19

Instantons and Solitons

In this Chapter we discuss the role of topology in Quantum Field Theory. Topology plays different roles in quantum field theory. One is in identifying classes of configurations of the Euclidean path integral that are topologically inequivalent. Such configurations are called *instantons* and describe *topological excitations* which play a key role in the non-perturbative definition of the path integral both in theories with global symmetries and in gauge theories, and in the theory of phase transitions. A closely related problem are *solitons*, classical configurations (and quantum states) with finite energy and non-trivial topology. Instantons and solitons are classified by a set of topological invariants, known as topological charges. When these topological invariants are included in the action of the theory (often as a consequence of quantum anomalies), they change the behavior of the theory in profound ways.

There are many reasons for considering topological excitations . One motivation is the study of mechanisms for quantum disorder, such as the physical origin of phases of a quantum field theory exhibiting confinement in the case of a gauge theory, and/or lack of long-range order in theories with global symmetries. This is also related to the problem of quantum tunneling processes in quantum field theory. In addition, this analysis leads to an understanding of the existence of topological excitations both in quantum field theory and in statistical physics. We will begin by defining what we mean by a topological excitation.

19.1 Instantons in Quantum Mechanics and Tunneling

Consider first a very simple problem, the quantum mechanics of a double well anharmonic oscillator. The potential $U(q)$ for the coordinate q has the form shown in Fig.19.1. The Lagrangian is

$$L = \frac{1}{2}\left(\frac{dq}{dt}\right)^2 - U(q) \tag{19.1}$$

where we set the mass $m = 1$. The potential is

$$U(q) = -\frac{\mu^2}{2}q^2 + \frac{\lambda}{4}q^4 \tag{19.2}$$

with λ small. This potential has two classical minima at $\pm q_0 = \pm\sqrt{\frac{\mu^2}{\lambda}}$, respectively. At the classical level, the symmetry $q \to -q$ is spontaneously broken. We will see that this symmetry is restored by tunneling processes involving instantons. Here we will follow closely the work by Polyakov (Polyakov, 1977), and the beautiful lectures by Coleman (Coleman, 1985).

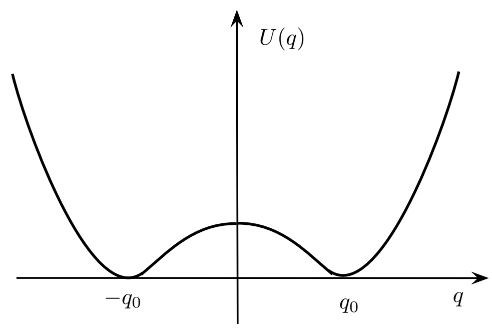

Figure 19.1 The potential $U(q)$ of a double-well anharmonic oscillator.

Suppose we want to compute the (imaginary time) time ordered correlator $C(\tau) = \langle q(0)q(\tau)\rangle$. By expanding in eigenstates the correlator becomes

$$\langle q(0)q(\tau)\rangle = \sum_n |\langle 0|q|n\rangle|^2 e^{-(E_n-E_0)\tau} \tag{19.3}$$

At long imaginary time, $\tau \to \infty$, due to tunneling processes the correlator decays exponentially

$$C(\tau) \sim e^{-\tau\Delta E} \tag{19.4}$$

where ΔE is the splitting of the two low-lying states, the symmetric and antisymmetric combination of the states on either well. As usual the imaginary time correlator is computed by the expression

$$\langle q(0)q(\tau)\rangle = \frac{1}{Z}\int \mathcal{D}q(\tau)\; q(0)q(\tau)\; e^{-\mathcal{E}[q]} \tag{19.5}$$

where $\mathcal{E}[q]$ is the action in imaginary time

$$\mathcal{E}[q] = \int_{-\infty}^{\infty} d\tau \left[\frac{1}{2} \left(\frac{dq}{d\tau} \right)^2 - \frac{\mu^2}{2} q^2 + \frac{\lambda}{4} q^4 \right] \tag{19.6}$$

and Z is the partition function

$$Z = \int \mathcal{D}q(\tau)\, e^{-\mathcal{E}[q]} \tag{19.7}$$

hence, this problem is equivalent to the computation of the correlation function on a problem in classical statistical mechanics in one dimension, where $\mathcal{E}[q]$ is the classical energy.

We will do this calculation sing the semiclassical expansion in imaginary time. We first seek the extrema of the Euclidean action $\mathcal{E}[q]$, which satisfy $\delta\mathcal{E} = 0$, which are the solutions of the Euler-Lagrange equation

$$\frac{d}{d\tau} \frac{\delta \mathcal{E}}{\delta \dot{q}} = \frac{\delta \mathcal{E}}{\delta q} \tag{19.8}$$

whose solutions will be denoted by $\bar{q}(\tau)$. For this system, the Euler-Lagrange equations are

$$\frac{d^2 \bar{q}}{d\tau^2} = -\mu^2 \bar{q} + \lambda \bar{q}^3 \tag{19.9}$$

which are the same as the classical equations of motion of a dynamical system $q(\tau)$ with the inverted potential, $-U(q) = \frac{\mu^2}{q}^2 - \frac{\lambda}{4} q^4$, shown in Fig.19.2a.

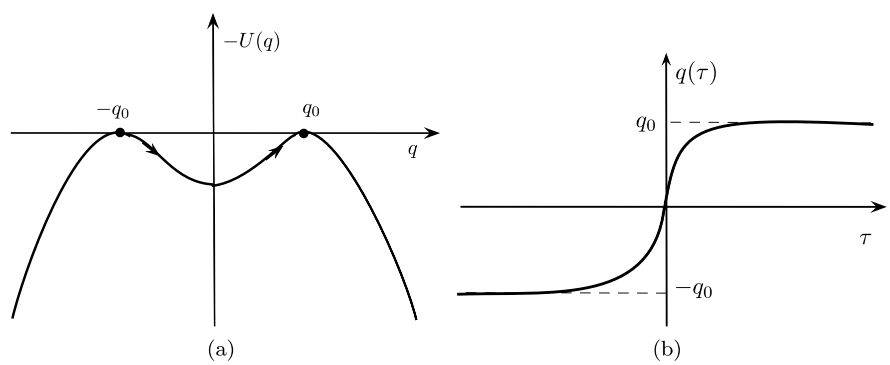

Figure 19.2 a) The inverted potential $-U(q)$, b) tunneling trajectory.

This system has a conserved "energy" E

$$E = \frac{1}{2} \left(\frac{d^2 \bar{q}}{d\tau^2} \right)^2 - U(\bar{q}) \tag{19.10}$$

which corresponds to the Lagrangian in real time.

The static solutions of the equation of motion, Eq(19.9), are the classical ground states, $\bar{q} = \pm\sqrt{\mu^2/\lambda}$, for which $\mathcal{E}(\bar{q}) = -\frac{\mu^4 T}{4\lambda}$, where T is the imaginary time span, and $T \to \infty$.

In addition, there is also a solution of Eq.(19.9) with $E = 0$, which obeys

$$\frac{d\bar{q}}{d\tau} = \pm\sqrt{2U[q]} \tag{19.11}$$

The solutions are

$$\bar{q}_c(\tau) = \pm\sqrt{\frac{\mu^2}{\lambda}}\tanh\left(\frac{\mu(\tau - a)}{\sqrt{2}}\right) \tag{19.12}$$

These solutions interpolate between q_0 and $-q_0$ (and viceversa) as τ goes from $-\infty$ to $+\infty$. These solutions represent the tunneling process. One such solution is depicted in Fig.19.2b. Their Euclidean action (or energy) is

$$\mathcal{E}[\bar{q}_c(\tau)] - \mathcal{E}[\pm q_0] = 2\sqrt{2}\frac{\mu^3}{\lambda} \tag{19.13}$$

which is finite.

Trajectories with finite Euclidean action are called *instantons*. This trajectory is topological in the sense that it interpolates smoothly between to inequivalent ground states (or vacua) at $\tau \to \pm\infty$, and that this trajectory cannot be smoothly deformed to the trivial configurations, at $\pm q_0$. In this particular problem, we see that we can either have an instanton, i.e. the trajectory that goes from $-q_0$ to $+q_0$, or an anti-instanton, that executes the reverse trajectory.

The instanton solution has an arbitrary parameter, a, and the Euclidean action does not depend on this parameter. Such parameters are called zero modes of instantons. This solution has a contribution to the partition function of the order of $\tau\exp(-2\sqrt{2}\mu^3/\lambda)$. This contribution becomes important at imaginary times long enough that $\tau > \exp(+2\sqrt{2}\mu^3/\lambda)$.

We will compute the correlator of Eq.(19.5) using the semiclassical expansion. This requires that we evaluate the contribution to the path integral presented in Eq.(19.5) at *each* classical solution. The full result is the sum of their individual contributions, separately to both the numerator and the denominator. As we will see below, the one-instanton approximation will fail at long times τ, and we will also need to include multi-instanton processes.

Let us compute first the contribution of one-instanton, Eq.(19.12). Since the instanton is a solution of the Euler-Lagrange equation, the first correction appears at quadratic order in the expansion of the imaginary time action $\mathcal{E}$. Thus, we need to examine the kernel

$$\left.\frac{\delta^2 \mathcal{E}}{\delta q(\tau)\delta q(\tau')}\right|_{\bar{q}_c(\tau)} = \left[-\frac{d^2}{d\tau^2} + (-\mu^2 + 3\lambda \bar{q}_c^2(\tau))\right]\delta(\tau - \tau') \tag{19.14}$$

where $\bar{q}_c(\tau)$ is the instanton solution of Eq.(19.12). Let $\{\psi_n(\tau)\}$ be a complete set of eigenstates of this operator

$$\left[-\frac{d^2}{d\tau^2} + (-\mu^2 + 3\lambda \bar{q}_c^2(\tau))\right]\psi_n(\tau) = \omega_n^2 \psi_n(\tau) \tag{19.15}$$

where $\{\omega_n^2\}$ is the spectrum of eigenvalues, and satisfy the boundary conditions $\psi_n(-\infty) = \psi_n(+\infty) = 0$. If the spectrum of eigenvalues of this equation were strictly positive, $\omega_n^2 > 0$, we could then use the eigenstates $\{\psi_n\}$ to parametrize the trajectories $q(\tau)$ and compute the path integral in the semiclassical approximation, as we have done in previous chapters.

However, in this case this is not possible since, although spectrum of the fluctuation kernel is non-negative, its lowest eigenstate has zero eigenvalue, i.e. it is an exact zero mode. The reason is that although the Euclidean action $\mathcal{E}$ is translationally invariant, the instanton solution is not, and has the parameter a. Since $\mathcal{E}[\bar{q}_c(\tau)]$ does not depend on a, then

$$\frac{\partial \mathcal{E}[\bar{q}_c(\tau)]}{\partial a} = 0 \tag{19.16}$$

Thus,

$$\frac{\delta}{\delta \bar{q}_c(\tau)} \frac{\partial \mathcal{E}[\bar{q}_c(\tau)]}{\partial a} = 0 \tag{19.17}$$

and

$$\frac{\delta^2 \mathcal{E}}{\delta \bar{q}_c(\tau)\delta \bar{q}_c(\tau')} \frac{d\bar{q}_c(\tau)}{da} = 0 \tag{19.18}$$

Therefore, we find that

$$\psi_0(\tau) = \frac{d\bar{q}_c(\tau)}{da} \tag{19.19}$$

is an exact eigenstate of the fluctuation kernel of Eq.(19.14) with eigenvalue $\omega_0^2 = 0$. In other words, it is an exact zero mode of the fluctuation kernel, and there is no restoring force along this direction in the space of functions.

The solution of this problem is to treat the parameter a as a collective coordinate of the solution, and it must be quantized exactly. We have dealt with a similar problem in Chapter 9 on the quantization of gauge theory.

There, we introduced the Faddeev-Popov gauge-fixing procedure. We will see that it is the way to solve our problem here as well.

The Euclidean action is invariant under translations, $q(\tau) \mapsto q_a(\tau) = q(\tau + a)$, $\mathcal{E}[q] = \mathcal{E}[q_a]$, and the integration measure is also invariant under translations, $\mathcal{D}q = \mathcal{D}q_a$. We will then follow the Faddeev-Popov procedure and define

$$1 = \int dF\ \delta(F[q_a]) = \int_{-\infty}^{+\infty} da\ \delta(F[q_a])\ \frac{\partial F}{\partial a} \tag{19.20}$$

which we insert into the expression of the partition function to obtain

$$\begin{aligned} Z &= \int \mathcal{D}q\ \exp(-\mathcal{E}[q]) \\ &= \int \mathcal{D}q \int_{-\infty}^{+\infty} da\ \delta(F[q_a])\ \frac{\partial F}{\partial a}\ \exp(-\mathcal{E}[q]) \end{aligned} \tag{19.21}$$

we now perform the change of variables, $q \to q_{-a}$, or $q(\tau) \to q(\tau - a)$, to find

$$Z = \int \mathcal{D}q_{-a} \int_{-\infty}^{+\infty} da\ \delta(F(q))\ D[q_{-a}, a]\ \exp(-\mathcal{E}[q_a]) \tag{19.22}$$

where we defined

$$D[q, a] = \frac{\partial F(a)}{\partial a} \tag{19.23}$$

which plays the role of the Faddeev-Popov determinant.

Using the translation invariance of the measure and of the action we obtain the result

$$Z = \int \mathcal{D}q \int_{-\infty}^{+\infty} da\ \delta(F(q))\ D[q_{-a}, a] \exp(-\mathcal{E}[q]) \tag{19.24}$$

We will now compute the Jacobian $D[q_{-a}, a]$ for the specific choice

$$F[q_a] \equiv \int_{-\infty}^{\infty} d\tau\ \frac{\partial \bar{q}_c}{\partial a}\bigg|_{a=0} \left(q(\tau + a) - \bar{q}_c(\tau)\big|_{a=0} \right) \tag{19.25}$$

Then,

$$\frac{\partial F}{\partial a}[q_a] = \int_{-\infty}^{\infty} d\tau\ \frac{\partial \bar{q}_c}{\partial a}\bigg|_{a=0} \frac{\partial q}{\partial a}(\tau + a) \tag{19.26}$$

Therefore,

$$D \equiv D[q_{-a}, a] = \int_{-\infty}^{\infty} d\tau\ \frac{\partial \bar{q}_c}{\partial a}\bigg|_{a=0} \frac{\partial q}{\partial a}(\tau)\bigg|_{a=0} \tag{19.27}$$

We will now proceed to parametrize the histories $q(\tau)$ as

$$q(\tau) = \bar{q}_c(\tau, a) + \sum_{n\neq 0} \xi_n \psi_n(\tau - a) \tag{19.28}$$

where $\{\psi_n(\tau)\}$ are the eigenstates of the fluctuation kernel with strictly positive eigenvalues, $\omega_n^2 > 0$, and $\bar{q}_c(\tau, a) = \bar{q}_c(\tau - a)$. Using that

$$\frac{\partial \bar{q}_c(\tau, a)}{\partial a} = -\frac{\partial \bar{q}_c(\tau)}{\partial \tau}, \qquad \frac{\partial \psi_n(\tau - a)}{\partial a} = -\frac{\partial \psi_n(\tau)}{\partial \tau} \tag{19.29}$$

we can write

$$\left.\frac{\partial q(\tau)}{\partial a}\right|_{a=0} = -\left[\frac{\partial \bar{q}_c(\tau)}{\partial \tau} + \sum_{n\neq 0} \xi_n \frac{\partial_n(\tau)}{\partial \partial \tau}\right]_{a=0} \tag{19.30}$$

Then, we find that the Jacobian D is

$$D = A + \sum_{n\neq 0} \xi_n r_n \tag{19.31}$$

Here we defined,

$$A = \int_{-\infty}^{\infty} d\tau \left(\frac{\partial \bar{q}_c}{\partial \tau}\right)^2, \qquad r_n = \int_{-\infty}^{\infty} d\tau \left.\frac{\partial \bar{q}_c}{\partial \tau}\right|_{a=0} \left.\frac{\partial \psi_n}{\partial \tau}\right|_{a=0} \tag{19.32}$$

With these results, the integration measure becomes

$$\mathcal{D}q(\tau) = da \left(A + \sum_{n\neq 0} \xi_n r_n\right) \prod_{n\neq 0} d\xi_n \simeq A\, da \prod_{n\neq 0} d\xi_n \tag{19.33}$$

where the approximate form is accurate in the limit $\lambda \ll \mu^3$, with A given in Eq.(19.32).

Therefore, to this level of approximation, the one-instanton contribution to the partition function is

$$\begin{aligned} Z_1 = &\int \mathcal{D}\tilde{q}(\tau) \int_{-\infty}^{\infty} da\, A\, \exp(-\mathcal{E}[\bar{q}_c]) \\ &\times \exp\left(-\frac{1}{2}\int d\tau \int d\tau' \left.\frac{\delta^2 \mathcal{E}}{\delta q(\tau)\delta q(\tau')}\right|_{\bar{q}_c(\tau)} \tilde{q}(\tau)\tilde{q}(\tau')\right) \end{aligned} \tag{19.34}$$

where

$$\tilde{q}(\tau) = q(\tau) - \bar{q}_c(\tau, a) = \sum_{n\neq 0} \xi_n \psi_n(\tau - a) \tag{19.35}$$

which only involves the eigenstates with non-vanishing eigenvalue. By doing the integrals over the non-zero modes, we find that the one-instanton contribution to the partition function is

$$Z_1 = A\left(\int_{-\infty}^{\infty} da\right) \exp(-\mathcal{E}[\bar{q}_c]) \prod_{n\neq 0} \omega_n^{-1} \tag{19.36}$$

while the contribution of the trivial saddle point, $\bar{q} = q_0$ to the partition function is

$$Z_0 = \prod_{n\neq 0} \omega_{n,0}^{-1} \tag{19.37}$$

where $\omega_{n,0}$ are the eigenvalues for the trivial saddle point.

