0 H^3 + H^4) \; . \tag{4.132}$$

The most striking feature about this result is that the Goldstone mode χ has completely disappeared from the Lagrangian and we are left with just the real, massive scalar H and the vector field A'_μ. However, the vector field is now massive with mass

$$m(A') = qv_0 \; , \tag{4.133}$$

end, hence, has three degrees of freedom as opposed to just two for a massless vector field. This explains the disappearance of the scalar χ: It has been "absorbed" by A'_μ to provide the additional (longitudinal) degree of freedom necessary for a massive vector field. This can also be seen from the transformation of A_μ in Eq. (4.129). Hence, we have learned that a spontaneously broken local symmetry leads to a mass for the associated vector boson and the conversion of the Goldstone field into the longitudinal mode of the vector. This is also called the *Higgs effect*. The same phenomenon but in its generalisation to non-Abelian gauge groups occurs in the breaking of the electro-weak $\mathrm{SU}(2) \times \mathrm{U}_Y(1) \to \mathrm{U}(1)$ gauge group to $\mathrm{U}(1)$ of electromagnetism. The three $W^\pm$ and Z vector bosons of the broken part of the electro-weak symmetry receive a mass (4.133) proportional to the symmetry breaking scale v_0 and absorb the three Goldstone bosons (see Section 4.3.4) which arise. A detailed discussion of electro-weak symmetry breaking requires introducing non-Abelian gauge symmetries which is beyond the scope of this lecture.

4.5 Further reading

- Group theory and its application to physics is discussed in B. G. Wybourne, *Classical Groups for Physicists* and T.-P. Cheng, L.-F. Li, *Gauge theory of elementary particle physics*, Chapter 4.
- The material on classical field theory covered in this chapter can be found in most quantum field theory books, including D. Bailin and A. Love, *Introduction to Gauge Theory*, Chapter 3 and 13; P. Ramond, *Field Theory: A Modern Primer*, Chapter 1; M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Chapter 2.

As usual, be warned that these references were not written with undergraduate readers in mind.

[7] One can think of Eq. (4.80) as a linearized version of the more accurate parameterization (4.128).

Chapter 5

Canonical Quantization

5.1 The general starting point

We have seen in Chapter 1 how to formulate quantum mechanics in terms of path integrals . This has led to an intuitive picture of the transition between classical and quantum physics. Later, we will show how to apply path integrals to the quantisation of field theories. In this chapter, we will to focus on the more traditional method of *canonical quantization*. Let us first recall how canonical quantization works for classical mechanics.

Start with a system in classical machanics, described by set of canonical coordinates $\{p_i(t), q_i(t)\}$ and a Hamiltonian $H = H(p_i, q_i)$. The prescription on how to perform the transition to the associated quantum mechanics consists of, firstly, replacing the canonical coordinates by operators [1], that is $p_i(t) \to \hat{p}_i(t)$ and $q_i(t) \to \hat{q}_i(t)$, and then imposing the *canonical commutation relations* [2]

$$[\hat{q}_i(t), \hat{q}_j(t)] = [\hat{p}_i(t), \hat{p}_j(t)] = 0 \ , \quad [\hat{p}_i(t), \hat{q}_j(t)] = -i\delta_{ij} \ . \tag{5.1}$$

The dymanics of this system is governed by the operator version of Hamilton's equations

$$\frac{d\hat{q}_i(t)}{dt} = i[\hat{H}, \hat{q}_i(t)] \ , \quad \frac{d\hat{p}_i(t)}{dt} = i[\hat{H}, \hat{p}_i(t)] \ . \tag{5.2}$$

How can we transfer this quantization procedure to field theory? Consider a classical field theory with a generic set of fields $\phi_a = \phi_a(x)$, conjugate momenta $\pi_a(x)$, Hamiltonian density $\mathcal{H}$ and Hamiltonian $H = \int d^3x\, \mathcal{H}$, as introduced in Section 4.2.1. On a discrete space consisting of lattice points $\{\mathbf{x}_i\}$, these fields can be represented by sets of canonical coordinates $\phi_{ai}(t) = \phi_a(t, \mathbf{x}_i)$ and $\pi_{ai}(t) = \pi_a(t, \mathbf{x}_i)$. On these we can simply impose the canonical commutation relations (5.1) which leads to

$$[\hat{\phi}_{ia}(t), \hat{\phi}_{bj}(t)] = [\hat{\pi}_{ai}(t), \hat{\pi}_{bj}(t)] = 0 \ , \quad [\hat{\pi}_{ai}(t), \hat{\phi}_{bj}(t)] = -i\delta_{ab}\delta_{ij} \ . \tag{5.3}$$

The continuum version of these equations is obtained by replacing $i \to \mathbf{x}$, $j \to \mathbf{y}$, $\delta_{ij} \to \delta^3(\mathbf{x} - \mathbf{y})$, $\hat{\phi}_{ai}(t) \to \hat{\phi}(t, \mathbf{x})$ and similarly for the conjugate momenta. This results in

$$[\hat{\phi}_a(t, \mathbf{x}), \hat{\phi}_b(t, \mathbf{y})] = [\hat{\pi}^a(t, \mathbf{x}), \hat{\pi}^b(t, \mathbf{y})] = 0 \ , \quad [\hat{\pi}^a(t, \mathbf{x}), \hat{\phi}_b(t, \mathbf{y})] = -i\delta^a_b\delta^3(\mathbf{x} - \mathbf{y}) \tag{5.4}$$

Note that these commutators are taken at equal time but at generally different points in space. The canonical commutation relations (5.4) together with the continuum version

$$\partial_0\hat{\phi}_a(t, \mathbf{x}) = i[\hat{H}, \hat{\phi}_a(t, \mathbf{x})] \ , \quad \partial_0\hat{\pi}^a(t, \mathbf{x}) = i[\hat{H}, \hat{\pi}^a(t, \mathbf{x})] \ , \tag{5.5}$$

of Hamilton's equations (5.2) provide the starting point for the canonical quantization of field theories. We will now investigate the consequences of this quantization procedure for the simplest case, a single real, free scalar field. To avoid cluttering the notation, we will drop hats on operators from now on. It will usually be clear from the context whether we refer to a classical object or its operator version.

[1] We are using operators in the Heisenberg picture.

[2] More formally, the transition between classical and quantum mechanics can also be understood as a replacement of canonical coordinates with operators and Poisson brackets with commutator brackets.

5.2 Canonical quantization of a real, free scalar field

Recap of theory and canonical quantisation conditions
From Sec. 4.3.1, the Lagrange density $\mathcal{L}$, Hamiltonian density $\mathcal{H}$ and the conjugate momentum π for a free real scalar field ϕ with mass m are given by

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right) \;, \quad \mathcal{H} = \frac{1}{2}\left(\pi^2 + (\boldsymbol{\nabla}\phi)^2 + m^2\phi^2)\right) \;, \quad \pi = \frac{\partial\mathcal{L}}{\partial(\partial_0\phi)} = \partial_0\phi \;. \tag{5.6}$$

and the associated equation of motion is the Klein-Gordon equation

$$\Box\phi + m^2\phi = 0 \;. \tag{5.7}$$

It should now be thought of as an operator equation for the field operator ϕ. The canonical commutation relations (5.4) in this case read

$$[\phi(t,\mathbf{x}),\phi(t,\mathbf{y})] = [\pi(t,\mathbf{x}),\pi(t,\mathbf{y})] = 0 \;, \quad [\pi(t,\mathbf{x}),\phi(t,\mathbf{y})] = -i\delta^3(\mathbf{x}-\mathbf{y}) \;. \tag{5.8}$$

Oscillator expansion of general solution
We can solve the free scalar field theory by writing down the most general solution of the operator Klein-Gordon equation. This can be done by starting with the general classical solution (4.56) and by promoting the coefficients $a(k)$ to operators. Hence, we have

$$\phi(x) = \int d^3\tilde{k}\left(a(k)e^{-ikx} + a^\dagger(k)e^{ikx}\right) \;, \quad \pi(x) = -i\int d^3\tilde{k}\, w_{\mathbf{k}}\left(a(k)e^{-ikx} - a^\dagger(k)e^{ikx}\right) \;. \tag{5.9}$$

What do the canonical commutation relations for ϕ and π imply for the commutators of $a(k)$ and $a^\dagger(k)$? To answer this question, we would like to express $a(k)$ in terms of $\phi(x)$ and $\pi(x)$ by inverting the above relations. We start by applying $2w_{\mathbf{q}}\int d^3x\, e^{-i\mathbf{q}\cdot\mathbf{x}}$ to the equation for $\phi(x)$. After carrying out the integration over $\mathbf{k}$ [3] one finds

$$2w_{\mathbf{q}}\int d^3x\, e^{-i\mathbf{q}\cdot\mathbf{x}}\phi(x) = a(q)e^{-iw_{\mathbf{q}}t} + a^\dagger(-q)e^{iw_{\mathbf{q}}t} \;. \tag{5.10}$$

Analogously, applying $2i\int d^3x\, e^{-i\mathbf{q}\cdot\mathbf{x}}$ to the equation for $\pi(x)$ results in

$$2i\int d^3x\, e^{-i\mathbf{q}\cdot\mathbf{x}}\pi(x) = a(q)e^{-iw_{\mathbf{q}}t} - a^\dagger(-q)e^{iw_{\mathbf{q}}t} \tag{5.11}$$

Adding the last two equations we find the desired expression for $a(q)$ and its conjugate

$$a(q) = \int d^3x\, e^{iqx}(w_{\mathbf{q}}\phi(x) + i\pi(x)) \;, \quad a^\dagger(q) = \int d^3x\, e^{-iqx}(w_{\mathbf{q}}\phi(x) - i\pi(x)) \;. \tag{5.12}$$

Combining these results and the canonical commutation relations (5.8) one finds for the commutators of $a(q)$ and $a^\dagger(q)$

$$[a(k),a(q)] = [a^\dagger(k),a^\dagger(q)] = 0 \;, \quad [a(k),a^\dagger(q)] = (2\pi)^3 2w_{\mathbf{k}}\delta^3(\mathbf{k}-\mathbf{q}) \;. \tag{5.13}$$

For each $k = q$ these equations are reminiscent of the commutation relations for the creation and annihilation operators of a harmonic oscillator. This should not come as a surprise given a single plain wave Ansatz $\phi(t,\mathbf{x}) = \phi_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}$ turns the Hamiltonian (5.6) (or, more precisely, as the Ansatz is complex, its complex counterpart (4.75)) into the Hamiltonian $\mathcal{H}_{\mathbf{k}} = \pi_{\mathbf{k}}^2 + w_{\mathbf{k}}^2\phi_{\mathbf{k}}^2$ for a single harmonic oscillator with frequency $w_{\mathbf{k}} = \sqrt{m^2+\mathbf{k}^2}$. We should, therefore, think of the free scalar field as an infinite collection of (decoupled) harmonic oscillators labelled by three-momentum $\mathbf{k}$ and with frequency $w_{\mathbf{k}}$.

The Fock space
The interpretation of $a^\dagger(k)$ and $a(k)$ as creation and annihilation operators also suggests a method for the construction of the space of states for this theory, the so-called *Fock space*. In close analogy with the simple harmonic oscillator, we first define the vacuum $|0\rangle$, with normalisation $\langle 0|0\rangle = 1$, as the state which is annihilated by all $a(k)$, that is

$$a(k)|0\rangle = 0 \;. \tag{5.14}$$

[3] Recall that $\int d^3x\, e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{x}} = (2\pi)^3\delta^3(\mathbf{k}-\mathbf{q})$.

Single particle states $|k\rangle$ are obtained by acting on this vacuum state with a single creation operator, so that

$$|k\rangle = a^\dagger(k)|0\rangle \ . \tag{5.15}$$

Using the commutation relations (5.13) and Eq. (5.14), one finds for their normalization

$$\langle k|q\rangle = \langle 0|a(k)a^\dagger(q)|0\rangle = \langle 0|[a(k), a^\dagger(q)]|0\rangle = (2\pi)^3 2w_{\mathbf{k}}\delta^3(\mathbf{k}-\mathbf{q}) \ . \tag{5.16}$$

We note that the RHS of this relation is the "covariant delta function" (4.55). A basis set of states is provided by all n particle states $|k_1, \ldots, k_n\rangle$, obtained by acting on the the vacuum with n creation operators

$$|k_1, \ldots, k_n\rangle = a^\dagger(k_1)\ldots a^\dagger(k_n)|0\rangle \ . \tag{5.17}$$

The number operator

The *number operator* N is defined by

$$N = \int d^3\tilde{k}\, a^\dagger(k)a(k) \ , \tag{5.18}$$

and from the commutation relations (5.13) one finds

$$[N, a^\dagger(q)] = \int d^3\tilde{k}\, [a^\dagger(k)a(k), a^\dagger(q)] = \int d^3\tilde{k}\, a^\dagger(k)[a(k), a^\dagger(q)] = \int d^3\tilde{k}\, a^\dagger(k)(2\pi)^3 2w_{\mathbf{k}}\delta^3(\mathbf{k}-\mathbf{q}) = a^\dagger(q) \ . \tag{5.19}$$

On an n-particle state the number operator acts as

$$\begin{aligned} N|k_1, \ldots, k_n\rangle &= Na^\dagger(k_1)\ldots a^\dagger(k_n)|0\rangle = \left\{a^\dagger(k_1)N + [N, a^\dagger(k_1)]\right\} a^\dagger(k_2)\ldots a^\dagger(k_n)|0\rangle & (5.20)\\ &= a^\dagger(k_1)Na^\dagger(k_2)\ldots a^\dagger(k_n)|0\rangle + |k_1, \ldots, k_n\rangle \ . & (5.21)\end{aligned}$$

We can repeat this procedure and commute N with all creation operators, picking up at each step the term $|k_1, \ldots, k_n\rangle$. In the last step we use $N|0\rangle = 0$ and find

$$N|k_1, \ldots, k_n\rangle = n|k_1, \ldots, k_n\rangle \ . \tag{5.22}$$

Hence, N indeed counts the number of particles in a state.

Four-momentum and normal ordering

We should now compute the conserved four-momentum (4.44) for the above solution of the free scalar theory. From the stress energy tensor (4.59) with $V = \frac{1}{2}m^2\phi^2$ we find (in the case of P_0 after a partial integration and using the Klein-Gordon equation)

$$H = P_0 = \frac{1}{2}\int d^3x\, (\pi^2 - \phi\partial_0\pi) \ , \quad \mathbf{P} = \int d^3x\, \pi\boldsymbol{\nabla}\phi \ . \tag{5.23}$$

Inserting Eqs. (5.9) results in

$$\begin{aligned} H &= \frac{1}{2}\int d^3x\, d^3\tilde{k}\, d^3\tilde{q}\, \big[-w_{\mathbf{k}}w_{\mathbf{q}}\left(a(k)e^{-ikx} - a^\dagger(k)e^{ikx}\right)\left(a(q)e^{-iqx} - a^\dagger(q)e^{iqx}\right) & (5.24)\\ &\qquad + w_{\mathbf{k}}^2\left(a(k)e^{-ikx} + a^\dagger(k)e^{ikx}\right)\left(a(q)e^{-iqx} + a^\dagger(q)e^{iqx}\right)\big)\big] & (5.25)\\ &= \frac{1}{2}\int d^3\tilde{k}\ w_{\mathbf{k}}\left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right] \ , & (5.26)\end{aligned}$$

and

$$\begin{aligned} \mathbf{P} &= -\int d^3x\, d^3\tilde{k}\, d^3\tilde{q}\, \mathbf{q}\, w_{\mathbf{k}}\left(a(k)e^{-ikx} - a^\dagger(k)e^{ikx}\right)\left(a(q)e^{-iqx} - a^\dagger(q)e^{iqx}\right) & (5.27)\\ &= \frac{1}{2}\int d^3\tilde{k}\, \mathbf{k}\left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right] \ . & (5.28)\end{aligned}$$

The above Hamiltonian corresponds to an integral over harmonic oscillator Hamiltonians with frequence $w_{\mathbf{k}}$ and labelled by three-momenta $\mathbf{k}$. This is in line with our interpretation of the free scalar field as a collection of decoupled harmonic oscillators. When trying to write the integrand in Eq. (5.26) in the standard harmonic oscillator

form $\mathcal{H}_{\mathbf{k}} = w_{\mathbf{k}}(a^\dagger(k)a(k) + \frac{1}{2})$ using the commutator (5.13) one encounters an infinite zero point energy, proportional to $\delta^3(0)$, which can be interpreted as the energy of the vacuum state $|0\rangle$. This is one of the many infinities in quantum field theory and it should not come as a surprise: After all, we are summing over an infinite number of harmonic oscillators each with finite zero-point energy $w_{\mathbf{k}}/2$. To deal with this infinity, we define the concept of *normal ordering* of operators. The normal ordered version $:\mathcal{O}:$ of an operator $\mathcal{O}$ is obtained by writing all creation operators to the left of all annihilation operators. So, for example, $:a(k)a^\dagger(q):=a^\dagger(q)a(k)$. An operator and its normal-ordered counterpart differ by a (usually infinite) number, often referred to as a *c-number*. With this definition we have

$$:H:=\int d^3\tilde{k}\, w_{\mathbf{k}}\, a^\dagger(k)a(k)\;,\quad :\mathbf{P}:=\int d^3\tilde{k}\,\mathbf{k}\, a^\dagger(k)a(k)\quad\text{or}\quad :P_\mu:\;=\int d^3\tilde{k}\, k_\mu\, a^\dagger(k)a(k)\;. \tag{5.29}$$

A short calculation shows that

$$[:P_\mu:,a^\dagger(k)]=\int d^3\tilde{q}\, q_\mu[a^\dagger(q)a(q),a^\dagger(k)]=\int d^3q\, q_\mu a^\dagger(q)\delta^3(\mathbf{k}-\mathbf{q})=k_\mu a^\dagger(k)\;. \tag{5.30}$$

For a single particle state $|k\rangle$ this implies

$$:H:|k\rangle=w_{\mathbf{k}}|k\rangle\;,\quad :\mathbf{P}:|k\rangle=\mathbf{k}|k\rangle\;, \tag{5.31}$$

and, hence, these states are eigenstates of $:P_\mu:$ with energy $w_{\mathbf{k}}$ and spatial momentum $\mathbf{k}$. Likewise, multi-particle states $|k_1,\ldots,k_n\rangle$ are eigenstates of $:P_\mu:$ with eigenvalue $\sum_{i=1}^n k_{i\mu}$. To summarise, we have constructed a basis of Fock space for the free real scalar field and, by looking at the action of conserved operators on these states, we have found an interpretation of this basis as n particle states with definite four-momentum.

Commutation of field operators and micro causality

Finally, we should look at the commutation properties of two field operators $\phi(x)$ and $\phi(x')$. From the canonical commutation relations we certainly know that these two operators commute for equal time, $t=t'$. However, the unequal time commutator $[\phi(x),\phi(x')]$ does not vanish. From the oscillator expansion (5.8) we anticipate that this commutator is given by a c-number and, hence, we can write $[\phi(x),\phi(x')]=\langle 0|[\phi(x),\phi(x')]|0\rangle$. By replacing ϕ with Eq. (5.9) we find

$$\begin{aligned}
[\phi(x),\phi(x')] &= \langle 0|[\phi(x),\phi(x')]|0\rangle\\
&= \langle 0|\int d^3\tilde{k}\, d^3\tilde{q}\left\{\left(a(k)e^{-ikx}+a^\dagger(k)e^{ikx}\right)\left(a(q)e^{-iqx'}+a^\dagger(q)e^{iqx'}\right)-(x\leftrightarrow x')\right\}|0\rangle\\
&= \langle 0|\int d^3\tilde{k}\, d^3\tilde{q}\left\{[a(k),a^\dagger(q)]e^{iqx'-ikx}-(x\leftrightarrow x')\right\}|0\rangle\\
&= \int d^3\tilde{k}\left\{e^{-iw_{\mathbf{k}}(t-t')}e^{ik(\mathbf{x}-\mathbf{x}')}-e^{iw_{\mathbf{k}}(t-t')}e^{ik(\mathbf{x}-\mathbf{x}')}\right\}\\
&= \int\frac{d^4k}{(2\pi)^3 2w_{\mathbf{k}}}\left\{\delta(k_0-w_{\mathbf{k}})+\delta(k_0+w_{\mathbf{k}})\right\}\epsilon(k_0)e^{-ik(x-x')}\\
&= \int\frac{d^4k}{(2\pi)^3}\,\delta(k^2-m^2)\epsilon(k_0)e^{-ik(x-x')}\equiv\Delta(x-x')\;,
\end{aligned} \tag{5.32}$$

where $\epsilon(k_0)=k_0/|k_0|$. All elements in the final integral are Lorentz invariant except for the function $\epsilon(k_0)$. However, if we restrict to Lorentz transformations Λ which preserve the sign of the time-component of four vectors (these are the orthochronos Lorentz transformations, see Table 4.1) then $\epsilon(k_0)$ remains unchanged and we have $\Delta(x)=\Delta(\Lambda x)$. We already know that $\Delta(0,\mathbf{x})=0$ since this corresponds to an equal-time commutator. Now consider a space-like vector x, that is $x^2<0$. With a suitable orthochronos Lorentz transformation Λ this vector can be written in the form $x=\Lambda(0,\mathbf{y})$ for some vector $\mathbf{y}$ and, hence, $\Delta(x)=\Delta(\Lambda(0,\mathbf{y}))=\Delta(0,\mathbf{y})=0$. We conclude that $\Delta(x)$ vanishes for all space-like vectors x. This means that the commutator $[\phi(x),\phi(x')]$ vanishes whenever x and x' are space-like separated. This fact is also referred to as *micro causality*. Two field operators at points with space-like separation should not be able to causally effect one another and one would, therefore, expect their commutator vanishes. We have just shown that this is indeed the case.

This concludes the canonical quantization of the free real scalar field and we move on to the next, more complicated case, the free complex scalar field.

5.3 Canonical quantization of the free complex scalar field

Recap of theory and canonical quantisation conditions
In Sec. (4.3.2) we have introduced a complex scalar field theory with a global $\mathrm{U}(1)$ symmetry. We would now like to quantise the free version of this theory following the same steps as for the free real scalar in the previous section. The field content of this theory consists of two real scalars ϕ_1 and ϕ_2 which are combined into a complex scalar $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$. Recall that the Lagrangian density, Hamiltonian density and conjugate momentum are given by

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial_\mu \phi - m^2 \phi^\dagger \phi \; , \quad \mathcal{H} = \pi^\dagger \pi + \boldsymbol{\nabla}\phi^\dagger \boldsymbol{\nabla}\phi + m^2 \phi^\dagger \phi \; , \quad \pi = \frac{\partial L}{\partial(\partial_0 \phi)} = \partial_0 \phi^\dagger \tag{5.33}$$

and the field equation is the Klein-Gordon equation

$$\Box\phi + m^2 \phi = 0 \; . \tag{5.34}$$

Canonical quantisation of this system can be performed by simply imposing the general commutation relations (5.4) on the two real scalars ϕ_1 and ϕ_2 and their associated conjugate momenta $\pi_1 = \partial_0 \phi_1$ and $\pi_2 = \partial_0 \phi_2$. The only non-zero commutation relations are then

$$[\pi_1(t,\mathbf{x}), \phi_1(t,\mathbf{y})] = -i\delta^3(\mathbf{x}-\mathbf{y}) \; , \quad [\pi_2(t,\mathbf{x}), \phi_2(t,\mathbf{y})] = -i\delta^3(\mathbf{x}-\mathbf{y}) \; . \tag{5.35}$$

Using $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ and $\pi = (\pi_1 + i\pi_2)/\sqrt{2}$ they can be easily translated to the complex field and its conjugate with the only non-zero commutation relations

$$[\pi(t,\mathbf{x}), \phi(t,\mathbf{y})] = [\pi^\dagger(t,\mathbf{x}), \phi^\dagger(t,\mathbf{y})] = -i\delta(\mathbf{x}-\mathbf{y}) \; . \tag{5.36}$$

Oscillator expansion and Fock space
The general classical solution to the Klein-Gordon equation for a complex scalar field has already been given in Eq. (4.53). The operator version of this solution reads

$$\phi(x) = \int d^3\tilde{k} \left(a_+(k) e^{-ikx} + a_-^\dagger(k) e^{ikx} \right) \; , \quad \phi^\dagger(x) = \int d^3\tilde{k} \left(a_-(k) e^{-ikx} + a_+^\dagger(k) e^{ikx} \right) \; . \tag{5.37}$$

As for the real scalar field below, we can invert these relations and compute the commutators of $a_\pm(k)$ and $a_\pm^\dagger(k)$ from the canonical commutation relations (5.36). The only non-zero commutators one finds in this way are

$$[a_\pm(k), a_\pm^\dagger(q)] = (2\pi)^3 2 w_{\mathbf{k}} \delta^3(\mathbf{k}-\mathbf{q}) \; . \tag{5.38}$$

This shows that we have two sets, $\{a_+^\dagger(k), a_+(k)\}$ and $\{a_-^\dagger(k), a_-(k)\}$, of creation and annihilation operators and, hence, two different types of one-particle states

$$|(k,+)\rangle = a_+^\dagger(k)|0\rangle \; , \quad |(k,-)\rangle = a_-^\dagger(k)|0\rangle \; . \tag{5.39}$$

As usual, the vacuum $|0\rangle$ is defined by $a_\pm(k)|0\rangle = 0$ and $\langle 0|0\rangle = 1$. Multi-particle states $|(k_1,\epsilon_1),\ldots,(k_n,\epsilon_n)\rangle = a_{\epsilon_1}^\dagger(k_1)\ldots a_{\epsilon_n}^\dagger(k_n)|0\rangle$ are now labelled by n momenta k_i and, in addition, n signs $\epsilon_i \in \{+1,-1\}$ to distinguish the two types of quanta.

Number operators and four-momentum
For each type of quanta we can introduce a number operator

$$N_\pm = \int d^3\tilde{k}\, a_\pm^\dagger(k) a_\pm(k) \; . \tag{5.40}$$

It is easy to show from the commutation relations (5.38) that

$$[N_\pm, a_\pm^\dagger(k)] = a_\pm^\dagger(k) \; , \quad [N_\pm, a_\mp^\dagger(k)] = 0 \; . \tag{5.41}$$

Hence N_+ (N_-) acting on a multi-particle state counts the number of quanta of $+$ type ($-$ type). The (normal-ordered) conserved four-momentum can be computed as for the real scalar field and one finds

$$:P_\mu: = \int d^3\tilde{k}\, k_\mu \left(a_+^\dagger(k) a_+(k) + a_-^\dagger(k) a_-(k) \right) \; , \tag{5.42}$$

where $k_0 = w_{\mathbf{k}}$, as usual. The relation

$$[P_\mu, a_\pm^\dagger(k)] = k_\mu a_\pm^\dagger(k) \tag{5.43}$$

shows that $|(k,\epsilon)\rangle$ can be interpreted as a state with four-momentum k_μ irrespective of ϵ and that a multi-particle state $|(k_1,\epsilon_1),\ldots,(k_n,\epsilon_n)\rangle$ has total four-momentum $\sum_{i=1}^n k_{i\mu}$.

Conserved $\mathrm{U}(1)$ charge
So far, the structure has been in complete analogy with the one for the real scalar field. However, the free complex scalar theory has one additional feature, namely the conserved $\mathrm{U}(1)$ current (4.77). The operator version of the associated charge can be written as

$$Q = iq \int d^3x \, (\phi^\dagger \pi^\dagger - \pi\phi) \; . \tag{5.44}$$

Inserting the expansions (5.37) one finds after some calculation

$$:Q: = q \int d^3\tilde{k} \, \left(a_+^\dagger(k) a_+(k) - a_-^\dagger(k) a_-(k) \right) = q(N_+ - N_-) \; . \tag{5.45}$$

In particular, this shows that the states $|(k,+)\rangle$ have charge $+q$ and the states $|(k,-)\rangle$ have charge $-q$. Therefore, it is sensible to identify $+$ states with particles and $-$ states with anti-particles. With this terminology, we see from Eq. (5.37) that $\phi(x)$ annihilates a particle or creates an anti-particle while $\phi^\dagger(x)$ creates a particle or annihilates an anti-particle.