We can similarly compute the contributions to the numerator of the correlation function $\langle q(0)q(\tau)\rangle$, Eq.(19.5). Putting it all together, we find that, up to multi-instanton contributions that will be discussed shortly, the correlator is given by

$$\begin{aligned}\langle q(0)q(\tau)\rangle =& \frac{\frac{\mu^2}{\lambda} + A\,K\,\left(\int_{-\infty}^{\infty} da\; \bar{q}_c(0,a)\bar{q}_c(\tau,a)\right)\exp(-\mathcal{E}[\bar{q}_c(\tau)])}{1 + A\,K\,\left(\int_{-\infty}^{\infty} da\right)\exp(-\mathcal{E}[\bar{q}_c]}\\ \simeq& \frac{\mu^2}{\lambda} + A\exp(-\mathcal{E}[\bar{q}_c])\;K\;\int_{-\infty}^{\infty} da\,\left(\bar{q}_c(0,a)\bar{q}_c(\tau,a) - \frac{\mu^2}{\lambda}\right)\end{aligned} \tag{19.38}$$

where K is the ratio

$$K = \frac{\prod_{n\neq 0}\omega_n^{-1}}{\prod_{n\neq 0}\omega_{n,0}^{-1}} = \frac{\text{Det}'\left[-\frac{d^2}{d\tau^2} + (-\mu^2 + 3\lambda\bar{q}_c^2(\tau))\right]^{-1/2}}{\text{Det}'\left[-\frac{d^2}{d\tau^2} + 4\mu^2\right]^{-1/2}} \tag{19.39}$$

where Det' denotes the determinant without the zero modes. The ratio of determinants K can be calculated using different methods, e.g see Section 5.2.1. In particular, the numerator of K is the determinant for the Schrödinger operator for the Pöschl-Teller potential, and the denominator for the linear harmonic oscillator.

It remains to compute A given in eq.(19.32), to find

$$A = \mathcal{E}[\bar{q}_c] = 2\sqrt{2}\frac{\mu^3}{\lambda} \tag{19.40}$$

and

$$\int_{-\infty}^{\infty} da\,\left[\bar{q}_c(0,a)\bar{q}_c(\tau,a) - \frac{\mu^2}{\lambda}\right] = -\frac{2\mu^2}{\lambda}\frac{\tau}{\tanh(\mu\tau/\sqrt{2})} \simeq -\frac{2\mu^2}{\lambda}\tau \tag{19.41}$$

where used the approximate form for long times, $\tau \gg \sqrt{2}/\mu$.

Therefore, at this level of approximation, the correlator at long times $\tau \gg \sqrt{2}/\mu$ becomes

$$\langle q(0)q(\tau)\rangle = \frac{\mu^2}{\lambda} + KA\exp(-\mathcal{E}[\bar{q}_c])\frac{2\mu^2}{\lambda}\tau \tag{19.42}$$

Thus, we find that, as expected, the instanton contribution is exponentially small. However, as anticipated, it fails at sufficiently long times τ

$$\tau \gtrsim \exp(\mathcal{E}[\bar{q}_c])/(2KA) \tag{19.43}$$

where $\mathcal{E}[\bar{q}_c] = 2\sqrt{2}\mu^3/\lambda$.

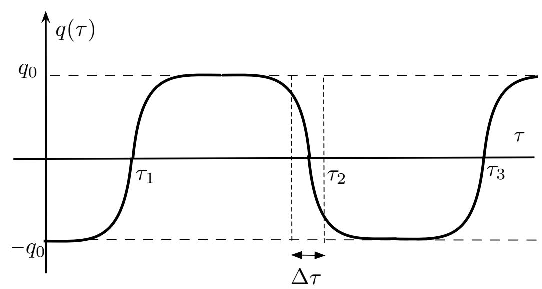

Figure 19.3 A multi-instanton process as a sequence of instantons and anti-instantons. Here τ_1, τ_2, τ_3, etc., are the location of an infinite sequence of instantons and anti-instantons in imaginary time τ. Here $\Delta\tau \sim 1/\mu$ is the width of the instanton.

We will now see that the solution of this problem is to include multi-instanton processes, which will amount to show that the series, whose first two terms are given by Eq.(19.42), exponentiates. If we note that if λ is small, the failure of the approximation occurs at times much longer than the width of an instanton $\Delta\tau \sim 1/\mu$, we see that we should be able to regard the sum over multi-instanton processes as a dilute gas of (exponentially) weakly interacting instantons. Let is denote a multi-instanton process by

$$\bar{q}_c(\tau) \simeq \sqrt{\frac{\mu^2}{\lambda}} \prod_{j=1}^{N} \text{sign}(\tau - a_j) \tag{19.44}$$

where $\{a_j\}$ are the locations of the instantons (and anti-instantons). The Euclidean action of an N instanton configuration is

$$\mathcal{E}_c^N \simeq N\frac{2\sqrt{2}\mu^3}{\lambda} \tag{19.45}$$

plus exponentially weak interaction terms.

In this limit, the problem becomes essentially identical to the low-temperature expansion of the classical one-dimensional Ising model. Indeed, we can picture the instantons as the domain walls of the Ising model. In this interpretation, the restoration of the $\mathbb{Z}_2$ symmetry $q \to -q$, is just the Landau-Peierls proof of the absence of spontaneous symmetry breaking in one-dimensional classical statistical mechanics in systems with a discrete global symmetry.

We can now do this calculation explicitly:

$$\langle q(0)q(\tau)\rangle = \frac{1}{Z}\frac{\mu^2}{\lambda}\sum_{N=0}^{\infty} C(\tau)^N \exp(-2\sqrt{2}\mu^3/\lambda) \times \int_{a_1<a_2<\ldots<a_N} da_1 \ldots da_N \prod_{j=1}^{N} \mathrm{sign}(\tau - a_j) \tag{19.46}$$

where

$$Z = \sum_{N=0}^{\infty} C(\tau)^N \exp(-2\sqrt{2}\mu^3/\lambda) \int_{a_1<a_2<\ldots<a_N} da_1 \ldots da_N \tag{19.47}$$

and

$$C(\tau) = \frac{2AK\tau}{\tanh(\mu\tau/\sqrt{2})} \simeq 2AK\tau \tag{19.48}$$

where we used the behavior at $\tau \gg 1/\mu$. Also

$$\int_{a_1<a_2<\ldots<a_N} da_1 \ldots da_N = \frac{T^N}{N!} \tag{19.49}$$

where $T \to \infty$ is the total span in imaginary time. Upon doing the sums, we find the result

$$\langle q(0)q(\tau)\rangle = \frac{\mu^2}{\lambda}\exp(-(\Delta E)\tau) \tag{19.50}$$

where

$$\Delta E \simeq 2AK \exp(-2\sqrt{2}\mu^3/\lambda) \tag{19.51}$$

Thus, the ground state and the first excited state of the double well are split by an amount ΔE, which has an essential singularity in λ, showing its non-perturbative character.

This rather elaborate presentation is the prototype of an instanton calculation. It is a sum of a large number of contributions each being exponentially small, and with a weight that has an essential singularity in the coupling constant, λ in the case at hand. This approach will succeed in theories for which the Euclidean action of the instanton is finite, and the instanton solution is labeled only by the location of the instanton. We will see below that we can relax the finite action condition up to a logarithmically divergence. However, in many cases of interest, such in theories which are classically scale-invariant, the instanton has zero modes associated with its scale, which will complicate the analysis.

19.2 Solitons in 1+1-dimensional ϕ^4 theory

Similarly, we can look at ϕ^4 theory in 1+1-dimensional Minkowski space-time. The Lagrangian density is

$$\mathcal{L} = \frac{1}{2}\left(\partial_t \phi\right)^2 - \frac{1}{2}\left(\partial_x \phi\right)^2 - U(\phi) \tag{19.52}$$

whose classical Hamiltonian density is

$$\mathcal{H} = \frac{1}{2}\Pi^2 + \frac{1}{2}\left(\partial_x \phi\right)^2 + U(\phi) \tag{19.53}$$

We can now *finite energy* static classical configurations which extremize the Hamiltonian $\mathcal{H}$. These extremal static configurations obey the same equations in the coordinate x as the instanton solutions of the double-well problem in imaginary time, that we just discussed. These finite energy classical solutions are known as *solitons*. The soliton solutions obey *boundary conditions* corresponding to distinct uniform classical solutions. In other words, the theory has a global $\mathbb{Z}_2$ symmetry and the soliton configuration interpolates between the two classically degenerate states at $\pm\phi_0$. The can be viewed as a domain wall in one space dimension.

Since we are interested in the broken symmetry state, it will be convenient to use the potential

$$U(\phi) = \frac{\lambda}{4!}(\phi_0^2 - \phi(x)^2)^2 \tag{19.54}$$

for which $m_0^2 = -\lambda\phi_0^2/12 < 0$. Its global minima are at $\pm\phi_0$ and their energy is zero. The finite energy solution $\phi(x)$ that extremizes the Hamiltonian, and interpolates between the two global minimum of the energy, obeys the same equation in the x coordinate as the instanton of the double-well potential in imaginary time τ. The classical soliton solution is

$$\phi(x) = \pm\phi_0 \tanh\left(\frac{(x - x_0)}{\sqrt{2}\xi}\right) \tag{19.55}$$

where $\xi = \left(\frac{\lambda\phi_0^2}{6}\right)^{-1/2} = |m_0|^{-1}$ is the correlation length. The energy of the classical soliton is

$$E_{\text{soliton}} = \frac{2}{3\sqrt{3}}\phi_0^3\sqrt{\lambda} \tag{19.56}$$

Furthermore, it is straightforward to show that there is also a boosted soliton solution of the form $\phi(x \pm vt)$ representing a soliton that moves at speed $v < 1$ (in units of the speed of light that we have set $c = 1$). Thus, these classical soliton solutions can be regarded as particles that have finite energy.

It is possible to promote these classical soliton solutions to a quantum state using semiclassical quantization. While in quantum mechanics the semiclassical approximation is WKB, the equivalent (in a broad sense) in quantum field theory is substantially more subtle and technical. This approach is not only viable but yields the exact answer in a class of special theories that have an infinite number of conservation laws, and are quantum integrable theories. All known theories of this type are in 1+1 spacetime dimensions. In higher dimensions, the only known theories that are exact at the semiclassical level involve supersymmetry. We will not pursues this problem here.

It is easy to realize that the soliton of ϕ^4 theory is the analog of the kink that we found in the context on the quantum one-dimensional Ising model in chapter 14. Such kinks (or soliton) solutions are domain walls separating regions with distinct broken symmetry states. Solutions of this type exist in the spontaneously broken phase of any theory in one space dimension with a discrete global symmetry group.

In the next sections we will generalize the concept of instant and soliton to other theories.

19.3 Vortices

Let us consider now theories with global continuous symmetries. For simplicity, we will discuss first the case in which the group is $U(1)$. In this case, the matter field is a complex scalar field $\phi(x) \in \mathbb{C}$. We will be interested in the phase in which the global $U(1)$ symmetry is spontaneously broken at the classical level. Deep in this phase, we can approximate the complex scalar field by a complex field of fixed modulus, i.e. a $U(1)$ non-linear sigma-model, and write

$$\phi(x) = \phi_0 \, e^{i\theta(x)} \tag{19.57}$$

where ϕ_0 is the vacuum expectation value in the broken symmetry phase, and $\theta(x) \in [0, 2\pi)$ is a (compact) phase field, the Goldstone boson of the spontaneously broken $U(1)$ symmetry.

The configurations of systems with compact symmetry groups (global or local) are classified by homotopy groups. This subject can be quite formal and heavily mathematical. We will work at a more physical and intuitive level, at the price of mathematical rigor. There are many excellent reviews and textbooks on the subject, such as the book by Nash and Sen (Nash and Sen, 1983), the book by Coleman (Coleman, 1985), and Polyakov's book (Polyakov, 1987). There are two outstanding reviews, one by Mermin (Mermin, 1979), which focuses on topological defects in ordered matter, and by Eguchi, Gilkey and Hanson (Eguchi et al., 1980), geared mostly towards gauge theory.

19.3.1 Topology in $d = 1$ dimension

Let us consider first the case of a $U(1)$ soliton in $D = 1 + 1$ spacetime dimensions. We will work with periodic boundary conditions in space. Hence, we have effectively compactified (wrapped) the spatial line coordinate onto a circumference of radius $R = L/(2\pi)$, where L is the total length. Thus, we are assuming that the spacetime manifold is $S^1 \times \mathbb{R}$, where the circle S^1 is space and $\mathbb{R}$ is time. The analysis that we will do now also applies to a quantum system, i.e. a field theory in 0+1 dimensions, with a $U(1)$ symmetry, such as a planar rigid rotor. In this context, the tunneling process from state $|0\rangle$ to state $|2\pi\rangle$ is described by an instanton.

As in the ϕ^4 soliton, we will seek a finite energy static solution of the classical equations of motion. A classical static configuration of the field $\phi(x)$ is a mapping (an assignment) of every point of the spatial manifold, which we will refer to as the *base space*, to a value of the phase θ, the *target space*. Since the phase $\theta(x)$ is defined mod 2π, the target space is topologically isomorphic to a circumference S^1. Therefore, in this case, the static classical configurations are maps of the S^1 base pace onto the S^1 target space (as shown in Fig.19.4), i.e.

$$\phi : S^1 \longrightarrow S^1 \tag{19.58}$$

Since the phases $\theta(x)$ are additive (mod 2π), smooth configurations can also be added. Hence, under addition (composition), smooth mappings of S^1 onto S^1 form a group. An example is the constant field configuration which is a mapping of the entire base space $x \in S^1$ to one particular point of the target space S^1, labeled by the value θ_0 of the phase of the field. On the other hand, there are many configurations that can be obtained by a smooth deformation of the constant field configuration. In terms of the mapping, these configurations can regarded as closed "curves" on the S^1 target space that begin and end at the same value θ_0 of the phase. Such configurations can be trivially deformed back onto the constant field configuration, θ_0. We will say that such configurations are "contractible" to each other.

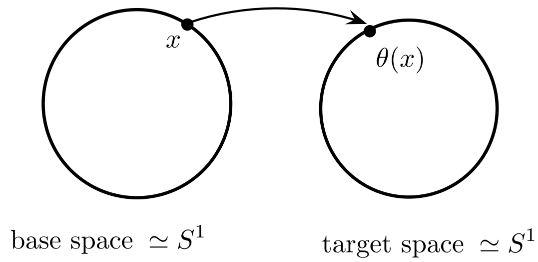

Figure 19.4 A configuration of a complex field as a mapping.

In topology, smooth mappings are called *homotopies.* Two configurations (mappings) can be deformed smoothly into each other we will say that they are *homotopic* to each other. In other words, configurations (mappings) that are homotopic to each are, in a topological sense, equivalent. Therefore, the operation of smooth deformation of mappings defines an equivalence relation between mappings, and we will say that mappings that can be deformed smoothly into each other belong to the same *equivalence class.* For example, in Fig.19.5 four configurations, labeled a, b, c and d, respectively are shown. Configurations a and b can be smoothly deformed into each other, and, hence, are homotopic to each other. So are configurations c and d. However, configurations a and b are not homotopic to configurations c and d, since a and b are periodic on $[0, L)$, but c and d jump by 2π at L, which is invisible for the complex field ϕ

The existence of an equivalence relation implies that the mappings can be classified. The question now is, how many distinct equivalence classes are there of mappings of S^1 base onto S^1 target? For any smooth mapping $\theta(x)$ we can define an integer-valued quantity known as the *winding number*, defined as the total change of the phase field accross S^1 base (in units of 2π),

$$N = \frac{(\Delta\theta)_L}{2\pi} = \frac{1}{2\pi}\int_0^L dx\ \partial_x\theta(x) = \frac{1}{2\pi}\int_0^L dx\ e^{i\theta(x)} i\partial_x e^{-i\theta(x)} \qquad (19.59)$$

where we used the notation $(\Delta\theta)_L = \theta(L) - \theta(0)$. Since the configurations are required to obey periodic boundary conditions mod 2π on S^1, i.e. $\theta(x+L) = \theta(x) + 2\pi k$, with $k \in \mathbb{Z}$, it follows that the winding number N is an integer.

On the other hand, under any smooth deformation of a configuration $\theta(x)$, the value of the winding number cannot be changed continuously since it is an integer. In this sense, the winding number is a *topological invariant*, which classifies the configurations into classes. We will summarize this statement by

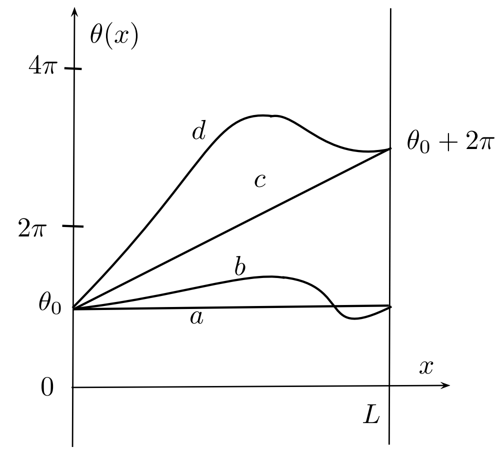

Figure 19.5 Homotopies of $S^1 \to S^1$: a) a constant configuration, b) a configuration that is homotopic to a, c) a configuration with winding number $+1$, and d) a configuration homotopic to c.

saying that the homotopy classes of mappings of S^1 to S^1 are isomorphic the the integers. This statement means that the homotopy classes for a group, known as the *homotopy group*, in this case $\pi_1(S^1)$. Here the subindex denotes that the base is S^1 and the argument is the target. We will represent this statement by the equation

$$\pi_1(S^1) \cong \mathbb{Z} \tag{19.60}$$

It is now natural to ask if (and when) an extension of this analysis always holds other continuous symmetries and to other dimensions. The general answer to this question is no. To see this, let us consider a theory in 1+1 spacetime dimensions, with a classically spontaneously broken $O(3)$ global symmetry, such as the non-linear sigma model, and ask if it has classical static soliton solutions.

Since the target space of the $O(3)$ non-linear sigma model is the two-sphere, S^2, the configurations are now mappings of the S^1 base space to the S^2 target space. Therefore, we can represent all configurations as a closed curve Γ on the two-sphere S^2, shown in Fig.19.6. However, all closed smooth curves on S^2 are contractible, meaning that they can be smoothly deformed to a point, such as the arbitrary point $\boldsymbol{n}_0$ shown in the same figure, and consequently all configurations are topologically trivial. The same fact holds for mappings of S^1 to the n-sphere, S^n (with $n > 1$) and are all topologically trivial. We express this statement by saying that their homotopy groups are

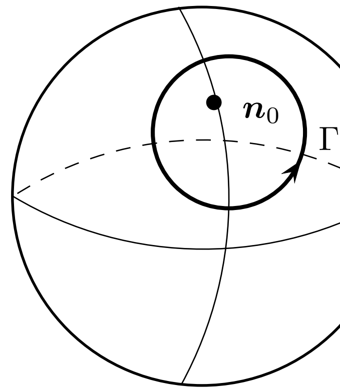

Figure 19.6 The homotopies of S^1 into S^2 are topologically trivial.

trivial

$$\pi_1(S^n) = 0 \tag{19.61}$$

Therefore, we conclude that in 1+1 dimensions $O(n)$ non-linear sigma models, with $n > 1$, do not have topologically non-trivial classical static soliton solutions

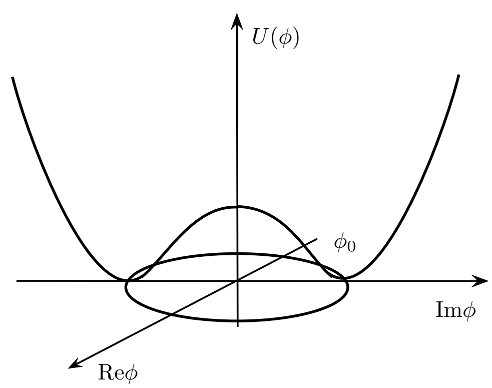

Figure 19.7 The potential $U(\phi)$ in the broken symmetry phase.

19.3.2 Vortices in $D = 2$ dimensions

The preceding analysis does not imply that that the topology is necessarily trivial in $d > 1$ dimensions, far from it! To see how non-trivial topology arises in $d > 1$, we will consider the case of a complex field $\phi(x)$ with a spontaneously broken global $U(1)$ symmetry in $d = 2$ dimensions. Again, $d = 2$ here will be either the dimension of an Euclidean spacetime, in which case the configurations with non-trivial topology are instantons, or the space dimension of a 2+1 dimensional spacetime and the configurations are solitons.

19.3.3 The complex scalar field

For definiteness we will consider a theory whose Euclidean action in $D = 2$ spacetime dimensions (or energy in $d = 2$ space dimensions) is

$$S = \int d^2x \left(\frac{1}{2} \left|\partial_\mu \phi\right|^2 + U(\phi) \right) \tag{19.62}$$

where the potential $U(\phi)$ has the form shown in Fig.19.7, e.g. $U(\phi) = u(|\phi|^2 - \phi_0^2)^2$, with $u > 0$. It is invariant under the global $U(1)$ symmetry, $\phi(x) \to \exp(i\varphi)\phi(x)$ and has a minimum at $|\phi(x)| = \phi_0$.

We will parametrize the complex field in terms of two real fields, the amplitude $\rho(x)$ and the phase $\theta(x)$, such that

$$\phi(x) = \rho(x) \exp(i\theta(x)) \tag{19.63}$$

Since the $U(1)$ symmetry is broken we will assume that, at large distances, the amplitude $\rho(x)$ approaches the limit

$$\lim_{|x|\to\infty} \rho(x) = \phi_0 \tag{19.64}$$

where, as before, ϕ_0 is the vacuum expectation value in the broken symmetry phase.

Let us now consider a circumference $C(R)$ of radius R, large enough so that we can approximate $\rho(x) \simeq \phi_0$, with any desired accuracy. While the amplitude $\rho(x)$ is essentially fixed to the value ϕ_0 for large enough R, the phase $\theta(x)$ may change on $C(R)$. In this limit, the complex field takes the asymptotic form $\phi(x) \mapsto \phi_0 \exp(i\theta(x))$ where $x \in C(R)$. Therefore, the value of the field $\phi(x)$ *on the circumference* $C(R)$ defines a map of S^1 base space, the points on the circumference $C(R)$ labeled by the azimuthal angle $\varphi \in [0, 2\pi)$, to the S^1 target space, the values of the phase θ of the field ϕ. From our previous analysis, we see that these mappings can also be classified by the homotopy group $\pi_1(S_1) \cong \mathbb{Z}$, with each class being classified by a *winding number* called the *vorticity*,

$$n = \frac{(\Delta\theta)_C}{2\pi} = \int_0^{2\pi} \frac{d\varphi}{2\pi} e^{i\theta(\varphi)} i\partial_\varphi e^{-i\theta(\varphi)} \tag{19.65}$$

where $(\Delta\theta)_C$ is the total change of the phase of the field ϕ on one turn on the circumference $C(R)$.

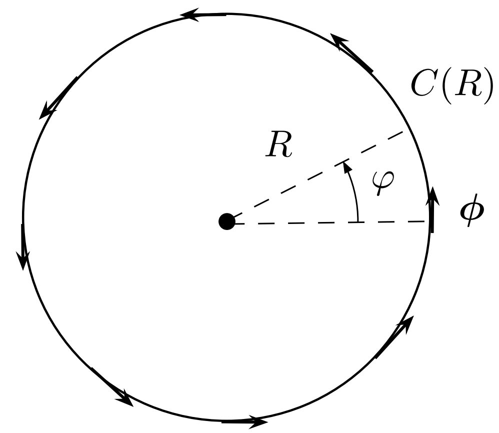

Figure 19.8 A vortex in $d = 2$ dimensions. Here R is the radius of the large circumference $C(R)$. The phase θ of the complex field $\phi(x)$ winds as the vector $\boldsymbol{\phi} = (\text{Re}\phi, \text{Im}\phi)$ rotates around the circumference by the azimuthal angle φ.

The configuration defined by this topology is called a *vortex*. Clearly, if the configuration $\phi(x)$ has a non-vanishing vorticity N_v, the field itself must vanish somewhere in the interior of the circle whose boundary is $C(R)$, say at the center, since otherwise this configuration will have a singularity. In addition, the amplitude $\rho(r)$ must vanish fast enough as $r \to 0$ so that the gradient terms in the action are small enough that the Euclidean action either does not diverge, or diverges at most logarithmically.

To see how this works we will assume that the potential $U(\phi)$ is sufficiently steep that the amplitude $\rho \simeq \phi_0$ essentially for all values of x except inside a small core or radius a, where it vanishes fast enough that there is no short distance singularity, at the price of a finite contribution to the action cut off by the size of the core. Thus, for all values of points outside the core, $|x| > a$, we can set $\phi(x) = \phi_0 \exp(i\theta(x))$.

What is the action of a configuration of vortices of this type? Outside the core of the vortex the gradient of the field simply becomes

$$\partial_\mu \phi(x) \simeq \phi_0 e^{i\theta(x)}\, i\partial_\mu \theta(x) \;\Rightarrow\; \left|\partial_\mu \phi\right|^2 \simeq \phi_0^2 \left(\partial_\mu \theta(x)\right)^2 \tag{19.66}$$

and the Euclidean action of the configuration is

$$S = \frac{\phi_0^2}{2} \int d^2x \left(\partial_\mu \theta(x)\right)^2 \tag{19.67}$$

A configuration of the field $\phi(x)$ has a local current density j_μ which, in the approximations that we are using becomes $j_\mu = \phi_0 \partial_\mu \theta(x)$. By analogy with a fluid, we define the *vorticity* $\omega(x)$ as the curl of the current,

$$\omega(x) = \epsilon_{\mu\nu} \partial_\mu j_\nu(x) = \phi_0 \epsilon_{\mu\nu} \partial_\mu \partial_\nu \theta(x) \tag{19.68}$$

which vanishes everywhere except at the locations x_j of the vortices, i.e. the singularities of the phase field $\theta(x)$. Let us assume that we have vortices at a set of points $\{x_j\}$ with vorticities $\{n_j\}$ so that the local vorticity is

$$\omega(x) = \sum_j n_j \delta^2(x - x_j) \tag{19.69}$$

where the phase field $\theta(x)$ satisfies Eq.(19.68). We will see now that $\theta(x)$ is given by the distribution of vortices as

$$\theta(x) = \sum_j n_j \mathrm{Im} \ln(z - z_j) \tag{19.70}$$

where $z = x_1 + ix_2$ are complex coordinates of the plane. Hence, the phase field is multivalued and has a branch cut ending at each vortex.

To see how this comes about we define $\vartheta(x)$ as the (Cauchy-Riemann) dual of the phase field $\theta(x)$, i.e.

$$\partial_\mu \vartheta(x) = \epsilon_{\mu\nu} \partial_\nu \theta(x) \tag{19.71}$$

which satisfies the Poisson Equation

$$-\partial^2 \vartheta(x) = \omega(x) \tag{19.72}$$

In terms of the field $\vartheta(x)$ the action takes the "2D electrostatic" form

$$S = \frac{\phi_0^2}{2} \int d^2x \, (\partial_\mu \vartheta)^2 = -\frac{\phi_0^2}{2} \int d^2x \, \vartheta \, \partial^2 \vartheta = \frac{\phi_0^2}{2} \int d^2x \, \omega(x) \vartheta(x) \tag{19.73}$$

Solving Poisson's Equation, Eq.(19.72),

$$\vartheta(x) = \int d^2y \, G(x - y) \omega(y) \tag{19.74}$$

in terms of the two-dimensional Green function $G(x - y)$, which satisfies

$$-\partial^2 G(x - y) = \delta^2(x - y) \tag{19.75}$$

whose solution is

$$G(x - y) = \int \frac{d^D p}{(2\pi)^D} \frac{e^{ip\cdot(x-y)}}{p^2} = \frac{\Gamma\left(\frac{D}{2} - 1\right)}{4\pi^D |x - y|^{D-2}} \tag{19.76}$$

We will define

$$G(0) = \lim_{a \to 0} G(a) \tag{19.77}$$

in terns of which, as $D \to 2$, we obtain

$$G(|x-y|) - G(a) = \frac{1}{2\pi} \ln\left(\frac{a}{|x-y|}\right) \tag{19.78}$$

which diverges (logarithmically) as $|x-y| \to \infty$.

Then, we can write

$$\begin{aligned} S =& \frac{\phi_0^2}{2} \int d^2x \, \omega(x)\vartheta(x) \\ =& \frac{\phi_0^2}{2} \int d^2x \int d^2y \, \omega(x) \, G(x-y) \, \omega(y) \\ =& \frac{\phi_0^2}{2} \sum_{i,j} n_i n_j G(x_i - x_j) \\ =& \frac{\phi_0^2}{2} \Big(\sum_j n_j \Big)^2 G(0) + \phi_0^2 \sum_{i>j} n_i n_j \left[G(x_i - x_j) - G(0) \right] \end{aligned} \tag{19.79}$$

Since $G(0)$ diverges as $a \to 0$, the action will diverge unless the *total vorticity* of the configuration vanishes, $\sum_j n_j = 0$.

Thus, the action of a collection of vortices (with zero total vorticity) is

$$S[n] = \frac{\phi_0^2}{2\pi} \sum_{i>j} n_i n_j \ln\left(\frac{a}{|x_i - x_j|}\right) \tag{19.80}$$

In particular, for a vortex-antivortex pair, with $n_1 = -n_2 = 1$, separated at a distance $R \gg a$, is $S[1,-1;R] = \left(\frac{\phi_0^2}{2\pi}\right) \ln(R/a)$. Therefore, the Euclidean action of a set of vortices is the same as the energy of a set of classical electrical charges with a logarithmic interaction.

In effect, we are rewriting the full partition function as

$$Z \propto \sum_{\{n_j\}} \prod_j \int d^2x_j \; \delta\Big(\sum_j n_j \Big) \exp(-S[n]) \tag{19.81}$$

where x_j are the coordinates of vortices and anti-vortices. Here, we neglected an uninteresting prefactor. In other words, we have mapped the problem to the thermodynamics of a neutral two-dimensional Coulomb gas at temperature $T = \pi/\phi_0^2$.

We will see below that, even though these vortices have a logarithmically divergent action. Although this is a violation of our criterion that instantons must have finite action, they play a key role on this theory. We will show below that, precisely due to the logarithmic interaction, we will be able recast the partition function of the theory as a partition function of vortices and anti-vortices.

To see how this may work, we will estimate the free energy cost of a vortex. The free energy is $F = U - TS$ where U is the energy, S is the entropy, and $T = \pi/\phi_0^2$. An argument due to Kosterlitz and Thouless (Kosterlitz and Thouless, 1973) estimates the contribution to the free energy of a single vortex with $n = 1$ to be

$$F_{\text{vortex}} = \ln(a/L) - T\ln(L/a)^2 \tag{19.82}$$

where L is the linear size of the system. The first term is the logarithmically divergent self-energy of a single vortex (cutoff by L). The second term is the entropy of a vortex which counts the logarithm of the number of places where vortex can be located. For ϕ_0^2 large enough, the energy wins over the entropy, $F_{\text{vortex}} > 0$, and free vortices are suppressed. On the other hand, vortex-anti-vortex pairs, i.e. dipoles with vanishing total vorticity, have a finite energy and logarithmic entropy. Thus, in this regime there will be a finite density of such dipoles.

However, for T large enough, or what is the same ϕ_0^2 small enough, the entropy will win over the entropy, $F_{\text{vortex}} < 0$, and the system becomes unstable against the proliferation of vortices (and anti-vortices). Hence, there should be a phase transition between a regime in which vortices and anti-vortices can only occur in bound pairs, to another state in which vortices and anti-vortices unbind and proliferate, leading to a state best described as a neutral plasma. This simple argument predicts that the critical temperature is $T_c = \frac{\pi}{2}$ (in units of $1/\phi_0^2$). In another section we will provide a more precise version of this Kosterlitz-Thouless transition.

19.3.4 The Abelian-Higgs model

We will now consider the theory of a complex scalar field minimally coupled coupled to a gauge field with a Maxwell action. In this theory, the abelian-Higgs model, or equivalently a superconductor coupled to a gauge field, the symmetry is local. This will result in a finite action for the vortex. To this end we consider the Euclidean Lagrangian

$$\mathcal{L} = \frac{1}{2}\left|D_\mu\phi\right|^2 + u(\phi_0^2 - |\phi|^2)^2 + \frac{1}{4}F_{\mu\nu}^2 \tag{19.83}$$

where u is the coupling constant, $D_\mu = \partial_\mu - ieA_\mu$ is the covariant derivative (with $\hbar = c = 1$), and $j_\mu = i(\phi^* D_\mu\phi - (D_\mu\phi)^*\phi)$ is the current.

We will seek a vortex solution of the same form as before, $\phi \to \phi_0 \exp(i\theta)$. However, now for the action to be finite we will require that $|D_\mu \phi| \to 0$ as $r \to \infty$. In this limit, the covariant derivative becomes

$$D_\mu \phi \to i(\partial_\mu \theta + eA_\mu)\phi_0 \ \exp(i\theta) \tag{19.84}$$

Therefore, as $r \to \infty$,

$$|D_\mu \phi|^2 \to 0 \Leftrightarrow \partial_\mu \theta - eA_\mu \to 0 \tag{19.85}$$

and the gauge field becomes, asymptotically, a pure gauge. Nevertheless, on an arbitrary closed contour Γ that encloses a vortex, the phase of the matter field has a winding number (the vorticity)

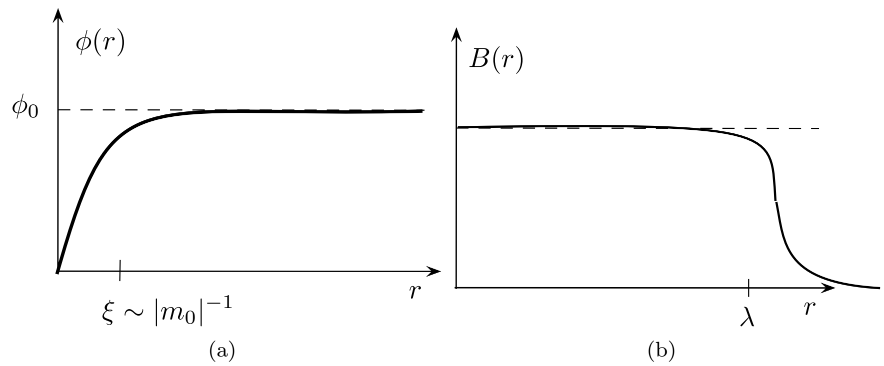

Figure 19.9 Schematic behavior of the vortex solution: a) Configuration of the amplitude of the complex scalar field, and b) of the magnetic field $B(x) = \epsilon_{\mu\nu}\partial_\mu A_\nu(x)$. Here $\xi = m_\phi^{-1}$ is the correlation length and $\lambda = m_v^{-1}$ is the penetration depth.

$$n = \frac{(\Delta\theta)_\Gamma}{2\pi} = \frac{e}{2\pi}\oint_\Gamma dx_\mu A_\mu(x) \tag{19.86}$$

However, using Stokes Theorem,

$$\oint_\Gamma dx_\mu A_\mu(x) = \int_\Sigma dS_\mu B_\mu = \Phi \tag{19.87}$$

where Σ is the region of the plane with boundary Γ, $B(x)$ is the magnetic field, and Φ is the magnetic flux through Σ. Therefore, we find that there is a relation between the vorticity and the flux

$$n = \frac{\Phi}{\phi_0} \tag{19.88}$$

where, upon restoring conventional units, $\phi_0 = 2\pi\hbar c/e$ is the flux quantum (not to be confused with the asymptotic value of the field!) and e is the electric charge. A configuration that obeys this flux quantization condition is known as an Abrikosov vortex (Abrikosov, 1957) , also known as the Nielsen-Olsen vortex (Nielsen and Olesen, 1973).