5.4 Time-ordered products, propagators and Wick's theorem (again)

Time-order products of operators
In this section we will develop some of the tools necessary to deal with interacting fields. While the methods presented apply to all types of fields (subject to straightforward modifications) we will concentrate on the real scalar field, for simplicity. We have seen that a field operator such as in Eq. (5.9) can be thought of as a superposition of creation and annihilation operators. Later, we will see that interaction processes can be understood as sequences of such creation and annihilation operations and can, therefore, be quantitatively described by products of field operators. More precisely, the operator products involved are ordered in time reflecting the time-line of the physical process. We are, therefore, interested in the *time-ordered product* of field operators defined by

$$T\left(\phi(t_1,\mathbf{x}_1)\ldots\phi(t_n,\mathbf{x}_n)\right) = \phi(t_{i_n},\mathbf{x}_{i_n})\ldots\phi(t_{i_1},\mathbf{x}_{i_1}) \quad \text{where} \quad t_{i_1} \leq t_{i_2} \leq \cdots \leq t_{i_n} \; . \tag{5.46}$$

and their vacuum expectation values. The above equation means time ordering T rearranges fields operators so that time increases from the right to the left.

The Feynman propagator
Of particular importance is the vacuum expectation value of a time ordered product of two field operators which is called the *Feynman propagator* Δ_F. From this definition the Feynman propagator can we written as

$$\Delta_F(x-y) = \langle 0|T(\phi(x)\phi(y))|0\rangle = \theta(x_0-y_0)\langle 0|\phi(x)\phi(y)|0\rangle + \theta(y_0-x_0)\langle 0|\phi(y)\phi(x)|0\rangle \; , \tag{5.47}$$

where the Heaviside function θ is defined by $\theta(x_0) = 1$ for $x_0 \geq 0$ and $\theta(x_0) = 0$ for $x_0 < 0$. To compute the Feynman propagator we first evaluate the above expression for the case $x_0 > y_0$. Inserting the field expansion (5.9) we have

$$\begin{aligned}
\Delta_F(x-y) &= \langle 0|\phi(x)\phi(y)|0\rangle & (5.48)\\
&= \langle 0| \int d^3\tilde{k}\, d^3\tilde{q} \, \left(a(k)e^{-ikx} + a^\dagger(k)e^{ikx}\right)\left(a(q)e^{-iqy} + a^\dagger(q)e^{iqy}\right)|0\rangle & (5.49)\\
&= \langle 0| \int d^3\tilde{k}\, d^3\tilde{q}\, [a(k), a^\dagger(q)] e^{iqy-ikx}|0\rangle = \int d^3\tilde{k}\, e^{-ik(x-y)} \; . & (5.50)
\end{aligned}$$

Analogously, we find for the case $x_0 < y_0$ that $\Delta_F(x-y) = \int d^3\tilde{k}\, e^{ik(x-y)}$. Combining this, the Feynman

propagator can be written as

$$\Delta_F(x-y) = \int \frac{d^3k}{(2\pi)^3 2w_{\mathbf{k}}} \left\{\theta(x_0-y_0)e^{-iw_{\mathbf{k}}(x_0-y_0)} + \theta(y_0-x_0)e^{iw_{\mathbf{k}}(x_0-y_0)}\right\} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \quad (5.51)$$

$$= \int \frac{d^4k}{(2\pi)^4} \frac{i}{(k_0 - w_{\mathbf{k}} + i\tilde{\epsilon})(k_0 + w_{\mathbf{k}} - i\tilde{\epsilon})} e^{-ik(x-y)} \quad (5.52)$$

$$= \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} e^{-ik(x-y)} \, . \quad (5.53)$$

The small quantity $\tilde{\epsilon} > 0$ in the second integral is to indicate that the pole at $k_0 = w_{\mathbf{k}} - i\tilde{\epsilon}$ is slightly below the real k_0 axis and the pole at $k_0 = -w_{\mathbf{k}} + i\tilde{\epsilon}$ is slightly above (The quantity $\epsilon > 0$ in the final integral serves the same purpose). With this understanding about the position of the poles the equality of the first and second line above can be shown by a contour integration as indicated in Fig. 5.1. The above result provides us with a simple

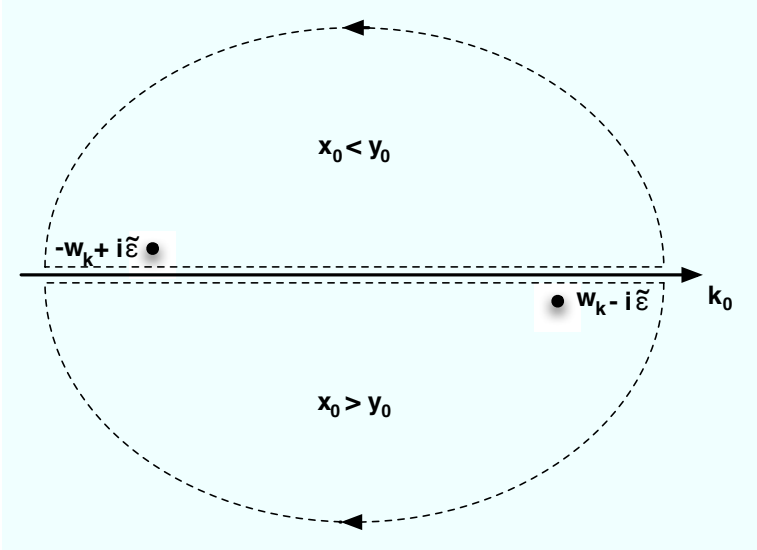

Figure 5.1: Location of poles for the Feynman propagator and contours to prove the equality between Eqs. (5.51) and (5.52). For $x_0 > y_0$ the integration along the real k_0 axis can be closed for $\mathrm{Im}(k_0) < 0$ since the real part of the exponent in Eq. (5.52) is negative in this case. Only the pole at $k_0 = w_{\mathbf{k}} - i\tilde{\epsilon}$ constributes and leads to the first term in Eq. (5.51). Analogously, for $x_0 < y_0$ the contour can be closed for $\mathrm{Im}(k_0) > 0$. Only the pole at $k_0 = -w_{\mathbf{k}} + i\tilde{\epsilon}$ contributes and leads to the second term in Eq. (5.51)

representation of the Feynman propagator and tells us that its Fourier transform $\tilde{\Delta}_F$ is given by

$$\tilde{\Delta}_F(k) = \frac{i}{k^2 - m^2 + i\epsilon} \, . \quad (5.54)$$

The Feynman propagator, particularly in the momentum space form (5.54), is central in the formulation of Feynman rules for the perturbation theory of interacting fields as we will see later.

Propagators as solutions to Klein-Gordon equation
Another interesting property of the Feynman propagator is that it solves the Klein-Gordon equation

$$(\Box_x + m^2)\Delta_F(x-y) = -i\delta^4(x-y) \, . \quad (5.55)$$

with a delta-function source. Functions with this property are also called *Green functions*. There are other Green functions of the Klein-Gordon equation which are given by an integral such as in Eq. (5.53) but with the poles in a different position relative to the real k_0 axis. For example, the case where both poles are below the real k_0 axis leads to the so-called *retarded Green function*. It vanishes for $x_0 < y_0$ since the upper contour in Fig. 5.1 contains no poles in this case.

Let us now prove Eq. (5.55). From Eq. (5.47) we have

$$\begin{aligned}\frac{\partial^2}{\partial x_0^2}\Delta_F(x-y) &= \frac{\partial}{\partial x_0}\left\{\delta(x_0-y_0)\langle 0|[\phi(x),\phi(y)]|0\rangle + \langle 0|T(\pi(x),\phi(y))|0\rangle\right\}\\ &= \delta(x_0-y_0)\langle 0|[\pi(x),\phi(y)]|0\rangle + \langle 0|T(\partial_0^2\phi(x),\phi(y))|0\rangle\\ &= -i\delta^4(x-y) + \langle 0|T(\partial_0^2\phi(x),\phi(y))|0\rangle\ ,\end{aligned}$$

where we used that $\frac{\partial}{\partial x_0}\theta(x_0)=\delta(x_0)$ and $\delta(x_0-y_0)\langle 0|[\phi(x),\phi(y)]|0\rangle = \delta(x_0-y_0)\Delta(0,\mathbf{x}-\mathbf{y})=0$ (see the discussion at the end of Section 5.2). It follows that

$$(\Box_x+m^2)\Delta_F(x-y) = \left(\frac{\partial^2}{\partial x_0^2}-\frac{\partial^2}{\partial \mathbf{x}^2}+m^2\right)\Delta_F(x-y) \tag{5.56}$$

$$= -i\delta^4(x-y) + \langle 0|T\left((\Box_x+m^2)\phi(x)\phi(y)\right)|0\rangle\ . \tag{5.57}$$

The field $\phi(x)$ satisfies the (free) Klein-Gordon equation and, hence, the second term vanishes. This completes the proof.

Evaluating time-ordered operator products, Wick's theorem

We would now like to understand how to evaluate time-ordered products and their vacuum expectation values more generally. To do so it is useful to split the field $\phi(x)$ into its positive and negative frequency parts $\phi_+(x)$ and $\phi_-(x)$ as

$$\phi(x)=\phi_+(x)+\phi_-(x)\ ,\quad \phi_+(x)=\int d^3\tilde{k}\, a(k)e^{-ikx}\ ,\quad \phi_-(x)=\int d^3\tilde{k}\, a^\dagger(k)e^{ikx}\ . \tag{5.58}$$

Since $\phi_+(x)$ only contains annihilation operators it is clear that $\phi_+(x)|0\rangle=0$. Likewise, $\langle 0|\phi_-(x)=0$. We begin with the time ordered product of two fields $\phi(x)$ and $\phi(y)$ for the case $x_0>y_0$ and write

$$\begin{aligned}T\left(\phi(x)\phi(y)\right) &= \phi_+(x)\phi_+(y)+\phi_+(x)\phi_-(y)+\phi_-(x)\phi_+(y)+\phi_-(x)\phi_-(y)\\ &= \phi_+(x)\phi_+(y)+\phi_-(x)\phi_+(y)+\phi_-(x)\phi_+(y)+\phi_-(x)\phi_-(y)+[\phi_+(x),\phi_-(y)]\\ &= :\phi(x)\phi(y): +\langle 0|[\phi_+(x),\phi_-(y)]|0\rangle\\ &= :\phi(x)\phi(y): +\langle 0|\phi(x)\phi(y)|0\rangle\end{aligned}$$

The point about introducing the commutator in the second line is that the first four terms have the creation operators to the left of the annihilation operators and, hence, correspond to the normal ordering of the field product. A similar calculation for the case $x_0<y_0$ leads to $T\left(\phi(x)\phi(y)\right)=:\phi(x)\phi(y):+\langle 0|\phi(y)\phi(x)|0\rangle$. Combining these two results we have

$$T\left(\phi(x)\phi(y)\right) =: \phi(x)\phi(y)+\Delta_F(x-y):\ . \tag{5.59}$$

Hence, we see that time and normal ordering are related via a Feynman propagator. When two field operators in a time-ordered product are combined into a Feynman propagator as in the above equation they are said to have been *contracted*. With this terminology we can formulate *Wick's theorem* as

$$T\left(\phi(x_1)\dots\phi(x_n)\right) =: \phi(x_1)\dots\phi(x_n) + \text{all possible contractions}: \tag{5.60}$$

By "all possible contraction" we refer to all possible ways of pairing up the field operators $\phi(x_1),\dots,\phi(x_n)$ into Feynman propagators, including partial pairings. When taking the vacuum expectation value of the above equation the first term and all partially contracted terms vanish due to normal ordering. This shows that the vacuum expectation of an odd number of time-ordered fields vanishes. For an even number of fields we are left with

$$\langle 0|T\left(\phi(x_1)\dots\phi(x_n)\right)|0\rangle = \sum_{\text{pairings}\,p}\Delta_F(x_{i_1}-x_{i_2})\dots\Delta_F(x_{i_{n-1}}-x_{i_n})\ , \tag{5.61}$$

where the sum runs over all pairings $p=\{(i_1,i_2),\dots,(i_{n-1},i_n)\}$ of the numbers $1,\dots,n$. This is precisely the same structure as the one we encountered in the context of n-point functions for Gaussian multiple integrals and Gaussian functional integrals. This is of course not an accident and the precise relation will become clear when we discuss the path integral quantisation of field theories.

Sketch of proof for Wick's theorem

We still need to prove Wick's theorem in its general form (5.60). For the case of two fields, $n=2$, we have already

done this in Eq. (5.59). The general case can be proven by induction in n. Rather than presenting this general argument it is probably more instructive to consider the example for $n = 3$. After a suitable relabelling we can arrange that $t_1 \geq t_2 \geq t_3$. With the short-hand notation $\phi_i = \phi(x_i)$ and $\Delta_{ij} = \Delta_F(x_i - x_j)$ we have

$$\begin{aligned} T(\phi_1\phi_2\phi_3) &= \phi_1\phi_2\phi_3 = \phi_1 \,:\phi_2\phi_3: + \Delta_{23}\, : \\ &= (\phi_{1+} + \phi_{1-}) \,:\phi_{2+}\phi_{3+} + \phi_{2-}\phi_{3+} + \phi_{3-}\phi_{2+} + \phi_{2-}\phi_{3-}: + \Delta_{23}\, : \\ &= :\phi_{1-}\phi_2\phi_3 + \phi_{1+}\phi_{2+}\phi_{3+} + \phi_{2-}\phi_{1+}\phi_{3+} + \phi_{3-}\phi_{1+}\phi_{2+} + \phi_{2-}\phi_{3-}\phi_{1+}: + \Delta_{23}\phi_1 \,: \\ &\quad + [\phi_{1+}, \phi_{2-}]\phi_{3+} + [\phi_{1+}, \phi_{3-}]\phi_{2+} + [\phi_{1+}, \phi_{2-}\phi_{3-}] \\ &= :\phi_1\phi_2\phi_3 + \Delta_{12}\phi_{3+} + \Delta_{13}\phi_{2+} + \Delta_{12}\phi_{3-} + \Delta_{13}\phi_{2-}: + \Delta_{23}\phi_1 \,: \\ &= :\phi_1\phi_2\phi_3 + \Delta_{12}\phi_3 + \Delta_{13}\phi_2 + \Delta_{23}\phi_1 : \end{aligned}$$

In the first line we have simply used Wick's theorem (5.59) for a product of two fields, applied to ϕ_2 and ϕ_3. In the second line, we need to move ϕ_{1+} and ϕ_{1-} into the normal ordering. This is easy for ϕ_{1-} since it consists of creation operators and, hence, has to be on the left of a product anyway. The annihilation part ϕ_{1+} of the field, on the other hand, has to be commuted to the right of any component ϕ_{i-}. The key is that the commutators which arise in this way precisely lead to the necessary contractions of ϕ_1 with ϕ_2 and ϕ_3. The proof for general n is analogous to the above calculation. It applies the induction assumption (that is, the validity of Wick's theorem for $n-1$ fields) to $\phi_2, \ldots, \phi_n$. Then, moving ϕ_1 into the normal ordering and commuting ϕ_{1+} with all negative frequency parts ϕ_{i-} generates all the missing contractions of ϕ_1 with the other fields.

Time-ordered product of complex scalar fields
We should briefly discuss time-ordered products for a free complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, described by the Lagrangian density (4.72) with $\lambda = 0$. In terms of the two real fields ϕ_1 and ϕ_2 the Lagrangian (4.72) splits into a sum of Lagrangians for two free real scalars with the same mass m. Hence, each of ϕ_1 and ϕ_2 has an oscillator expansions as in Eq. (5.9) and it immediately follows that $\langle 0|T(\phi_1(x)\phi_1(y))|0\rangle = \langle 0|T(\phi_2(x)\phi_2(y))|0\rangle = \Delta_F(x-y)$ and $\langle 0|T(\phi_1(x)\phi_2(y))|0\rangle = 0$. This implies for the complex scalar ϕ that

$$\langle 0|T\left(\phi(x)\phi^\dagger(y)\right)|0\rangle = \Delta_F(x-y)\,, \quad \langle 0|T(\phi(x)\phi(y))|0\rangle = \langle 0|T\left(\phi^\dagger(x)\phi^\dagger(y)\right)|0\rangle = 0\,. \tag{5.62}$$

For a product of operators ϕ and $\phi^\dagger$ Wick's theorem can be applied straightforwardly, but the above equations tell us that only contractions of ϕ with $\phi^\dagger$ need to be taken into account.

5.5 Canonical quantization of a vector field

Recap of theory and canonical quantisation conditions
We recall from Section 4.4.1 that the Lagrangian density for a vector field A_μ with associated field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is given by

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{5.63}$$

and that this Lagrangian is invariant under the gauge transformations

$$A_\mu \to A_\mu + \partial_\mu \Lambda\,. \tag{5.64}$$

From Eq. (4.99), the canonical momenta π^μ are

$$\pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = F^{\mu 0}\,. \tag{5.65}$$

To quantize this theory we interpret the field A_μ as a collection of four scalar fields (which happen to be labelled by a space-time index) and then follow the canonical quantization procedure. We should then impose the canonical commutation relations (5.4) which take the form

$$[\pi^\mu(t,\mathbf{x}), A_\nu(t,\mathbf{y})] = -i\delta^\mu_\nu \delta^3(\mathbf{x}-\mathbf{y}) \tag{5.66}$$

However, Eq. (5.65) implies that the canonical momentum π^0 vanishes and this is inconsistent with the $\mu = \nu = 0$ part of Eq. (5.66). Clearly, viewing A_μ as four scalar fields is too simple. In fact, while typical kinetic terms for four scalar fields A_μ would be of the form $\sum_\nu \partial_\mu A_\nu \partial^\mu A_\nu$ and, hence, depend on the symmetric and anti-symmetric parts of $\partial_\mu A_\nu$, the Lagrangian (5.63) only depends on the anti-symmetric part, that is, on the field

strength $F_{\mu\nu}$. This special form of the Maxwell Lagrangian is of course responsible for the existence of the gauge symmetry (5.64) as well as for the vanishing of the conjugate momentum π^0 and it links these two features. In essence, gauge symmetry is the crucial difference to the scalar field theory. There are various viable methods to quantise a gauge theory, but here we will follow the most obvious approach of fixing a gauge before quantisation.

Gauge fixing

Since we would like to manifestly preserve covariance we use the Lorentz gauge condition (4.102), which we impose on the theory by means of a Lagrange multiplier λ. This means that, instead of the Maxwell Lagrangian (5.63), we start with

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{\lambda}{2}(\partial_\mu A^\mu)^2 \,, \tag{5.67}$$

For the conjugate momenta we now find

$$\pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = F^{\mu 0} - \lambda \eta^{\mu 0}\partial_\nu A^\nu \,, \tag{5.68}$$

and, hence, in particular $\pi^0 = -\lambda\partial_\nu A^\nu$ is no longer zero. The equation of motion (4.101) for A_μ is now modified to

$$\Box A_\mu - (1-\lambda)\partial_\mu\partial_\nu A^\nu = 0 \,, \tag{5.69}$$

and, in addition, we have to impose the gauge condition $\partial_\mu A^\mu = 0$ (which formally arises from the action (5.67) as the λ equation of motion). However, clearly we should not impose this condition as an operator equation since this would lead us back to a situation where $\pi^0 = 0$. Instead, it will later be imposed as a condition on physical states. At any rate, the obstruction to imposing canonical quantisation conditions has been removed and we require the canonical commutation relations (5.66) for A_μ and the conjugate momenta (5.68).

Oscillator expansion

We should now solve the equation of motion (5.69) to work out the properties of the creation and annihilation operators. While this can be done for arbitrary λ, we adopt the so-called *Feynman gauge* [4], $\lambda = 1$, which simplifies the equation of motion (5.67) to $\Box A_\mu = 0$. The general solution to this equation has already been written down in Section 4.4.1 and is given by

$$A_\mu(x) = \sum_{\alpha=0}^{3}\int d^3\tilde{k}\,\epsilon_\mu^{(\alpha)}(k)\left(a^{(\alpha)}(k)e^{-ikx} + a^{(\alpha)\dagger}(k)e^{ikx}\right) \,. \tag{5.70}$$

Here, we have used the polarisation vectors $\epsilon_\mu^{(\alpha)}(k)$ defined before Eq. (4.109) and the four-momentum is given by $(k_\mu) = (w_{\mathbf{k}}, \mathbf{k})$ with $w_{\mathbf{k}} = |\mathbf{k}|$. We recall from our discussion in Section 4.4.1 that classically $\epsilon_\mu^{(\alpha)}(k)$ for $\alpha = 1, 2$ are the two transversal, physical polarisations while the other two polarisations can be removed through gauge transformations.

Inserting Eq. (5.70) into Eq. (5.68) gives the expansion for the conjugate momenta. Similar to the scalar field case (see below Eq. (5.9)), we can now invert these expansions to express the oscillators $a^{(\alpha)}(k)$ and their hermitian conjugates in terms of the field A_μ and the momenta π^μ. As for scalar fields, these results can be inserted into the quantisation condition (5.66) to determine the commutation relations of the oscillators. After a straightforward (but slightly lengthy) calculation one finds that

$$[a^{(\alpha)}(k), a^{(\alpha')}(q)] = [a^{(\alpha)\dagger}(k), a^{(\alpha')\dagger}(q)] = 0 \,, \quad [a^{(\alpha)}(k), a^{(\alpha')\dagger}(q)] = -\eta^{\alpha\alpha'}(2\pi)^3 2w_{\mathbf{k}}\delta^3(\mathbf{k}-\mathbf{q}) \tag{5.71}$$

The Fock space

As comparison with Eq. (5.13) shows the above commutation relations are very similar to the one for four real scalar fields, however with one crucial difference: The commutator $[a^{(0)}(k), a^{(0)\dagger}(q)]$ has the opposite sign to that in Eq. (5.13). Naively, the Fock space of this theory is spanned by states created from the vacuum $|0\rangle$ by acting with any combination and any number of operators $a^{(\alpha)\dagger}(k)$, where $\alpha = 0, 1, 2, 3$. Let us call this space $\mathcal{F}_0$. In particular, $\mathcal{F}_0$ contains the state $|(k,0)\rangle = a^{(0)\dagger}(k)|0\rangle$ which satisfies

$$\langle(k,0)|(k,0)\rangle = \langle 0|[a^{(0)}(k), a^{(0)\dagger}(k)]|0\rangle < 0 \,, \tag{5.72}$$

[4] Although common terminology this is somewhat misleading since the choice of λ is clearly not a gauge choice in the usual sense. We also remark that independence of the quantisation procedure on the choice of λ can be explicitly demonstrated.

and, hence, has negative norm. Clearly this is physically unacceptable and indicates that the space $\mathcal{F}_0$ still contains unphysical states and cannot be the proper Fock space of the theory. A related problem emerges when looking at the conserved four-momentum. Calculating the four-momentum for the theory (5.67) from the general formalism in Section 4.2.3 and inserting the field expansion (5.70) one finds after some calculation

$$: P_\mu := \int d^3\tilde{k}\, k_\mu \left[\sum_{\alpha=1}^{3} a^{(\alpha)\dagger}(k)a^{(\alpha)}(k) - a^{(0)\dagger}(k)a^{(0)}(k)\right] . \tag{5.73}$$

The negative sign in front of the last term means that the state $|(k,0)\rangle$ has "negative energy".

The existence of unphysical states is not at all surprising given that we have not yet imposed the Lorentz gauge condition. We have seen above that requiring the operator equation $\partial_\mu A^\mu = 0$ is too strong and leads to problems with quantisation. Instead we define a space $\mathcal{F}_1 \subset \mathcal{F}_0$ of physical states on which the gauge condition is satisfied, that is we require

$$\langle\tilde{\Phi}|\partial_\mu A^\mu|\Phi\rangle = 0 . \tag{5.74}$$

between two physical states $|\Phi\rangle, |\tilde{\Phi}\rangle \in \mathcal{F}_1$. To guarantee this condition is satsified it is sufficient that $(\partial_\mu A^\mu)^{(+)}|\Phi\rangle = 0$ where the annihilation part $(\partial_\mu A^\mu)^{(+)}$ of $\partial_\mu A^\mu$ is proportional to

$$\int d^3\tilde{k}\, k_0 \left(a^{(0)}(k) - a^{(3)}(k)\right) e^{-ikx} . \tag{5.75}$$

To obtain this last result, we have taken the ∂_μ derivative of the first term in Eq. (5.70) and used that $k^\mu \epsilon_\mu^{(\alpha)}(k) = 0$ for $\alpha = 1, 2$ and $k^\mu \epsilon_\mu^{(0)}(k) = -k^\mu \epsilon_\mu^{(3)}(k) = k_0$ (see Section 4.4.1). This means physical states $|\Phi\rangle \in \mathcal{F}_1$ are defined by the condition

$$b_-(k)|\Phi\rangle = 0 . \tag{5.76}$$

where we have defined a new basis $b_\pm(k) = (a^{(3)}(k) \pm a^{(0)}(k))/\sqrt{2}$ for the non-transversal operators. Clearly all transversal states, that is, states created from the vacuum by acting only with transversal creation operators $a^{(\alpha)\dagger}(k)$, where $\alpha = 1, 2$ along with their linear combinations, satisfy the condition (5.76) and, hence, are elements of $\mathcal{F}_1$. It would be nice if $\mathcal{F}_1$ consisted of such states only but things are not quite so simple. To analyse the condition (5.76) for non-transversal states we note that

$$[b_-(k), b_-^\dagger(q)] = 0 , \tag{5.77}$$

while $[b_-(k), b_+^\dagger(q)] \neq 0$. This means, in addition to $a^{(\alpha)\dagger}(k)$, where $\alpha = 1, 2$, physical states can also contain $b_-^\dagger(k)$ but not $b_+^\dagger(k)$. This means $\mathcal{F}_1$ is spanned by states created from the vacuum acting with any number of operators $a^{(\alpha)\dagger}(k)$, where $\alpha = 1, 2$ and $b_-^\dagger(k)$. Note that the commutators of these operators with their associated annihilation operators are all non-negative and, hence, no states with negative norm are left in $\mathcal{F}_1$. However, if a state contains at least one operator $b_-^\dagger(k)$ its norm vanishes, as is clear from the vanishing of the commutator (5.77). Physically, we should discard such zero norm states and the formal way of doing this is to identify each two states in $\mathcal{F}_1$ if their difference has zero norm. In this way, we obtain the proper Fock space $\mathcal{F}_2$, whose elements are the classes of states obtained from this identification. In particular, in each class there is a "representative" with only transverse oscillators. In conclusion, we see that the proper Fock space $\mathcal{F}_2$ can be thought of as spanned by states of the form $|(k_1,\alpha_1),\ldots,(k_n,\alpha_n)\rangle = a^{(\alpha_1)\dagger}(k_1)\ldots a^{(\alpha_n)\dagger}(k_n)|0\rangle$, where $\alpha_i = 1, 2$ are transverse oscillators only.

Independence of physical quantities on Fock space representative

One remaining point which needs checking is that physical quantities are independent on which representative for a class in $\mathcal{F}_2$ is being used. Let us verify this for the case of the four-momentum (5.73) which can also be written in the form

$$: P_\mu := \int d^3\tilde{k}\, k_\mu \left[\sum_{\alpha=1}^{2} a^{(\alpha)\dagger}(k)a^{(\alpha)}(k) + b_+^\dagger(k)b_-(k) + b_-^\dagger(k)b_+(k)\right] . \tag{5.78}$$

Taken between a physical state $|\Phi\rangle \in \mathcal{F}_1$, that is $\langle\Phi| : P_\mu : |\Phi\rangle$, the last two terms in Eq. (5.78) do not contribute (using that all a type operators commutate with all b type operators as well as Eqs. (5.76), (5.77)). This means

$$\langle\Phi| : P_\mu : |\Phi\rangle = \langle\Phi| \int d^3\tilde{k}\, k_\mu \sum_{\alpha=1}^{2} a^{(\alpha)\dagger}(k)a^{(\alpha)}(k)|\Phi\rangle . \tag{5.79}$$

This shows that the four-momentum only depends on the transverse modes and, since every class in $\mathcal{F}_2$ has exactly one representative with transverse modes only, the desired independence on the choice of representative.