Since the gauge field becomes a pure gauge at infinity, the magnetic field must vanish as infinity, $B(x) \to 0$ are $r \to \infty$. Similarly, the complex scalar field must vanish as $r \to 0$, and $B(x)$ remains finite as $r \to 0$.

The vortex solution, of topological charge n, of the Euclidean equations of motion, obey the asymptotic conditions

$$\lim_{r\to\infty} \phi(r,\varphi) = \phi_1, \qquad \lim_{r\to\infty} A_i(r,\varphi) = n\partial_i\varphi \tag{19.89}$$

and

$$\epsilon_{ij}F_{ij} = 4\pi n\delta^2(\boldsymbol{x}), \qquad \frac{1}{4\pi}\oint dx_i A_i = n \in \mathbb{Z} \tag{19.90}$$

Here and below (r,φ) are the polar coordinates of the plane. The solutions can be found using the ansatz

$$\phi(x_1,x_2) = f(r)g_n(\varphi), \qquad A_i = -ia(r)g_n^{-1}(\varphi)\partial_i g_n(\varphi) \tag{19.91}$$

and obey the boundary conditions

$$f(0) = a(0) = 0, \qquad \lim_{r\to\infty} f(r) = \phi_0, \lim_{r\to\infty} a(r) = 1 \tag{19.92}$$

These equations do not have any explicit analytic solutions and, in general, are solved numerically. The asymptotic behavior at large distances, $r \to \infty$, of the vortex solutions is

$$a(r) = -\frac{1}{e} + m_v r K_1(m_v r), \qquad f(r) = \phi_0 + O(\exp(-m_\phi r) \tag{19.93}$$

where $K_1(z)$ is the Bessel function, and the masses m_v for the gauge field, and m_ϕ for the complex scalar field are, respectively

$$m_v = e\phi_0, \qquad m_\phi = 2\sqrt{2u}\phi_0 \tag{19.94}$$

The two scales, m_ϕ and m_v, are, respectively, the inverses of the correlation length $\xi \sim M_\phi^{-1}$ of the scalar field and of the penetration depth $\lambda \sim m_v^{-1}$ (here we are using the terminology of superconductivity) of the gauge field, see Fig. 19.9. It is well known form the theory of superconductivity that, if $\xi > \lambda$ (a type-I superconductor), vortices with topological charge of the *same sign* have an attractive interaction, and, conversely, in the type II regime, $\xi < \lambda$, they repel each other. The case of a neutral complex scalar field (i.e. without a dynamical gauge field) corresponds to the limit $\lambda \to \infty$, where the vortices repel each other, as was shown above. It can be shown that the

Euclidean action for a vortex solution with topological charge n obeys the Bogomol'nyi bound $S_n \geq \pi |n| \phi_0^2$.

At the special value $m_v = m_\phi$, or equivalently $e^2 = 8u$, the crossover point between type I and type II, the vortices do not interact with each other, at least classically. At this special point, the Bogomol'nyi point, the vortex equations can be reduces to two first order partial differential equations, which have a similar structure to the instanton self-dual equations that we will discuss below. In the field theory literature, these equations are known as the BPS equations, for Bogomol'nyi, Prasad and Sommerfield (Bogomol'nyi, 1976; Prasad and Sommerfield, 1975). The BPS solutions saturate the Bogomol'nyi bound, $S_n^{BPS} = \pi |n| \phi_0^2$.

The effects of instantons in the two-dimensional abelian Higgs model has been studied in detail (Callan, Jr. et al., 1976), at the level of the effects of semiclassical fluctuations, including the computation of the fluctuation determinants for vortex solutions (Schaposnik, 1978). The result is that the the partition function of the model can be expressed in terms of the partition function of a dilute gas of vortices and anti-vortices, of the form

$$Z = \sum_{\{n_j\}} \frac{z^N}{N!} \prod_j \int d^2 x_j \exp\left(-\frac{2\pi m_v^2}{e^2} \sum_{i>j} n_i n_j K_0(m_v |\boldsymbol{x}_i - \boldsymbol{x}_j|) \right) \tag{19.95}$$

where $N = \sum_j n_j$ is the total vorticity, z is a fugacity that accounts for effects of short-distance singularities, and $K_0(m_v|x|)$ is the Bessel function.

This expression differs from the one we found in the case of the complex scalar field in several important ways. One is that the interaction is proportional to the Bessel function $K_0(m_v|x|)$. This interaction decays exponentially fast at distances longer than the penetration depth $\lambda \sim 1/m_v$, and crosses over to a logarithmic interaction at short distances. Hence, the self-energy of a vortex is finite, instead of being logarithmically divergent. As a result, the free energy of a single vortex is always negative, $F_{\text{vortex}} < 0$, signaling a proliferation instability.

Consequently, the Kosterlitz-Thouless argument now implies that the vortices (and anti-vortices) always proliferate, and vortices and anti-vortices are always in a plasma phase. We will see below that in this state in which the instantons proliferate, external test charges are confined. Thus, this example provides a scenario for how instantons can provide a mechanism for confinement. However, it is important to stress that this works because the only zero modes of the instantons (the vortices) are their locations. We will see that in classically scale invariant theories with instantons have are additional zero modes related to their scale, which will complicate the analysis. The important result is that instantons lead to an area law for a Wilson loop carrying a charge that is a fraction of the charge of the the Higgs field, and hence confinement of these static external sources, while Wilson loops with the charge of the Higgs field have a perimeter law. We will see in chapter 18 that a lattice version of this theory show that this theory is confining even in the latter case since the only allowed states are neutral bound states.

19.4 Instantons and solitons of non-linear sigma models

19.4.1 The $O(3)$ non-linear sigma model

We will now consider the $O(3)$ non-linear sigma model in two dimensions and show that it has instanton solutions. The Euclidean action of the non-linear sigma model is

$$S = \int d^2x \, \frac{1}{2}(\partial_\mu \boldsymbol{n})^2, \quad \text{with, } \boldsymbol{n}^2 = 1 \tag{19.96}$$

and its partition function is

$$Z = \int \mathcal{D}\boldsymbol{n} \, \exp(-\frac{1}{g}S[\boldsymbol{n}]) \tag{19.97}$$

where g is the coupling constant. Here $\boldsymbol{n}$ has three real components and obeys the constraint $\boldsymbol{n}^2 = 1$.

In order to derive the Euclidean equations of motion we will implement the constraint using a Lagrange multiplier field $\lambda(x)$, in terms of which the action becomes

$$S = \int d^2x \, \frac{1}{2}\Big[(\partial_\mu \boldsymbol{n})^2 - \lambda(x)(\boldsymbol{n}(x)^2 - 1)\Big] \tag{19.98}$$

The Euler-Lagrange Equations are

$$\frac{\delta S}{\delta n_a(x)} - \partial_\mu \frac{\delta S}{\delta \partial_\mu n_a(x)} = 0, \qquad \frac{\delta S}{\delta \lambda(x)} = 0 \tag{19.99}$$

The second equation simply implies that the constraint $\boldsymbol{n}^2 = 1$ is obeyed everywhere, while the first equation yields the condition

$$-\partial^2 n_a(x) = \lambda(x) n_a(x) \tag{19.100}$$

Upon taking the inner product with the field $\boldsymbol{n}(x)$, the last equation becomes

$$\lambda(x) = -\boldsymbol{n} \cdot \partial^2 \boldsymbol{n} \tag{19.101}$$

where we used the constraint. Upon plugging in this result in the first equation of Eq.(19.99), the equation of motion becomes

$$\partial^2 n^a(x) = n^a(x)\boldsymbol{n}(x)\cdot\partial^2\boldsymbol{n}(x) \tag{19.102}$$

which are non-linear and non-trivial.

To extremize (minimize) the Euclidean action, a smooth configuration $\boldsymbol{n}(x)$, which satisfies the constraint $\boldsymbol{n}^1 = 1$, must be such that

$$\lim_{r\to 0} r||\partial_\mu \boldsymbol{n}||^2 = 0 \tag{19.103}$$

in order for the action to be finite. Hence, we will have to require that the field $\boldsymbol{n}(x)$ must approach a fixed (but arbitrary) value $\boldsymbol{n}_0$ as $r\to\infty$. Thus, the requirement that Euclidean action is finite implies that, for the allowed configurations, the base space, the plane $\mathbb{R}^2$, has been compactified to a sphere S^2, at least in a topological sense. This can be done, for instance, using the stereographic projection shown in Fig.19.10.

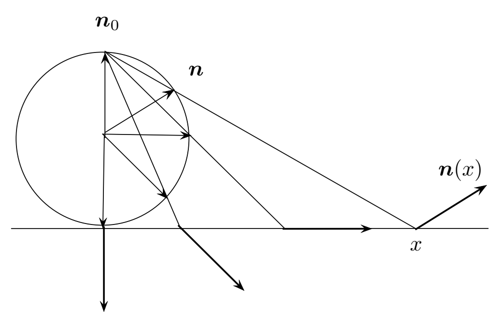

Figure 19.10 The stereographic projection of a field configuration.

On the other hand, the target space of the $O(3)$ non-linear sigma model is also the two sphere S^2. Thus, the finite action solutions are smooth maps

$$\boldsymbol{n}: S^2 \mapsto S^2 \tag{19.104}$$

We will now see that these maps are classified by the homotopy group $\pi_2(S^2)$, and that the classes are labeled by an integer $\pi_2(S^2)\cong\mathbb{Z}$, which we will call the *topological charge* Q. Let Q be defined by the expression

$$Q = \frac{1}{8\pi}\int_{S^2} d^2x\, \epsilon_{\mu\nu}\boldsymbol{n}(x)\cdot\partial_\mu\boldsymbol{n}(x)\times\partial_\nu\boldsymbol{n}(x) \tag{19.105}$$

We will show that Q is a integer, a topological invariant that labels the topological class.

Let ξ_1 and ξ_2 be the two Euler angles of the S^2 target sphere. The infinitesimal oriented area element sS_a^{target} (shown in Fig.19.11) is given by

$$d\boldsymbol{S}^{\text{target}} = \frac{1}{2}\left(\frac{\partial \boldsymbol{n}}{\partial \xi_1} \times \frac{\partial n}{\partial \xi_2} - \frac{\partial n}{\partial \xi_2} \times \frac{\partial n}{\partial \xi_1}\right) \tag{19.106}$$

Then, we can write Q as follows

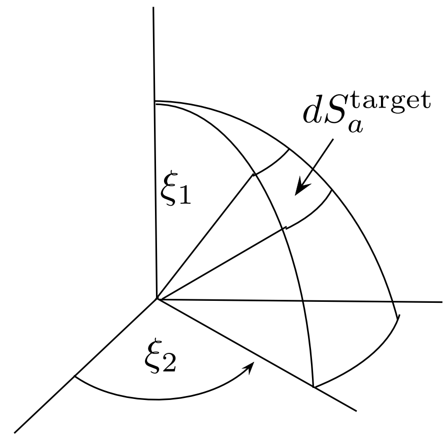

Figure 19.11 The infinitesimal oriented area element of S^2.

$$Q = \frac{1}{8\pi}\int_{S^2_{\text{base}}} d^2x\ \epsilon_{\mu\nu}\epsilon_{abc} n_a(x) \frac{\partial n_b}{\partial x_\mu}\frac{\partial n_c}{\partial x_\nu} \tag{19.107}$$

The configuration $\boldsymbol{n}(x)$ maps a point x of S^2_{base} to a point ξ of S^2_{target}. Regarded as a change of variables, the mapping induces the change in the measure (a Jacobian)

$$\epsilon_{rs} d^2\xi = \epsilon_{\mu\nu}\frac{\partial \xi_r}{\partial x_\mu}\frac{\partial \xi_s}{\partial x_\nu} \tag{19.108}$$

Upon performing the change of variables, Q becomes

$$\begin{aligned} Q =& \frac{1}{8\pi}\int_{S^2_{\text{target}}} d^2\xi\ \epsilon_{rs}\epsilon_{abc} n_a \frac{\partial n_b}{\partial \xi_r}\frac{\partial n_c}{\partial \xi_s} \\ =& \frac{1}{4\pi}\int_{S^2_{\text{target}}} d\boldsymbol{S}_{\text{target}} \cdot \boldsymbol{n} \\ =& \frac{1}{4\pi}\int_{S^2_{\text{target}}} |dS_{\text{target}}| \end{aligned} \tag{19.109}$$

where we used the fact that $\boldsymbol{n}$ and $d\boldsymbol{S}_{\text{target}}$ are parallel. However, the last integral is just the area of S^2, which is equal to 4π.

Therefore, Q is an integer that counts how many times is the two-sphere S^2_{target} is swept as the configuration $\boldsymbol{n}(x)$ spans the entire compactified plane, the base sphere S^2_{base}. This result does not change if the configuration $\boldsymbol{n}(x)$ is changed smoothly. We conclude that $Q \in \mathbb{Z}$ is a topological invariant of the class of mappings, and that the mappings are classified by the homotopy group

$$\pi_2(S^2) \cong \mathbb{Z} \tag{19.110}$$

We will now show that the topological charge (or, rather, its absolute value) places a lower bound to the Euclidean action of a configuration of the field $\boldsymbol{n}$ in a given topological class, labeled by Q. To see this, we consider the trivial identity

$$\left(\partial_\mu \boldsymbol{n} \pm \boldsymbol{n} \times \partial_\nu \boldsymbol{n}\right)^2 \geq 0 \tag{19.111}$$

However, since

$$(\boldsymbol{n} \times \partial_\nu \boldsymbol{n})^2 = \boldsymbol{n}^2 (\partial_\nu \boldsymbol{n})^2 - (\boldsymbol{n} \cdot \partial_\nu \boldsymbol{n})^2 = (\partial_\nu \boldsymbol{n})^2 \tag{19.112}$$

and $\boldsymbol{n}^2$ and $\boldsymbol{n} \cdot \partial_\mu \boldsymbol{n} = 0$, it follows that

$$(\partial_\mu \boldsymbol{n} \pm \epsilon_{\mu\nu} \boldsymbol{n} \times \partial_\nu \boldsymbol{n})^2 = 2(\partial_\mu \boldsymbol{n})^2 \pm 2\epsilon_{\mu\nu} \boldsymbol{n} \cdot \partial_\mu \boldsymbol{n} \times \partial_\nu \boldsymbol{n} \geq 0 \tag{19.113}$$

which implies the bound

$$(\partial_\mu \boldsymbol{n})^2 \geq \epsilon_{\mu\nu} \boldsymbol{n} \cdot \partial_\mu \boldsymbol{n} \times \partial_\nu \boldsymbol{n} \tag{19.114}$$

Hence, the action $S[\boldsymbol{n}]$ of a configuration of the fields has a lower bound

$$S[\boldsymbol{n}] = \frac{1}{2} \int d^2x \, (\partial_\mu \boldsymbol{n})^2 \geq \frac{1}{2} \int d^2x \, \epsilon_{\mu\nu} \boldsymbol{n} \cdot \partial_\mu \boldsymbol{n} \times \partial_\nu \boldsymbol{n} \tag{19.115}$$

which is to say

$$S[\boldsymbol{n}] \geq 4\pi |Q| \tag{19.116}$$

There is a class of configurations $\boldsymbol{n}$ that saturate the bound, $S[\boldsymbol{n}] = 4\pi|Q|$. To do that, they must satisfy the identity

$$(\partial_\mu \boldsymbol{n} \pm \epsilon_{\mu\nu} \boldsymbol{n} \times \partial_\nu \boldsymbol{n})^2 = 0 \tag{19.117}$$

which implies that these special configurations obey the self dual (and anti-self-dual) equation

$$\partial_\mu \boldsymbol{n} = \pm \epsilon_{\mu\nu} \boldsymbol{n} \times \partial_\nu \boldsymbol{n} \tag{19.118}$$

together with the constraint $\boldsymbol{n}^2 = 1$. The solutions of the self-dual equation, Eq.(19.118) are the instantons (and anti-instantons) of the 1+1-dimensional non-linear sigma model. These solutions are also the static configurations of the solitons of the 2+1-dimensional non-linear sigma-model, where they are known as skyrmions.

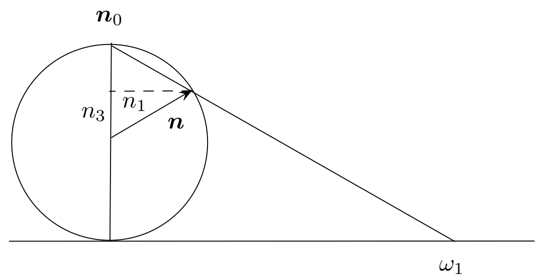

Figure 19.12 Stereographic projection of the target space.

We will solve the self-dual equations using the stereographic projection of S^2 target space onto a (target) plane, $\mathbb{R}^2$, with coordinates (ω_1, ω_2) (see Fig.19.12):

$$\omega_1 = \frac{2n_1}{1-n_3}, \qquad \omega_2 = \frac{2n_2}{1-n_3} \tag{19.119}$$

It will be convenient to define complex coordinates ω,

$$\omega = \omega_1 + i\omega_2 = 2\frac{n_1 + in_2}{1-n_3} \tag{19.120}$$

Similarly, we also define $n = n_1 + in_2$. Using that

$$\partial_1 \omega = \frac{2}{(1-n_3)^2}\left(\partial_1 n + n \overleftrightarrow{\partial_1} n_3\right) \tag{19.121}$$

we can rewrite the self-dual equations as

$$\partial_1 n = \mp in \overleftrightarrow{\partial_2} n_3, \qquad \partial_2 n = \mp in \overleftrightarrow{\partial_1} n_3 \tag{19.122}$$

or, more compactly

$$\partial_1 \omega = \pm i \partial_2 \omega \tag{19.123}$$

which is equivalent to say that ω_1 and ω_2 obey the Cauchy-Riemann equations

$$\frac{\partial \omega_1}{\partial x_1} = \pm \frac{\partial \omega_2}{\partial x_2}, \qquad \frac{\partial \omega_1}{\partial x_2} = \mp \frac{\partial \omega_2}{\partial x_1} \tag{19.124}$$

Therefore, $\omega(z)$ must be an analytic function of $z = x_1 + ix_2$ (but not an entire function!). As such, $\omega(z)$ may have zeros and poles but not branch cuts.