Feynman propagator
Finally, we should look at the Feynman propagator for a vector field, defined, as usual, as the vacuum expectation value $\langle 0|T(A_\mu(x)A_\nu(y))|0\rangle$ of a time-ordered product of two fields. Inserting the field expansion (5.70) and using the commutation relations (5.71) we can perform a calculation completely analogous to the one in Section 5.4 to obtain the vector field propagator in Feynman gauge, $\lambda = 1$. This leads to

$$D_{F,\mu\nu}(x-y) \equiv \langle 0|T(A_\mu(x)A_\nu(y))|0\rangle = -\eta_{\mu\nu}\Delta_{F,0}(x-y)\ , \tag{5.80}$$

where $\Delta_{F,0}$ is the scalar field Feynman propagator (5.53) for $m = 0$.

5.6 Further reading

Canonical quantisation is covered in most standard text books on quantum field theory, including

- J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, vol 2, chapters 11, 12, 14.
- C. Itzykson and J.-B. Zuber, *Quantum Field Fields*, chapters 3.1, 3.2.
- M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, chapter 2.

Chapter 6

Interacting Quantum Fields

In this chapter, we will develop the formalism for perturbatively interacting quantum field theory. The main goal is to derive the Feynman rules for bosonic field theories from first principles and show how to calculate matrix elements and cross sections. To do this we will first introduce the S-matrix and then show how cross sections and decay rates are expressed in terms of S-matrix elements. We then need to understand how to compute S-matrix elements from field theory. The first step is to derive the reduction formula which shows how to compute S-matrix elements from vacuum expectation values of time-ordered products of interacting fields. These products of interacting fields are then re-written in terms of free fields using the evolution operator. From there, Wick's theorem will lead to the Feynman rules. Whenever the formalism needs to be developed with reference to a particular theory we will first focus on the real scalar field theory (4.47) with $\lambda\phi^4$ interaction for simplicity and subsequently generalise to more complicated theories.

6.1 The S-matrix

Consider an interacting field theory [1] with field $\phi(x)$, where $(x) = (t, \mathbf{x})$. We would like to describe a scattering process, that is a process where interactions are only important for a finite time around $t \simeq 0$ and the field ϕ asymptotically approaches free fields in the limits $t \to \pm\infty$. For $t \to -\infty$ we denote the free-field limit of ϕ by $\phi_{\rm in}$ and for $t \to +\infty$ by $\phi_{\rm out}$. To these asymptotically free fields we can associate creation and annihilation operators $a_{\rm in}$, $a_{\rm out}$ and $a^\dagger_{\rm in}$, $a^\dagger_{\rm out}$ in the way discussed in the previous chapter and define "in" and "out" states as

$$|k_1, \ldots, k_n\rangle_{\rm in} = a^\dagger_{\rm in}(k_1)\ldots a^\dagger_{\rm in}(k_n)|0\rangle \ , \quad |k_1, \ldots, k_n\rangle_{\rm out} = a^\dagger_{\rm out}(k_1)\ldots a^\dagger_{\rm out}(k_n)|0\rangle \tag{6.1}$$

Given an initial state $|i\rangle_{\rm in} = |k_1, \ldots, k_n\rangle_{\rm in}$ with n particles and momenta $k_1, \ldots, k_n$ and a final state $|f\rangle_{\rm out} = |q_1, \ldots, q_m\rangle$ with m particles and momenta $q_1, \ldots, q_m$, we are interested in computing the amplitude

$${}_{\rm out}\langle f|i\rangle_{\rm in} \tag{6.2}$$

which provides the probability for a transition from $|i\rangle_{\rm in}$ to $|f\rangle_{\rm out}$. With the operator S defined by

$$|q_1, \ldots, q_m\rangle_{\rm out} = S^\dagger|q_1, \ldots, q_m\rangle_{\rm in} \tag{6.3}$$

this amplitude can be written as

$$S_{fi} \equiv {}_{\rm in}\langle f|S|i\rangle_{\rm in} = {}_{\rm out}\langle f|i\rangle_{\rm in} \ . \tag{6.4}$$

The matrix S_{fi} is called the ***S-matrix*** and it encodes the basic physical information we wish to calculate. Assuming that both the in states $|i\rangle_{\rm in}$ and the out states $|f\rangle_{\rm out}$ form a complete set of states on the Fock space we have

$$\mathbf{1} = \sum_f |f\rangle_{\rm out}\,{}_{\rm out}\langle f| = \sum_f S^\dagger|f\rangle_{\rm in}\,{}_{\rm in}\langle f|S = S^\dagger S \tag{6.5}$$

and, hence, the S matrix is unitary. We also require that the vacuum state is invariant and unique so that

$$|0\rangle_{\rm in} = |0\rangle_{\rm out} \equiv |0\rangle \tag{6.6}$$

[1] For simplicity we consider a single field but the generalisation to multiple fields should be obvious.

It is customary to write the S-matrix as

$$S = \mathbf{1} + iT \tag{6.7}$$

where the unit operator corresponds to the free evolution from "in" to "out" states and T encodes the non-trivial scattering. In a theory which conserves four-momentum the matrix elements of T take the general form

$$T_{fi} \equiv {}_{\rm in}\langle f|T|i\rangle_{\rm in} = (2\pi)^4\delta^4(k_i - q_f)\mathcal{M}(i \to f) \,, \tag{6.8}$$

where k_i and q_f are the total four-momenta of the initial and final state and $\mathcal{M}$ is the *invariant matrix element*. We need to understand how to compute $\mathcal{M}$ from the underlying field theory and, to make contact with experimentally measurable quantities, we need to express cross sections and decay rates in terms of $\mathcal{M}$. We start with the latter.

6.2 Cross sections and decay rates

Wave packets and transition probability

The incoming state in a scattering experiment is of course not an exact momentum eigenstate. More realistically, an n particle incoming state can be described by a wave packet

$$|\tilde{i}\rangle_{\rm in} = \int \left(\prod_{a=1}^{n} d^3\tilde{p}_a\, \tilde{f}_a(p_a)\right) |p_1, \ldots, p_n\rangle_{\rm in} \,, \quad f_a(x) = \int d^3\tilde{p}\, \tilde{f}_a(p) e^{-ipx} \,, \tag{6.9}$$

where $p_0 = w_{\mathbf{p}}$ in the second integral, so that $f_a(x)$ are solutions of the Klein-Gordon equation. By writing $f_a(x) = e^{-ik_a x}F_a(x)$ with slowly varying functions $F_a(x)$ we ensure that these solutions are "close" to momentum eigenstates with momenta k_a. The probability currents $j_{a\mu}$ for these solutions are given by the usual formula in relativistic quantum mechanics

$$j_{a\mu} \equiv i\left(f_a^\star(x)\partial_\mu f_a(x) - f_a(x)\partial_\mu f_a^\star(x)\right) \simeq 2k_{a\mu}|f_a(x)|^2 \,. \tag{6.10}$$

They provide us with expressions for the number of particles per volume and the flux per volume which will be needed in the calculation of the cross section.

In a first instance, we are interested in the transition probability

$$W(\tilde{\imath} \to f) = |{}_{\rm in}\langle f|S|\tilde{i}\rangle_{\rm in}|^2 \tag{6.11}$$

from the initial state $|\tilde{i}\rangle_{\rm in}$ as defined above to some final state $|f\rangle_{\rm out}$. Inserting Eqs. (6.7), (6.8) and (6.9) into this expression [2] and keeping only the non-trivial scattering part, related to the T matrix, one finds

$$\begin{aligned} W(\tilde{\imath} \to f) &= (2\pi)^8 \int \left(\prod_{a=1}^{n} d^3\tilde{p}_a\, \tilde{f}_a^\star(p_a)\right)\left(\prod_{b=1}^{n} d^3\tilde{p}'_b\, \tilde{f}_b(p'_b)\right)\delta^4\left(\sum_a p_a - \sum_b p'_b\right) && (6.12)\\ &\times \quad \delta^4\left(q_f - \sum_a p_a\right)\mathcal{M}(p_1, \ldots, p_n \to f)^\star \mathcal{M}(p'_1, \ldots, p'_n \to f) \,. && (6.13) \end{aligned}$$

The wave functions $\tilde{f}_a$ are peaked around momenta k_a and we can, hence, approximate $\mathcal{M}(p_1, \ldots, p_n \to f) \simeq \mathcal{M}(p'_1, \ldots, p'_n \to f) \simeq \mathcal{M}(k_1, \ldots, k_n \to f)$ in the above integral. Together with the integral representation

$$(2\pi)^4\delta^4\left(\sum_a p_a - \sum_a p'_a\right) = \int d^4x\, \exp\left(-ix\left(\sum_a p'_a - \sum_a p_a\right)\right) \tag{6.14}$$

of the delta function this implies

$$W(\tilde{\imath} \to f) = (2\pi)^4 \int d^4x \left(\prod_{a=1}^{n} |f_a(x)|^2\right)\delta^4\left(q_f - \sum_a k_a\right)|\mathcal{M}(k_1, \ldots, k_n \to f)|^2 \,. \tag{6.15}$$

[2] When doing this, we have to keep in mind that the "in" states $|i\rangle_{\rm in}$ in the definition (6.8) of the matrix element are the exact momentum eigenstates $|k_1, \ldots, k_1\rangle_{\rm in}$, whereas the "in" states $|\tilde{i}\rangle_{\rm in}$ which enter the transition probability (6.11) are the wave packets (6.9).

We have obtained our first important result, the transition probability from state $|\tilde{i}\rangle_{\rm in}$ to state $|f\rangle_{\rm out}$ per unit time and volume

$$\frac{dW(\tilde{i}\to f)}{dVdt} = \left(\prod_{a=1}^{n}|f_a(x)|^2\right)(2\pi)^4\delta^4\left(q_f-\sum_a k_a\right)|\mathcal{M}(k_1,\ldots,k_n\to f)|^2\,. \tag{6.16}$$

Decay rate
Let us now be more specific and first consider a single particle with mass M, momentum $k=(M,\mathbf{0})$ and wave function $f(x)$ in the initial state, that is, we study the decay of a particle at rest. Let us define the decay rate $\Gamma(k\to f)$ as the transition probability per unit time and volume (6.16) divided by the number density of initial particles and integrated over the momenta $q_1,\ldots,q_m$ of the final state $|f\rangle_{\rm out}$ with energy resolution Δ. From Eq. (6.10), the number density of initial particles is given by $2M|f(x)|^2$ and, hence, from Eq. (6.16) we find for the decay rate into m particles

$$\Gamma(k\to f) = \frac{1}{2M}\int_\Delta\left(\prod_{b=1}^{m}d^3\tilde{q}_b\right)(2\pi)^4\delta^4\left(k-q_f\right)|\mathcal{M}(k\to f)|^2\,. \tag{6.17}$$

Cross section for two-particle scattering
As the next example, we discuss the scattering of two particles with momenta k_1 and k_2 and masses m_1 and m_2 in the initial state. For simplicity, we consider particles of type 2 at rest in the laboratory frame, that is we consider them as target particles, and particles of type 1 as the incident ones. We then define the cross section $\sigma(k_1,k_2\to f)$ as the transition probability per unit time and volume (6.16) divided by the number density of target particles, divided by the incident flux density and integrated over the momenta $q_1,\ldots,q_m$ of the final state $|f\rangle_{\rm out}$ with energy resolution Δ. From Eq. (6.10) the number density of target particles is $2m_2|f_2(x)|^2$ and the incident flux density is $2|\mathbf{k}_1||f_1(x)|^2$. Dividing Eq. (6.16) by these densities and integrating over the final state momenta the cross section for a $2\to m$ scattering turns out to be

$$\sigma(k_1,k_2\to f) = \frac{1}{4\sqrt{(k_1\cdot k_2)^2-m_1^2m_2^2}}\int_\Delta\left(\prod_{b=1}^{m}d^3\tilde{q}_b\right)(2\pi)^4\delta^4\left(k_1+k_2-q_f\right)|\mathcal{M}(k_1,k_2\to f)|^2\,. \tag{6.18}$$

Here, we have re-written the original kinematical pre-factor $1/(4m_2|\mathbf{k}_1|)$ in a covariant way using the identity

$$m_2|\mathbf{k}_1| = m_2\sqrt{k_{10}^2-m_1^2} = \sqrt{(k_1\cdot k_2)^2-m_1^2m_2^2}\,. \tag{6.19}$$

In this way, the above expression for the cross section is manifestly covariant.

Cross section for $2\to 2$ scattering
We would now like to specialise this result somewhat further to a two particle final state with momenta q_1 and q_2, that is, we consider a $2\to 2$ scattering. We carry out the discussion in the *center of mass frame*, the frame in which the total initial three-momentum vanishes, that is $\mathbf{k}_1+\mathbf{k}_2=0$. Three-momentum conservation of course then implies that $\mathbf{q}_1+\mathbf{q}_2=0$. The kinematic situation for $2\to 2$ scattering in the center of mass frame is summarised in Fig. 6.1. For simplicity, we also assume that the masses of all four particles are equal to m, so in particular we have $k_1\cdot k_2 = q_1\cdot q_2$. This means the kinematical pre-factor in Eq. (6.18) can be re-written as

$$\sqrt{(k_1\cdot k_2)^2-m^4} = \sqrt{(q_1\cdot q_2)^2-m^4} = \sqrt{(q_{10}q_{20}+\mathbf{q}_1^2)^2-(q_{10}^2-\mathbf{q}_1^2)(q_{20}^2-\mathbf{q}_1^2)} = E|\mathbf{q}_1|\,, \tag{6.20}$$

where $E=k_{10}+k_{20}=q_{10}+q_{20}$ is the total *center of mass energy*. We write the integral in Eq. (6.18) as

$$\begin{aligned}\int_\Delta d^3\tilde{q}_1d^3\tilde{q}_2\,(2\pi)^4\delta^4(k_1+k_2-q_1-q_2) &= \int_\Delta\frac{d^3\mathbf{q}_1}{(2\pi)^3 2q_{10}}\frac{d^3\mathbf{q}_2}{(2\pi)^3 2q_{20}}(2\pi)^4\delta(E-q_{10}-q_{20})\delta^3(\mathbf{q}_1+\mathbf{q}_2)\\ &= \frac{1}{16\pi^2}\int_\Delta\frac{d\Omega\,|\mathbf{q}_1|^2d|\mathbf{q}_1|}{q_{10}q_{20}}\delta(E-q_{10}-q_{20})\end{aligned} \tag{6.21}$$

where we have carried out the $\mathbf{q}_2$ integral in the last step, so that $q_{10}=q_{20}=\sqrt{m^2+\mathbf{q}_1^2}$. Also $d\Omega=\sin\theta\,d\theta\,d\phi$ is the solid-angle differential for $\mathbf{q}_1$. Since

$$\frac{d}{d|\mathbf{q}_1|}(q_{10}+q_{20}) = \frac{|\mathbf{q}_1|}{q_{10}}+\frac{|\mathbf{q}_1|}{q_{20}} = \frac{|\mathbf{q}_1|E}{q_{10}q_{20}} \tag{6.22}$$

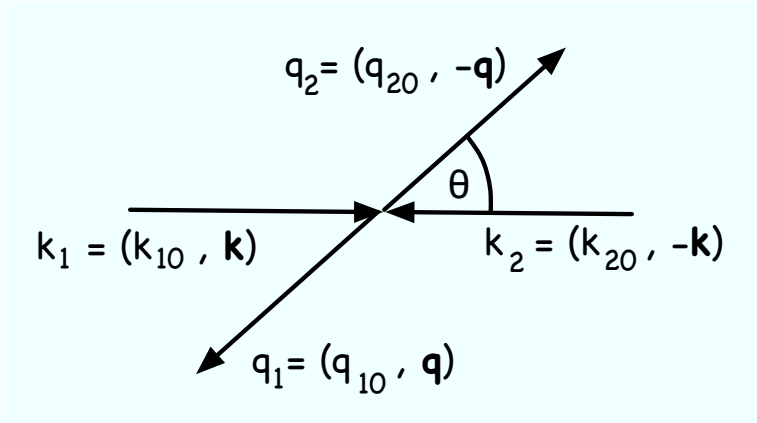

Figure 6.1: Kinematics of $2 \to 2$ scattering in the center of mass frame. The total center of mass energy is given by $E = k_{10} + k_{20} = q_{10} + q_{20}$.

we can use the delta function to replace the integral over $|\mathbf{q}_1|$ in Eq. (6.21) by the inverse of this factor. Inserting Eq. (6.21) in this form together with Eq. (6.20) into Eq. (6.18) and assuming good angular resolution, so that the integral $\int_\Delta$ can be dropped we have

$$\frac{d\sigma(k_1, k_2 \to q_1, q_2)}{d\Omega} = \frac{1}{64\pi^2 E^2} |\mathcal{M}(k_1, k_2 \to q_1, q_2)|^2 \ . \tag{6.23}$$

To summarise, this expression provides the differential cross section for a $2 \to 2$ scattering where all four particles have the same mass. Note that E is the total center of mass energy. Since we have integrated out the delta-function in Eq. (6.18) the matrix element should be evaluated for conserved four-momentum.

6.3 The reduction formula

Removing a particle from the in state

Having expressed decay rates and cross sections in terms of S-matrix elements we now need to understand how to compute these S-matrix elements from quantum field theory. The first step is to derive the so-called *LSZ reduction formula* (after Lehmann, Symanzik and Zimmermann) which allows one to express S-matrix element in terms of vacuum expectation values of (time-ordered) field operator products. Focusing on a real scalar field ϕ, we now derive this formula for an initial state $|i, k\rangle_{\rm in}$ with a particle of momentum k and an arbitrary set of other particles, collectively denoted as i, and a final state $|f\rangle$. The relevant S-matrix element can then be written as

$$\begin{aligned} {}_{\rm out}\langle f|i, k\rangle_{\rm in} &= {}_{\rm out}\langle f|a^\dagger_{\rm in}(k)|i\rangle = {}_{\rm out}\langle f|a^\dagger_{\rm out}(k)|i\rangle_{\rm in} + {}_{\rm out}\langle f|(a^\dagger_{\rm in}(k) - a^\dagger_{\rm out}(k))|i\rangle_{\rm in} & (6.24)\\ &= {}_{\rm out}\langle f - k|i\rangle_{\rm in} - {}_{\rm out}\langle f|i \int d^3x\, e^{-ikx} \overleftrightarrow{\partial_0} (\phi_{\rm in}(x) - \phi_{\rm out}(x)|i\rangle_{\rm in} & (6.25) \end{aligned}$$

The creation operators have been expressed in terms of "in" and "out" fields by means of Eq. (5.12) which can also be written in the form

$$a^\dagger_{\rm in/out}(k) = -i \int d^3x\, e^{-ikx} \overleftrightarrow{\partial_0} \phi_{\rm in/out}(x) \ , \tag{6.26}$$

where $k_0 = w_{\mathbf{k}}$, as usual. Here, we have used the short-hand notation $f(t) \overleftrightarrow{\partial_t} g(t) = f(t)\partial_t g(t) - \partial_t f(t) g(t)$. The state $|f - k\rangle_{\rm out}$ refers to the "out" state $|f\rangle_{\rm out}$ with a particle of momentum k removed (or zero if there is no such particle in $|f\rangle_{\rm out}$). Hence, the first term in Eq. (6.25) either vanishes or corresponds to a forward scattering contribution where one of the momenta has not changed. We will usually drop this term in the subsequent calculation since we are interested in the non-trivial scattering part of the amplitude. Taking time limits $x_0 \to \pm\infty$

we can convert the "in" and "out" fields above into a full interacting field ϕ by writing

$$
\begin{aligned}
{}_{\rm out}\langle f|i,k\rangle_{\rm in} &= {}_{\rm out}\langle f|\left(\lim_{x_0\to\infty}-\lim_{x_0\to-\infty}\right) i\int d^3x\, e^{-ikx}\overleftrightarrow{\partial_0}\phi(x)|i\rangle_{\rm in} &(6.27)\\
&= {}_{\rm out}\langle f|i\int d^4x\,\partial_0\left(e^{-ikx}\overleftrightarrow{\partial_0}\phi(x)\right)|i\rangle_{\rm in} &(6.28)\\
&= {}_{\rm out}\langle f|i\int d^4x\,\left(e^{-ikx}\partial_0^2\phi(x)-\partial_0^2e^{-ikx}\phi(x)\right)|i\rangle_{\rm in} &(6.29)
\end{aligned}
$$

The wave function e^{-ikx} (with $k_0=w_{\mathbf{k}}$) satisfies the Klein-Gordon equation, so that $\partial_0^2e^{-ikx}=(\nabla^2-m^2)e^{-ikx}$. Using this in the second term in Eq. (6.29) and subsequently integrating by parts in the spatial directions one finds

$$
{}_{\rm out}\langle f|i,k\rangle_{\rm in}=i\int d^4x\,e^{-ikx}(\Box_x+m^2)\,{}_{\rm out}\langle f|\phi(x)|i\rangle_{\rm in}\ . \tag{6.30}
$$

Hence, we have succeeded in removing one particle from the "in" state of our S-matrix element and replacing it by a field operator $\phi(x)$. The idea is to now repeat this process until all particles, both from the "in" and the "out" state are removed and one is left with a vacuum expectation value of field operators.

Removing one particle from the out state

To further illustrate this we perform one more step explicitly, namely the reduction of a particle with momentum q from the "out" state which we write as $|f\rangle_{\rm out}=|\tilde f,q\rangle_{\rm out}$. For the S-matrix element which appears in Eq. (6.30) this leads to

$$
{}_{\rm out}\langle\tilde f,q|\phi(x)|i\rangle_{\rm in}={}_{\rm out}\langle\tilde f|a_{\rm out}(q)\phi(x)|i\rangle_{\rm in}={}_{\rm out}\langle\tilde f|\phi(x)|i-q\rangle_{\rm in}+{}_{\rm out}\langle\tilde f|\left(a_{\rm out}(q)\phi(x)-\phi(x)a_{\rm in}(q)\right)|i\rangle_{\rm in}\ . \tag{6.31}
$$

As before, we discard the forward scattering term and replace the annihilation operators with fields using Eq. (6.26). Taking into account that

$$
{}_{\rm out}\langle\tilde f|\left(\phi_{\rm out}(y)\phi(x)-\phi(x)\phi_{\rm in}(y)\right)|i\rangle_{\rm in}=\left(\lim_{y_0\to\infty}-\lim_{y_0\to-\infty}\right){}_{\rm out}\langle\tilde f|T\left(\phi(y)\phi(x)\right)|i\rangle_{\rm in}\ , \tag{6.32}
$$

we find following the same steps as before that

$$
{}_{\rm out}\langle\tilde f,q|\phi(x)|i\rangle=i\int d^4y\,e^{iqy}(\Box_y+m^2)\,{}_{\rm out}\langle\tilde f|T\left(\phi(y)\phi(x)\right)|i\rangle_{\rm in} \tag{6.33}
$$

This is precisely the same structure as in the first reduction (6.30) apart from the opposite sign in the exponent e^{iqy}. This sign difference arises because we have reduced a particle in the "out" state as opposed to a particle in the "in" state. In addition, we learn that products of field operators which arise in this way arrange themselves into time-ordered products.

The general reduction formula for real scalar fields

We can now iterate Eqs. (6.30) and Eq. (6.33) and eliminate all particles from an "in" state $|k_1,\ldots,k_n\rangle_{\rm in}$ and an "out" state $|q_1,\ldots,q_m\rangle_{\rm out}$. If we assume that $k_i\neq q_j$ for all i and j, then forward scattering terms are absent and we have

$$
\begin{aligned}
{}_{\rm out}\langle q_1,\ldots,q_m|k_1,\ldots,k_n\rangle_{\rm in} &= i^{m+n}\int\left(\prod_{a=1}^n d^4x_a\,e^{-ik_ax_a}(\Box_{x_a}+m^2)\right)\left(\prod_{b=1}^m d^4y_b\,e^{iq_by_b}(\Box_{y_b}+m^2)\right)\\
&\quad\times\langle0|T\left(\phi(y_1)\cdots\phi(y_m)\phi(x_1)\cdots\phi(x_n)\right)|0\rangle\ , &(6.34)
\end{aligned}
$$

The vacuum expectation values of time-ordered products which arise in the above formula are also referred to as *N-point Green functions*

$$
\mathcal{G}^{(N)}(z_1,\ldots,z_N)=\langle0|T(\phi(z_1)\ldots\phi(z_N))|0\rangle\ . \tag{6.35}
$$

Eq. (6.34) shows that the Green functions determine the S-matrix elements and are the crucial objects we need to calculate. It also displays a remarkable property, the so-called *crossing symmetry*: All that changes in Eq. (6.34) if a particle with momentum k is shifted from the "in" to the "out" state is the sign of the momentum k in the wave function e^{-kx}. This means, for example, that a $2\to2$ scattering and a $1\to3$ decay related by transferring one

"in" particle to an "out" particle are described by the same four-point Green function $\mathcal{G}^{(4)}$.

Reduction formula for complex scalar fields
We should now briefly discuss the reduction formula for other types of fields, starting with a complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$. In the free-field case, it is easy to see (by comparing Eqs. (5.9) and (5.37)) that $a_+^\dagger(k) = (a_1^\dagger(k) - ia_2^\dagger(k))/\sqrt{2}$ and $a_-^\dagger(k) = (a_1^\dagger(k) + ia_2^\dagger(k))/\sqrt{2}$. Applying Eq. (6.26) to ϕ_1 and ϕ_2 and forming appropriate linear combinations this implies

$$a^\dagger_{+,\mathrm{in/out}}(k) = -i\int d^3x\, e^{-ikx} \overleftrightarrow{\partial_0} \phi^\dagger_{\mathrm{in/out}}(x)\ , \quad a^\dagger_{-,\mathrm{in/out}}(k) = -i\int d^3x\, e^{-ikx} \overleftrightarrow{\partial_0} \phi_{\mathrm{in/out}}(x)\ . \tag{6.36}$$

The reduction formula for a complex scalar can now be derived exactly as above. This leads to Eq. (6.34) with one modification: Positively charged "in" particle, generated by $a^\dagger_{+,\mathrm{in}}(k)$, lead to an operator $\phi^\dagger$ in the time-ordered product, while negatively charged "in" states, generated by $a^\dagger_{-,\mathrm{in}}(k)$, lead to ϕ. For "out" states the situation is reversed, with operators ϕ for positive particles and operators $\phi^\dagger$ for negative particles.

Reduction formula for vector fields
Finally, for gauge fields, the above calculation can be repeated using the expansion (5.70) for a free gauge field and its inversion in terms of creation operators $a^{(\alpha)\dagger}(k)$. Again, the result is very similar to the previous one, except that the polarisation $\epsilon_\mu^{(\alpha)}(k)$ of the photon must be taken into account. For a photon with momentum k and polarisation $\epsilon_\mu^{(\alpha)}(k)$ in the "in" state $|i,(k,\alpha)\rangle_{\mathrm{in}}$ one finds in analogy with Eq. (6.30) that

$${}_{\mathrm{out}}\langle f|i,(k,\alpha)\rangle_{\mathrm{in}} = i\int d^4x\, \epsilon_\mu^{(\alpha)}(k) e^{-ikx} \Box_x\, {}_{\mathrm{out}}\langle f|A^\mu(x)|i\rangle_{\mathrm{in}}\ , \tag{6.37}$$

while for a photon with momentum q and polarisation $\epsilon_\mu^{(\alpha)}(k)$ in the "out" state we have in analogy with Eq. (6.33)

$${}_{\mathrm{out}}\langle f,(q,\alpha)|i\rangle_{\mathrm{in}} = i\int d^4x\, \epsilon_\mu^{(\alpha)}(q) e^{iqx} \Box_x\, {}_{\mathrm{out}}\langle f|A^\mu(x)|i\rangle_{\mathrm{in}}\ . \tag{6.38}$$

Repeated reduction leads to time-ordered operator product as for scalar fields. It should also be clear that for theories with different types of fields, for example for theories with scalar and vector fields, the various reduction formulae above can be combined and applied successively until all particles are reduced.

6.4 Perturbative evaluation of Green functions and the evolution operator

Schrödinger, Heisenberg and interaction pictures
We have managed to write S-matrix elements in terms of Green functions, that is, vacuum expectation values of time ordered products of interacting fields. We do know how to evaluate time-ordered products of free fields using Wick's theorem but this method does not apply to interacting fields so easily. What we need to do is to express interacting fields and Green functions in terms of free-fields so that Wick's theorem can be applied. We begin by splitting up the full Hamiltonian H of a field theory as

$$H = \int d^3x\, \mathcal{H} = H_0 + H_1 \tag{6.39}$$

where H_0 is the free Hamiltonian (quadratic in the fields and conjugate momenta) and $H_1 = \int d^3x\, \mathcal{H}_1$ contains the interactions. For example, for our simple toy model, a real scalar field ϕ, we have

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right) - \frac{\lambda}{4!}\phi^4\ , \quad H_1 = \int d^3x\, \mathcal{H}_1 = \frac{\lambda}{4!}\int d^3x\, \phi(x)^4\ . \tag{6.40}$$

So far, we have worked in the *Heisenberg picture* where the field operator $\phi(x)$ is time-dependent and Fock space states $|a\rangle$ are time-independent. In the *Schrödinger picture*, on the other hand, the field operator $\phi_{\mathrm{S}}(\mathbf{x})$ is time-independent while the corresponding Fock space states $|a,t\rangle_{\mathrm{S}}$ are time-dependent. The two pictures are related by the time-evolution operator e^{iHt} in the usual way

$$\phi_{\mathrm{S}}(\mathbf{x}) = e^{-iHt}\phi(t,\mathbf{x})e^{iHt}\ , \quad |a,t\rangle_{\mathrm{S}} = e^{-iHt}|a\rangle\ . \tag{6.41}$$

For the purpose of perturbation theory it is useful to introduce a third picture, the interaction picture, with fields and states denoted by $\phi_{\rm I}(x)$ and $|a,t\rangle_{\rm I}$. It is intermediate between the two previous pictures and both operators and states depend on time. In terms of the Schrödinger picture, it is defined through time evolution with the free Hamiltonion H_0, that is

$$\phi_{\rm I}(t,\mathbf{x}) = e^{iH_0t}\phi_{\rm S}(\mathbf{x})e^{-iH_0t} \ , \quad |a,t\rangle_{\rm I} = e^{iH_0t}|a,t\rangle_{\rm S} \ . \tag{6.42}$$

Definition of the evolution operator
Combining the above relations it is clear that interaction and Heisenberg picture are related by

$$\phi_{\rm I}(t,\mathbf{x}) = U(t,0)\phi(t,\mathbf{x})U^{-1}(t,0) \ , \quad |a,t\rangle_{\rm I} = U(t,0)|a\rangle \ , \tag{6.43}$$

where

$$U(t,0) = e^{iH_0t}e^{-iHt} \tag{6.44}$$

is called the *evolution operator*. It encodes the difference in time evolution between interaction and Heisenberg picture due to the interaction Hamiltonian H_1. More generally, we can define the evolution operator $U(t,t_0)$ between two times t_0 and t by $|a,t\rangle_{\rm I} = U(t,t_0)|a,t_0\rangle_{\rm I}$. This means from Eq. (6.43) that

$$U(t,t_0) = U(t,0)U^{-1}(t_0,0) \ , \tag{6.45}$$

so, in particular, $U^{-1}(t,t_0) = U(t_0,t)$ and $U(t_0,t_0) = 1$. For the composition of two evolution operators we have the rule

$$U(t,t')U(t',t_0) = U(t,0)U^{-1}(t',0)U(t',0)U^{-1}(t_0,0) = U(t,0)U^{-1}(t_0,0) = U(t,t_0) \ . \tag{6.46}$$

From Eqs. (6.45) and (6.44), the time derivative of the evolution operator is given by

$$i\frac{\partial}{\partial t}U(t,t_0) = i\left(\frac{\partial}{\partial t}U(t,0)\right)U^{-1}(t_0,0) = \left(e^{iH_0t}He^{-iH_0t} - H_0\right)U(t,t_0) \ . \tag{6.47}$$

and, hence, $U(t,t_0)$ satisfies the simple differential equation

$$i\frac{\partial}{\partial t}U(t,t_0) = H_{1,\rm I}(t)U(t,t_0) \quad \text{where} \quad H_{1,\rm I} = e^{iH_0t}H_1e^{-iH_0t} \ . \tag{6.48}$$

Note that $H_{1,\rm I}$, the interaction Hamiltonian written in the interaction picture, has the same form as H_1 but with ϕ replaced by $\phi_{\rm I}$.

Perturbative solution for evolution operator
Let us pause for a moment and see what we have achieved. The field $\phi_{\rm I}$ in the interaction picture evolves with the free Hamiltonion, as is clear from Eq. (6.42), and we should, hence, think of it as a free field. The evolution operator relates this free field to the full, interacting field ϕ in the Heisenberg picture via Eq. (6.43). Therefore, if we can find an explicit expression for the evolution operator we have succeeded in writing the interacting field in terms of a free field. To do this, we need to solve the differential equation (6.48) subject to the initial condition $U(t_0,t_0) = 1$. It turns out that the solution is given by

$$U(t,t_0) = T\,\exp\left(-i\int_{t_0}^t dt_1\, H_{1,\rm I}(t_1)\right) = 1 + \sum_{p=1}^{\infty}\frac{(-i)^p}{p!}\int_{t_0}^t d^4x_1\,\ldots\int_{t_0}^t d^4x_p\, T\left(\mathcal{H}_{1,\rm I}(x_1)\ldots\mathcal{H}_{1,\rm I}(x_p)\right) \ . \tag{6.49}$$

The exponential form of this solution is as expected for a differential equation of the form (6.48) and the initial condition is obviously satisfied. However, the appearance of the time-ordering operator is perhaps somewhat surprising. To verify the precise form of this solution, we focus on the terms up to second order in $\mathcal{H}_{1,\rm I}$. This will be sufficient to illustrate the main idea and justify the appearance of the time-ordering. While the term linear in $\mathcal{H}_{1,\rm I}$ in Eq. (6.49) contains a single integral $\int_{t_0}^t dt_1$ and is easy to differentiate with respect to t the situation is more complicated for the quadratic term which contains two integrals $\int_{t_0}^t dt_1\int_{t_0}^t dt_2$. We re-write this double-integral by splitting the integration region into two parts, for $t_1 \geq t_2$ and $t_1 < t_2$, so that $\int_{t_0}^t dt_1\int_{t_0}^t dt_2 =$

$\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 + \int_{t_0}^{t} dt_2 \int_{t_0}^{t_2} dt_1$. Then we have

$$\begin{aligned} i\frac{\partial}{\partial t}U(t,t_0) &= i\frac{\partial}{\partial t}\left(1 - i\int_{t_0}^{t} dt_1\, H_{1,\mathrm{I}}(t_1) - \frac{1}{2}\int_{t_0}^{t} dt_1 \int_{t_0}^{t} dt_2\, T(H_{1,\mathrm{I}}(t_1)H_{1,\mathrm{I}}(t_2)) + \mathcal{O}(3)\right) \\ &= H_{1,\mathrm{I}}(t) - \frac{i}{2}\frac{\partial}{\partial t}\left(\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H_{1,\mathrm{I}}(t_1)H_{1,\mathrm{I}}(t_2) + \int_{t_0}^{t} dt_2 \int_{t_0}^{t_2} dt_1\, H_{1,\mathrm{I}}(t_2)H_{1,\mathrm{I}}(t_1) + \mathcal{O}(3)\right) \\ &= H_{1,\mathrm{I}}(t) - \frac{i}{2}H_{1,\mathrm{I}}(t)\left(\int_{t_0}^{t} dt_2\, H_{1,\mathrm{I}}(t_2) + \int_{t_0}^{t} dt_1\, H_{1,\mathrm{I}}(t_1) + \mathcal{O}(2)\right) \\ &= H_{1,\mathrm{I}}(t)\left(1 - i\int_{t_0}^{t} dt_1\, H_{1,\mathrm{I}}(t_1)\right) + \mathcal{O}(3) = H_{1,\mathrm{I}}(t)U(t,t_0) + \mathcal{O}(3) \end{aligned}$$

The main point is that the order of the interaction Hamiltonians in the two integrals in the second line is reversed as a consequence of time-ordering. This allows us, after differentiating in the third line, to factor $H_{1,\mathrm{I}}(t)$ to the left for both terms in the bracket. Without time-ordering $H_{1,\mathrm{I}}(t)$ would have been to the left of the first term in the bracket and to the right of the second. Since interaction Hamiltonians at different times do not necessarily commute this would have been a problem. The general proof is a straightforward generalisation of the above calculation to all orders.

Perturbative calculation of Green functions

Armed with this solution for the evolution operator, we now return to our original problem, the perturbative calculation of Green functions. Using Eqs. (6.43) and (6.46) one can write

$$\begin{aligned} \mathcal{G}^{(N)}(z_1,\ldots,z_n) &= \langle 0|T\left(\phi(z_1)\ldots\phi(z_N)\right)|0\rangle = \langle 0|T\left(U^{-1}(t,0)\phi_{\mathrm{I}}(z_1)\ldots\phi_{\mathrm{I}}(z_N)U(t,0)\right)|0\rangle \\ &= \langle 0|T\left(U^{-1}(t,0)\phi_{\mathrm{I}}(z_1)\ldots\phi_{\mathrm{I}}(z_N)U(t,-t)U(-t,0)\right)|0\rangle \\ &= \lim_{t\to\infty}\langle 0|U^{-1}(t,0)T\left(\phi_{\mathrm{I}}(x_1)\ldots\phi_{\mathrm{I}}(x_n)U(t,-t)\right)U(-t,0)|0\rangle\,. \end{aligned} \tag{6.50}$$

The vacuum should be invariant under the action of the evolution operator up to a constant so we have

$$U(-t,0)|0\rangle = \beta_-|0\rangle\,, \quad U(t,0)|0\rangle = \beta_+|0\rangle\,, \tag{6.51}$$

for complex numbers $\beta_\pm$. They can be determined as

$$\begin{aligned} \beta_+^\star\beta_- &= \langle 0|U(-t,0)|0\rangle\langle 0|U^{-1}(t,0)|0\rangle = \sum_a \langle 0|U(-t,0)|a\rangle\langle a|U^{-1}(t,0)|0\rangle \\ &= \langle 0|U(-t,0)U^{-1}(t,0)|0\rangle = \langle 0|U(-t,0)|0\rangle = \left(\langle 0|U(t,-t)|0\rangle\right)^{-1} \end{aligned} \tag{6.52}$$

so we finally find for the Green function

$$\mathcal{G}^{(N)}(z_1,\ldots,z_N) = \lim_{t\to\infty}\frac{\langle 0|T\left(\phi_{\mathrm{I}}(z_1)\ldots\phi_{\mathrm{I}}(z_N)U(t,-t)\right)|0\rangle}{\langle 0|U(t,-t)|0\rangle}\,. \tag{6.53}$$

We can now insert the explicit solution (6.49) for the evolution operator in the Taylor expanded form into this result. This leads to a formula for the Green functions which only depends on vacuum expectation values of time-ordered free-field operator product, each of which can be evaluated using Wick's theorem. As in Section 1.1.4 each term which arises in this way can be represented by a Feynman diagram. Feynman diagrams which arise form the numerator of Eq. (6.53) have N external legs, Feynman diagrams from the denominator are vacuum bubbles. It turns out that the denominator in Eq. (6.53) precisely cancels all Feynman diagrams from the numerator which contain disconnected vacuum bubbles. In fact, we have seen an explicit example of this in Eq. (1.79). With this additional information we can write

$$\mathcal{G}^{(N)}(z_1,\ldots,z_N) = \sum_{p=0}^{\infty}\frac{(i)^p}{p!}\int d^4y_1\ldots d^4y_p\,\langle 0|T\left(\phi_{\mathrm{I}}(z_1)\ldots\phi_{\mathrm{I}}(z_N)\mathcal{L}_{\mathrm{int}}(y_1)\ldots\mathcal{L}_{\mathrm{int}}(y_p)\right)|0\rangle_{\text{no bubbles}} \tag{6.54}$$

with the interaction Lagrangian $\mathcal{L}_{\mathrm{int}} = -\mathcal{H}_{1,\mathrm{I}}$. The interaction Lagrangian is proportional to a coupling constant (λ in the case of our scalar field theory example), so the above expression for the Green functions can be seen as an expansion in this coupling constant. If its value is sufficiently small it should be a good approximation to only

compute a finite number of terms up to a certain power in the coupling. Only in this case is Eq. (6.54) of direct practical use. In the above derivation we have used notation appropriate for scalar field theory but it is clear that the basic structure of Eq. (6.54) remains valid for other types of fields. To summarise, combining the reduction formula (6.34) (and its generalisations to other types of fields) with the perturbation expansion (6.54) and Wick's theorem provides us with a practical way of calculating S-matrix elements and, via Eqs. (6.17) and (6.18) decay rates and cross sections, in terms of Feynman diagrams.

Green functions in momentum space
For practical calculations, it is more convenient to formulate Feyman rules in momentum space and to this end we introduce the Fourier transforms $\tilde{\mathcal{G}}^{(N)}$ of Green functions by

$$(2\pi)^4\delta^4(p_1+\cdots+p_N)\tilde{\mathcal{G}}^{(N)}(p_1,\ldots,p_N)=\int\left(\prod_{A=1}^{N}d^4z_A\,e^{-ip_Az_A}\right)\mathcal{G}^{(N)}(z_1,\ldots,z_N)\,, \tag{6.55}$$

where we denote "in" momenta k_a and "out" momenta q_b collectively by $(p_A)=(k_a,-q_b)$. To see where this leads we should rewrite the LSZ reduction formula (see Eq. (6.34))

$${}_{\rm out}\langle q_1,\ldots,q_m|k_1,\ldots,k_n\rangle_{\rm in}=i^N\int\prod_A d^4z_A e^{-p_Az_A}(\Box_{z_A}+m^2)\mathcal{G}^{(N)}(z_1,\ldots,z_N) \tag{6.56}$$

in momentum space, using the inversion

$$\mathcal{G}^{(N)}(z_1,\ldots,z_N)=\int\prod_B\frac{d^4Q_B}{(2\pi)^4}e^{iQ_Bz_B}(2\pi)^4\delta^4(Q_1+\cdots+Q_N)\tilde{\mathcal{G}}^{(N)}(Q_1,\ldots,Q_N) \tag{6.57}$$

of Eq. (6.55). Inserting this into the LSZ formula (6.56) we can explicitly carry out the $(\Box_{z_A}+m^2)$ operations which now only act on the exponentials in Eq. (6.57) and produce factors $-i/\tilde{\Delta}_F(Q_B)$ of inverse Feynman propagators, one for each external leg of the Green function. It is, therefore, useful to introduce *amputated Green functions*

$$\tilde{\mathcal{G}}^{(N)}_{\rm amp}(p_1,\ldots,p_N)=\frac{\tilde{\mathcal{G}}^{(N)}(p_1,\ldots,p_N)}{\prod_{A=1}^{N}\tilde{\Delta}_F(p_A)}\,, \tag{6.58}$$

which are related to the ordinary Green functions by removing the propagators for the external legs. The remaining integerations over z_A and Q_B can then trivially be carried out and we remain with

$${}_{\rm out}\langle q_1,\ldots,q_m|k_1,\ldots,k_n\rangle_{\rm in}=(2\pi)^4\delta^4\left(\sum_a k_a-\sum_b q_b\right)\tilde{\mathcal{G}}^{(N)}_{\rm amp}(k_1,\ldots,k_n,-q_1,\ldots,-q_m)\,. \tag{6.59}$$

Comparison with Eq. (6.8) shows that for real scalar fields the amputated Green function equals the matrix element $\mathcal{M}$. For other types of fields there are slight modifications. For complex scalar fields, we have to consider Green functions defined with the appropriate number of fields ϕ and their conjugates $\phi^\dagger$, depending on the number of particles and anti-particles in the "in" and "out" state. Apart from this the above formulae remain valid. For each vector particle with momentum k and polarisation $\epsilon^\mu(k)$ in the "in" or "out" state, the Green function carries a Lorentz index μ and from Eqs. (6.37) and (6.38) this should be contracted into the corresponding polarisation vector $\epsilon^\mu(k)$ to obtain the matrix element $\mathcal{M}$. With this small modification, the above formulae also apply to vector fields.

6.5 Feynman rules and examples

What remains to be done is to explicitly derive the Feynman rules for calculating the amputated Green functions by applying Wick's theorem to the perturbative expansion (6.54).

6.5.1 The real scalar field

Four-point function
As usual, we begin with the real scalar field with Lagrangian density (6.40) and interaction Lagrangian density [3]

[3] For simplicity of notation, we drop the subscript "I" from hereon and assume that fields are in the interaction picture.

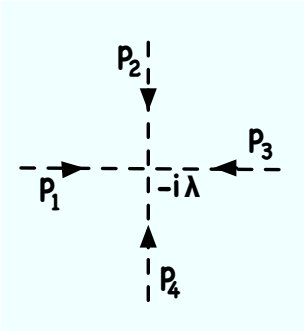

Figure 6.2: Feynman diagram for 4-point function in real scalar field theory with $\lambda\phi^4/4!$ interaction to order λ.

$\mathcal{L}_{\rm int} = -\lambda\phi^4/4!$. Let us first calculate some examples before formulating the general Feynman rules. We start with the 4-point function to order λ. Dropping disconnected parts, we have from Eq. (6.54)

$$\begin{aligned}
\mathcal{G}^{(4)}(z_1,z_2,z_3,z_4) &= -\frac{i\lambda}{4!}\int d^4y\,\langle 0|T\left(\phi(z_1)\phi(z_2)\phi(z_3)\phi(z_4)\phi(y)^4\right)|0\rangle \\
&= -i\lambda\int d^4y\,\Delta_F(z_1-y)\Delta_F(z_2-y)\Delta_F(z_3-y)\Delta_F(z_4-y) \\
&= -i\lambda\int \frac{d^4k_1}{(2\pi)^4}\cdots\frac{d^4k_4}{(2\pi)^4}\int d^4y\,e^{iy(k_1+\cdots+k_4)}e^{-i(k_1z_1+\cdots+k_4z_4)}\tilde\Delta_F(k_1)\ldots\tilde\Delta_F(k_4) \\
&= -i\lambda(2\pi)^4\int \frac{d^4k_1}{(2\pi)^4}\cdots\frac{d^4k_4}{(2\pi)^4}\delta^4(k_1+\cdots+k_4)e^{-i(k_1z_1+\cdots+k_4z_4)} \\
&\qquad\qquad \times\tilde\Delta_F(k_1)\ldots\tilde\Delta_F(k_4)\,, \qquad (6.60)
\end{aligned}$$

From the first to the second line we have used Wick's theorem which only leads to one type of term (although with multiplicity 24) which can be represented by the Feynman diagram in Fig. 6.2. Then we have used the representation (5.53) of the Feynman propagator to re-write the expression in momentum space. From this form it is easy to see that the amputated momentum space Green function is

$$\tilde{\mathcal{G}}^{(4)}_{\rm amp}(p_1,\ldots,p_4) = -i\lambda\,. \qquad (6.61)$$

Inserting this into Eq. (6.23) we find the cross section for $2\to 2$ scattering

$$\frac{d\sigma}{d\Omega} = \frac{\lambda^2}{64\pi^2E^2}\,,\quad \sigma = \frac{\lambda^2}{16\pi E^2}\,, \qquad (6.62)$$

where E is the total center of mass energy.

Two-point function
Next, we discuss the 2-point function to order λ. From Eq. (6.54) we have

$$\begin{aligned}
\mathcal{G}^{(2)}(z_1,z_2) &= \langle 0|T\left(\phi(z_1)\phi(z_2)\right)|0\rangle - \frac{i\lambda}{4!}\int d^4y\,\langle 0|T\left(\phi(z_1)\phi(z_2)\phi(y)^4\right)|0\rangle + \mathcal{O}(\lambda^2) \\
&= \Delta_F(z_1-z_2) - \frac{i\lambda}{2}\Delta_F(0)\int d^4y\,\Delta_F(z_1-y)\Delta_F(z_2-y) + \mathcal{O}(\lambda^2)
\end{aligned}$$

The two corresponding Feynman diagrams are depicted in Fig. 6.3. It is straightforward to Fourier transform this expression (introducing integration variables $z = z_1 - z_2$ and $\tilde z = (z_1+z_2)/2$ in the first term and $\tilde z_1 = z_1 - y$ and $\tilde z_2 = z_2 - y$ in the second term). Dividing by the external propagators one then finds for the amputated Green function

$$\begin{aligned}
\tilde{\mathcal{G}}^{(2)}_{\rm amp}(p,-p) &= \tilde\Delta_F(p)^{-1} - \frac{i\lambda}{2}\Delta_F(0) + \mathcal{O}(\lambda^2) \\
&= -i\left[p^2 - m^2 + \frac{\lambda}{2}\int d^4k\,\frac{i}{k^2-m^2+i\epsilon}\right] + \mathcal{O}(\lambda^2) \qquad (6.63)
\end{aligned}$$

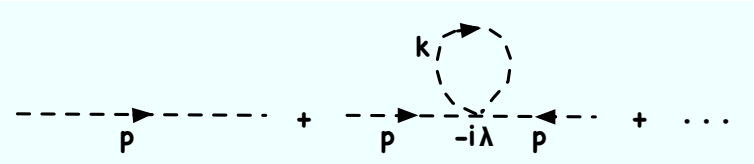

Figure 6.3: Feynman diagrams for 2-point function in real scalar field theory with $\lambda\phi^4/4!$ interaction to order λ.

The diagram at order λ is a loop diagram and, as a result, we have to carry out an integration over the internal loop momentum k in Eq. (6.63).

Feynman rules

Before we discuss this result, we would like to summarise our experience so far and formulate the Feynman rules for the amputated Green functions in real scalar field theory.

- For the (connected part of the) N point function at order λ^p draw all possible (connected) Feynman graphs with N external legs and p four-vertices. Assign a directed momentum to each line in those Feyman graphs such that momentum is conserved at each vertex. Then, to each graph associate an expression obtained by:
- writing down a Feynman propagator $\tilde{\Delta}_F(k)$ for each internal line with momentum k
- writing down $-i\lambda$ for each vertex
- writing down an integral $\int d^4k\,/(2\pi)^4$ for each momentum k which is not fixed by momentum conservation at the vertices
- dividing the expression by a symmetry factor which equals the number of permutations of internal lines one can make for fixed vertices.

The first two of these rules are obvious from Eq. (6.54) and Wick's theorem. The fact that momentum needs to be conserved at each vertex can be seen from Eq. (6.60): The calculation which led to this result would be exactly the same for a single vertex within a larger Feynman diagram. The appearance of the delta function in Eq. (6.60) then signals momentum conservation at this vertex. Our result (6.61) for the amputated 4-point function shows that $-i\lambda$ is the correct factor to include for a vertex. Finally, the precise symmetry factor for each graph follows from the number of pairings which arise in Wick's theorem.

Discussion of two-point function, regularisation and renormalization

Let us now return to a discussion of the 2-point function. To carry out the integral in Eq. (6.63) we recall the position of the poles in the Feynman propagator as indicated in Fig. 5.1. We deform the contour of integration by a 90 degrees counter-clockwise rotation from the real to the imaginary k_0 axis so that we are not crossing any poles. Then, the integration goes over the imaginary k_0 axis and to convert this back to an integration over a real axis from $-\infty$ to $+\infty$ we introduce the "Euklidean momentum" $\bar{k} = (\bar{k}_0, \bar{\mathbf{k}}) = (-ik_0, \mathbf{k})$ and denote its Euklidean length by κ. The integral in Eq. (6.63) can then be re-written as

$$\int d^4k\,\frac{i}{k^2 - m^2 + i\epsilon} = \int d\Omega_3 \int_0^\infty d\kappa\,\frac{\kappa^3}{\kappa^2 + m^2} = 2\pi^2 \int_0^\infty d\kappa\,\frac{\kappa^3}{\kappa^2 + m^2} \tag{6.64}$$

Clearly, the last integral is quadratically divergent as $\kappa \to \infty$. Recall that in the context of free field theory we have found infinities due to summing up an infinite number of zero point energies. This problem was resolved by the process of "normal ordering". Here we encounter another and more serious type of singularity in field theories. They arise from loops in Feynman diagrams at high momentum and are, therefore, also referred to as *ultraviolet singularities*. Their appearance is of course unphysical as, for example, each of the external legs of the 4-point function can acquire an additional loop at order λ^2. This would make the 4-point function and, hence, the associated cross section for $2 \to 2$ scattering infinite. Dealing with and removing such singularities is difficult and discussing the full procedure is beyond the scope of this lecture. Here, we would just like to outline the main idea. It turns out that in certain classes of theories, so called *renormalizable theories*, only a finite number of different types of such singularities arise. For renormalizable theories, these singularities can then be absorbed into (infinite) redefinitions of the parameters (masses and couplings), of the theory. Once physical quantities are expressed in terms of these redefined parameters they turn out to be finite. A useful rule of thumb is that a theory is renormalizable when it does not contain any parameters with negative mass dimension. From this criterion, the real

scalar field theory with $\lambda\phi^4$ interaction should be renormalizable and this is indeed the case. Let us briefly discuss how one might go about dealing with the above singularity. The first step is to *regularise* the amplitude. This refers to some sort of prescription which allows one to assign a finite value to the diverging integral in question. There are many different ways of doing this but perhaps the simplest and most intuitive one is to introduce a *cut-off*, that is, to modify the upper limit of the integral (6.64) to a finite value Λ. If we do this we find

$$\tilde{G}^{(2)}_{\rm amp}(p,-p) = -i\left[p^2 - m^2 + \frac{\pi^2\lambda}{2}\left(\Lambda^2 - m^2\ln(1+\Lambda^2/m^2)\right)\right] . \tag{6.65}$$

We can now define a "renormalized" mass m_R by

$$m_R^2 = m^2 - \frac{\pi^2\lambda}{2}\left(\Lambda^2 - m^2\ln(1+\Lambda^2/m^2)\right) . \tag{6.66}$$

The main idea is that expressed in terms of this mass (and the other remormalized parameters of the theory) physical quantities are finite provided the theory is renormalizable.

Eq. (6.66) reveals another interesting feature which is characteristic for renormalizable scalar field theories: the appearance of divergencies *quadratic* in the cut-off Λ. It turns out that in renormalizable theories without scalar fields divergencies are at most logarithmic. Quadratic divergencies in scalar field theory lead to a serious "naturalness" problem called the *hierarchy problem*. Physically, we can think about the cut-off Λ as the energy scale above which the theory in question ceases to be valid and new physics becomes relevant. Assume that we have a theory with a very high cut-off scale Λ but a small physical scalar mass m_R. Then Eq. (6.66) requires a very precise cancelation between the two terms on the right-hand side. For example, in the standard model of particle physics, a theory shown to be renormalizable, the mass of the Higgs particle should be of the order of the electroweak symmetry breaking scale, so a few hundred GeV. If the standard model was valid all the way up to the Planck scale, $\sim 10^{18}$ GeV, this would require a 32 digit cancellation in Eq. (6.66). The hierarchy problem may, therefore, be seen as a reason for why we should expect new physics not far above the electroweak scale. Supersymmetry is one of the main candidates for such physics and it is indeed capable of resolving the hierarchy problem.