In terms of the function $\omega(z)$, the action for an instanton takes the form

$$S = \int d^2x \frac{\left|\frac{d\omega}{dz}\right|}{(1 + \frac{|\omega|^2}{4})^2} \tag{19.125}$$

such that

$$|Q| = \frac{S}{4\pi} \tag{19.126}$$

A solution of these equations is an analytic function $\omega(z)$ that has a *zero* of order $p \in \mathbb{Z}$

$$\omega(z) = \text{const.} \left(\frac{z - z_0}{\lambda}\right)^p \tag{19.127}$$

where z_0 is a complex number and λ is real. This instanton solution has topological charge $Q = p$. A solution with a *pole* of order p is an anti-instanton with topological charge $Q = -p$.

In both cases, these solutions have two arbitrary parameters, known as the zero modes of the instanton: the location of the instanton, represented by the complex number z_0 on the plane, and the scale of the instanton (or of the anti-instanton) set by λ. Notice that the vortices that we discussed above have only one zero mode, their location while the instantons have, in addition, an arbitrary scale. This is a consequence of the classical scale invariance of the two-dimensional non-linear sigma model.

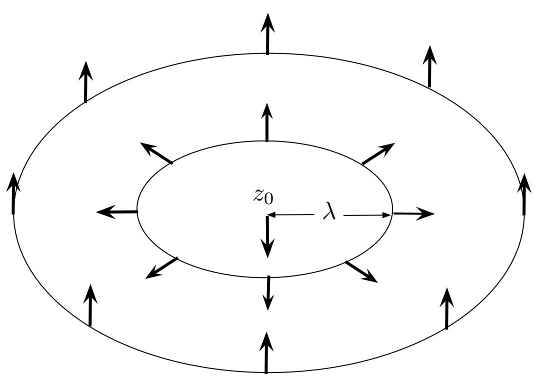

Figure 19.13 Instanton with topological charge $Q = 1$. The origin z_0 and the scale (radius) λ are zero modes of the instanton.

The general solution has the form

$$\omega(z) = \prod_i \left(\frac{z - z_i}{\lambda}\right)^{m_i} \prod_j \left(\frac{\lambda}{z - z_j}\right)^{n_j} \tag{19.128}$$

whose topological charge is

$$Q = \sum_i (m_i - n_i) \tag{19.129}$$

where instantons have topological charges m_i and anti-instantons $-n_j$.

Since the action for this solution saturates the bound, $S = 4\pi|Q|$, it depends only on the total topological charge Q and not on where the instantons and anti-instantons may be located. In other words, the instantons (and anti-instantons) do not interact with each other. We already encountered this feature at the Bogomol'nyi point of the abelian Higgs model. We will find the same feature in other theories with instanton solutions.

The instanton with topological charge $Q = 1$ located at the origin, $z_0 = 0$, is the solution

$$\omega(z) = \frac{z}{\lambda} \tag{19.130}$$

where λ is the scale, or radius, of the instanton. Upon retracing the steps of this construction, we find that this corresponds to the configuration of the field $\boldsymbol{n}(\boldsymbol{x})$

$$n_3(\boldsymbol{x}) = \frac{\boldsymbol{x}^2 - 4\lambda^2}{\boldsymbol{x}^2 + 4\lambda^2}, \qquad n_i(\boldsymbol{x}) = \frac{4\lambda x_i}{\boldsymbol{x}^2 + 4\lambda^2} \tag{19.131}$$

where $i = 1, 2$ labels the transverse components. This solution is shown qualitatively in Fig.19.13.

The construction of the instantons of the $O(3)$ non-linear sigma model is very elegant and beautiful. However, it poses a number of problems. In the case of the abelian-Higgs (or superconductor) model, the the vortex solutions have only one arbitrary parameter, the location of the vortex. We will see below, following the work of Kosterlitz and Thouless and its generalizations, that it is possible to compute the partition function of the model, essentially exactly, in terms of a sum over the configurations of vortices and anti-vortices. In fact, if vortices are ignored the partition function is trivial.

However, it has turned out to be very hard to recast the partition function of the non-linear sigma model in terms of a sum over instanton and anti-instanton configurations. As we saw, in addition to the location of the instantons, as a consequence of the scale invariance of the classical theory, the solution has an arbitrary parameter, the scale λ. Thus, in addition to summing over all possible locations of instantons and anti-instantons, one must also sum over instantons of all possible sizes, ranging from the UV to the IR, leading to serious infrared problems.

As we know form the renormalization group analysis that was discussed in Chapter 15, the classical fixed point is unstable to quantum fluctuations, and the effective coupling constant flows to strong coupling. One would have expected that the partition function of the instantons would have provided a physical picture based on how the $O(3)$ symmetry is restored at long scales (and scale invariance is broken) in terms of a gas (or plasma) of instantons and anti-instantons. To this date, this program has only been completed successfully in theories with large enough supersymmetry that the fluctuation determinant for an instanton can be computed exactly and there are no corrections to the semiclassical approximation.

19.5 Coset non-linear sigma models

Here we focused on the case of the $O(3)$ non-linear sigma model and showed that it has topologically non-trivial configurations classified by the homotopy group $\pi_2(S^2) = \mathbb{Z}$. However, this is a special property of the group $O(3)$. For the group $O(N)$, with $N > 3$, all configurations are trivial since $\pi_2(S^N) = 0$. This poses the question whether there are other theories that will more generically have instantons.

We will now see that there are are theories that generically (for "all N") have instantons. These are non-linear sigma models whose target spaces are cosets of the form G/H where G is a simply connected Lie group and H has at least one $U(1)$ subgroup. Since G is simply connected, the configurations of the non-linear sigma model can be smoothly deformed to the identity $I \in G$. Hence G it is topologically trivial, $\pi_2(G) = 0$. Hence, a principal chiral non-linear sigma model whose field is a group element, $g(x) \in G$, with action

$$S = \frac{1}{2u^2} \int d^2x \operatorname{tr}(\partial_\mu g(x) \partial_\mu g^{-1}(x)) \tag{19.132}$$

does not have instantons (here u^2 is a dimensionless coupling constant).

But chiral theories on the coset G/H do have instantons. Ti see that this is true we will follow Polyakov's construction, and define a field $\varphi_a(x)$

$$\varphi_a(x) = g_{ab}(x)\varphi_b^{(0)} \tag{19.133}$$

where $g(x) \in G$, and $\varphi^{(0)}$ is a constant field that is invariant under the action of the subgroup H, i.e. for all $h \in H$

$$h_{ab}\varphi_b^{(0)} = \varphi_a^{()} \tag{19.134}$$

We should note that, with these definitions, the field $g(x)$ does not have to be continuous.

Let $g^N(x)$ be the matrix-valued fields defined on the Northern hemisphere of S^2_{base}, and $g^S(x)$ defined on the Southern hemisphere of S^2_{base}. We will require that at the Equator, the common boundary of the two hemispheres and is isomorphic to S^1, the field $\varphi(x)$ be continuous on the whole S^2_{base}. Let us consider configurations of the field $g(x)$ that are discontinuous at the Equator, i.e. such that on points x on the Equator of S^2_{base}

$$g^N(x) = g^S(x)\, h(x) \tag{19.135}$$

where $x \in S^1$ and $h(x) \in H$. Clearly $\varphi(x)$ is continuous at the Equator since its restriction to each hemisphere

$$\varphi^N(x) = g^N(x)\varphi^{())}, \qquad \text{and } \varphi^S(x) = g^S(x)\varphi^{(0)} \tag{19.136}$$

Then, at the Equator

$$\varphi^N(x) = g^N(x)\varphi^{(0)} = g^S(x)h(x)\varphi^{(0)} \equiv g^S(x)\varphi^{(0)} = \varphi^S(x) \tag{19.137}$$

where $h(x) \in H$.

Hence, the field $\varphi(x)$ is continuous at the Equator. Moreover, its field configurations define a mapping of S^2_{base} to G/H. Such mappings can be classified according to the maps from the Equator, S^1, onto H. If the subgroup H contains at least an $U(1)$ subgroup,

$$H = U(1) \times \text{something trivial} \tag{19.138}$$

then, we can use the winding number of the maps of S^1 to $U(1)$ to classify the maps of S^2 to G/H. In other words, we have shown that, even though $\pi_2(G) = 0$, now we have

$$\pi_2(G/H) = \pi_1(H) \tag{19.139}$$

Now, since for $H \simeq U(1)$, $\pi_1(H) = \mathbb{Z}$, we conclude that

$$\pi_2(G/H) = \mathbb{Z} \tag{19.140}$$

19.6 The $\mathbb{CP}^{N-1}$ instanton

One example of such a non-linear sigma model is the $\mathbb{CP}^{N-1}$ model, discussed in sections 16.5.1 and 17.3 where we solved it in the large-N limit. The $\mathbb{CP}^{N-1}$ model is a non-linear sigma model of an N-component complex field $z_a(x)$ (with $a = 1, \ldots, N$) which transforms in the fundamental representation of $SU(N)$, with a gauged $U(1)$ subgroup. The field obeys the constraint $\sum_{a=1}^N |z_a(x)|^2 = 1$, everywhere. The classical Lagrangian is

$$\mathcal{L} = \frac{1}{g^2}|(\partial_\mu + iA_\mu)z_a|^2 \tag{19.141}$$

where $A_\mu(x)$ is a dynamical $U(1)$ gauge field, and g is the dimensionless coupling constant. In this case, the coset is

$$\mathbb{CP}^{N-1} \cong \frac{SU(N)}{SU(N-1) \otimes U(1)} \tag{19.142}$$

which has instantons for all values of N. Recall that the $\mathbb{CP}^2$ model is equivalent to the $O(3)$ non-linear sigma model. However, we will see that the $\mathbb{CP}^{N-1}$ model has instantons for all values of N.

At the classical level, this theory breaks the $SU(N)$ symmetry spontaneously down to an unbroken $SU(N-1)$ subgroup and the $U(1)$ gauged subgroup. As in the case of the $O(3)$ non-linear sigma model, the requirement that the action be finite implies, in this case, that the covariant derivative vanishes at long distances,

$$\lim_{|x|\to\infty} (\partial_\mu + iA_\mu) z_a = 0 \tag{19.143}$$

which can be achieved by the condition

$$\lim_{|x|\to\infty} \boldsymbol{z}(x) = \boldsymbol{z}_0 \, e^{i\theta(x)} \tag{19.144}$$

where $\boldsymbol{z}_0$ is arbitrary and constant (but with $\boldsymbol{z}_0 \cdot \boldsymbol{z}_0 = 1$) and $\theta(x) \in [0, 2\pi)$ is an arbitrary phase. Thus, the winding of the phase θ is an-integer valued topological invariant. In terms of the $\mathbb{CP}^{N-1}$ field, the topological charge is

$$Q = \frac{1}{2\pi i} \int_\Sigma d^2x \; \partial_\mu (\epsilon_{\mu\nu} z_a^* \partial_\nu z^a) = \frac{1}{2\pi} \int_\Sigma d^2x \; \epsilon_{\mu\nu} \partial_\mu A_\nu = \frac{1}{2\pi} \oint_{\partial\Sigma} dx_\mu \partial_\mu \theta \tag{19.145}$$

where Σ is a large disk with boundary $\partial\Sigma$, and where we used the identification of te gauge field A_μ in terms of the $\mathbb{CP}^{N-1}$ field.

Much as in the case of the $O(3)$ non-linear sigma model, the topological charge Q of the $\mathbb{CP}^{N-1}$ model sets a lower bound for the action of a configuration. To show this, we define the field C_μ^a (with $a = 1, \ldots, N$),

$$C_\mu^a(x) = \partial_\mu z^a(x) - z^a (z_b^* \partial_\mu z_b) \tag{19.146}$$

which obeys the obvious inequality

$$(C_\mu^a(x) \pm i\epsilon_{\mu\nu} C_\nu^a(x))^2 \geq 0 \tag{19.147}$$

which gives the condition that

$$|\partial_\mu z_a|^2 + (z_a^* \partial_\mu z_a)^2 \geq \pm i\epsilon_{\mu\nu} \partial_\mu (z_a^* \partial_\nu z_a) \tag{19.148}$$

(repeated indices are summed over) where we recognize the left hand side as the Lagrangian of the $\mathbb{CP}^{N-1}$ model, after integrating-out the gauge field A_μ). Therefore, we find that the action obeys the now familiar inequality

$$S[z_a] \geq 2\pi |Q| \tag{19.149}$$

The instantons of the $\mathbb{CP}^{N-1}$ model are the configurations which saturate this bound, and satisfy the self-dual (and anti-self-dual) equation

$$C_\mu^a(x) = \pm\, i\epsilon_{\mu\nu} C_\nu^a(x) \tag{19.150}$$

The solutions of these first order partial differential equations yields the instanton solutions in terms of rational functions of the complex plane, much in the same way as what we did for the $O(3)$ model. In terms of the gauge field A_μ, the $Q = \pm n$ instanton (and anti-instanton) at the origin, $x = 0$, is

$$A_\mu^\pm = \pm n\; \epsilon_{\mu\nu} \frac{x_\nu}{x^2 + \lambda^2} \tag{19.151}$$

where λ is an arbitrary scale.

Naively, we expect that since the instantons have finite Euclidean action, their contribution to the partition function would be exponentially small of the order of $\exp(-S_E/g^2)$, up to an "entropic" prefactor which is (naively) a subleading contribution. this expectation is naive since, in addition to summing over the locations of the instantons (which the case of vortices can be computed), in the non-linear sigma models one has to sum (integrate) over instantons of all possible scales. Thus, it is far from obvious that this naive argument is actually correct.

19.7 The 't Hooft-Polyakov Magnetic Monopole

We will now turn to the case of non-abelian gauge theories. We will begin by constructing the analog of the $U(1)$ vortex. This is the Dirac magnetic monopole.

In section 19.3.2 we showed that the complex scalar field in 1+1 Euclidean dimensions has singular vortex configurations of the phase field $\theta(x)$. These configurations are multivalued and have branch cuts. In 1931 Dirac proposed to describe a magnetic monopole as a configuration of magnetic fields created the current I flowing through an infinitely long and infinitesimally thin solenoid. Let us assume that the infinitesimally thin solenoid, the "Dirac string", runs along the x_3 axis from $x_3 \to -\infty$ and ends at the origin, $\boldsymbol{x} = (0,0,0)$. Then, the end of the long solenoid acts as a positive magnetic pole. The magnetic flux through the solenoid is $2\pi q$, where q is the magnetic charge. Outside the solenoid there is an isotropic magnetic field $\boldsymbol{B}(\boldsymbol{x})$, radiating outwards from the end of the solenoid. In other words, the magnetic field is

$$B_i(\boldsymbol{x}) = \frac{q}{2}\frac{x_i}{|\boldsymbol{x}|^2} - 2\pi q \delta_{i,3}\delta(x_1)\delta(x_2)\theta(-x_3) \tag{19.152}$$

The first term is the (magnetic) "Coulomb" field of the magnetic monopole of magnetic charge q. The second (and singular) term is the solenoid. Notice that, if the solenoid is included, this is an allowed configuration of Maxwell's equations. Thus, this is a magnetic monopole of charge q at the origin with its singular Dirac string attached. However, the solenoid makes this configuration singular, much in the same way as the vortex. Indeed, phase field configuration (in $d = 2$) of a vortex of topological charge n at the origin can be written as

$$\partial_i\theta(\boldsymbol{x}) = n\epsilon_{ij}\frac{x_i}{|\boldsymbol{x}|^2} - 2\pi n\delta(x_1)\theta(-x_2) \tag{19.153}$$

The second term is singular and similar to a Dirac string, and represents the branch cut of the phase field.

On the other hand, we saw that in the abelian Higgs model, a theory of a complex scalar field coupled to the a $U(1)$ gauge field, has vortex classical regular vortex solutions in the classically spontaneously broken phase of this theory. G. 't Hooft ('t Hooft, 1976) and A. M. Polyakov (Polyakov, 1975) showed that a regular configuration which at long distances becomes a Dirac magnetic monopole exists in the Higgs sector of the Georgi-Glashow model (Georgi and Glashow, 1974), that we already discussed in section 18.11.2.

The Georgi-Glashow model has a three-component real scalar field $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$, that transforms under the adjoint representation of $SU(2)$, and a Yang-Mills gauge field taking values in the algebra of the gauge subgroup $G = SU(2)$ associated with the weak interactions of a grand unified gauge theory with gauge group $SU(5)$. In the spontaneously broken phase, the Lagrangian in $D = 2 + 1$-dimensional Euclidean space is

$$\mathcal{L} = \frac{1}{2}(D_\mu\boldsymbol{\phi})^2 - \frac{m^2}{2}\boldsymbol{\phi}^2 + \frac{\lambda}{4!}(\boldsymbol{\phi}^2)^2 + \frac{1}{4}\mathrm{tr}F_{\mu\nu}^2 \tag{19.154}$$

where $D_\mu\boldsymbol{\phi} = \partial_\mu\boldsymbol{\phi} + g\boldsymbol{A}_\mu \times \boldsymbol{\phi}$ is the covariant derivative in the adjoint representation of $SU(2)$. In this phase, the $SU(2)$ gauge symmetry is spontaneously broken down to its $U(1)$ subgroup. As a subgroup of the compact Lie group $SU(2)$, the unbroken $U(1)$ subgroup is compact.

The classical (Euclidean) equations of motion are (with $a = 1, 2, 3$ and $i, j = 1, 2, 3$)

$$D_i F^{aij} = g\epsilon^{abc}(D_j\phi_b)\phi_c \tag{19.155}$$

and

$$D_i D_i \phi_a = -\lambda \boldsymbol{\phi}^2 \phi_a + \lambda f^2 \phi_a \tag{19.156}$$

where $\lambda f^2 = |m^2|$. The Euclidean action of the $D = 3$ dimensional instanton (or the energy of the $D = 3$ soliton) is

$$S_E = \int d^3x \left[\frac{1}{4} F_{ij}^a F^{aij} + \frac{1}{2} D_i \phi_a D_i \phi_a + \frac{\lambda}{4!} (\boldsymbol{\phi}^2 - f^2)^2 \right] \tag{19.157}$$

Let us first look at the zero energy solutions, the classical vacuum. They are $A_i^a = 0$, $\boldsymbol{\phi}^2 = f^2$, and $D_i \boldsymbol{\phi} = 0$. In this case, the latter reduces to $\partial_i \boldsymbol{\phi} = 0$ and $\boldsymbol{\phi}$ is a constant vector.