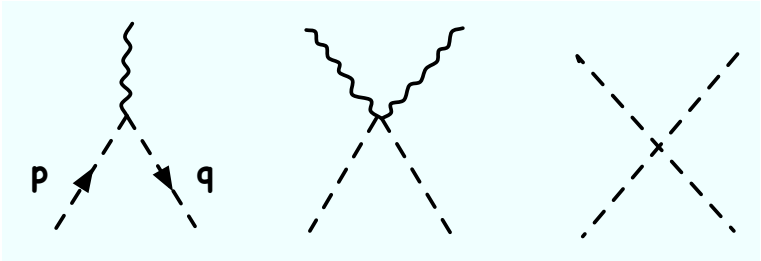

Figure 6.4: Vertices for scalar electrodynamics. From left to right the corresponding vertex Feynman rules are $-ie(p_\mu + q_\mu)$, $2iq^2\eta_{\mu\nu}$ and $-i\lambda$.

6.5.2 Scalar electrodynamics

From Eq. (4.127) we recall that the Lagrangian density of scalar electrodynamics (plus gauge fixing term in Feynman gauge) for a complex scalar ϕ with charge e under a vector field A_μ is given by

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}(\partial_\mu A^\mu)^2 + [(\partial_\mu + ieA_\mu)\phi]^\dagger\,[(\partial^\mu + ieA^\mu)\phi] - m^2\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2 , \tag{6.67}$$

so we have for the interaction Lagrangian

$$\mathcal{L}_{\rm int} = ieA^\mu\left(\phi\partial_\mu\phi^\dagger - \phi^\dagger\partial_\mu\phi\right) + e^2A_\mu A^\mu\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2 . \tag{6.68}$$

So, in addition to a quartic scalar field vertex similar to what we have encountered in real scalar field theory, we expect a triple vertex coupling a vector field to two scalars and another quartic vertex coupling two vectors to two

scalars. To find the Feynman rules for these vertices let us compute the appropriate Green functions. Using the first, cubic interaction in Eq. (6.68) we find for the 3-point function

$$\begin{aligned}\mathcal{G}^{(3)}_{\mu}(z_1,z_2,z_3) &= -e\int d^4y\,\langle 0|T\left(\phi(z_1)\phi^\dagger(z_2)A_\mu(z_3)A^\nu(y)(\phi(y)\partial_\nu\phi^\dagger(y)-\phi^\dagger(y)\partial_\nu\phi(y))\right)|0\rangle \\ &= -ie(2\pi)^4\int\frac{d^4p}{(2\pi)^4}\frac{d^4q}{(2\pi)^4}\frac{d^4k}{(2\pi)^4}\delta^4(p-q+k)e^{-i(pz_1-qz_2+kz_3)} \\ &\qquad \times\tilde{\Delta}_F(p)\tilde{\Delta}_F(q)\tilde{\Delta}_{F,0}(k)(p_\mu+q_\mu)\,, \end{aligned} \tag{6.69}$$

where $\tilde{\Delta}_{F,0}(k) = i/(k^2+i\epsilon)$ is the Feynman propagator for zero mass. For the amputated Green function this means

$$\tilde{\mathcal{G}}^{(3)}_{\mathrm{amp},\mu}(p,-q,k) = -ie(p_\mu+q_\mu) \tag{6.70}$$

and this is precisely the expression for the triple vertex. From a very similar calculation using two external vector fields, two external scalars and the second interaction term in (6.68) one finds

$$\tilde{\mathcal{G}}^{(4)}_{\mathrm{amp},\mu\nu}(p_1,p_2,p_3,p_4) = 2ie^2\eta_{\mu\nu}\,. \tag{6.71}$$

Finally, the quartic scalar field interaction from the last term in (6.68) comes with a vertex factor $-i\lambda$, just as in the case of a real scalar field. These Feynman rules for the interactions in scalar electrodynamics are summarised in Fig. 6.4. Finally, the propagators for internal scalar and vector field lines are

$$\tilde{\Delta}_F(k) = \frac{i}{k^2-m^2+i\epsilon}\,,\quad \tilde{D}_F(k)_{\mu\nu} = \frac{-i}{k^2+i\epsilon}\eta_{\mu\nu}\,. \tag{6.72}$$

Space-time indices on internal vector field propagators and vertices have to be contracted in the obvious way. This completes the Feynman rules for the amputated Green functions in scalar electrodynamics. To obtain the matrix element the μ index for each external photon with momentum k and polarisation α has to be contracted into $\epsilon^{(\alpha)}_\mu(k)$.

6.6 Further reading

Perturbation theory in the canonical formalism is covered in most standard text books on quantum field theory, including

- J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, vol 2, chapters 16,17.
- C. Itzykson and J.-B. Zuber, *Quantum Fields*, chapters 5.1, 6.1.
- M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, chapter 4.

Chapter 7

Path Integrals in Quantum Field Theory

In the previous chapters we have followed the "traditional", canonical approach to quantise field theories. We would now like to show how field theories can be quantised, arguably in a more elegant way, by using path integrals. We begin by recalling some features of path integrals in quantum mechanics.

7.1 Quantum mechanics flashback

We have shown in Section 1.2 how quantum mechanics can be formulated in terms of path integrals. In particular, we have found the expression (1.86)

$$\langle x,t|x',t'\rangle = \langle x|e^{-iH(t-t')}|x'\rangle \sim \int \mathcal{D}x \exp\left\{i\int_{t'}^{t} d\tau\, \mathcal{L}(x,\dot{x})\right\} \tag{7.1}$$

for the matrix element between two position eigenstates. Here, $|x,t\rangle = e^{iHt}|x\rangle_S$ are Heisenberg picture states which coincide with Schrödinger picture states $|x(t)\rangle_{\rm S}$ at a given, fixed time t. The central objects to compute in quantum field theory are vacuum expectation values of time-ordered field operator product. The analogous quantities in quantum mechanics, for simplicity written for just two operators, are [1] $\langle x_f,t_f|T(\hat{x}(t_1)\hat{x}(t_2))|x_i,t_i\rangle$. Focusing on the case $t_1 > t_2$ and inserting complete sets of states we have

$$\begin{aligned}\langle x_f,t_f|T(\hat{x}(t_1)\hat{x}(t_2))|x_i,t_i\rangle &= \langle x_f|e^{-iH(t_i-t_1)}\hat{x}_S e^{-iH(t_1-t_2)}\hat{x}_S e^{-iH(t_2-t_f)}|x_i\rangle \\ &= \int dx_1 dx_2\, \langle x_f|e^{-iH(t_i-t_1)}|x_1\rangle\langle x_1|\hat{x}_S e^{-iH(t_1-t_2)}|x_2\rangle\langle x_2|\hat{x}_S e^{-iH(t_2-t_f)}|x_i\rangle\end{aligned}$$

Using $\hat{x}|x\rangle = x|x\rangle$, replacing every expectation value by Eq. (7.1) and combining the three path integrals, together with the integrations over x_1 and x_2 into a single path integral this can be written as

$$\langle x_f,t_f|T(\hat{x}(t_1)\hat{x}(t_2))|x_i,t_i\rangle \sim \int \mathcal{D}x\, x(t_1)x(t_2) \exp\left\{i\int_{t_i}^{t_f} d\tau\, \mathcal{L}(x,\dot{x})\right\}\ . \tag{7.2}$$

For $t_2 > t_1$ the result is actually the same: time-ordering is automatic in path integrals. It is also clear that the above argument can be repeated for a product of an arbitrary number of operators $\hat{x}(t)$. Further, it can be shown that $\lim_{t_i\to-\infty,t_f\to\infty}\langle x_f,t_f|T(\hat{x}(t_1)\ldots\hat{x}(t_N))|x_i,t_i\rangle \sim \langle 0|T(\hat{x}(t_1)\ldots\hat{x}(t_N))|0\rangle$, so we have the final result

$$\langle 0|T(\hat{x}(t_1)\ldots\hat{x}(t_N))|0\rangle \sim \int \mathcal{D}x\, x(t_1)\ldots x(t_N) e^{iS[x]}\ . \tag{7.3}$$

7.2 Basics of field theory path integrals

As in previous sections, we focus on the simple toy example of a real scalar field ϕ with Lagrangian density $\mathcal{L} = \mathcal{L}(\partial_\mu\phi,\phi)$ and action $S = \int d^4x\,\mathcal{L}$ although the formalism applies more generally. As discussed in the previous chapter, the central objects in quantum field theory which carry all the relevant physical information are the Green functions

$$\mathcal{G}^{(N)}(x_1,\ldots,x_N) = \langle 0|T(\hat{\phi}(x_1)\ldots\hat{\phi}(x_N))|0\rangle\ . \tag{7.4}$$

[1] In this chapter, we will use hats to distinghuish operators from their classical counterparts which appear in the path integral.

In analogy with Eq. (7.3) we can now express these Green functions in terms of a path integral as

$$\mathcal{G}^{(N)}(x_1,\ldots,x_N) = \mathcal{N}\int \mathcal{D}\phi\,\phi(x_1)\ldots\phi(x_N)e^{iS[\phi]}\,, \tag{7.5}$$

where $\mathcal{N}$ is a normalization to be fixed shortly. Analogous to what we did in Chapter 1, it is useful to introduce a generating functional

$$W[J] = \mathcal{N}\int \mathcal{D}\phi\,\exp\left\{i\int d^4x\,\left[\mathcal{L}(\partial_\mu\phi,\phi)+J(x)\phi(x)\right]\right\} \tag{7.6}$$

for the Green functions such that

$$i^N\mathcal{G}^{(N)}(x_1,\ldots,x_N) = \left.\frac{\delta^N W[J]}{\delta J(x_1)\ldots\delta J(x_N)}\right|_{J=0}\,. \tag{7.7}$$

To fix the normalization $\mathcal{N}$ we require that $W[J]|_{J=0} = \langle 0|0\rangle = 1$. For the purpose of explicit calculations it is useful to introduce a Eulkidean or *Wick rotated* version of the generating functional. To do this we define Euklidean four-vectors by $\bar{x} = (\bar{x}_0,\bar{\mathbf{x}}) = (ix_0,\mathbf{x})$, associated derivatives $\bar{\partial}_\mu = \frac{\partial}{\partial\bar{x}_\mu}$ and a Euklidean version of the Lagrangian density $\mathcal{L}_E = \mathcal{L}_E(\bar{\partial}_\mu\phi,\phi)$. We can then re-write the generating functional $W[J]$ and obtain its Euklidean counterpart

$$W_E[J] = \mathcal{N}\int \mathcal{D}\phi\exp\left\{\int d^4\bar{x}\,\left[\mathcal{L}_E(\bar{\partial}_\mu\phi,\phi)+J(\bar{x})\phi(\bar{x})\right]\right\} \tag{7.8}$$

and associated Euklidean Green functions

$$\mathcal{G}_E^{(N)}(\bar{x}_1,\ldots,\bar{x}_N) = \left.\frac{\delta^N W_E[J]}{\delta J(\bar{x}_1)\ldots\delta J(\bar{x}_N)}\right|_{J=0}\,. \tag{7.9}$$

In analogy with Chapter 1 (see Eq. (1.81) we also define the generating functional $Z[J]$ by

$$W[J] = e^{iZ[J]}\,. \tag{7.10}$$

and its associated Green functions

$$i^N G^{(N)}(x_1,\ldots,x_N) = i\left.\frac{\delta^N Z[J]}{\delta J(x_1)\ldots\delta J(x_N)}\right|_{J=0} \tag{7.11}$$

which correspond to connected Feynman diagrams. We will refer to $G^{(n)}$ as *connected Green functions* and to $Z[J]$ as the generating functional for connected Green functions.

The full information about the quantum field theory is now encoded in the generating functional, so this is the primary object to compute. We begin by doing this in the simplest case, the free theory.

7.3 Generating functional and Green functions for free fields

The Lagrangian density of a free, real scalar field is given by $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2)$ and the Euklidean version reads $\mathcal{L}_E = -\frac{1}{2}(\bar{\partial}_\mu\phi\bar{\partial}_\mu\phi + m^2\phi^2)$. We write the kinetic term as

$$\int d^4\bar{x}\,\bar{\partial}_\mu\phi(\bar{x})\bar{\partial}_\mu\phi(\bar{x}) = \int d^4\bar{x}\,d^4\bar{y}\,\phi(\bar{y})\left(\bar{\partial}^y_\mu\bar{\partial}^x_\mu\delta^4(\bar{x}-\bar{y})\phi(\bar{x})\right)\,, \tag{7.12}$$

so that the Euklidean generating functional takes the form

$$W_E[J] = \mathcal{N}\int\mathcal{D}\phi\exp\left\{-\frac{1}{2}\int d^4\bar{x}\,d^4\bar{y}\,\phi(\bar{y})A(\bar{y},\bar{x})\phi(\bar{x}) + \int d^4\bar{x}\,J(\bar{x})\phi(\bar{x})\right\}\,. \tag{7.13}$$

with the operator

$$A(\bar{y},\bar{x}) = \left(\bar{\partial}^y_\mu\bar{\partial}^x_\mu + m^2\right)\delta^4(\bar{x}-\bar{y})\,. \tag{7.14}$$

This is a Gaussian path integral with a source J of precisely the type we have discussed in Section 1. From Eq. (1.42) we find

$$W_E[J] = \tilde{\mathcal{N}}\exp\left\{\frac{1}{2}\int d^4\bar{x}\,d^4\bar{y}\,J(\bar{y})\Delta_E(\bar{y}-\bar{x})J(\bar{x})\right) \tag{7.15}$$

with the inverse $\Delta_E(\bar{y}-\bar{x}) = A^{-1}(\bar{y},\bar{x})$ and a suitable normalization $\tilde{\mathcal{N}}$. How do we compute the inverse of the operator A? With the representation $\delta^4(\bar{x}-\bar{y}) = \int \frac{d^4\bar{k}}{(2\pi)^4} e^{i\bar{k}(\bar{x}-\bar{y})}$ of the delta function, we write A as a Fourier transform

$$A(\bar{y},\bar{x}) = \int \frac{d^4\bar{k}}{(2\pi)^4}(\bar{k}^2+m^2)e^{i\bar{k}(\bar{x}-\bar{y})} \; . \tag{7.16}$$

Then, we invert A by taking the inverse inside the Fourier transform, that is

$$\Delta_E(\bar{y}-\bar{x}) = \int \frac{d^4\bar{k}}{(2\pi)^4} e^{i\bar{k}(\bar{x}-\bar{y})} \frac{1}{\bar{k}^2+m^2} \; . \tag{7.17}$$

Now, we would like to revert to Minkowski space with coordinates x and y. To obtain a Minkowski product in the exponent in Eq. (7.17) we also need to introduce the Minkowski momentum $k = (k_0, \mathbf{k}) = (i\bar{k}_0, \bar{\mathbf{k}})$. By inserting all this into Eq. (7.15) we find for the generating functional in Minkowski space

$$W[J] = \exp\left\{-\frac{1}{2}\int d^4x\, d^4y\, J(x)\Delta_F(x-y)J(y)\right\} \; , \tag{7.18}$$

where Δ_F is the Feynman propagator precisely as introduces earlier (see Eq. (5.53). We have also chosen $\tilde{\mathcal{N}} = 1$ so that $W[J]|_{J=0} = 1$. Eq. (7.18) is the general result for the generating functional of the free scalar field theory. All the free Green functions can now be calculated from Eq. (7.7) and we already know from Chapter 1 that the result can be obtained by applying Wick's theorem. In particular, we have for the 2-point function

$$\mathcal{G}^{(2)}(x,y) = -\left.\frac{\delta^2 W[J]}{\delta J(x)\delta J(y)}\right|_{J=0} = \Delta_F(x-y) \; . \tag{7.19}$$

Wick's theorem appears both in the context of canonical quantisation (see Eq. (5.61)) and path integral quantisation. Applying it in either case shows that both types of Green functions are indeed identical, at least for the case of free fields. This confirms that canonical and path integral quanitsation are equivalent.

For later reference, we note that from Eqs. (7.18) and (7.10) the generating functional Z for free fields is given by

$$Z[J] = \frac{i}{2}\int d^4x\, d^4y\, J(x)\Delta_F(x-y)J(y) \; . \tag{7.20}$$

7.4 The effective action

In Chapter 1 we have seen that the path integral formalism provides an intuitive picture for the transition between quantum and classical physics. We will now study this transition in more detail for the case of field theories. To this end we define the classical field ϕ_c by

$$\phi_c(x) = \frac{\delta Z[J]}{\delta J(x)} \; . \tag{7.21}$$

Then, from the definition (7.10) of the generating functional Z we have

$$\phi_c(x) = -\frac{i}{W[J]}\frac{\delta W[J]}{\delta J(x)} = \frac{\langle 0|\hat{\phi}(x)|0\rangle_J}{\langle 0|0\rangle_J} \; , \tag{7.22}$$

where we have suggestively defined the vacuum expectation values $\langle 0|0\rangle_J = W[J]$ and $\langle 0|\hat{\phi}(x)|0\rangle_J = -i\delta W[J]/\delta J(x)$ in the presence of the source J. Eq. (7.22) shows that ϕ_c is the suitably normalised vacuum expectation value of the field operator $\hat{\phi}$, so its interpretation as a classical field is sensible. Note that ϕ_c is a function of the source J. From Eq. (7.10), the generating functional Z is "on the same footing" as the exponential in the path integral and can, hence, be seen as some sort of effective action. However, it still contains the effect of the source term $J(x)\phi(x)$ in Eq. (7.6). To remove this source term we define the effective action Γ by a Legendre transform

$$\Gamma[\phi_c] = Z[J] - \int d^4x\, J(x)\phi_c(x) \; . \tag{7.23}$$

Differentiating the left-hand side of this equation by $\delta/\delta J(y)$ and using the definition (7.21) of the classical field it follows that Γ is independent of the source J, as the notation suggests.

To see that these definitions conform with our intuition, let us first discuss free fields. In Eq. (7.20) we have calculated the generating functional Z for free fields and inserting this into Eq. (7.21) we find for the classical field

$$\phi_c(x) = i \int d^4y \, \Delta_F(x-y) J(y) \; . \tag{7.24}$$

Since the Feynman propagator satisfies the equation $(\Box + m^2)\Delta_F(x) = -i\delta^4(x)$ (see Eq. (5.55)) we find for the above classical field that

$$(\Box + m^2)\phi_c(x) = J(x) \; , \tag{7.25}$$

so it is a solution to the Klein-Gordon equation with source $J(x)$, as one would expect for a classical field coupled to a source J. Inserting (7.20) and (7.24) into the effective action (7.23) it follows

$$\Gamma[\phi_c] = -\frac{1}{2}\int d^4x \, \phi_c(x) J(x) = -\frac{1}{2}\int d^4x \, \phi_c(x)(\Box + m^2)\phi_c(x) = \frac{1}{2}\int d^4x \, \left[\partial_\mu \phi_c \partial^\mu \phi_c - m^2 \phi_c^2\right] \tag{7.26}$$

and, hence, the effective action coincides with the classical action as one would expect for a free theory.

For interacting theories, the generating functional can of course not be calculated exactly. However, we can proceed to evaluate the path integral

$$W[J] = \mathcal{N} \int \mathcal{D}\phi \exp\left\{ iS[\phi] + i\int d^4x \, J(x)\phi(x) \right\} \tag{7.27}$$

in the saddle point approximation as discussed in Chapter 1. The solution ϕ_0 to the classical equations of motion is determined from the classical equations of motion

$$\frac{\delta S}{\delta \phi(x)}[\phi_0] = -J(x) \; . \tag{7.28}$$

Then, to leading order in the saddle point approximation we find

$$W[J] \sim \exp\left\{ iS[\phi_0] + i\int d^4x J(x)\phi_0(x) \right\} \; , \quad Z[J] = S[\phi_0] + \int d^4x \, J(x)\phi_0(x) \; . \tag{7.29}$$

With this result for Z and Eqs. (7.21), (7.23) and (7.28) we immediately conclude that $\phi_c = \phi_0$ and $\Gamma[\phi_c] = S[\phi_c]$. Hence, in the lowest order saddle point approximation the effective action Γ is simply the classical action. Beyond this leading order, the effective action of course receives corrections due to quantum effects and differs from the classical action. This leads to a systematic approach to calculate these quantum corrections to the effective action the details of which are beyond the scope of the lecture.

The above formalism also sheds light on another point which we have glossed over so far. Our discussion of spontaneous symmetry breaking in Chapter 4 has been carried in the context of classical fields and it has not been obvious what its status should be in the quantum theory. Spontaneous symmetry breaking of a quantum theory should be analysed using the above effective action (or, more precisely, the effective potential, which is the scalar potential of the effective action). Hence, we see that the results of Chapter 4 make sense in quantum theory, but have to be viewed as a leading order approximation.

7.5 Feynman diagrams from path integrals

To develop perturbation theory in the path integral formalism we split the Lagrangian density as $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\rm int}$ into the free Lagrangian $\mathcal{L}_0$ and the interaction piece $\mathcal{L}_{\rm int}$. The free generating functional associated with $\mathcal{L}_0$ is denoted by $W_0[J]$ and the full generating functional by $W[J]$. We have

$$\begin{aligned} W[J] &= \mathcal{N} \int \mathcal{D}\phi \, \exp\left\{ i\int d^4x \, (\mathcal{L}_0 + \mathcal{L}_{\rm int} + J(x)\phi(x) \right\} \\ &= \mathcal{N} \exp\left\{ i\int d^4x \, \mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(x)}\right)\right\} W_0[J] \; . \\ &= \mathcal{N}\left[1 + \sum_{p=1}^{\infty} \frac{i^p}{p!} \int d^4y_1 \ldots d^4y_p \, \mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(y_1)}\right) \ldots \mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(y_p)}\right)\right] W_0[J] \; , \end{aligned} \tag{7.30}$$

where

$$\mathcal{N}^{-1} = \exp\left\{ i \int d^4x \, \mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(x)}\right)\right\} W_0[J]\Big|_{J=0} \tag{7.31}$$

to ensure that $W[J]|_{J=0} = 1$. This is a perturbative series for the full generating functional in terms of the free one, $W_0[J]$. We recall from Eq. (7.18) that the free generating functional is given by

$$W_0[J] = \exp\left\{ -\frac{1}{2}\int d^4x \, d^4y \, J(x)\Delta_F(x-y)J(y)\right\} \ , \tag{7.32}$$

so all the functional differentiations in Eq. (7.30) can be carried out explicitly and lead to Feynman propagators. The Green functions (7.7) can then written as

$$\begin{aligned}\mathcal{G}^{(N)}(z_1,\ldots,z_N) &= \mathcal{N}\frac{\delta}{\delta J(z_1)}\cdots\frac{\delta}{\delta J(z_1)} \\ &\quad\times\left[1+\sum_{p=1}^{\infty}\frac{i^p}{p!}\int d^4y_1\ldots d^4y_p\,\mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(y_1)}\right)\ldots\mathcal{L}_{\rm int}\left(-i\frac{\delta}{\delta J(y_p)}\right)\right] W_0[J]\Big|_{J=0}\end{aligned}$$

We know from Chapter 1 (see Eq. (1.76) that this expression can be worked out using Wick's theorem. The result is a sum over products of Feynman propagators, suitably integrated, and each term can be associated to a Feynman diagram. This is precisely the same structure as for the perturbative Green function (6.54) obtained from canonical quantization (after diagrams with disconnected vacuum bubbles are cancelled due to the normalization factor $\mathcal{N}$). We have, therefore, explicitly verified that canonical and path integral approach lead to the same perturbative Green functions. From hereon, working out the Green functions explicitly and calculating decay rates and cross sections works exactly like in the canonical formalism: We calculate the space-time Green functions from the above formula and then derive the Fourier transformed and amputated Green functions from Eqs. (6.55) and (6.58), respectively. Is is clear that this also leads to the same set of Feynman rules, so there is no need to repeat their derivation. Compared to the hard work in the canonical approach it is remarkable how relatively easily the path integral formalism delivers the same results.

7.6 Further reading

Path integrals in quantum field theory and path integral quantization of the scalar field theory are covered in most standard text books on quantum field theory, including

- P. Ramond, *Field Theory: A Modern Primer*
- D. Bailin and A. Love, *Introduction to Gauge Field Theory*
- S. Weinberg, *The Quantum Theory of Fields*, vol. 1

Chapter 8

Many-Particle Quantum Systems

There are two routes which take us to problems in many-particle quantum physics. One is to start from classical field theory and quantise, as in Chapter 5 of these notes. The other is to start from one-body or few-body quantum mechanics, and consider the special aspects which become important when many particles are involved. Clearly, the first route is the natural one to take if, for example, we want to begin with Maxwell's equations and arrive at a description of photons. Equally, the second route is the appropriate one if, for instance, we want to begin with a model for liquid ^{4}He and arrive at an understanding of superfluidity. In this chapter we will set out the second approach and illustrate it using applications from condensed matter physics. Although the problems we cover can all be formulated using functional integrals, we will use Hamiltonians, operators and operator transformations instead. This choice is made partly for simplicity, and partly in order to introduce a useful set of techniques.

8.1 Identical particles in quantum mechanics

Many-particle quantum systems are always made up of many *identical particles*, possibly of several different kinds. Symmetry under exchange of identical particles has very important consequences in quantum mechanics, and the formalism of many-particle quantum mechanics is designed to build these consequences properly into the theory. We start by reviewing these ideas.

Consider a system of N identical particles with coordinates $\mathbf{r}_1, \ldots \mathbf{r}_N$ described by a wavefunction $\psi(\mathbf{r}_1 \ldots \mathbf{r}_N)$. For illustration, suppose that the Hamiltonian has the form

$$\mathcal{H} = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}V(\mathbf{r}_i) + \sum_{i<j}U(\mathbf{r}_i - \mathbf{r}_j)\ .$$

Here there are three contributions to the energy: the kinetic energy of each particle (∇_i^2 operates on the coordinates $\mathbf{r}_i$); the one-body potential energy $V(\mathbf{r})$; and the two-particle interaction potential $U(\mathbf{r}_i - \mathbf{r}_j)$. To discuss symmetry under exchange of particles, we define the exchange operator $\mathcal{P}_{ij}$ via its action on wavefunctions:

$$\mathcal{P}_{ij}\psi(\ldots\mathbf{r}_i\ldots\mathbf{r}_j\ldots) = \psi(\ldots\mathbf{r}_j\ldots\mathbf{r}_i\ldots)\ .$$

Since $[\mathcal{H}, \mathcal{P}_{ij}] = 0$, we can find states that are simultaneous eigenstates of $\mathcal{H}$ and $\mathcal{P}_{ij}$. Moreover, a system that is initially in an eigenstate of $\mathcal{P}_{ij}$ will remain in one under time evolution with $\mathcal{H}$. For these reasons we examine the eigenvalues of $\mathcal{P}_{ij}$. Since $(\mathcal{P}_{ij})^2 = 1$, these are $+1$ and -1. Now, it is an observational fact (explained in relativistic quantum field theory by the spin-statistics theorem) that particles come in two kinds and that particles of a given kind are always associated with the same eigenvalue of the exchange operator: $+1$ for bosons and -1 for fermions.

8.1.1 Many particle basis states

In a discussion of many-particle quantum systems we should restrict ourselves to wavefunctions with the appropriate symmetry under particle exchange. We can do this by using a set of basis states that has the required symmetry. As a starting point, suppose that we have a complete, orthonormal set of single-particle states $\phi_1(\mathbf{r}), \phi_2(\mathbf{r}) \ldots$.

Next we would like to write down a wavefunction representing an N-particle system with one particle in state l_1, one in state l_2 and so on. The choice

$$\phi_{l_1}(\mathbf{r})\phi_{l_2}(\mathbf{r})\dots\phi_{l_N}(\mathbf{r})$$

is clearly unsatisfactory because for general $l_1, l_2 \dots$ it has no particular exchange symmetry. Instead we take

$$\psi(\mathbf{r}_1 \dots \mathbf{r}_N) = \mathcal{N} \sum_{\text{distinct perms.}} (\pm 1)^P \phi_{k_1}(\mathbf{r}_1)\dots\phi_{k_N}(\mathbf{r}_N) \; . \tag{8.1}$$

Several aspects of the notation in Eq. (8.1) require comment. The sign inside the brackets in $(\pm 1)^P$ is $+1$ for bosons and -1 for fermions. The set of labels $\{k_1 \dots k_N\}$ is a permutation of the set $\{l_1 \dots l_N\}$. The permutation is called *even* if it can be produced by an even number of exchanges of adjacent pairs of labels, and is *odd* otherwise; the integer P is even or odd accordingly. The sum is over all *distinct* permutations of the labels. This means that if two or more of the labels l_n are the same, then permutations amongst equal labels do not appear as multiple contributions to the sum. Finally, $\mathcal{N}$ is a normalisation, which we determine next.

To normalise the wavefunction, we must evaluate

$$\int \mathrm{d}^d\mathbf{r}_1 \dots \int \mathrm{d}^d\mathbf{r}_N \, \psi^*(\mathbf{r}_1 \dots \mathbf{r}_N)\psi \;\; (\mathbf{r}_1 \dots \mathbf{r}_N) \; .$$

Substituting from Eq. (8.1), we obtain a double sum (over permutations $k_1 \dots k_N$ and $h_1 \dots h_N$) of terms of the form

$$\int \mathrm{d}^d\mathbf{r}_1 \, \phi^*_{k_1}(\mathbf{r})\phi_{h_1}(\mathbf{r}_1) \dots \int \mathrm{d}^d\mathbf{r}_N \, \phi^*_{k_N}(\mathbf{r})\phi_{h_N}(\mathbf{r}_1) \; .$$

These terms are zero unless $k_1 = h_1$, $k_2 = h_2$, and $\dots k_N = h_N$, in which case they are unity. Therefore only the diagonal terms in the double sum contribute, and we have

$$\int \dots \int |\psi|^2 = |\mathcal{N}|^2 \sum_{\text{dist. perms.}} (\pm 1)^{2P} = |\mathcal{N}|^2 \, \frac{N!}{n_1!\, n_2! \, \dots}$$

where the n_1, $n_2 \dots$ are the numbers of times that each distinct orbital appears in the set $\{l_1 \dots l_N\}$, and the ratio of factorials is simply the number of distinct permutations. Hence we normalise the wavefunction to unity by taking

$$\mathcal{N} = \left(\frac{n_1! \; n_2! \; \dots}{N!} \right)^{1/2} \; .$$

8.1.2 Slater determinants

For fermion wavefunctions we can get the correct signs by thinking of Eq. (8.1) as a determinant

$$\psi(\mathbf{r}_1 \dots \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{l_1}(\mathbf{r_1}) & \dots & \phi_{l_1}(\mathbf{r_N}) \\ & \dots & \\ \phi_{l_N}(\mathbf{r_1}) & \dots & \phi_{l_N}(\mathbf{r_N}) \end{vmatrix} \; . \tag{8.2}$$

Note that this determinant is zero either if two orbitals are the same ($l_i = l_j$) or if two coordinates coincide ($\mathbf{r}_i = \mathbf{r}_j$), so the Pauli exclusion principle is correctly built in. Note also that, since the sign of the determinant is changed if we exchange two adjacent rows, it is necessary to keep in mind a definite ordering convention for the single particle orbitals $\phi_l(\mathbf{r})$ to fix the phase of the wavefunction.

For bosons, we should use an object similar to a determinant, but having all terms combined with a positive sign: this is known as a *permanent*.

8.1.3 Occupation numbers

We can specify the basis states we have constructed by giving the number of particles n_l in each orbital l. Clearly, for fermions $n_l = 0$ or 1, while for bosons $n_l = 0, 1, \dots$. These occupation numbers are used within Dirac notation as labels for a state: $|n_1, n_2, \dots\rangle$.

8.1.4 Fock space

Combining states $|n_1, n_2, \ldots\rangle$ with all possible values of the occupation numbers, we have basis vectors for states with any number of particles. This vector space is known as *Fock space*. Using it, we can discuss processes in which particles are created or annihilated, as well as ones with fixed particle number, described by wavefunctions of the form $\psi(\mathbf{r}_1 \ldots \mathbf{r}_N)$.

8.1.5 The vacuum state

It is worth noting that one of the states in Fock space is the vacuum: the wavefunction for the quantum system when it contains no particles, written as $|0\rangle$. Clearly, in recognising this as a quantum state we have come some way from the notation of single-body and few-body quantum mechanics, with wavefunctions written as functions of particle coordinates. Of course, $|0\rangle$ is different from 0, and in particular $\langle 0|0\rangle = 1$.

8.1.6 Creation and annihilation operators

Many of the calculations we will want to do are carried out most efficiently by introducing creation operators, which add particles when they act to the right on states from Fock space. Their Hermitian conjugates are annihilation operators, which remove particles. Their definition rests on the set of single particle orbitals from which we built Fock space: $c_l^\dagger$ adds particles to the orbital $\phi_l(\mathbf{r})$. More formally, we define

$$c_{l_1}^\dagger c_{l_2}^\dagger \ldots c_{l_N}^\dagger |0\rangle \tag{8.3}$$

to be the state with coordinate wavefunction

$$\psi(\mathbf{r}_1, \ldots \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_{\text{all perms}} (\pm 1)^P \phi_{k_1}(\mathbf{r}_1) \ldots \phi_{k_N}(\mathbf{r}_N) = (n_1!\, n_2!\, \ldots)^{1/2} |n_1, n_2 \ldots\rangle \; . \tag{8.4}$$

A detail to note is that the sum in Eq. (8.4) is over all permutations, while that in Eq. (8.1) included only distinct permutations. The difference (which is significant only for bosons, since it is only for bosons that we can have $n_l > 1$), is the reason for the factor $(n_1!\, n_2!\, \ldots)^{1/2}$ appearing on the right of Eq. (8.4). This choice anticipates what is necessary in order for boson creation and annihilation operators to have convenient commutation relations.

Annihilation operators appear when we take the Hermitian conjugate of Eq. (8.3), obtaining $\langle 0|\, c_{l_N} \ldots c_{l_2} c_{l_1}$. Let's examine the effect of creation and annihilation operators when they act on various states. Since $c_l^\dagger|0\rangle$ is the state with coordinate wavefunction $\phi_l(\mathbf{r})$, we know that $\langle 0|c_l\, c_l^\dagger|0\rangle = 1$, but for any choice of the state $|\phi\rangle$ other than the vacuum, $c_l^\dagger|\phi\rangle$ contains more than one particle and hence $\langle 0|c_l\, c_l^\dagger|\phi\rangle = 0$. From this we can conclude that

$$c_l\, c_l^\dagger|0\rangle = |0\rangle \; ,$$

demonstrating that the effect of c_l is to remove a particle from the state $|n_l{=}1\rangle \equiv c_l^\dagger|0\rangle$. We also have for any $|\phi\rangle$ the inner products $\langle 0|c_l^\dagger|\phi\rangle = \langle\phi|c_l\,|0\rangle = 0$, and so we can conclude that

$$c_l\,|0\rangle = \langle 0|c_l^\dagger = 0 \; .$$

8.1.7 Commutation and anticommutation relations

Recalling the factor of $(\pm 1)^P$ in Eq. (8.4), we have for any $|\phi\rangle$

$$c_l^\dagger c_m^\dagger|\phi\rangle = \pm c_m^\dagger c_l^\dagger|\phi\rangle \; ,$$

where the upper sign is for bosons and the lower one for fermions. From this we conclude that boson creation operators commute, and fermion creation operators anticommute: that is, for bosons

$$[c_l^\dagger, c_m^\dagger] = 0$$

and for fermions

$$\{c_l^\dagger, c_m^\dagger\} = 0 \; ,$$

where we use the standard notation for an anticommutator of two operators A and B: $\{A, B\} = AB+BA$. Taking Hermitian conjugates of these two equations, we have for bosons

$$[c_l, c_m] = 0$$

and for fermions

$$\{c_l, c_m\} = 0 \ .$$

Note for fermions we can conclude that $(c_l)^2$=$(c_l^\dagger)^2$=0, which illustrates again how the Pauli exclusion principle is built into our approach.

Finally, one can check that to reproduce the values of inner products of states appearing in Eq. (8.4), we require for bosons

$$[c_l\,, c_m^\dagger] = \delta_{lm}$$

and for fermions

$$\{c_l\,, c_m^\dagger\} = \delta_{lm} \ .$$

To illustrate the correctness of these relations, consider for a single boson orbital the value of $|[(c^\dagger)^n|0\rangle]|^2$. From Eq. (8.4) we have $|[(c^\dagger)^n|0\rangle]|^2 = n!$. Let's recover the same result by manipulating commutators: we have

$$\begin{aligned}
\langle 0|(c)^n(c^\dagger)^n|0\rangle &= \langle 0|(c)^{n-1}([c, c^\dagger] + c^\dagger c)(c^\dagger)^{n-1}|0\rangle \\
&= m\langle 0|(c)^{n-1}(c^\dagger)^{n-1}|0\rangle + \langle 0|c^\dagger(c)^{n-m}c^\dagger c(c)^m(c^\dagger)^{n-1}|0\rangle \\
&= n\langle 0|(c)^{n-1}(c^\dagger)^{n-1}|0\rangle + \langle 0|c^\dagger(c)^{n-1}(c^\dagger)^{n-1}|0\rangle \\
&= n(n-1)\ldots(n-l)\langle 0|(c^\dagger)^{n-l}(c)^{n-l}|0\rangle \\
&= n!\,\langle 0|0\rangle \ .
\end{aligned}$$

Of course, manipulations like these are familiar from the theory of raising and lowering operators for the harmonic oscillator.

8.1.8 Number operators

From Eq. (8.4) as the defining equation for the action of creation operators in Fock space we have

$$c_l^\dagger|n_1 \ldots n_l \ldots\rangle = (\pm 1)^{n_1+\ldots+n_{l-1}}\sqrt{n_l+1}|n_1 \ldots n_l + 1 \ldots\rangle \ ,$$

or zero for fermions if n_l=1. Similarly, by considering the Hermitian conjugate of a similar equation, we have

$$c_l|n_1 \ldots n_l \ldots\rangle = (\pm 1)^{n_1+\ldots+n_{l-1}}\sqrt{n_l}|n_1 \ldots n_l - 1 \ldots\rangle \ ,$$

or zero for both bosons and fermions if n_l=0. In this way we have

$$c_l^\dagger c_l\,|\ldots n_l \ldots\rangle = n_l|\ldots n_l \ldots\rangle$$

where the possible values of n_l are n_l=0, 1, 2 ... for bosons and n_l=0, 1 for fermions. Thus the combination $c_l^\dagger c_l$, which we will also write as $\hat{n}_l$, is the number operator and counts particles in the orbital ϕ_l.

8.1.9 Transformations between bases

In the context of single-particle quantum mechanics it is often convenient to make transformations between different bases. Since we used a particular set of basis functions in our definition of creation and annihilation operators, we should understand what such transformations imply in operator language.

Suppose we have two complete, orthonormal sets of single-particle basis functions, $\{\phi_l(\mathbf{r})\}$ and $\{\rho_\alpha(\mathbf{r})\}$. Then we can expand one in terms of the other, writing

$$\rho_\alpha(\mathbf{r}) = \sum_l \phi_l(\mathbf{r})U_{l\alpha} \tag{8.5}$$

with $U_{l\alpha} = \langle\phi_l|\rho_\alpha\rangle$. Note that U is a unitary matrix, since

$$\begin{aligned}(\mathrm{U}\mathrm{U}^\dagger)_{ml} &= \sum_\alpha \langle\phi_m|\rho_\alpha\rangle\langle\rho_\alpha|\phi_l\rangle \\ &= \langle\phi_m|\phi_l\rangle \qquad \text{since} \qquad \sum_\alpha |\rho_\alpha\rangle\langle\rho_\alpha| = \mathbf{1} \\ &= \delta_{ml}\ .\end{aligned}$$

Now let $c_l^\dagger$ create a particle in orbital $\phi_l(\mathbf{r})$, and let $d_\alpha^\dagger$ create a particle in orbital $\rho_\alpha(\mathbf{r})$. We can read off from Eq. (8.5) an expression for $d_\alpha^\dagger$ in terms of $c_l^\dagger$:

$$d_\alpha^\dagger = \sum_l c_l^\dagger U_{l\alpha}\ .$$

From the Hermitian conjugate of this equation we also have

$$d_\alpha = \sum_l U_{l\alpha}^* c_l = \sum_l (\mathrm{U}^\dagger)_{\alpha l} c_l\ .$$

Effect of transformations on commutation relations

We should verify that such transformations preserve commutation relations. For example, suppose that c_l and $c_l^\dagger$ are fermion operators, obeying $\{c_l\,, c_m^\dagger\} = \delta_{lm}$. Then

$$\{d_\alpha, d_\beta^\dagger\} = \sum_{lm} U_{l\alpha}^* U_{m\beta}\,\{c_l\,, c_m^\dagger\} = (\mathrm{U}^\dagger\mathrm{U})_{\alpha\beta} = \delta_{\alpha\beta}\,.$$

Similarly, for boson operators commutation relations are preserved under unitary transformations.

8.1.10 General single-particle operators in second-quantised form

To continue our programme of formulating many-particle quantum mechanics in terms of creation and annihilation operators, we need to understand how to transcribe operators from coordinate representation or first-quantised form to so-called second-quantised form. In the first instance, we examine how to do this for one-body operators – those which involve the coordinates of one particle at a time. An example is the kinetic energy operator. Suppose in general that $A(\mathbf{r})$ represents such a quantity for a single-particle system. Then for a system of N particles in first-quantised notation we have

$$\hat{A} = \sum_{i=1}^N A(\mathbf{r}_i)\ .$$

We want to represent $\hat{A}$ using creation and annihilation operators. As a first step, we can characterise $A(\mathbf{r})$ by its matrix elements, writing

$$A_{lm} = \int \phi_l^*(\mathbf{r}) A(\mathbf{r}) \phi_m(\mathbf{r}) \mathrm{d}^d\mathbf{r}\ .$$

Then

$$A(\mathbf{r})\phi_m(\mathbf{r}) = \sum_l \phi_l(\mathbf{r}) A_{lm}\ . \tag{8.6}$$

The second-quantised representation is

$$\hat{A} = \sum_{pq} A_{pq} c_p^\dagger c_q\ . \tag{8.7}$$

To justify this, we should verify that reproduces the correct matrix elements between all states from the Fock space. We will simply check the action of $\hat{A}$ on single particles states. We have

$$\hat{A}|\phi_m\rangle = \sum_{pq} A_{pq} c_p^\dagger c_q c_m^\dagger |0\rangle\ .$$

Now, taking as an example bosons,

$$c_p^\dagger c_q c_m^\dagger |0\rangle = c_p^\dagger([c_q, c_m^\dagger] + c_m^\dagger c_q)|0\rangle = c_p^\dagger \delta_{qm}|0\rangle$$

so

$$\hat{A}|\phi_m\rangle = \sum_p |\phi_p\rangle A_{pm} \ ,$$

reproducing Eq. (8.6), as required.

8.1.11 Two-particle operators in second-quantised form

It is important to make this transcription for two-body operators as well. Such operators depend on the coordinates of a pair of particles, an example being the two-body potential in an interacting system. Writing the operator in first-quantised form as $A(\mathbf{r}_1, \mathbf{r}_2)$, it has matrix elements which carry four labels:

$$A_{lmpq} = \int \phi_l^*(\mathbf{r}_1)\phi_m^*(\mathbf{r}_2)A(\mathbf{r}_1,\mathbf{r}_2)\phi_p(\mathbf{r}_2)\phi_q(\mathbf{r}_1)\mathrm{d}^d\mathbf{r}_1\mathrm{d}^d\mathbf{r}_2 \ .$$

Its second-quantised form is

$$\hat{A} \equiv \sum_{ij} A(\mathbf{r}_i,\mathbf{r}_j) = \sum_{lmpq} A_{lmpq} c_l^\dagger c_m^\dagger c_p c_q \ . \tag{8.8}$$

Again, to justify this one should check matrix elements of the second-quantised form between all states in Fock space. We will content ourselves with matrix elements for two-particle states, evaluating

$$\langle A\rangle = \langle 0|c_y c_x \hat{A} c_a^\dagger c_b^\dagger|0\rangle$$

by two routes. In a first-quantised calculation with $\pm$ signs for bosons and fermions, we have

$$\begin{aligned}
\langle A\rangle &= \frac{1}{2}\int\int [\phi_x^*(\mathbf{r}_1)\phi_y^*(\mathbf{r}_2) \pm \phi_x^*(\mathbf{r}_2)\phi_y^*(\mathbf{r}_1)]\cdot[A(\mathbf{r}_1,\mathbf{r}_2)+A(\mathbf{r}_2,\mathbf{r}_1)]\cdot[\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2)\pm\phi_a(\mathbf{r}_2)\phi_b(\mathbf{r}_1)]\mathrm{d}^d\mathbf{r}_1\mathrm{d}^d\mathbf{r}_2 \\
&= \frac{1}{2}[A_{xyba} \pm A_{xyab} + A_{yxab} \pm A_{yxba} + A_{xyba} \pm A_{xyab} + A_{yxab} \pm A_{yxba}] \\
&= (A_{xyba} + A_{yxab}) \pm (A_{xyab} + A_{yxba}) \ .
\end{aligned} \tag{8.9}$$

Using the proposed second-quantised form for $\hat{A}$, we have

$$\langle A\rangle = \sum_{lmpq} A_{lmpq}\langle 0|c_y c_x c_l^\dagger c_m^\dagger c_p c_q c_a^\dagger c_b^\dagger|0\rangle \ .$$

We can simplify the vacuum expectation value of products of creation and annihilation operators such as the one appearing here by using the appropriate commutation or anticommutation relation to move annihilation operators to the right, or creation operators to the left, whereupon acting on the vacuum they give zero. In particular

$$c_p c_q c_a^\dagger c_b^\dagger|0\rangle \ = (\delta_{aq}\delta_{bp} \pm \delta_{ap}\delta_{bq})|0\rangle$$

and

$$\langle 0|c_y c_x c_l^\dagger c_m^\dagger = \langle 0|(\delta_{ym}\delta_{xl} \pm \delta_{yl}\delta_{xm}) \ .$$

Combining these, we recover Eq. (8.9).

8.2 Diagonalisation of quadratic Hamiltonians

If a Hamiltonian is quadratic (or, more precisely, bilinear) in creation and annihilation operators we can diagonalise it, meaning we can reduce it to a form involving only number operators. This is an approach that applies directly to Hamiltonians for non-interacting systems, and also to Hamiltonians for interacting systems when interactions are treated within a mean field approximation.

8.2.1 Number-conserving quadratic Hamiltonians

Such Hamiltonians have the form

$$\mathcal{H} = \sum_{ij} H_{ij} a_i^\dagger a_j \ .$$

Note that in order for the operator $\mathcal{H}$ to be Hermitian, we require the matrix H to be Hermitian. Since the matrix H is Hermitian, it can be diagonalised by unitary transformation. Denote this unitary matrix by U and let the eigenvalues of H be ε_n. The same transformation applied to the creation and annihilation operators will diagonalise $\mathcal{H}$. The details of this procedure are as follows. Let

$$\alpha_l^\dagger = \sum_i a_i^\dagger U_{il} \ .$$

Inverting this, we have

$$\sum \alpha_l^\dagger (\mathrm{U}^\dagger)_{lj} = a_j^\dagger$$

and taking a Hermitian conjugate

$$\sum_l U_{jl} \alpha_l = a_j \ .$$

Substituting for $a^\dagger$'s and a's in terms of $\alpha^\dagger$'s and α's, we find

$$\mathcal{H} = \sum_{lm} \alpha_l^\dagger (\mathrm{U}^\dagger \mathrm{H} \mathrm{U})_{lm} \alpha_m = \sum_n \varepsilon_n \alpha_n^\dagger \alpha_n \equiv \sum_n \varepsilon_n \hat{n}_n \ .$$

Thus the eigenstates of $\mathcal{H}$ are the occupation number eigenstates in the basis generated by the creation operators $\alpha_n^\dagger$.

8.2.2 Mixing creation and annihilation operators: Bogoliubov transformations

There are a number of physically important systems which, when treated approximately, have bilinear Hamiltonians that include terms with two creation operators, and others with two annihilation operators. Examples include superconductors, superfluids and antiferromagnets. These Hamiltonians can be diagonalised by what are known as *Bogoliubov transformations*, which mix creation and annihilation operators, but, as always, preserve commutation relations. We now illustrate these transformations, discussing fermions and bosons separately.

Fermions

Consider for fermion operators the Hamiltonian

$$\mathcal{H} = \epsilon(c_1^\dagger c_1 + c_2^\dagger c_2) + \lambda(c_1^\dagger c_2^\dagger + c_2 c_1) \ ,$$

which arises in the BCS theory of superconductivity. Note that λ must be real for $\mathcal{H}$ to be Hermitian (more generally, with complex λ the second term of $\mathcal{H}$ would read $\lambda c_1^\dagger c_2^\dagger + \lambda^* c_2 c_1$). Note as well the opposite ordering of labels in the terms $c_1^\dagger c_2^\dagger$ and $c_2 c_1$, which is also a requirement of Hermiticity.

The fermionic Bogoliubov transformation is

$$\begin{aligned} c_1^\dagger &= u d_1^\dagger + v d_2 \\ c_2^\dagger &= u d_2^\dagger - v d_1 \ , \end{aligned} \tag{8.10}$$

where u and v are c-numbers, which we can in fact take to be real, because we have restricted ourselves to real λ. The transformation is useful only if fermionic anticommutation relations apply to both sets of operators. Let us suppose they apply to the operators d and $d^\dagger$, and check the properties of the operators c and $c^\dagger$. The coefficients of the transformation have been chosen to ensure that $\{c_1^\dagger, c_2^\dagger\} = 0$, while

$$\{c_1^\dagger, c_1\} = u^2 \{d_1^\dagger, d_1\} + v^2 \{d_2^\dagger, d_2\}$$

and so we must require $u^2 + v^2 = 1$, suggesting the parameterisation $u = \cos\theta$, $v = \sin\theta$.

The remaining step is to substitute in $\mathcal{H}$ for $c^\dagger$ and c in terms of $d^\dagger$ and d, and pick θ so that terms in $d_1^\dagger d_2^\dagger + d_2 d_1$ have vanishing coefficient. The calculation is perhaps clearest when it is set out using matrix notation. First, we can write $\mathcal{H}$ as

$$\mathcal{H} = \frac{1}{2} \begin{pmatrix} c_1^\dagger & c_2 & c_2^\dagger & c_1 \end{pmatrix} \begin{pmatrix} \epsilon & \lambda & 0 & 0 \\ \lambda & -\epsilon & 0 & 0 \\ 0 & 0 & \epsilon & -\lambda \\ 0 & 0 & -\lambda & -\epsilon \end{pmatrix} \begin{pmatrix} c_1 \\ c_2^\dagger \\ c_2 \\ c_1^\dagger \end{pmatrix} + \epsilon$$

where we have used the anticommutator to make substitutions of the type $c^\dagger c = 1 - c\, c^\dagger$.

For conciseness, consider just the upper block

$$\begin{pmatrix} c_1^\dagger & c_2 \end{pmatrix} \begin{pmatrix} \epsilon & \lambda \\ \lambda & -\epsilon \end{pmatrix} \begin{pmatrix} c_1 \\ c_2^\dagger \end{pmatrix}$$

and write the Bogoliubov transformation also in matrix form as

$$\begin{pmatrix} c_1 \\ c_2^\dagger \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} d_1 \\ d_2^\dagger \end{pmatrix} .$$

We pick θ so that

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \epsilon & \lambda \\ \lambda & -\epsilon \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \tilde{\epsilon} & 0 \\ 0 & -\tilde{\epsilon} \end{pmatrix} ,$$

where $\tilde{\epsilon} = \sqrt{\epsilon^2 + \lambda^2}$. Including the other 2×2 block of $\mathcal{H}$, we conclude that

$$\mathcal{H} = \tilde{\epsilon}(d_1^\dagger d_1 + d_2^\dagger d_2) + \epsilon - \tilde{\epsilon} .$$

Bosons

The Bogoliubov transformation for a bosonic system is similar in principle to what we have just set out, but different in detail. We are concerned with a Hamiltonian of the same form, but now written using boson creation and annihilation operators:

$$\mathcal{H} = \epsilon(c_1^\dagger c_1 + c_2^\dagger c_2) + \lambda(c_1^\dagger c_2^\dagger + c_2 c_1) .$$

We use a transformation of the form

$$\begin{aligned} c_1^\dagger &= u d_1^\dagger + v d_2 \\ c_2^\dagger &= u d_2^\dagger + v d_1 . \end{aligned}$$

Note that one sign has been chosen differently from its counterpart in Eq. (8.10) in order to ensure that bosonic commutation relations for the operators d and $d^\dagger$ imply the result $[c_1^\dagger, c_2^\dagger] = 0$. We also require

$$[c_1, c_1^\dagger] = u^2[d_1, d_1^\dagger] - v^2[d_2, d_2^\dagger] = 1$$

and hence $u^2 - v^2 = 1$. The bosonic Bogoliubov transformation may therefore be parameterised as $u = \cosh\theta$, $v = \sinh\theta$.

We can introduce matrix notation much as before (but note some crucial sign differences), with

$$\mathcal{H} = \frac{1}{2} \begin{pmatrix} c_1^\dagger & c_2 & c_2^\dagger & c_1 \end{pmatrix} \begin{pmatrix} \epsilon & \lambda & 0 & 0 \\ \lambda & \epsilon & 0 & 0 \\ 0 & 0 & \epsilon & \lambda \\ 0 & 0 & \lambda & \epsilon \end{pmatrix} \begin{pmatrix} c_1 \\ c_2^\dagger \\ c_2 \\ c_1^\dagger \end{pmatrix} - \epsilon ,$$

where for bosons we have used the commutator to write $c^\dagger c = c\, c^\dagger - 1$. Again, we focus on one 2×2 block

$$\begin{pmatrix} c_1^\dagger & c_2 \end{pmatrix} \begin{pmatrix} \epsilon & \lambda \\ \lambda & \epsilon \end{pmatrix} \begin{pmatrix} c_1 \\ c_2^\dagger \end{pmatrix}$$

and write the Bogoliubov transformation also in matrix form as

$$\begin{pmatrix} c_1 \\ c_2^\dagger \end{pmatrix} \begin{pmatrix} u & v \\ v & u \end{pmatrix} \begin{pmatrix} d_1 \\ d_2^\dagger \end{pmatrix} .$$

Substituting for c and $c^\dagger$ in terms of d and $d^\dagger$, this block of the Hamiltonian becomes

$$\begin{pmatrix} d_1^\dagger & d_2 \end{pmatrix} \begin{pmatrix} u & v \\ v & u \end{pmatrix} \begin{pmatrix} \epsilon & \lambda \\ \lambda & \epsilon \end{pmatrix} \begin{pmatrix} u & v \\ v & u \end{pmatrix} \begin{pmatrix} d_1 \\ d_2^\dagger \end{pmatrix} .$$

In the fermionic case the matrix transformation was simply an orthogonal rotation. Here it is not, and so we should examine it in more detail. We have

$$\begin{pmatrix} u & v \\ v & u \end{pmatrix} \begin{pmatrix} \epsilon & \lambda \\ \lambda & \epsilon \end{pmatrix} \begin{pmatrix} u & v \\ v & u \end{pmatrix} = \begin{pmatrix} \epsilon[u^2+v^2]+2\lambda uv & 2\epsilon uv + \lambda[u^2+v^2] \\ 2\epsilon uv + \lambda[u^2+v^2] & \epsilon[u^2+v^2]+2\lambda uv \end{pmatrix} .$$

It is useful to recall the double angle formulae $u^2+v^2 = \cosh 2\theta$ and $2uv = \sinh 2\theta$. Then, setting $\tanh 2\theta = -\lambda/\epsilon$ we arrive at

$$\mathcal{H} = \tilde{\epsilon}(d_1^\dagger d_1 + d_2^\dagger d_2) - \epsilon + \tilde{\epsilon} .$$

with

$$\tilde{\epsilon} = \sqrt{\epsilon^2 - \lambda^2}. \tag{8.11}$$

Note that in the bosonic case the transformation requires $\epsilon > \lambda$: if this is not the case, $\mathcal{H}$ is not a Hamiltonian for normal mode oscillations about a stable equilibrium, but instead represents a system at an unstable equilibrium point.