We will now seek finite action solutions. As in our analysis of the instantons of the non-linear sigma model, the finite action requirement implies that the fields should approach the vacuum solution sufficiently fast as $r \to \infty$ (here $r = |\boldsymbol{x}|$ is the radial coordinate). In $D = 3$ dimensions the required asymptotic behavior is that, as $r to \infty$, $r^{3/2} D_i \phi^a \to 0$, and $\boldsymbol{\phi}^2 \to f^2$. In spherical coordinates (r, θ, φ), the θ component of the covariant derivative is

$$D_\theta \phi^a = \frac{1}{r} \frac{\partial \phi^a}{\partial \theta} + g \epsilon^{abc} A_\theta^b \phi^c \tag{19.158}$$

Hence, $D_\theta \phi^a \to 0$ (as $r \to \infty$), provided $A_\theta^b \sim \frac{1}{r}$, also as $r \to \infty$. the same asymptotic behavior holds for the other components. On the other hand, if $A_i \sim 1/r$, then, by dimensional counting, the field strength $F \sim 1/r^2$,a nd $F^2 \sim 1/r^4$ (again, as $r \to \infty$), which is integrable at large r in $D = 3$ dimensions.

This analysis shows that, unlike the case of the $D = 2$ non-linear sigma models, the finite action solutions of the Georgi-Glashow model may have fields ϕ^a which are not equivalent at spatial infinity since they are allowed to asymptotically point in different directions. In particular, the asymptotic behavior for the scalar field only requires that at spatial infinity, i.e. at the surface of a large sphere S^2 of radius $r \to \infty$, the magnitude of the field must be fixed but not its direction.

Thus, the finite Euclidean action solutions are maps of the base space, the large sphere S^2 with large radius r, to the S^2 target space of the scalar field with fixed norm. In other words, the topology is the same as in the case of the $D = 2$ non-linear sigma model, and the configurations are also classified by the homotopy group $\pi_2(S^2) \cong \mathbb{Z}$, and the classes are labeled by the topological charge $Q \in \mathbb{Z}$ of the $D = 2$ $O(3)$ non-linear sigma model shown in Eq.(19.105). For instance, the configuration of the scalar field $\boldsymbol{\phi}$

on the large sphere at spatial infinity S^2 with topological charge $Q = 1$ is the hedgehog configuration (a "hairy ball") shown in Fig.19.14.

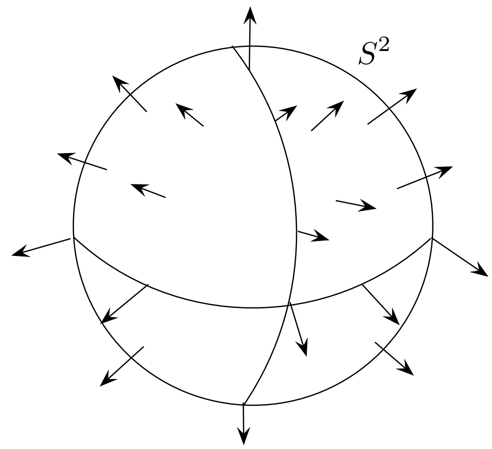

Figure 19.14 Hedgehog configuration of the $O(3)$ field $\boldsymbol{\phi}$ at the large sphere S^2 at spatial infinity.

How is this related to the magnetic monopole? To see the relation we will use a gauge-invariant formulation due to 't Hooft. Let us define the gauge-invariant field strength tensor $F_{\mu\nu}$,

$$F_{\mu\nu} \equiv \widehat{\phi}_a F^a_{\mu\nu} - \frac{1}{g}\epsilon^{abc}\widehat{\phi}_a D_\mu \widehat{\phi}_b D_\nu \widehat{\phi}_c \tag{19.159}$$

where we defined $\widehat{\phi}^a = \phi^a/||\boldsymbol{\phi}||$. For a topologically trivial configuration such as $\widehat{\boldsymbol{\phi}} = (0,0,1)$, this field strength reduces to

$$F_{\mu\nu} = \partial_\mu A^3_\nu - \partial_\nu A^3_\mu \tag{19.160}$$

In $D = 3$ dimensions, we can define the dual (pseudovector) $F^*_\mu = \frac{1}{2}\epsilon_{\mu\nu\lambda}F^{\nu\lambda}$, and in $D = 4$ dimensions, the dual tensor $F^*_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}F^{\lambda\rho}$. Then, for topologically trivial configurations, the dual tensors satisfy the Bianchi identities, $\partial^\mu F^*\mu = 0$ (in $D = 3$) and $\partial^\mu F^*_{\mu\nu} = 0$, just as in Maxwell's theory. Hence, configurations whose scalar fields have trivial configurations at infinity do not have magnetic monopoles.

On the other hand, we can use the definition of the field tensor $F_{\mu\nu}$ of Eq.(19.159) to compute the divergence of its dual and find, in $D = 4$,

$$\partial^\nu F^*_{\mu\nu} = \frac{1}{2g}\epsilon_{\mu\nu\lambda\rho}\epsilon_{abc}\partial^\nu\widehat{\phi}^a\partial^\lambda\widehat{\phi}^b\partial^\rho\widehat{\phi}^c \equiv \frac{4\pi}{g}j_\mu \tag{19.161}$$

where j_μ is the topological current. In $D = 3$, we find

$$\partial^\mu F^*_\mu = \frac{4\pi}{g} j_0(x) \tag{19.162}$$

where j_0 is the topological charge density. Its integral over all space is given by

$$\begin{aligned}\int d^3x \; j_0(x) =& \frac{1}{8\pi} \int d^3x \epsilon_{ijk} \epsilon^{abc} \partial_i \widehat{\phi}^a \partial_j \widehat{\phi}^b \partial_k \widehat{\phi}^c \\ =& \frac{1}{8\pi} \int d^3x \epsilon_{ijk} \epsilon^{abc} \partial_i \left(\widehat{\phi}^a \partial_j \widehat{\phi}^b \partial_k \widehat{\phi}^c \right) \\ =& \frac{1}{8\pi} \int_{S^2} d^2S_i \epsilon_{ijk} \epsilon^{abc} \widehat{\phi}^a \partial_j \widehat{\phi}^b \partial_k \widehat{\phi}^c \end{aligned} \tag{19.163}$$

which we recognize as the integer-valued topological charge Q that classifies the configurations of the field $\widehat{\boldsymbol{\phi}}$ at the sphere S^2 at spatial infinity. Therefore, in $D = 3$ dimensions the Bianchi identity now yields the topological charge density $j_0(x)$, whose integral over all space is the topological charge. The same line of argument yields, in $D = 4$, the result that the Bianchi identity is equal to the topological current j_μ.

We conclude that in $D = 3$ the instantons are point-like instantons, where as in $D = 4$ they are currents. On the other hand, in $D = 4$ the monopoles are point-like finite-energy topological solitons. It follows that, since the magnetic field is $B_i = \frac{1}{2}\epsilon_{ijk} F_{jk}$, the divergence of the field is

$$\partial_i B_i = \frac{4\pi}{g} j_0(x) \tag{19.164}$$

Hence, the magnetic charge of a monopole of topological charge Q is

$$m = \int d^3x \; \partial_i B_i(x) = \frac{Q}{g} \tag{19.165}$$

Let us work out the monopole with topological charge $Q = 1$. We will use the spherically-symmetric ansatz

$$\begin{aligned}\phi^a(x) =& \delta_{ia} \frac{x^i}{r} F(r), \\ A_i^a(x) =& \epsilon_{aij} \frac{x^i}{r} W(r)\end{aligned} \tag{19.166}$$

where the functions $F(r)$ and $W(r)$ of the radius $r = |x|$ have the asymptotic behavior

$$\lim_{r\to\infty} F(r) = f, \qquad \lim_{r\to\infty} grW(r) = 1 \tag{19.167}$$

In this solution, at large distances $r \to \infty$, the magnetic field points outwards and isotropically away from the origin, and has the asymptotic behavior of a magnetic monopole of charge $1/g$,

$$\boldsymbol{B}(\boldsymbol{x}) \sim \frac{\boldsymbol{x}}{gr^3}, \qquad \text{as } r \to \infty \tag{19.168}$$

Upon introducing the functions $K(r)$ and $H(r)$,

$$K(r) \equiv 1 - grW(r), \qquad H(r) \equiv grF(r) \tag{19.169}$$

the field equations become

$$\begin{aligned} r^2 \frac{d^2K}{dr^2} =& K(K^2-1) + H^2 K, \\ r^2 \frac{d^2H}{dr^2} =& 2HK^2 + \lambda\left(\frac{H^2}{g^2} - r^2 f^2\right) H \end{aligned} \tag{19.170}$$

As in the other cases we discussed, these equations are, in general quite difficult to solve, except in the BPS limit, $\lambda \to 0$ where

$$K(r) = \frac{gfr}{\sinh(gfr)}, \qquad H(r) = \frac{gfr}{\tanh(gfr)} - 1 \tag{19.171}$$

Notice that both $K(r)$ and $H(r)$ are regular functions as $r \to 0$, where $H \to 0$ and $F \to 1$. Hence, the 'tHooft-Polyakov monopole is regular near the origin, and the potential singularity is smeared at distances shorter than the length scale $\xi \sim 1/(gf)$.

It can be shown that these solutions satisfy (and saturate) the Bogomol'nyi bound

$$E \geq 4\pi \frac{Qf}{g} \tag{19.172}$$

and, hence, have finite action. They also satisfy the Bogomol'nyi equation

$$F_{ij}^a = \epsilon_{ijk} D_k \phi^a \tag{19.173}$$

In summary, the Georgi-Glashow model has monopole solutions, the 't Hooft-Polyakov monopole, instantons in 2+1 dimensions and solitons in 3+1 dimensions, with a quantized magnetic charge. However, unlike the Dirac monopole, the 't Hooft-Polyakov monopole has finite Euclidean action (or energy) and is not a singular configuration. Shortly we will see that in 2+1 dimensions these monopole instantons lead to confinement of static sources charged under the unbroken $U(1)$ subgroup of $SU(2)$.

19.8 The Yang-Mills instanton in $D = 4$ dimensions

Let us now discuss the instanton solutions of pure Yang-Mills gauge theory in $D = 4$ (Euclidean) dimensions. Let G be a simple and compact gauge group. The Euclidean Yang-Mills action is

$$S = \frac{1}{4} \int d^4x \, \mathrm{tr} F_{\mu\nu}^2 \tag{19.174}$$

As before, we will seek finite Euclidean action solutions, $S < \infty$. This requires that, at long distances, the field strength vanishes as

$$F_{\mu\nu} \sim O(1/r^2), \qquad \text{as } r \to \infty \tag{19.175}$$

Since the field strength vanishes at long distances, in the same asymptotic limit, the gauge field must approach a pure gauge transformation labeled by $g(x) \in G$, i.e.

$$A_\mu \sim g^{-1} \partial_\mu g + O(1/r), \qquad \text{as } r \to \infty \tag{19.176}$$

Therefore, on a sphere S^3 of large radius R, the gauge field configurations A_μ are mapped onto the gauge transformations labeled by the group elements $g \in G$. In other words, the finite action solutions are in one-to-one correspondence with the smooth mappings of the large sphere S^3 onto the gauge group G,

$$g(x) : S^3 \mapsto G \tag{19.177}$$

In the case of $G = SU(2)$, then we can write the group elements as

$$g(x) = n_4(x) \, I + i\boldsymbol{n}(x) \cdot \boldsymbol{\sigma}, \qquad g^{-1} = g^\dagger \tag{19.178}$$

where I is the 2×2 identity matrix and σ_i are the three Pauli matrices. Since in $SU(2)$, $\det g = 1$, the four-component real vector $(\boldsymbol{n}, n_4)$ must satisfy the constraint

$$\boldsymbol{n}^2 + n_4^2 = 1 \tag{19.179}$$

This implies, that $SU(2) \cong S^3$. Therefore, the maps of $S^3 \mapsto G$ are maps of $S^3 \mapsto S^3$. These mappings are classified by the homotopy group

$$\pi_3(SU(2)) \cong \pi_3(S^3) \cong \mathbb{Z} \tag{19.180}$$

The topological charge , or winding number, that classifies these maps is the Pontryagin index which counts the number of times that S^3 covers S^3. As in the $\pi_2(S^2)$ case, the Pontryagin index is the Jacobian of the map $g(x)$. Moreover, since very simple and compact group G has an $SU(2)$ subgroup, the Pontryagin index classifies all these maps, and $\pi_3(G) = \mathbb{Z}$.

In general, for $g \in G$, the Pontryagin index is

$$Q = \frac{1}{24\pi^2} \int_{S^3} \epsilon_{\mu\nu\lambda} \text{tr}(L_\mu L_\nu L_\lambda) \tag{19.181}$$

where $L_\mu = g^{-1}\partial_\mu g$, and S^3 is the boundary of four-dimensional Euclidean spacetime. In the case of $SU(2)$, the Pontryagin index is

$$Q = \frac{1}{32\pi^2} \int_{S^3} d^3x \, \epsilon^{abcd} \epsilon_{\mu\nu\lambda} \, n^a \partial_\mu n^b \partial_\nu n^c \partial_\lambda n^d \tag{19.182}$$

After some algebra, it can be shown that the topological charge Q can be written as an integral in four-dimensional Euclidean spacetime Ω, whose boundary is S^3, of a total derivative,

$$Q = \frac{1}{32\pi^2} \int_\Omega d^4x \, \epsilon^{\mu\nu\lambda\rho} \text{tr}\left(F_{\mu\nu} F_{\lambda\rho}\right) \equiv \frac{1}{8\pi^2} \int d^4x \, \text{tr} F \wedge F^* \tag{19.183}$$

where $F^*_{\mu\nu} = \frac{1}{4}\epsilon_{\mu\nu\lambda\rho} F^{\lambda\rho}$, is the dual tensor. Therefore, the Pontryagin index Q is given by

$$Q = \frac{1}{8\pi^2} \int d^4x \, \text{tr}\left(F^{\mu\nu} F^*_{\mu\nu}\right) \tag{19.184}$$

which labels the topological classes of all compact simply connected gauge groups.

We will now show that the Pontryagin index places a lower bound to the Euclidean action of gauge fields belonging to a topological class. The argument is similar to the one we used for the instantons of the non-linear sigma model. Thus, we rewrite the Yang-Mills Euclidean action as

$$\begin{aligned} S =& \frac{1}{4g^2} \int d^4x \, \text{tr} F_{\mu\nu}^2 \\ =& \frac{1}{8g^2} \int d^4x \, \text{tr}\left(F_{\mu\nu} - F_{\mu\nu^*}\right)^2 + \frac{1}{4g^2} \int d^4x \, \text{tr}\left(F^{\mu\nu} F^*_{\mu\nu}\right)^2 \\ =& \frac{8\pi^2}{g^2} Q + \frac{1}{8g^2} \int d^4x \, \text{tr}\left(F_{\mu\nu} - F_{\mu\nu^*}\right)^2 \end{aligned} \tag{19.185}$$

Since the last term is manifestly positive, we find the lower bound for the Euclidean action

$$S \geq \frac{8\pi^2}{g^2} Q \tag{19.186}$$

Once again, we can seek the configurations that saturate the bound, which satisfy the self-duality ("Cauchy-Riemann") equation

$$F_{\mu\nu} = F^*_{\mu\nu} \tag{19.187}$$

for which

$$S = \frac{8\pi^2}{g^2} Q \tag{19.188}$$

are instantons with $Q > 0$.

Similarly, the gauge fields that satisfy the anti-self-dual equation

$$F_{\mu\nu} = -F^*_{\mu\nu} \tag{19.189}$$

satisfy

$$S = \frac{8\pi^2}{g^2} |Q| \tag{19.190}$$

and are anti-instantons with $Q < 0$.

In addition, if a gauge field is elf-dual, $F_{\mu\nu} = F^*_{\mu\nu}$, it also satisfies the equation $D_\mu F^{\mu\nu} = 0$. An example of this solution is the $Q = 1$ instanton of the $SU(2)$ Yang-Mills theory, for which the gauge field is given by

$$A^a_\mu = -\eta_{a\mu\nu} \frac{(x_\nu - a_\nu)}{(\boldsymbol{x} - \boldsymbol{a})^2 + \rho^2} \tag{19.191}$$

where we introduced the tensor $\eta_{abc} = \epsilon_{abc}$, $\eta_{ab0} = \delta_{ab}$, etc. Here ρ is the arbitrary scale of the instanton (reflecting the classical scale invariance of four-dimensional Yang-Mills theory), and a_μ is an (also arbitrary) location of the instanton. The field strength of this solution is

$$F^a_{\mu\nu} = -4\eta_{a\mu\nu} \frac{\rho^2}{(\boldsymbol{x} - \boldsymbol{a})^2 + \rho^2} \tag{19.192}$$

19.9 Vortices and the Kosterlitz-Thouless transition

We will now discuss two specific examples of theories in which topological excitations (instantons) either drive their phase transitions, or lead to a non-perturbative phase at all values of the coupling constant. We will work with the lattice formulation of these theories where the results can be obtained more simply. All of the cases that we will consider have a (compact) $U(1)$ symmetry, either global or local, the $U(1)$ non-linear sigma model in $D = 2$ dimensions, known in Statistical Mechanics as the XY model, and Polyakov's compact electrodynamics in $D = 3$ dimensions. Much of what we will do can be extended to theories with more general abelian symmetry groups, both continuous or discrete, but not to non-abelian symmetries.

We will consider a system on $D = 2$-dimensional square lattice with a $U(1)$ degree of freedom at very site $\boldsymbol{r}$, a two-component vector of unit length,

$\boldsymbol{n}(\boldsymbol{r}) = (\cos\theta(\boldsymbol{r}), \sin\theta(\boldsymbol{r}))$, where the phase is defined in the interval $0 \leq \theta(\boldsymbol{r}) < 2\pi$. The energy (or lattice Euclidean action) is (here $\mu = 1, 2$ are the two directions on the square lattice)

$$E = -J \sum_{\boldsymbol{r},j} \boldsymbol{n}(\boldsymbol{r}) \cdot \boldsymbol{n}(\boldsymbol{r} + \boldsymbol{e}_j) = -J \sum_{\boldsymbol{r},\mu} \cos(\Delta_\mu \theta(\boldsymbol{r})) \tag{19.193}$$

where $\Delta_\mu \theta(\boldsymbol{r}) = \theta(\boldsymbol{r} + \boldsymbol{r}_\mu) - \theta(\boldsymbol{r})$ is the lattice difference operator, and $\boldsymbol{e}_\mu$ are the two orthonormal vectors of the square lattice, i.e. $\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = \delta_{\mu\nu}$.

The partition function of this model is

$$Z = \prod_{\boldsymbol{r}} \int_0^{2\pi} \frac{d\theta(\boldsymbol{r})}{2\pi} \exp\Big(-\frac{J}{T} \sum_{\boldsymbol{r},\mu} \cos(\Delta_\mu \theta(\boldsymbol{r}))\Big) \tag{19.194}$$

In the low temperature regime, $T \ll J$, we can naively take the continuum limit and write

$$Z \simeq \int \mathcal{D}\theta \; \exp(-\frac{1}{2g} \int d^2x \, (\nabla\theta)^2) \tag{19.195}$$

(with $g = Ta^2/J$, where a is the lattice spacing) which is formally a massless free field theory if it were not for the fact that the lattice model is invariant under local periodic shifts of the phase

$$\theta(\boldsymbol{r}) \mapsto \theta(\boldsymbol{r}) + 2\pi n(\boldsymbol{r}), \qquad \text{with } n(\boldsymbol{r}) \in \mathbb{Z} \tag{19.196}$$

This local symmetry requires that the only allowed observables must obey the periodicity condition. In other words, the field θ, even in the continuum, is compactified, and the global symmetry of the theory is $U(1)$ and not $\mathbb{R}$.

In section 19.3.3 we discussed this continuum theory. There we saw that this theory has singular vortex configurations (instantons). The lattice model also has vortices, but the lattice definition makes them regular. We will now show how the vortices arise in the lattice model and what role they play.

Our main tool will be a *duality transformation*, a generalization of a method first introduced by Kramers and Wannier in the context of the two-dimensional classical Ising model, and closely related to topology. In general, this duality transformation has two ingredients. One is a geometric duality, or duality of forms. Geometric duality says that in D dimensions a p-form is dual to a $D-p$ form. In $D = 2$ dimensions this means that a theory with a global symmetry, and hence defined on the links of the lattice (one forms), is dual to a theory defined on the links of the *dual* lattice, and hence also on one-forms. Therefore, in $D = 2$ dimensions, a theory with a global symmetry is dual to another theory also with a global symmetry, defined on the dual lattice. We will see next that in $D = 3$ dimensions the dual is a gauge theory. On the other hand, the second ingredient is that the dual of a theory is defined on the representations of the group. If the group is abelian, its representations are one-dimensional and define a group, the dual group. This is where this approach breaks down in non-abelian theories since their representations are, in general, not one-dimensional and do not form a group.

Let us see how this works out in the context of a theory with a $U(1)$ symmetry. To simplify the notation, from now on we will set $J = 1$ (or, what is the same, that T is measured in units of J).

By inspection of the partition function of Eq.(19.194) we see that the Gibbs weight of a configuration can be expressed as a product over links,

$$\exp\Big(-\frac{1}{T}\sum_{\boldsymbol{r},\mu} \cos(\Delta_\mu \theta(\boldsymbol{r}))\Big) = \prod_{\boldsymbol{r},\mu=1,2} \exp\Big(-\frac{i}{T}\cos(\Delta_\mu \theta(\boldsymbol{r}))\Big) \tag{19.197}$$

Consider now an expression of the form $\exp(V[\theta])$, which we will require to be a periodic function of θ. Therefore, it can be expanded in a Fourier series

$$\exp(V[\theta]) = \sum_{\ell\in\mathbb{Z}} V_\ell e^{i\ell\theta} \tag{19.198}$$

since $e^{i\theta} \in U(1)$, and $e^{i\ell\theta}$ are vectors in the ℓth irreducible representation of $U(1)$, then since $e^{i\theta_1} \in U(1)$ and $e^{\theta_2} \in U(1)$, then $e^{i(\theta-1+\theta_2)} \in U(1)$. Similarly, $e^{i\ell_1\theta} \in U(1)$, and $e^{i\ell_2\theta} \in U(1)$, then $e^{i(\ell_1+\ell_2)\theta} \in U(1)$. Then, the representations of $U(1)$ form a group under the addition. In other words, the representations form a group, the dual group, which is isomorphic to the group of integers, $\mathbb{Z}$.

In particular,

$$\exp(\beta\cos\theta) = \sum_{\ell\in\mathbb{Z}} I_\ell(\beta) e^{i\ell\theta} \tag{19.199}$$

where

$$I_\ell(\beta) = \int_0^{2\pi} \frac{d\theta}{2\pi} \exp(\beta\cos\theta - i\ell\theta) \tag{19.200}$$

is a Bessel function of imaginary argument. Here, we have set $\beta = 1/T$.

Here we will modify the Gibbs weight of Eq.(19.194) using the more tractable expressions (which has the same symmetries)

$$Z = \prod_{\boldsymbol{r}} \int_0^{2\pi} \frac{d\theta(\boldsymbol{r})}{2\pi} \sum_{\{\ell_\mu(\boldsymbol{r})\}} \exp\Big(-\sum_{\boldsymbol{r},\mu} \frac{\ell_\mu^2(\boldsymbol{r})}{2\beta} + i\sum_{\boldsymbol{r},\mu} \ell_\mu(\boldsymbol{r})\Delta_\mu\theta(\boldsymbol{r})\Big) \tag{19.201}$$

where we defined a set of integer-valued variables, $\ell_\mu(\boldsymbol{r}) \in \mathbb{Z}$ for each link $(\boldsymbol{r},\mu)$ of the square lattice.

There is an equivalent formulation of what we have done which is quite instructive. We note that what we done is equivalent to make the replacement (on every link)

$$\exp(\beta\cos\Delta_\mu\theta(\boldsymbol{r})) \mapsto \text{const.} \sum_{\ell_\mu(\boldsymbol{r})} \exp\Big(-\frac{\beta}{2}(\Delta_\mu\theta(\boldsymbol{r}) - 2\pi\ell_\mu(\boldsymbol{r}))^2\Big) \quad (19.202)$$

where on the right hand side the variable θ now takes all possible real values. Periodicity is recovered by the local discrete transformation

$$\theta(\boldsymbol{r}) \mapsto \theta(\boldsymbol{r}) + 2\pi k_\mu(\boldsymbol{r}), \qquad \ell_\mu(\boldsymbol{r}) + k_\mu(\boldsymbol{r}) \quad (19.203)$$

where $k_\mu(\boldsymbol{r})$ is an integer-valued local gauge transformation. Therefore the variables $\ell_\mu(\boldsymbol{r})$ play the role of a gauge field whose gauge group is $\mathbb{Z}$. In this picture, the variable θ is not compact and it does not have any vorticity. The field is compactified by the discrete gauge symmetry. Vortices are created by Dirac-strings of the integer-valued gauge fields, such as in the example of Fig.19.15, that shows a Dirac string on the open path Γ of the dual lattice. On the links pierced by Γ we have set $\ell_\mu = m$, and zero on all other links. The circulation on a closed path C of the discrete gauge field is $\oint_C = m$ if the dual site $\boldsymbol{R}$ is enclosed by C and zero otherwise. The Dirac string forces a jump of $2\pi m$ on θ accross the Dirac string. Thus, this configuration represents a vortex of topological charge m at $\boldsymbol{R}$.

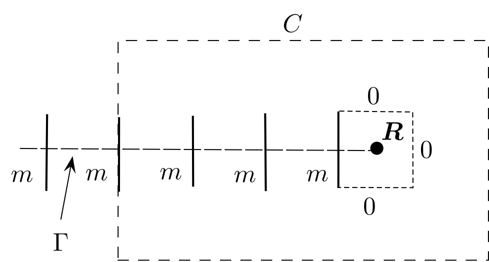

Figure 19.15 A Dirac string Γ is a configuration of the discrete gauge fields ℓ_μ that creates a vortex at the site $\boldsymbol{R}$ of the dual lattice. Here $\ell_\mu = m$ on all the links pierced by the Dirac string Γ, and 0 on all other links. The circulation of the integer-valued gauge field ℓ_μ on any closed contour C, $\oint_C \ell_\mu = m$ if C encloses site $\boldsymbol{R}$, and zero otherwise.

We now return to the duality transformation by integrating out the periodic $\theta(\boldsymbol{r})$ field in the partition function of Eq.(19.201). Since

$$\int_0^{2\pi} \frac{d\theta(\boldsymbol{r})}{2\pi} \exp(i\theta(\boldsymbol{r})\Delta_\mu\ell_\mu(\boldsymbol{r})) = \delta(\Delta_\mu\ell_\mu(\boldsymbol{r})) \quad (19.204)$$

where we defined the lattice divergence,

$$\Delta_\mu \ell_\mu(\boldsymbol{r}) = \sum_{\mu=1,2} \Big[\ell_\mu(\boldsymbol{r}) - \ell_\mu(\boldsymbol{r} - \boldsymbol{e}_\mu) \Big] \tag{19.205}$$

the partition function takes the equivalent form of a sum over configurations on the integer-valued link variables, $\ell_\mu(\boldsymbol{r})$,

$$Z = \sum_{\{\ell_\mu(\boldsymbol{r})\}} \prod_{\boldsymbol{r}} \delta(\Delta_\mu \ell_\mu(\boldsymbol{r})) \exp\Big(-\sum_{\boldsymbol{r},\mu} \frac{\ell_\mu^2(\boldsymbol{r})}{2\beta} \Big) \tag{19.206}$$

Formally, this partition function is a sum of configurations of closed loops. If $\beta \ll 1$, i.e. $T \gg J$, the partition function is dominated by the smallest loops: this is the high temperature expansion.

The constraint that the loops are closed, $\Delta_\mu \ell_\mu(\boldsymbol{r}) = 0$ at every site $\boldsymbol{r}$, is solved by

$$\ell_\mu(\boldsymbol{r}) = \epsilon_{\mu\nu} \Delta_\nu S(\boldsymbol{R}) \tag{19.207}$$

where $S(\tilde{\boldsymbol{r}}) \in \mathbb{Z}$ are a set of integer-valued variables defined on the sites $\boldsymbol{R}$ of the dual of the square lattice, i.e. the set of points at the center of the plaquettes of the square lattice. After solving the constraint, the partition function becomes that pof the discrete gaussian model

$$Z_{\rm DGM} = \sum_{\{S(\boldsymbol{R})\}} \exp\Big(-\frac{1}{2\beta} \sum_{\boldsymbol{R},\mu} (\Delta_\mu S(\boldsymbol{R}))^2 \Big) \tag{19.208}$$

Thus, we obtained the result that the dual of the XY model, which has a global $U(1)$ symmetry, is a theory defined on the dual lattice with integer-valued degrees of freedom, with a global symmetry, $S(\boldsymbol{R}) \to S(\boldsymbol{R}) + P$, where $P \in \mathbb{Z}$. This model is known as the discrete Gaussian model. As we will see is closely related to the vortices of the original problem. The configurations of the discrete variable $S(\boldsymbol{R})$ can be regarded as describing a discretized two-dimensional surface, such as the surface of a three-dimensional cubic crystal.

Therefore, the dual of the theory with a global $U(1)$ symmetry is a theory with a global $\mathbb{Z}$ symmetry. Moreover, under duality $\beta \to 1/\beta$. hence duality exchanges strong coupling with weak coupling. In summary, duality is the set of correspondences

$$U(1) \leftrightarrow \mathbb{Z}, \qquad \beta \leftrightarrow \frac{1}{\beta}, \qquad \text{direct lattice} \leftrightarrow \text{dual lattice} \tag{19.209}$$

It is instructive to see what is the dual of the order parameter field $\exp(i\theta(\boldsymbol{r}))$, and of its correlation function,

$$G(\boldsymbol{r}-\boldsymbol{r}') = \langle \exp(i\theta(\boldsymbol{r}))\exp(-i\theta(\boldsymbol{r}'))\rangle \tag{19.210}$$

Let $n = \pm 1$ at the sites $\boldsymbol{r}$ and $\boldsymbol{r}'$ where the operators are inserted, i.e.

$$n(\boldsymbol{x}) = \delta(\boldsymbol{x}-\boldsymbol{r}) - \delta(\boldsymbol{x}-\boldsymbol{r}') \tag{19.211}$$

In the presence of these insertions, the constraint on the discrete gauge fields in the partition function of Eq.(19.206) now becomes

$$\Delta_\mu \ell_\mu(\boldsymbol{r}) = n(\boldsymbol{r}) \tag{19.212}$$

The solution now is

$$\ell_\mu(\boldsymbol{r}) = \epsilon_{\mu\nu}\left(\Delta_\nu S(\boldsymbol{R}) + B_\nu(\boldsymbol{R})\right) \tag{19.213}$$

where $B_\mu(\boldsymbol{R})$ is any integer-valued vector field on the links of the dual lattice that obeys

$$\epsilon_{\mu\nu}\Delta_\mu B_\nu(\boldsymbol{R}) = n(\boldsymbol{x}) \tag{19.214}$$

Hence, under duality we have the correspondence

$$\text{charge} \leftrightarrow \text{flux} \tag{19.215}$$

which is a discrete version of electromagnetic duality. The correlator then is mapped to a ration of two partition functions in the discrete Gaussian model

$$\langle \exp(i\theta(\boldsymbol{r}))\exp(-i\theta(\boldsymbol{r}'))\rangle = \frac{Z_{\text{DGM}}[\boldsymbol{r},\boldsymbol{r}']}{Z_{\text{DGM}}} \tag{19.216}$$

where the numerator is the partition function of Eq.(19.208) with discrete fluxes ± 1 at $\boldsymbol{r}$ and $\boldsymbol{r}'$, respectively. The free energy cost of the fluxes, $\Delta F[\boldsymbol{r},\boldsymbol{r}']$, is

$$\Delta F[\boldsymbol{r},\boldsymbol{r}'] = -\frac{1}{\beta}\ln\langle \exp(i\theta(\boldsymbol{r}))\exp(-i\theta(\boldsymbol{r}'))\rangle \tag{19.217}$$

Hence, in the symmetric (disordered) phase where the correlator decays exponentially with distance, $G(\boldsymbol{r}-\boldsymbol{r}') \sim \exp(-|\boldsymbol{r}-\boldsymbol{r}'|/\xi)$, the free energy cost increases linearly with distance as $\Delta F[\boldsymbol{r},\boldsymbol{r}'] \sim \sigma|\boldsymbol{r}-\boldsymbol{r}'|$, with $\sigma = T/\xi$, and the fluxes are confined.

The discrete Gaussian model of Eq.(19.208) is a partition function over the integers. We will now use the Poisson summation formula to transform the partition function in two different and useful ways. The Poisson summation formula is the following identity for a series of a function $f(n)$, where $n \in \mathbb{Z}$,

$$\sum_{n\in\mathbb{Z}} f(n) = \lim_{y\to 0^+} \int_{-\infty}^{\infty} dx \sum_{m\in\mathbb{Z}} f(x)\, e^{2\pi i m x} e^{-ym^2} \tag{19.218}$$

where $y > 0$ plays the role of a convergence factor. Using the Poisson summation formula, the partition function Z_{DGM} of Eq.(19.208) becomes,

$$Z_{\text{DGM}} = \sum_{\{m(\boldsymbol{R})\}} \prod_{\boldsymbol{R}} \int_{-\infty}^{\infty} d\phi(\boldsymbol{R})\ \exp(-S_{\text{DGM}}(\phi, m)) \tag{19.219}$$

where

$$S_{\text{DGM}}(\phi, m) = \frac{1}{2\beta} \sum_{\boldsymbol{R},\mu} (\Delta_\mu \phi(\boldsymbol{R}))^2 + y \sum_{\boldsymbol{R}} m^2(\boldsymbol{R}) + 2\pi i \sum_{\boldsymbol{R}} m(\boldsymbol{R})\phi(\boldsymbol{R}) \tag{19.220}$$

We will now work on this result to obtain two useful mappings. The parameter y suppresses the contributions to the partition function of Eq.(19.219) of progressively large values on the variable $|m|$, we will make the approximation of keeping (for now) only the values $m = 0, \pm 1$. For y large enough, or $z = \exp(-y)$ small enough, we can approximate the sum over the m variables as

$$\begin{aligned}\sum_{m(\boldsymbol{R})\in\mathbb{Z}} \exp(-ym^2(\boldsymbol{R}) + 2\pi i m(\boldsymbol{R})\phi(\boldsymbol{R})) &\simeq 1 + 2z\cos(2\pi\phi(\boldsymbol{R})) + O(z^2) \\ &\simeq \exp(2z\cos(2\pi\phi(\boldsymbol{R})) + O(z^2))\end{aligned} \tag{19.221}$$

Then, the partition function of Eq.(19.219) reduces to

$$Z = \int_{-\infty}^{\infty} \prod_{\boldsymbol{R}} d\phi(\boldsymbol{R})\ \exp\Big(- \frac{1}{2\beta} \sum_{\boldsymbol{R},\mu} (\Delta_\mu\phi(\boldsymbol{R}))^2 + 2z \sum_{\boldsymbol{R}} \cos(2\pi\phi(\boldsymbol{R}))\Big) \tag{19.222}$$

which, in the formal continuum limit, $a \to 0$, becomes

$$Z \simeq\in \mathcal{D}\phi\ \exp(-\int d^2x\ \mathcal{L}_{\text{SG}}[\phi]) \tag{19.223}$$

where

$$\mathcal{L}_{\text{SG}} = \frac{T}{2}(\partial_\mu\phi)^2 - g\cos(2\pi\phi) \tag{19.224}$$

where $g = 4z/a^2$ (where a is the lattice spacing). Upon the change of variables $\varphi = \sqrt{T}\phi$, the Lagrangian takes the form

$$\mathcal{L} = \frac{1}{8\pi}(\partial_\mu \varphi)^2 - g\cos\left(\frac{\varphi}{R}\right) \tag{19.225}$$

This is the Lagrangian of the sine-Gordon theory in 1+1-dimensional Euclidean space-time, using the notation of section 21.6.1), with $R = \sqrt{T/\pi}$

The sine-Gordon theory is invariant under discrete uniform shifts $\phi(x) \mapsto \phi(x) + 2\pi n R$, where R is the compactification radius. This is a theory of a compactified boson, the φ field, with a periodic perturbation. The two-dimensional free compactified boson is a conformal field theory that we will discuss in some detail in Section 21.6.1.