8.3 Density correlations in ideal quantum gases

Having set up the machinery, we now apply it to some problems of physical interest. One of the simplest is a calculation of particle correlations in an ideal quantum gas. These correlations arise purely from quantum statistics, since in an ideal gas there are no interactions between particles and therefore no correlations at all at a classical level.

Consider N identical particles in a three-dimensional cubic box of side L with periodic boundary conditions. The single particle eigenstates are plane waves: we write

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{L^{3/2}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \qquad \text{with} \qquad \mathbf{k} = \frac{2\pi}{L}(l,m,n) , \qquad l,m,n \quad \text{integer} .$$

Introducing creation operators $c_{\mathbf{k}}^\dagger$ for particles in these orbitals, the creation operator for a particle at the point $\mathbf{r}$ is

$$c^\dagger(\mathbf{r}) = \frac{1}{L^{3/2}} \sum_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} c_{\mathbf{k}}^\dagger .$$

With particle coordinates denoted by $\mathbf{r}_i$, the density operator in first-quantised form is

$$\rho(\mathbf{r}) \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) .$$

In second quantised form it is simply the number operator at the point $\mathbf{r}$,

$$\rho(\mathbf{r}) = c^\dagger(\mathbf{r}) c(\mathbf{r}) = \frac{1}{L^3} \sum_{\mathbf{kq}} \mathrm{e}^{\mathrm{i}(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} c_{\mathbf{k}}^\dagger c_{\mathbf{q}} ,$$

as is confirmed by using Eq. (8.7).

We will calculate $\langle\rho(\mathbf{r})\rangle$ and $\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle$. As a prelude to the quantum calculation, it is worth considering the problem classically. In a classical system the meaning of the average $\langle\ldots\rangle$ is a normalised multiple integral of all particle coordinates over the volume of the box. Hence

$$\langle\rho(\mathbf{r})\rangle = \frac{N}{L^3} \int \mathrm{d}^3\mathbf{r}_i \, \delta(\mathbf{r} - \mathbf{r}_i) = \frac{N}{L^3}$$

and

$$\begin{aligned}\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle &= \frac{N}{L^3}\int \mathrm{d}^3\mathbf{r}_i\,\delta(\mathbf{r}-\mathbf{r}_i)\delta(\mathbf{r}_i) + \frac{N(N-1)}{L^6}\int\int \mathrm{d}^3\mathbf{r}_i\mathrm{d}^3\mathbf{r}_j\,\delta(\mathbf{r}-\mathbf{r}_i)\delta(\mathbf{r}_j) \\ &= \langle\rho(\mathbf{r})\rangle^2 + \frac{N}{L^3}\left[\delta(\mathbf{r}) - \frac{1}{L^3}\right] .\end{aligned}$$

In the large volume limit we have simply

$$\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle = \langle\rho(\mathbf{r})\rangle^2 + \langle\rho(\mathbf{r})\rangle\delta(\mathbf{r}) . \tag{8.12}$$

Now we move to the quantum calculation, taking the average $\langle\ldots\rangle$ to mean an expectation value in number eigenstates for the orbitals $\phi_{\mathbf{k}}(\mathbf{r})$, weighted by Boltzmann factors at finite temperature. To evaluate $\langle\rho(\mathbf{r})\rangle$ we need

$$\langle c^\dagger_{\mathbf{k}} c_{\mathbf{q}}\rangle = \delta_{\mathbf{k},\mathbf{q}}\langle n_{\mathbf{k}}\rangle ,$$

where $n_{\mathbf{k}}$ is the number operator for the orbital $\phi_{\mathbf{k}}(\mathbf{r})$. Since $\sum_{\mathbf{k}}\langle n_{\mathbf{k}}\rangle = N$, we find

$$\langle\rho(\mathbf{r})\rangle = \frac{N}{L^3}$$

as for a classical system. The two-point correlation function is more interesting. To evaluate $\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle$ we need averages of the form $\langle c^\dagger_{\mathbf{k}} c_{\mathbf{q}} c^\dagger_{\mathbf{l}} c_{\mathbf{p}}\rangle$. These are non-zero in two cases: (i) $\mathbf{k}=\mathbf{q}$ and $\mathbf{l}=\mathbf{p}$; or (ii) $\mathbf{k}=\mathbf{p}$ and $\mathbf{l}=\mathbf{q}$, with $\mathbf{q}\neq\mathbf{p}$ to prevent double-counting of terms included under (i). With $\mathbf{p}\neq\mathbf{q}$ we have $\langle c^\dagger_{\mathbf{q}} c_{\mathbf{q}} c^\dagger_{\mathbf{p}} c_{\mathbf{p}}\rangle = \langle n_{\mathbf{q}} n_{\mathbf{p}}\rangle$, and $\langle c^\dagger_{\mathbf{p}} c_{\mathbf{q}} c^\dagger_{\mathbf{q}} c_{\mathbf{p}}\rangle = \langle n_{\mathbf{p}}(1\pm n_{\mathbf{q}})\rangle$, where the upper sign is for bosons and the lower one for fermions. From this

$$\begin{aligned}\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle &= \frac{1}{L^6}\sum_{\mathbf{kqlp}} \mathrm{e}^{\mathrm{i}(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}}\langle c^\dagger_{\mathbf{k}} c_{\mathbf{q}} c^\dagger_{\mathbf{l}} c_{\mathbf{p}}\rangle \\ &= \frac{1}{L^6}\sum_{\mathbf{qp}}\langle n_{\mathbf{q}}\rangle\langle n_{\mathbf{p}}\rangle + \frac{1}{L^6}\sum_{\mathbf{qp}}\mathrm{e}^{\mathrm{i}(\mathbf{q}-\mathbf{p})\cdot\mathbf{r}}\langle n_{\mathbf{p}}\rangle(1\pm\langle n_{\mathbf{q}}\rangle) \\ &\quad + \frac{1}{L^6}\sum_{\mathbf{k}}\left[\langle n^2_{\mathbf{k}}\rangle - \langle n_{\mathbf{k}}\rangle^2 - \langle n_{\mathbf{k}}\rangle(1\pm\langle n_{\mathbf{k}}\rangle)\right] .\end{aligned}$$

The final term is negligible in the limit $L\to\infty$ with N/L^3 fixed, and we discard it. In this limit we can also make the replacement

$$\frac{1}{L^3}\sum_{\mathbf{p}} \to \frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{p}$$

so that we obtain finally

$$\langle\rho(\mathbf{r})\rho(\mathbf{0})\rangle = \langle\rho(\mathbf{r})\rangle^2 + \langle\rho(\mathbf{r})\rangle\delta(\mathbf{r}) \pm \left|\frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{p}\langle n_{\mathbf{p}}\rangle \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{r}}\right|^2 . \tag{8.13}$$

The final term on the right-hand side of Eq. (8.13) is the correction to our earlier classical result, Eq. (8.12), and represents the consequences of quantum statistics. Its detailed form as a function of $\mathbf{r}$ depends on the momentum distribution of particles, but the most important features are quite general: there is an enhancement in density correlations for bosons and a suppression for fermions, on a lengthscale of order $\hbar/\Delta p$, where Δp is a characteristic momentum for the gas. In summary: bosons bunch and fermions exclude.

8.4 Spin waves in a ferromagnet

We move now to a problem that illustrates how one can make approximations in order to obtain a simple description of excitations in an interacting system. The model we study is the quantum Heisenberg ferromagnet in the limit where the spin magnitude S is large. It represents an insulating ferromagnetic material, in which magnetic ions with well-defined magnetic moments occupy the sites of a lattice.

The three components of spin at site $\mathbf{r}$ are represented by operators $S^x_{\mathbf{r}}$, $S^y_{\mathbf{r}}$ and $S^z_{\mathbf{r}}$. Their commutation relations are the standard ones, and with $\hbar = 1$ take the form

$$[S^i_{\mathbf{r}_1}, S^j_{\mathbf{r}_2}] = \mathrm{i}\delta_{\mathbf{r}_1,\mathbf{r}_2}\epsilon_{ijk}S^k_{\mathbf{r}_1} .$$

We reproduce them in order to emphasise two points: first, they are more complicated than those for creation and annihilation operators, since the commutator is itself another operator and not a number; and second, spin operators acting at different sites commute. We will also make use of spin raising and lowering operators, defined in the usual way as $S^+ = S^x + \mathrm{i}S^y$ and $S^- = S^x - \mathrm{i}S^y$.

The Heisenberg Hamiltonian with nearest neighbour ferromagnetic exchange interactions of strength J is

$$\mathcal{H} = -J \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} \mathbf{S}_{\mathbf{r}} \cdot \mathbf{S}_{\mathbf{r}'} \equiv -J \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} \left[S^z_{\mathbf{r}} S^z_{\mathbf{r}'} + \frac{1}{2} \left(S^+_{\mathbf{r}} S^-_{\mathbf{r}'} + S^-_{\mathbf{r}} S^+_{\mathbf{r}'} \right) \right] . \tag{8.14}$$

Here $\sum_{\langle \mathbf{r}\mathbf{r}' \rangle}$ denotes a sum over neighbouring pairs of sites on the lattice, with each pair counted once. Thinking of the spins as classical, three component vectors with length S, the lowest energy states are ones in which all spins are parallel. The model is simple enough that we can write down the exact quantum ground states as well: for the case in which the spins are aligned along the positive z-axis, the ground state $|0\rangle$ is defined by the property $S^z_{\mathbf{r}}|0\rangle = S|0\rangle$ for all $\mathbf{r}$. Other ground states can be obtained by a global rotation of spin direction. Individual ground states in both the classical and quantum descriptions break the rotational symmetry of the Hamiltonian, and so we expect excitations which are Goldstone modes and therefore gapless. There are known as spin waves or magnons. Eigenstates with a single magnon excitation can in fact be found exactly, but in general to go further we need to make approximations. As a next step, we set out one approximation scheme.

8.4.1 Holstein Primakoff transformation

This transformation expresses spin operators in terms of boson operators. It provides an obvious way to build in the fact that spin operators at different sites commute. In a non-linear form it also reproduces exactly the commutation relations between two spin operators associated with the same site, but we will use a linearised version of the transformation which is approximate. At a single site we take the eigenvector of S^z with eigenvalue S to be the boson vacuum, and associate each unit reduction in S^z with the addition of a boson. Then

$$S^z = S - b^\dagger b \, .$$

From this we might guess $S^+ \propto b$ and $S^- \propto b^\dagger$. In an attempt to identify the proportionality constants we can compare the commutator $[S^+, S^-] = 2S^z$ with $[b, b^\dagger] = 1$. Since the commutator is an operator in the first case and a number in the second, our guessed proportionality cannot be exact, but within states for which $\langle S^z \rangle \approx S$ (meaning $\langle S^z \rangle - S \ll S$, which can be satisfied only if $S \gg 1$) we can take

$$S^+ \approx (2S)^{1/2} b \qquad \text{and} \qquad S^- \approx (2S)^{1/2} b^\dagger \, . \tag{8.15}$$

In an exact treatment, corrections to these expressions form a series in powers of $b^\dagger b/S$. Using this transformation and omitting the higher order terms, the Hamiltonian may be rewritten approximately as

$$\mathcal{H} = -J \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} S^2 - JS \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} \left[b^\dagger_{\mathbf{r}} b_{\mathbf{r}'} + b^\dagger_{\mathbf{r}'} b_{\mathbf{r}} - b^\dagger_{\mathbf{r}} b_{\mathbf{r}} - b^\dagger_{\mathbf{r}'} b_{\mathbf{r}'} \right] . \tag{8.16}$$

8.4.2 Approximate diagonalisation of Hamiltonian

Applying the approach of Section 8.2.1, we can diagonalise Eq. (8.16) by a unitary transformation of the creation and annihilation operators. In a translationally invariant system this is simply a Fourier transformation. Suppose the sites form a simple cubic lattice with unit spacing. Take the system to be a cube with with side L and apply periodic boundary conditions. The number of lattice sites is then $N = L^3$ and allowed wavevectors are

$$\mathbf{k} = \frac{2\pi}{L}(l, m, n) \qquad \text{with} \quad l, m, n \quad \text{integer} \qquad \text{and} \quad 1 \leq l, m, n \leq L \, .$$

Boson operators in real space and reciprocal space are related by

$$b_{\mathbf{r}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \, b_{\mathbf{k}} \qquad \text{and} \qquad b^\dagger_{\mathbf{r}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \, b^\dagger_{\mathbf{k}} \, .$$

We use these transformations, and introduce the notation $\mathbf{d}$ for vectors from a site to its nearest neighbours, and z for the coordination number of the lattice (the number of neighbours to a site: six for the simple cubic lattice), to

obtain

$$\begin{aligned}\mathcal{H} &= -JS^2Nz - JS\sum_{\mathbf{rd}}\sum_{\mathbf{kq}}\frac{1}{N}\mathrm{e}^{\mathrm{i}\mathbf{r}\cdot(\mathbf{k}-\mathbf{q})}[\mathrm{e}^{\mathrm{i}\mathbf{d}\cdot\mathbf{q}}-1]b^\dagger_{\mathbf{q}}b_{\mathbf{k}} \\ &= -JS^2Nz - JS\sum_{\mathbf{q}}\epsilon_{\mathbf{q}}b^\dagger_{\mathbf{q}}b_{\mathbf{q}}\ ,\end{aligned}$$

where

$$\epsilon_{\mathbf{q}} = 2JS(3-\cos q_x-\cos q_y-\cos q_z)\ .$$

In this way we have approximated the original Heisenberg Hamiltonian, involving spin operators, by one that is quadratic in boson creation and annihilation operators. By diagonalising this we obtain an approximate description of the low-lying excitations of the system as independent bosons. The most important feature of the result is the form of the dispersion a small wavevectors. For $q \ll 1$ we have $\epsilon_{\mathbf{q}} = JSq^2 + \mathcal{O}(q^4)$, illustrating that excitations are indeed gapless. The fact that dispersion is quadratic, and not linear as it is, for example for phonons, reflects broken time-reversal symmetry in the ground state of the ferromagnet.

8.5 Weakly interacting Bose gas

As a final example, we present a treatment of excitations in a Bose gas with repulsive interactions between particles, using an approximation that is accurate if interactions are weak. There is good reason for wanting to understand this problem in connection with the phenomenon of superfluidity: the flow of Bose liquids without viscosity below a transition temperature, as first observed below 2.1 K in liquid ^{4}He. Indeed, an argument due to Landau connects the existence of superfluidity with the form of the excitation spectrum, and we summarise this argument next.

8.5.1 Critical superfluid velocity: Landau argument

Consider superfluid of mass M flowing with velocity $\mathbf{v}$, and examine whether friction can arise by generation of excitations, characterised by a wavevector $\mathbf{k}$ and an energy $\epsilon(k)$. Suppose production of one such excitation reduces the bulk velocity to $\mathbf{v}-\Delta\mathbf{v}$. From conservation of momentum

$$M\mathbf{v} = M\mathbf{v} - M\Delta\mathbf{v} + \hbar\mathbf{k}$$

and from conservation of energy

$$\frac{1}{2}Mv^2 = \frac{1}{2}M|\mathbf{v}-\Delta\mathbf{v}|^2 + \epsilon(k)\ .$$

From these conditions we find at large M that $\mathbf{k}$, $\mathbf{v}$ and $\epsilon(k)$ should satisfy $\hbar\mathbf{k}\cdot\mathbf{v} = \epsilon(k)$. The left hand side of this equation can be made arbitrarily close to zero by choosing $\mathbf{k}$ to be almost perpendicular to $\mathbf{k}$, but it has a maximum for a given k, obtained by taking $\mathbf{k}$ parallel to $\mathbf{v}$. If $\hbar kv < \epsilon(k)$ for all k then the equality cannot be satisfied and frictional processes of this type are forbidden. This suggests that there should be a critical velocity v_{c} for superfluid flow, given by $v_{\mathrm{c}} = \min_k[\epsilon(k)/k]$. For v_{c} to be non-zero, we require a real, interacting Bose liquid to behave quite differently from the non-interacting gas, since without interactions the excitation energies are just those of individual particles, giving $\epsilon(k) = \hbar^2k^2/2m$ for bosons of mass m, and hence $v_{\mathrm{c}} = 0$. Reassuringly, we find from the following calculation that interactions have the required effect. For completeness, we should note also that while a critical velocity of the magnitude these arguments suggest is observed in appropriate experiments, in others there can be additional sources of friction that lead to much lower values of v_{c}.

8.5.2 Model for weakly interacting bosons

There are two contributions to the Hamiltonian of an interacting Bose gas: the single particle kinetic energy $\mathcal{H}_{\mathrm{KE}}$ and the interparticle potential energy $\mathcal{H}_{\mathrm{int}}$. We introduce boson creation and annihilation operators for plane wave states in a box with side L, as in Section 8.3. Then

$$\mathcal{H}_{\mathrm{KE}} = \sum_{\mathbf{k}}\frac{\hbar^2k^2}{2m}c^\dagger_{\mathbf{k}}c_{\mathbf{k}}\ .$$

Short range repulsive interactions of strength parameterised by u are represented in first-quantised form by

$$\mathcal{H}_{\rm int} = \frac{u}{2}\sum_{i\neq j}\delta(\mathbf{r}_i - \mathbf{r}_j)\ .$$

Using Eq. (8.8) this can be written as

$$\mathcal{H}_{\rm int} = \frac{u}{2L^3}\sum_{\mathbf{kpq}} c^\dagger_{\mathbf{k}} c^\dagger_{\mathbf{p}} c_{\mathbf{q}} c_{\mathbf{k+p-q}}\ .$$

With this, our model is complete, with a Hamiltonian $\mathcal{H} = \mathcal{H}_{\rm KE} + \mathcal{H}_{\rm int}$.

8.5.3 Approximate diagonalisation of Hamiltonian

In order to apply the techniques set out in Section 8.2.1 we should approximate $\mathcal{H}$ by a quadratic Hamiltonian. The approach to take is suggested by recalling the ground state of the non-interacting Bose gas, in which all particles occupy the $\mathbf{k} = \mathbf{0}$ state. It is natural to suppose that the occupation of this orbital remains macroscopic for small u, so that the ground state expectation value $\langle c^\dagger_{\mathbf{0}} c_{\mathbf{0}}\rangle$ takes a value N_0 which is of the same order as N, the total number of particles. In this case we can approximate the operators $c^\dagger_{\mathbf{0}}$ and $c_{\mathbf{0}}$ by the c-number $\sqrt{N_0}$ and expand $\mathcal{H}$ in decreasing powers of N_0. We find

$$\mathcal{H}_{\rm int} = \frac{uN_0^2}{2L^3} + \frac{uN_0}{2L^3}\sum_{\mathbf{k}\neq\mathbf{0}}\left[2c^\dagger_{\mathbf{k}}c_{\mathbf{k}} + 2c^\dagger_{-\mathbf{k}}c_{-\mathbf{k}} + c^\dagger_{\mathbf{k}}c^\dagger_{-\mathbf{k}} + c_{\mathbf{k}}c_{-\mathbf{k}}\right] + \mathcal{O}([N_0]^0)\ .$$

At this stage N_0 is unknown, but we can write an operator expression for it, as

$$N_0 = N - \sum_{\mathbf{k}\neq\mathbf{0}} c^\dagger_{\mathbf{k}}c_{\mathbf{k}}\ .$$

It is also useful to introduce notation for the average number density $\rho = N/L^3$. Substituting for N_0 we obtain

$$\mathcal{H}_{\rm int} = \frac{u\rho}{2}N + \frac{u\rho}{2}\sum_{\mathbf{k}\neq\mathbf{0}}\left[c^\dagger_{\mathbf{k}}c_{\mathbf{k}} + c^\dagger_{-\mathbf{k}}c_{-\mathbf{k}} + c^\dagger_{\mathbf{k}}c^\dagger_{-\mathbf{k}} + c_{\mathbf{k}}c_{-\mathbf{k}}\right] + \mathcal{O}([N_0]^0)$$

and hence

$$\mathcal{H} = \frac{u\rho}{2}N + \frac{1}{2}\sum_{\mathbf{k}\neq\mathbf{0}}\left[E(k)\left(c^\dagger_{\mathbf{k}}c_{\mathbf{k}} + c^\dagger_{-\mathbf{k}}c_{-\mathbf{k}}\right) + u\rho\left(c^\dagger_{\mathbf{k}}c^\dagger_{-\mathbf{k}} + c_{\mathbf{k}}c_{-\mathbf{k}}\right)\right] + \ldots \tag{8.17}$$

with

$$E(k) = \frac{\hbar^2k^2}{2m} + u\rho\ .$$

At this order we have a quadratic Hamiltonian, which we can diagonalise using the Bogoliubov transformation for bosons set out in Section 8.2.2. From Eq. (8.11), we find that the dispersion relation for excitations in the Bose gas is

$$\epsilon(k) = \left[\left(\frac{\hbar^2k^2}{2m} + u\rho\right)^2 - (u\rho)^2\right]^{1/2}\ .$$

At large k ($\hbar^2k^2/2m \gg u\rho$), this reduces to the dispersion relation for free particles, but in the opposite limit it has the form

$$\epsilon(k) \simeq \hbar v k \qquad \text{with} \quad v = \sqrt{\frac{u\rho}{m}}\ .$$

In this way we obtain a critical velocity for superfluid flow, which is proportional to the interaction strength u, illustrating how interactions can lead to behaviour quite different from that in a non-interacting system.

8.6 Further Reading

- R. P. Feynman *Statistical Mechanics* (Addison Wesley). Chapter 6 provides a straightforward introduction to the use of particle creation and annihilation operators.
- A. Altland and B. D. Simons *Condensed Matter Field Theory* (CUP). Chapters 1 and 2 offer a good overview of the material covered in this part of the course.
- J-P Blaizot and G. Ripka *Quantum Theory of Finite Systems* (MIT Press) is a useful, clear and complete advanced reference book.

Chapter 9

Phase Transitions

In this chapter we will examine how the statistical mechanics of phase transitions can be formulated using the language of field theory. As in relativistic applications, symmetry will be an important guide, but since phenomena in a condensed matter setting are not constrained by Lorentz invariance, we encounter a variety of new possibilities.

9.1 Introduction

To provide a context, we start by summarising some basic facts about phase transitions. Take first a substance such as water, that can exist as a solid, liquid or vapour. Consider its phase diagram in the plane of temperature and pressure, as sketched in Fig. 9.1. The solid is separated from both liquid and vapour by phase boundaries, and one

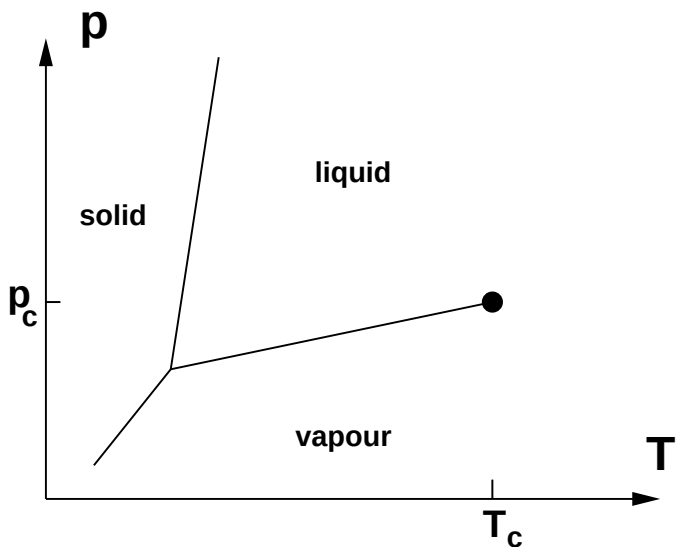

Figure 9.1: Schematic phase diagram in temperature T and pressure p for a material with solid, liquid and vapour phases. Line denote phase boundaries; the liquid-vapour critical point has coordinates $(T_{\rm c}, p_{\rm c})$.

cannot get from the solid to one of these other phases without crossing the phase boundary. By contrast, the liquid-vapour phase boundary ends at a critical point, and while some paths between the liquid and vapour cross this phase boundary, others do not. On crossing a phase boundary in this phase diagram, the system undergoes a discontinuous phase transition: properties such as density change discontinuously and there is a latent heat. On the other hand, if one follows a path between liquid and vapour that avoids the phase boundary by going around the critical point, properties vary smoothly along the path. As an intermediate case, we can consider a path between liquid and vapour that goes through the critical point. This turns out to involve a continuous (but sharp) phase transition, and it is partly behaviour at such critical points that will be focus of this chapter. To emphasise the distinction between discontinuous, or first order, transitions and continuous ones, it is useful to compare the behaviour of the heat capacity as a function of temperature in each case, as sketched in Fig. 9.2. At a first order transition there is a

latent heat, represented as a δ-function spike in the heat capacity, while at a continuous transition there is no latent heat, but the heat capacity shows either a cusp or a divergence, depending on the specific example.

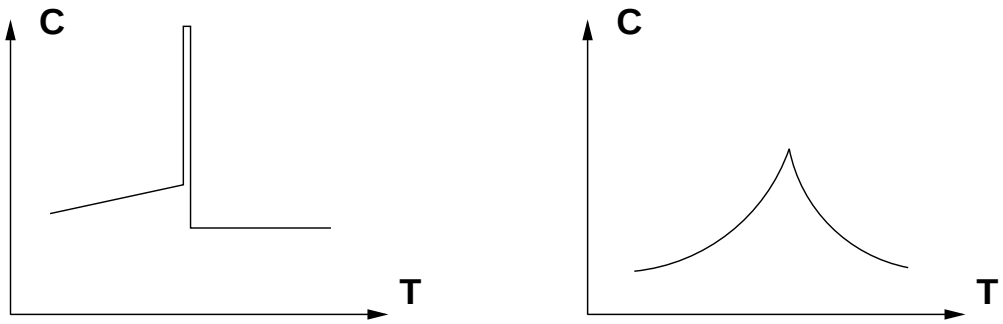

Figure 9.2: Schematic dependence of heat capacity on temperature: at a pressure $p < p_{\rm c}$ (left); and at $p = p_{\rm c}$ (right).

An alternative way of viewing the liquid-vapour transition is to examine behaviour as a function of density rather than pressure. A phase diagram of this kind is shown in Fig. 9.3. From this viewpoint the consequence of the first-order transition is that at temperatures $T < T_{\rm c}$ there is intermediate range of densities which the system can attain only as a mixture of two distinct, coexisting phases. One of these phases is the liquid, which has a higher density than that of the system on average, and the other is the vapour, with lower density than average. As the critical point is approached, the density difference between liquid and vapour reduces. Near the critical point, as the two phases become more similar, microscopic density fluctuations grow in size: roughly speaking, such fluctuations involve microscopic regions of vapour appearing within the liquid, or vice-versa. The correlation length represents the maximum size of such fluctuations, and diverges at the critical point. Sufficiently close to the critical point it is larger than the wavelength of light, and in these circumstance density fluctuations scatter light strongly, leading to a phenomenon known as *critical opalescence*: a cloudiness that appears in fluids close to their critical point.

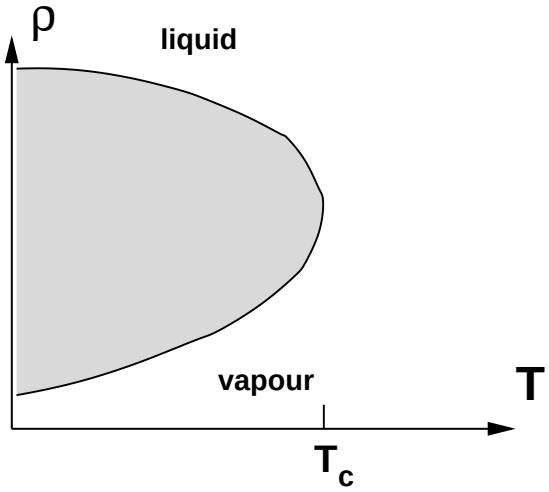

Figure 9.3: Phase diagram in the density-temperature plane. The region of two-phase coexistence is shaded; elsewhere the system exists in a single phase.

9.1.1 Paramagnet-ferromagnet transition

An important feature of our understanding of phase transitions is that there are close parallels between different examples. To illustrate this we next examine the transition between the paramagnetic and ferromagnetic phases of a magnetic material. The counterpart to the liquid-vapour phase diagram is shown in Fig. 9.4 (left), where we display

behaviour as a function of magnetic field (in place of pressure) and temperature. The properties of the system, including most importantly its magnetisation, vary smoothly with applied field above the critical temperature. Below the critical temperature, and neglecting hysteresis effects, there is a discontinuous change on sweeping applied field through zero, represented by a first-order phase boundary in the figure. In this case there is a symmetry under reversal of field and magnetisation that was not evident for the liquid vapour transition. Also shown in Fig. 9.4 are the two-phase coexistence region (centre) and the behaviour of the magnetic susceptibility (right). The latter diverges at the critical point: the divergence again reflects the presence of large fluctuations in the critical region, which are readily polarised by an applied field.

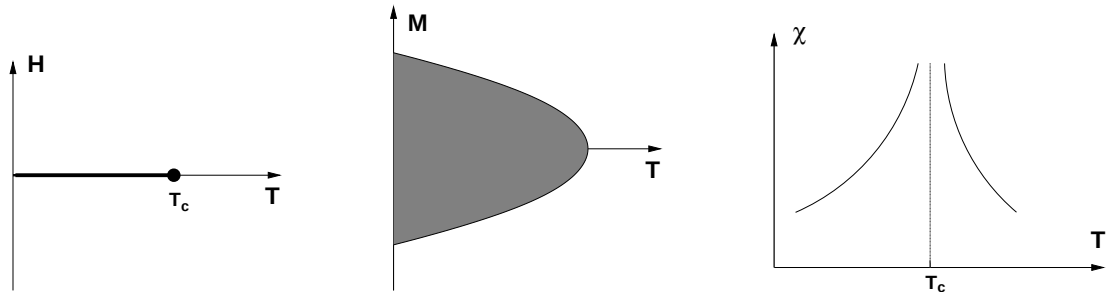

Figure 9.4: Left: phase diagram for a ferromagnet in the plane of field H vs temperature T, with T_c the critical point. Centre: the region of two-phase coexistence, in the plane of magnetisation M vs temperature. Right: behaviour of magnetic susceptibility χ.

9.1.2 Other examples of phase transitions

Other examples of continuous phase transitions include the normal to superfluid transition in liquid ^{4}He, and the superconducting transition in many metals and alloys. A further example, in which the transition is (for reasons we will examine in due course) first order, occurs in liquid crystals consisting of rod-like molecules. In their isotropic phase, these molecules are randomly orientated, while in the nematic phase they acquire an average orientation, despite remaining liquid and therefore positionally disordered.

9.1.3 Common features

The most important common feature of these phase transitions is the occurrence of spontaneous symmetry breaking. This is rather obscured in the case of the liquid-vapour transition, but is clear in our other examples. For instance, the ferromagnet might equally well acquire a magnetisation parallel or antiparallel to a given axis in the sample, and the two possibilities are related by time-reversal symmetry. In superfluids and superconductors, the condensate has a definite quantum-mechanical phase, whose value relative to that of another condensate can be probed if two condensates are coupled. And in a nematic liquid crystal the spontaneous choice of an orientation axis for molecules breaks the rotational symmetry displayed by the isotropic phase. We characterise symmetry breaking, both in magnitude and (in a generalised sense) orientation using an order parameter. For the ferromagnet, this is simply the magnetisation, m. For the superfluid, it can be taken to be the condensate amplitude ψ. And for the liquid-vapour transition, we take it to be the difference $\delta\rho = \rho - \rho_c$ in density ρ from its value ρ_c at the critical point.

A second feature common to continuous transitions is the existence of a lengthscale, the correlation length ξ, that diverges as the critical point is approached. For example, at the liquid-vapour transition density fluctuations are correlated over distances of order ξ and the correlation function $\langle \delta\rho(\mathbf{0})\delta\rho(\mathbf{r})\rangle$ falls to zero for $r \gg \xi$.

9.1.4 Critical behaviour

As a critical point is approached, many physical quantities show power-law behaviour that can be characterised by giving values of critical exponents. Their values turn out to be *universal* in the sense that they are determined by symmetries and a few other features of the system, but are insensitive to many microscopic details. The symbols used for the exponents are established by convention, and we now introduce them using the transition from paramagnet to a ferromagnet as an example. It is convenient to discuss behaviour as a function of reduced temperature,

defined as

$$t = \frac{T - T_{\rm c}}{T_{\rm c}} \, .$$

The cusp or divergence in the heat capacity C is represented by the exponent α, via $C \sim |t|^{-\alpha}$. The way in which the order parameter m decreases as the critical point is approached is described by the exponent β, with $m \sim |t|^{\beta}$ for $t < 0$, while the divergences in the susceptibility χ and the correlation length ξ are represented as $\chi \sim |t|^{-\gamma}$ and $\xi \sim |t|^{-\nu}$. Finally, at the critical temperature, the variation in the order parameter with field is written as $m \sim |h|^{1/\delta}$.

9.2 Field theory for phase transitions

There is a general approach to writing down a field theory, or continuum description, of a phase transitions. The first step is to identify an order parameter that characterises the nature of symmetry breaking at the transition. The value of the order parameter is determined by an average over a system as a whole. We introduce a field, defined as a function of position within the sample, that takes values in the order parameter space, and write the free energy density of the system as an expansion in powers of this field and its gradients, including all terms allowed by the symmetries of the high symmetry phase. We calculate the free energy associated with the degrees of freedom involved in the transition as a functional integral over all field configurations.

More explicitly, writing the order parameter field at point $\mathbf{r}$ as $\varphi(\mathbf{r})$, and the free energy density as $F(\varphi)$, we define a partition function

$$Z = \int D\varphi \, {\rm e}^{-\int F(\varphi) \, {\rm d}^d \mathbf{r}}$$

and obtain the free energy for the sample from this partition function in the usual way, as $-k_{\rm B}T \ln Z$. To take a specific example, consider again the paramagnet-ferromagnet transition, with a real, scalar order parameter field $\varphi(\mathbf{r})$ that represents the local magnetisation (which we take to have orientations only parallel or antiparallel to a preferred crystal axis). Time-reversal takes $\varphi(\mathbf{r})$ to $-\varphi(\mathbf{r})$, and we expect this and also spatial inversion to be symmetries. As a result, an expansion of the free energy density should contain only even powers of $\varphi(\mathbf{r})$ (apart from a linear coupling to an applied magnetic field h) and only even order derivatives of $\varphi(\mathbf{r})$. Hence we have

$$F(\varphi) = \frac{a}{2}\varphi^2 + \frac{b}{4}\varphi^4 + \frac{1}{2}|\nabla\varphi|^2 + \ldots - h\varphi \, . \tag{9.1}$$

What can we say about the coefficients a, b, ... that appear in this expansion? First, if it makes sense to truncate the expansion at order φ^4, we must have $b > 0$, so that $F(\varphi)$ is bounded below. (Alternatively, if $b < 0$, we would need to include a term in φ^6.) Secondly, we note that the minimum of $F(\varphi)$ is at $\varphi = 0$ for $a > 0$, and at non-zero φ for $a < 0$. This suggests that, to describe a phase transition, a should vary with temperature and change sign in the vicinity of the critical point. We therefore postulate that $a = At$, with t the reduced temperature and A an unimportant coefficient.

9.2.1 Mean field theory: the saddle-point approximation

Next, we should face the problem of evaluating the functional integral over field configurations. In general this is difficult and one can only hope to make progress by using approximations. The simplest approximation is a saddle-point one, which turns out to be equivalent to mean field theory. Starting from Eq. (9.1), we have from $\frac{\delta F}{\delta \varphi} = 0$ the saddlepoint equation

$$-\nabla^2\varphi(\mathbf{r}) + a\varphi(\mathbf{r}) + b\varphi^3(\mathbf{r}) - h = 0 \, . \tag{9.2}$$

Unless boundary conditions impose a spatially-varying solution, we expect φ to be independent of $\mathbf{r}$. Then with $h = 0$ Eq. (9.2) has the solutions

$$\varphi = 0 \qquad \text{and} \qquad \varphi = \pm\sqrt{\frac{-a}{b}} \, .$$

The first is the minimum of $F(\varphi)$ for $a > 0$ and a maximum for $a < 0$; the other solutions are real only for $a < 0$ when they are minima. From this we can conclude that the order parameter varies as

$$\varphi \propto |t|^{1/2} \qquad \text{for} \qquad t < 0 \, .$$

Thus the critical exponent β takes the value $\beta = 1/2$. At the critical point φ is non-zero only for non-zero h, with the dependence

$$\varphi \propto |h|^{1/3} .$$

From this we recognise the critical exponent value $\delta = 3$. To evaluate the susceptibility $\chi \equiv \partial\varphi/\partial h$ we should consider Eq. (9.2) for non-zero h. Rather than attempting to find the solution for φ explicitly as a function of h, it is more convenient to differentiate the equation directly, giving

$$(a + 3b\varphi^2)\frac{\partial\varphi}{\partial h} = 1$$

and hence

$$\chi = \begin{cases} \frac{1}{At} & t > 0 \\ -\frac{1}{2At} & t < 0 \end{cases}$$

Thus the exponent γ takes the value $\gamma = 1$.

9.2.2 Correlation function

To evaluate the correlation function $\langle\varphi(\mathbf{0})\varphi(\mathbf{r})\rangle$ we need to go beyond the saddle-point treatment. We will consider fluctuations within a Gaussian approximation, taking $a > 0$ for simplicity. Within this approximation, we would like to evaluate

$$\langle\varphi(\mathbf{0})\varphi(\mathbf{r})\rangle = \frac{\int D\varphi\ \varphi(\mathbf{0})\varphi(\mathbf{r})\ \mathrm{e}^{-\int \mathrm{d}^d\mathbf{r}\frac{a}{2}\varphi^2 + \frac{1}{2}|\nabla\varphi|^2}}{\int D\varphi\ \mathrm{e}^{-\int \mathrm{d}^d\mathbf{r}\frac{a}{2}\varphi^2 + \frac{1}{2}|\nabla\varphi|^2}} .$$

We do so by diagonalising the quadratic form in the argument of the exponentials: since our system is translation-invariant, this is done by Fourier transform. To be explicit about the Fourier transforms, we consider a system of linear size L in each direction with periodic boundary conditions. Then we can introduce wavevectors $\mathbf{k} = \frac{2\pi}{L}(l_1, \ldots l_d)$ with l_i's integer, and write

$$\varphi(\mathbf{r}) = \frac{1}{L^{d/2}}\sum_{\mathbf{k}} \varphi_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \qquad \text{and} \qquad \varphi_{\mathbf{k}} = \frac{1}{L^{d/2}}\int \mathrm{d}^d\mathbf{r}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} .$$

In these terms we have

$$\int \mathrm{d}^d\mathbf{r}\ \varphi^2(\mathbf{r}) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}\varphi_{-\mathbf{k}} \qquad \text{and} \qquad \int \mathrm{d}^d\mathbf{r}\ |\nabla\varphi(\mathbf{r})|^2 = \sum_{\mathbf{k}} k^2 \varphi_{\mathbf{k}}\varphi_{-\mathbf{k}} .$$

Also, we note from the definition of $\varphi_{\mathbf{k}}$ that $\varphi^*_{\mathbf{k}} = \varphi_{-\mathbf{k}}$, so that, as independent quantities we can take the real and imaginary parts of $\varphi_{\mathbf{k}}$ for one half of all wavevectors – say those with component $k_1 > 0$. This means we can write the functional integration as multiple integrals over just these components, with

$$\int D\varphi = \prod_{k_1>0} \int \mathrm{d}\mathrm{Re}\varphi_{\mathbf{k}} \int \mathrm{d}\mathrm{Im}\varphi_{\mathbf{k}} .$$

This leads us to the result

$$\langle\varphi_{\mathbf{k}}\varphi_{\mathbf{q}}\rangle = \delta_{\mathbf{k}+\mathbf{q},\mathbf{0}}\ \frac{1}{(a+k^2)} ,$$

and hence

$$\langle\varphi(\mathbf{0})\varphi(\mathbf{r})\rangle = \frac{1}{L^d}\sum_{\mathbf{k}} \frac{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}}{(a+k^2)} \sim \frac{1}{(2\pi)^d}\int \mathrm{d}^d\mathbf{k}\frac{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}}{(a+k^2)}$$

For general d one finds that this correlation function falls off on a scale set by the correlation length ξ given by $\xi^{-2} = a$. In $d = 3$ the integral gives

$$\langle\varphi(\mathbf{0})\varphi(\mathbf{r})\rangle = \frac{1}{2\pi r}\mathrm{e}^{-r/\xi} .$$

Since we have made the connection $\xi \propto t^{-1/2}$, we have obtained the value for another critical exponent: $\nu = 1/2$.

9.2.3 Consequences of symmetry

Without symmetry to exclude odd powers of the order parameter, we would have an expression for the free energy density of the form

$$F(\varphi) = \frac{a}{2}\varphi^2 + \frac{c}{3}\varphi^3 + \frac{b}{4}\varphi^4 + \frac{1}{2}|\nabla\varphi|^2 + \dots \ , \tag{9.3}$$

where we have omitted a linear term by choosing to expand around a minimum, but must allow all higher powers. In this case, the transition is generically a discontinuous one, as is seen most easily by considering the graphs of $F(\varphi)$ shown in Fig. 9.5.

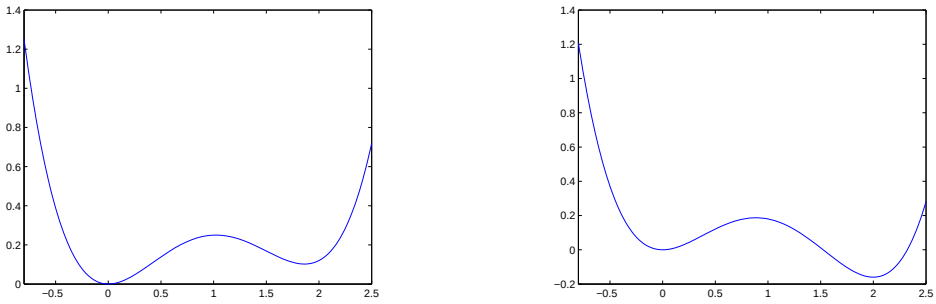

Figure 9.5: Sketch of Eq. (9.3). Left: for a large and positive. Right: for a small and positive. The minimum of $F(\varphi)$ jumps discontinuously from $\varphi = 0$ to a finite value of φ as a is reduced, representing a first-order transition.

9.2.4 Other examples of phase transitions

It is instructive to examine how one identifies an order parameter and constructs the continuum theory appropriate for other examples of phase transitions.

Liquid crystals: the isotropic-nematic transition

Liquid crystals are fluids of rod-like molecules. As fluids, the positions of the molecules are not ordered, and in the isotropic phase their orientations are also disordered. In the nematic phase, however, molecules spontaneously align along a common axis. To describe the transition between these two states we should identify a suitable order parameter and write down a free energy expansion using the symmetry of the isotropic phases to guide us. Suppose $\hat{n}$ is a unit vector aligned with a molecule. It is tempting to use the average $\langle\hat{n}\rangle$ itself as an order parameter, as we might if it were magnetisation. For the liquid crystal, however, this would not be correct, since the molecules are invariant under inversion, which takes $\hat{n}$ to $-\hat{n}$. To construct something from $\hat{n}$ that has the same invariance under inversion as the molecule itself, we first consider the tensor $n_i n_j$. This is still not quite ideal, since it is non-zero for an isotropically distributed $\hat{n}$ (taking the value $\langle n_i n_j\rangle = \frac{1}{3}\delta_{ij}$). Our final choice for an order parameter is therefore the tensor

$$Q_{ij} = \langle n_i n_j\rangle - \frac{1}{3}\delta_{ij} \ .$$

We expect it to vanish in the isotropic phase. By contrast, if molecules are fully aligned, say along the $\hat{z}$ direction, then

$$\mathbf{Q} = \begin{pmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix} \ .$$

The isotropic phase is symmetric under rotations of the molecules, with translates to rotations of the tensor $\mathbf{Q}$. Thus the free energy density $F(\mathbf{Q})$ should be invariant $\mathbf{Q} \to \mathbf{R}\mathbf{Q}\mathbf{R}^{-1}$, a property of traces of all powers of $\mathbf{Q}$. We therefore expect

$$F(\mathbf{Q}) = \frac{a}{2}\mathrm{Tr}[\mathbf{Q}^2] + \frac{c}{3}\mathrm{Tr}[\mathbf{Q}^3] + \frac{b}{4}\mathrm{Tr}[\mathbf{Q}^4] + \dots \ .$$

A significant feature of this expression is that it contains a cubic term. As a consequence, we expect the isotropic-nematic transition to be first-order, as it indeed is experimentally. It is a significant achievement to have reached an understanding of the nature of this transition, using only its symmetries.

The superfluid transition

To describe the transition (in, for example, liquid ^{4}He) between the normal and superfluid states, we use the condensate wavefunction amplitude ψ as an order parameter. It is a complex scalar, and $F(\psi)$ should be invariant under changes in its phase: $\psi \to \mathrm{e}^{\mathrm{i}\theta}\psi$, with θ real. Thus we have

$$F(\psi) = \frac{a}{2}|\psi|^2 + \frac{b}{4}|\psi|^4 + \frac{1}{2}|\nabla\psi|^2 + \dots \ .$$

The superconducting transition

The transition between the normal and superconducting states of a metal can also be described using a complex scalar order parameter that represents a condensate amplitude. In this case, however, the condensate is charged, and this means that we should consider how our expression for free energy varies under gauge transformations. Following arguments similar to those developed in Chapter 4, we arrive (with some suggestive notation for constants that are in fact phenomenological) at what is known as Landau-Ginzburg theory:

$$F(\psi) = \frac{a}{2}|\psi|^2 + \frac{b}{4}|\psi|^4 + \frac{1}{2m}\psi^*(-\mathrm{i}\hbar\nabla - q\mathbf{A})^2\psi + \dots \ .$$

9.3 Consequences of fluctuations

It is important to understand to what extent our treatment of these free energy densities, using mean field theory in the form of a saddle-point approximation to the functional integral, is correct. In fact, the approach can go wrong at two different levels, and behaviour depends on the dimensionality d of the system. The most dramatic possibility, which applies in low dimensions, is that fluctuations are so strong the system is disordered at any non-zero temperature. A less acute failing would be that there is a transition, but with critical behaviour that is not captured by mean field theory. Two borderline values of dimension, known as the lower and upper critical dimensions (d_l and d_u) separate these alternatives: for $d < d_l$, there is no finite-temperature phase transition; for $d_l < d < d_u$ there is a transition but critical behaviour is not well-described by mean field theory; and for $d_u < d$, critical properties follow mean field predictions.

9.3.1 The lower critical dimension

We have already seen from our discussion of statistical mechanics in one dimension (see Chapter 3) that one-dimensional systems with short range interactions do not have spontaneously broken symmetry at non-zero temperature. For the one-dimensional Ising model we arrived at this result both by an exact calculation, using transfer matrices, and more intuitively, by considering the energy and entropy associated with kinks or domain walls separating regions of parallel spins. We now examine the counterpart to this argument in two dimensions.

Systems with discrete broken symmetry: the Peierls argument

Consider a low-temperature configuration of a two-dimensional Ising model, as sketched in Fig. 9.6. We would like to find how frequent are domains of reversed spins, by estimating the free energy associated with the domain walls or boundaries separating spins of opposite orientation. Since with ferromagnetic exchange interactions of strength J, a pair of parallel spins has energy $-J$, and an antiparallel pair an energy $+J$, it costs an energy $2JL$ to introduce a domain wall of length L into the ground state. Such a domain wall may have many configurations. Counting these configurations exactly is a difficult exercise since domain walls should not intersect themselves, but we can make a reasonable estimate by considering a random walk on the square lattice that is constrained at each step not to turn back on itself. These means that three possibilities are open at each step (continuing straight ahead, turning left, or turning right), and the total number of configurations after L steps (ignoring the requirement for the domain wall to close on itself) is 3^L, implying an entropy of $k_\mathrm{B}L\ln 3$. The free energy of the domain wall is hence

$$F_L = L(2J - k_\mathrm{B}T\ln 3) \ .$$

Crucially, this is positive for $T < 2J/(k_\mathrm{B}\ln 3)$, and so domain walls are rare and long-range order is stable at low temperature.

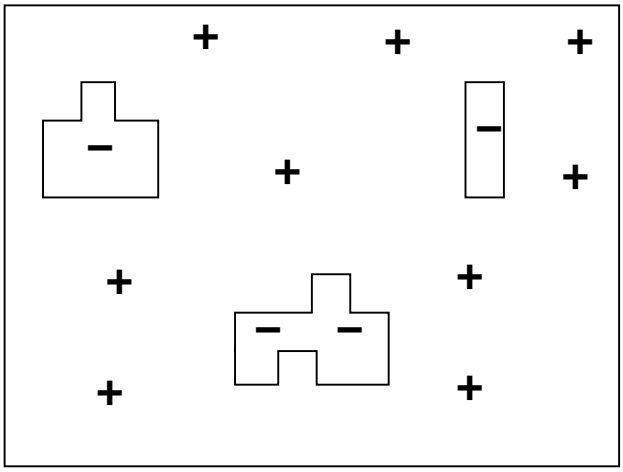

Figure 9.6: Typical low-temperature configuration of a two-dimensional Ising model. The system has positive magnetisation, but domains of spins with the opposite orientation appear as thermal excitations.

Systems with continuous broken symmetry: Goldstone modes

Our conclusions about order in the Ising model in fact also apply to other systems in which the broken symmetry is discrete. Systems with a continuous broken symmetry behave differently, however: their low energy excitations are long-wavelength Goldstone modes, which are more effective at disrupting long-range order than are sharp domain walls in the Ising model. To understand the effect of Goldstone modes on long-range order, we should calculate the correlation function for fluctuations $\varphi(\mathbf{r})$ in the order parameter. A calculation similar to the one set out above in Section 9.2.2 leads to the result

$$\langle \varphi(\mathbf{0}\varphi(\mathbf{r})\rangle \sim \frac{k_{\mathrm{B}}T}{(2\pi)^d J}\int \mathrm{d}^d\mathbf{k}\frac{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}}{k^2}\,. \tag{9.4}$$

If the notion that the system has long range order is to be self-consistent, fluctuations should be small. In particular a divergence in $\langle\varphi^2(\mathbf{0}\rangle$ would signal an instability. The integral on the right-hand side of Eq. (9.4) is divergent at small k in one and two dimensions, and this indicates the absence of spontaneously broken symmetry at non-zero temperature in one and two-dimensional systems. (The integral is divergent at large k is two and more dimensions, but this divergence does not have the same physical significance because in a condensed matter system there is always an upper limit to wavevectors, set by the inverse atomic spacing.)

Summarising, we have found that the value of the lower critical dimension is $d_l = 1$ for systems with discrete symmetry, and $d_l = 2$ for systems with continuous symmetry.

9.3.2 The upper critical dimension

For systems which are above their lower critical dimension and so have an ordering transition, we can ask whether it is reasonable to neglect fluctuations, by comparing the amplitude of fluctuations in the order parameter with its mean value. The approach leads to what is known as the Ginzburg criterion for the validity of mean field theory. Specifically, in a system with an order parameter field $\varphi(\mathbf{r})$, we compare its mean square fluctuation $\langle[\varphi(\mathbf{r}) - \langle\varphi\rangle]^2\rangle_\xi$, averaged over a region of linear size set by the correlation length, with $\langle\varphi\rangle^2$, the square of the order parameter itself. We have

$$\begin{aligned}\langle[\varphi(\mathbf{r}) - \langle\varphi\rangle]^2\rangle_\xi &= \xi^{-2d}\int_{\xi^d}\mathrm{d}^d\mathbf{r}\int_{\xi^d}\mathrm{d}^d\mathbf{r}'\,[\langle\varphi(\mathbf{r})\varphi(\mathbf{r}')-\langle\varphi\rangle^2]\\ &= \xi^{-d}\int\mathrm{d}^d\mathbf{r}\,[\langle\varphi(\mathbf{0})\varphi(\mathbf{r})-\langle\varphi\rangle^2]\\ &= \xi^{-d}\chi\,.\end{aligned} \tag{9.5}$$

The last equality in Eq. (9.5) follows from the definition of susceptibility, which gives

$$\begin{aligned}\chi = \left.\frac{\partial}{\partial h}\right|_{h=0} \langle\varphi(\mathbf{0})\rangle &= \left.\frac{\partial}{\partial h}\right|_{h=0} \frac{\int D\varphi\, \varphi(\mathbf{0})\, \mathrm{e}^{-F+h\int\varphi(\mathbf{r})}}{\int D\varphi\, \mathrm{e}^{-F+h\int\varphi(\mathbf{r})}} \\ &= \int \mathrm{d}^d\mathbf{r}\, [\langle\varphi(\mathbf{0})\varphi(\mathbf{r})\rangle - \langle\varphi(\mathbf{0})\rangle\langle\varphi(\mathbf{r})\rangle] \ .\end{aligned}$$

We are now in a position to express both the mean square fluctuations and the order parameter in terms of the reduced temperature t and the critical exponents. We have $\xi^{-d}\chi \sim |t|^{d\nu-\gamma}$ and $\langle\varphi\rangle^2 \sim |t|^{2\beta}$. If fluctuations close to the critical point are to be small compared to the mean, we require

$$\xi^{-d}\chi \sim |t|^{d\nu-\gamma} \ll \langle\varphi\rangle^2 \sim |t|^{2\beta}$$

for $|t|$ small. This is the case if $\nu d - \gamma > 2\beta$, which is to say

$$d > \frac{2\beta+\gamma}{\nu} = 4 \ ,$$

where the ratio of exponents has been evaluated using the results of mean field theory presented earlier in this chapter. Our conclusion, then, is that while mean field theory provides a qualitatively correct treatment of phase transitions, for systems in three dimensions it is not quantitatively accurate to neglect fluctuations. In fact, accurate calculations of exponent values for systems in fewer than four dimensions require a more serious treatment of interactions in field theory, using renormalisation group methods.

9.4 Further Reading

- J. M. Yeomans, *Statistical Mechanics of Phase Transitions* (OUP). Chapters 1 and 4 provide a straightforward introduction to the material covered in these lectures.
- P. M. Chaiken and T. C. Lubensky, *Principles of Condensed Matter Physics* (CUP). Chapters 1, 3 and 4 cover similar material to that in these lectures.
- N. Goldenfeld, *Lectures on Phase Transitions and The Renormalization Group* (Addison Wesley) Chapter 5 covers the same material as these lectures. The book also offers a readable introduction to more advanced ideas.