For the present purposes it will suffice to say that the cosine operator is a vertex operator with scaling dimension $\Delta = 1/R^2$. In Chapter 15 we saw that, in two spacetime dimensions, an operator is irrelevant if $\Delta > 2$, which means that it is irrelevant if $\frac{1}{R^2} = \frac{\pi}{T} > 2$, or, what is the same for $T < T_c = \frac{\pi}{2}$. On the other hand, at this value of T the cosine operator is not only marginal but actually it is marginally relevant. Therefore the coupling constant g runs to strong coupling under the RG to a regime in which the cosine operator becomes dominant. Furthermore, in this regime the fieldφ will be pinned at the minima of the cosine potential, and its fluctuations have a mass gap.

The value $T_c = \frac{\pi}{2}$ that we just obtained is precisely the same which the energy-entropy Kosterlitz-Thouless argument (discussed in section 19.3.2) predicts that vortices will proliferate. We will now see that this is not an accident.

In section 19.3.2 we showed that the partition function for a complex scalar field in two Euclidean dimensions can be rewritten as the partition function of a neutral Coulomb gas, in which the charges are the vortices of the theory. We will use duality to see how does that come about. Our starting point will be the partition function written in the form of Eq.(19.219), where the action is given in Eq.(19.220). Instead of doing the sums over the integer-valued variables $m(\boldsymbol{R})$, we will instead integrate the continuous variables $\phi(\boldsymbol{R})$. The resulting form of the partition function is

$$\begin{aligned} Z =& Z_0 \sum_{\{m(\boldsymbol{R})\}} \delta(\sum_{\boldsymbol{R}} m(\boldsymbol{R}) = 0) \\ &\times \exp\left(- \sum_{\boldsymbol{R}} y\, m^2(\boldsymbol{R}) - \frac{1}{2}\sum_{\boldsymbol{R},\boldsymbol{R}'} \left(\frac{2\pi}{T}\right)^2 m(\boldsymbol{R}) G_{\text{reg}}(\boldsymbol{R}-\boldsymbol{R}') m(\boldsymbol{R}')\right) \end{aligned} \tag{19.226}$$

where $G_{\text{reg}}(\boldsymbol{R}-\boldsymbol{R}') = G(\boldsymbol{R}-\boldsymbol{R}')-G(0)$, and $G(\boldsymbol{R}-\boldsymbol{R}')$ is the lattice Green function, i.e. the solution of the difference equation

$$-\Delta^2_{\boldsymbol{R}} G(\boldsymbol{R}-\boldsymbol{R}') = \delta(\boldsymbol{R}-\boldsymbol{R}') \tag{19.227}$$

and $G(0)$ is infrared divergent. Here Δ^2 is the lattice Laplacian operator, i.e. $\Delta^2 f(\boldsymbol{R}) = \sum_\mu \Delta_\mu f(\boldsymbol{R})$, and Z_0 is the trivial partition function

$$Z_0 = \int_{-\infty}^{\infty} \prod_{\boldsymbol{R}} d\phi(\boldsymbol{R}) \exp\Big(-\frac{1}{2\beta} \sum_{\boldsymbol{R},\mu} (\Delta_\mu \phi(\boldsymbol{R}))^2\Big) \tag{19.228}$$

At long distances, $|\boldsymbol{R}-\boldsymbol{R}'| \gg a$, the lattice Green function approaches the continuum result

$$G(\boldsymbol{R}-\boldsymbol{R}') \simeq \frac{1}{2\pi} \ln\left(\frac{|\boldsymbol{R}-\boldsymbol{R}'|}{a}\right) \tag{19.229}$$

Putting it all together we recover the result that up to an uninteresting prefactor, the partition function is that of a gas of particles with both types of charges with a logarithmic interaction, the two-dimensional neutral Coulomb gas of Eq.(19.80).

The moral of this analysis is that, in a theory in two dimensions with a $U(1)$ global symmetry, the phase transition to the disordered state is driven by a process of vortex proliferation. In this language, the mass gap is equivalent to the fact that at high temperatures the neutral Coulomb gas experiences Debye screening.

19.10 Monopoles and Confinement in Compact Electrodynamics

We will now discuss Polyakov's compact quantum electrodynamics in $D = 3$ Euclidean dimensions. This is an abelian lattice gauge theory with a compact abelian gauge group $U(1)$. In section 19.7 we saw that this theory can be regarded as the low-energy limit of the Georgi-Glashow model. The main difference between this theory and Maxwell's electrodynamics is that it has instanton magnetic monopoles. We will see that in this theory magnetic monopoles play a similar role to the vortices in two dimensions. However, unlike the case of the vortices, we will see that the monopoles always proliferate and that this leads to confinement.

Let is consider a $U(1)$ abelian lattice gauge theory on a cubic lattice. Thus, as before, the degrees of freedom are vector fields defined on the links

$(\boldsymbol{r}, \mu)$ of the lattice (with $\mu = 1, 2, 3$). the partition function is

$$Z = \prod_{\boldsymbol{r},\mu} \int_0^{2\pi} \frac{dA_\mu(\boldsymbol{r})}{2\pi} \, e^{-S[A_\mu]} \tag{19.230}$$

where

$$S[A_\mu] = -\frac{1}{2g^2} \sum_{\boldsymbol{r},\mu\nu} \cos F_{\mu\nu}(\boldsymbol{R}) \tag{19.231}$$

Here

$$F_{\mu\nu}\boldsymbol{r}) = \sum_{\text{plaquette}} A_\mu(\boldsymbol{r}) \equiv \Delta_\mu A_\nu(\boldsymbol{r}) - \Delta_\nu A_\mu(\boldsymbol{r}) \tag{19.232}$$

is the flux on the plaquette $(\boldsymbol{r}, \mu\nu)$.

Since the action is invariant under periodic shifts, $A_\mu(\boldsymbol{r}) \to A_\mu(\boldsymbol{r}) + 2\pi\ell_\mu(\boldsymbol{r})$, the Bianchi identity for the field strength $F_{\mu\nu}(\boldsymbol{r})$ can only be satisfied modulo 2π. However, the operator $\exp(iF_{\mu\nu})$ obeys the Bianchi identity

$$\prod_{\text{cube faces}} e^{iF_{\mu\nu}} = 1 \tag{19.233}$$

Thus, this theory has magnetic monopoles with quantized magnetic charge.

We can analyze this theory by an extension of the duality we just used in $D = 2$. Our first step is to express the partition function as

$$Z = \int_{-\infty}^{\infty} \prod_{\boldsymbol{r},\mu\nu} \frac{dA_\mu(\boldsymbol{r})}{2\pi} \sum_{\{m_{\mu\nu}(\boldsymbol{r})\}} \exp\left(-\frac{1}{2g^2} (F_{\mu\nu}(\boldsymbol{r}) + 2\pi m_{\mu\nu}(\boldsymbol{r}))^2 \right) \tag{19.234}$$

where the gauge fields are no longer periodic, and $\{m_{\mu\nu}\}$ are integer-valued two-form variables (Kalb-Ramond fields) defined on the plaquettes. This expression for the partition function is invariant under the replacements

$$\begin{aligned} A_\mu &\to A_\mu + 2\pi\ell_\mu \\ F_{\mu\nu} &\to F_{\mu\nu} + 2\pi(\Delta_\mu \ell_\nu - \Delta_\nu \ell_\mu) \\ m_{\mu\nu} &\to m_{\mu\nu} + \Delta_\mu \ell_\nu - \Delta_\nu \ell_\mu \end{aligned} \tag{19.235}$$

which enforces periodicity. In this form, $F_{\mu\nu}$ does not have monopoles and satisfies the Bianchi identity. However, the two-form integer-valued fields $m_{\mu\nu}$ do not satisfy the Bianchi identity. In fact, the quantity

$$N(\boldsymbol{R}) = \frac{1}{2} \epsilon_{\mu\nu\lambda} \Delta_\mu m_{\nu\lambda} \in \mathbb{Z} \tag{19.236}$$

defined on the sites of the dual lattice (i.e. the centers if the cubes), measures their integer-valued violation of the Bianchi identity. Thus, the integers

$\{m_{\mu\nu}\}$ defined on the plaquettes, represent Dirac strings of a configuration of monopoles.

The partition function now is

$$Z = \sum_{\{m_{\mu\nu}(\boldsymbol{r})\}} \int \mathcal{D}A_\mu \ \exp(-\frac{g^2}{4} \sum_{\boldsymbol{r},\mu\nu} m_{\mu\nu}^2(\boldsymbol{r}) + i \sum_{\boldsymbol{r},\mu} A_\mu(\boldsymbol{r}) \Delta_\nu m_{\mu\nu}(\boldsymbol{r})) \tag{19.237}$$

The integral over the gauge fields now leads to the constraint

$$\Delta_\mu m_{\mu\nu} = 0 \tag{19.238}$$

on every plaquette of the cubic lattice. We will solve this constraint in terms of the integer-valued field $S(\boldsymbol{R})$ defined on the sites of the dual lattice (the centers of the elementary cubes),

$$m_{\mu\nu}(\boldsymbol{r}) = \epsilon_{\mu\nu\lambda} \Delta_\lambda S(\boldsymbol{R}) \tag{19.239}$$

The partition function reduces to the following

$$\begin{aligned} Z &= \sum_{\{m_{\mu\nu}(\boldsymbol{r})\}} \prod_{\boldsymbol{r},\mu} \delta(\Delta_\mu m_{\mu\nu}) \exp(-\sum_{\boldsymbol{r},\mu\nu} \frac{g^2}{4} m_{\mu\nu}^2(\boldsymbol{r})) \\ &= \sum_{\{S(\boldsymbol{R})\}} \exp(-\frac{g^2}{2} \sum_{\boldsymbol{R},\mu} (\Delta_\mu S(\boldsymbol{R}))^2) \end{aligned} \tag{19.240}$$

Therefore, in close analogy with what we found for vortices in $D = 2$, we find that the dual of compact electrodynamics is the discrete Gaussian model, now instead in $D = 3$ dimensions!

We can now repeat almost verbatim what we just did in two dimensions, and conclude that compact electrodynamics is dual to the three-dimensional sine-Gordon theory and to the three-dimensional (charge-neutral) Coulomb gas, defined in terms of integer valued variables $m(\boldsymbol{R})$ defined on the centers of the cubes (the monopoles),

$$Z_{\rm CG} = \sum_{\{m(\boldsymbol{R})\} \in \mathbb{Z}} \exp\left(-\frac{1}{2} \sum_{\boldsymbol{R},\boldsymbol{R}'} \left(\frac{2\pi}{g}\right)^2 m(\boldsymbol{R}) G(\boldsymbol{R} - \boldsymbol{R}') m(\boldsymbol{R}') \right) \tag{19.241}$$

where $G(\boldsymbol{R} - \boldsymbol{R}')$ is the three-dimensional Green function which, at long distances, behaves as

$$G(\boldsymbol{R} - \boldsymbol{R}') = \frac{a}{4\pi |\boldsymbol{R} - \boldsymbol{R}'|} \tag{19.242}$$

where a is the lattice spacing. So, now we find a three-dimensional Coulomb gas.

However, the physics is very different. In contrast with the $D = 2$ case, the energy (Euclidean action) of a monopole is now *infrared finite*. The energy-entropy argument then implies that the entropy, which goes as $\ln L^3$ in $D = 3$, always wins over the energy which is finite. This means that the monopoles always proliferate and that they are in the plasma phase. This conclusion also follows form the sine-Gordon picture where one finds that the cosine operator is always relevant in all dimensions $D > 2$.

A straightforward calculation shows that the Wilson loop always the area law for all values of the coupling constant g. To see this we need to find the dual of the Wilson loop operator of a closed loop Γ which, for simplicity, we will take to be planar,

$$W[\Gamma] = \langle \exp(i \oint_\Gamma A_\mu) \rangle \tag{19.243}$$

The insertion of this operator modifies the constraint to

$$\Delta_\nu m_{\mu\nu}(\boldsymbol{r}) = j_\mu^\Gamma(\boldsymbol{r}) \tag{19.244}$$

where j_μ^Γ is the current, that defines the Wilson loop, on the closed contour Γ of the lattice (the boundary of an open surface Σ.) The solution to this constraint is

$$m_{\mu\nu}(\boldsymbol{r}) = \epsilon_{\mu\nu\lambda}(\Delta_\lambda S(\boldsymbol{R}) + B_\lambda(\boldsymbol{R})) \tag{19.245}$$

where $B_\lambda(\boldsymbol{R})$ is an integer-valued gauge field defined on the links of the dual lattice such that

$$\epsilon_{\mu\nu\lambda}\Delta_\nu B_\lambda(\boldsymbol{R}) = j_\mu^\Gamma(\boldsymbol{r}) \tag{19.246}$$

which has the form of a discrete version of Ampére's law. A solution of this constraint is to set $B_\lambda = 1$ on all links of the dual lattice piercing the surface Σ.

In the dual theory, the expectation value of the Wilson loop is

$$W[\Gamma] = \frac{1}{Z} \sum_{\{S(\boldsymbol{R})\}} \exp\left(-\frac{g^2}{2} \sum_{\boldsymbol{R},\mu} (\Delta_\mu S(\boldsymbol{R}) + B_\mu(\boldsymbol{R}))^2 \right) \tag{19.247}$$

where Z is the partition function of Eq.(19.240). This expectation value can be computed wither in the three-dimensional sine-Gordon formulation, or in the three-dimensional Coulomb gas. In the latter representation, the expectation value of the Wilson loop is given by

$$W[\Gamma] = W[\Gamma]_{\text{Maxwell}} \times \left\langle \exp\left(4\pi i \sum_{\boldsymbol{R},\boldsymbol{R}'} \Delta_\mu B_\mu(\boldsymbol{R}) G(\boldsymbol{R} - \boldsymbol{R}') m(\boldsymbol{R}')\right)\right\rangle_{\text{CG}} \tag{19.248}$$

where $\langle \mathcal{O}[m] \rangle_{\text{CG}}$ denotes the expectation value of the operator $\mathcal{O}[m]$ is the three-dimensional Coulomb gas (whose partition function is given in Eq.(19.241)), and

$$W[\Gamma]_{\text{Maxwell}} = \exp\left(-\frac{g^2}{2} \sum_{\boldsymbol{r},\boldsymbol{r}'} j_\mu^\Gamma(\boldsymbol{r}) G(\boldsymbol{r}-\boldsymbol{r}') j_\mu^\Gamma(\boldsymbol{r}') \right) \tag{19.249}$$

is the expectation value of the Wilson loop in the Maxwell theory, i.e. the abelian gauge theory without magnetic monopoles, which are represented by the integer-valued degrees of freedom $\{m(\boldsymbol{R}\}$. Recall that in section 9.7 we calculated $W[\Gamma]_{\text{Maxwell}}$, and showed that the effective interaction between two static sources obeys the Coulomb law, $\sim 1/R^{d-2}$, which in $2+1$ dimensions is a logarithmic interaction.

The Coulomb gas expectation value shown in Eq.(19.248) is the contribution of the monopoles. The quantity $\Delta_\mu B_\mu(\boldsymbol{R})$ is non-zero only on the links of the dual lattice piercing the surface Σ (where $\Gamma = \partial\Sigma$), and it is equal toe $+1$ on the sited of the dual lattice just above Σ, and -1 on the sites just below Σ. Hence, $\Delta_\mu B_\mu(\boldsymbol{R})$ is an uniform density of lattice dipoles perpendicular to Σ. In other words, this is essentially the negative of the self-energy of this uniform dipole distribution. A straightforward calculation shows that, for a very large Wilson loop, this expectation value behaves as $\exp(-\sigma\mathcal{A}[\Sigma])$, where $\mathcal{A}[\Sigma]$ is the minimal area spanned by the loop Γ, with a string tension $\sigma \propto g^2$.

Therefore, we conclude that in compact electrodynamics in 2+ dimensions, large Wilson loops obeys the area law and that small loops obey the Maxwell law. The crossover happens at a length scale ξ, the confinement scale, that in the Coulomb gas is the Debye screening length. On scales large compared to the screening length the discreteness of the Coulomb charges is negligible. In this sense, confinement is the result of the proliferation (or "condensation") of magnetic monopoles.

What is the role on monopoles in higher dimensions and in other theories? In $d = 4$ dimensions compact QED has monopole solitons with finite energy. In the Euclidean path-integral they are represented by closed loops, much in the same way as a vortex loops in $2+1$ dimensional theories with a global $U(1)$ symmetry, e.g. the $D = 3$ dimensional classical XY model. However, while the three-dimensional XY model has a continuous phase transition, the numerical evidence suggest the compact QED in four dimensions has a first order confinement transition, and hence, is not described by a fixed point. On the other hand, although non-abelian theories have monopole instantons in $D = 4$ dimensions, as we saw earlier in this chapter, aside from suggestive numerical evidence, it has not been possible to show that confinement is actually driven by a proliferation (or condensation) or magnetic monopoles.