Chapter 1

$\pi_0(\mathcal{Q})$ and solitons

A soliton is a classical solution of nonlinear field equations which (1) is non-singular, (2) has finite energy and (3) is localized in space. We will only consider static solitons. In this case the field equations can be obtained by varying a functional that we will call the static energy. In some cases, the solitons are local minima of the static energy, and are separated from the absolute minimum (the vacuum) by a finite energy barrier. Such solitons are called “nontopological solitons”. We will only be interested in another class of solitons, which either cannot be deformed continuously into the vacuum, or if they can, are separated from the vacuum by an infinite energy barrier. Such solitons are called “topological solitons”.

In order to make this concept mathematically more precise, it is convenient to think of a field theory as a mechanical system with an infinite dimensional configuration space. Let us define the classical configuration space of the theory, $\mathcal{Q}$, to be the space of smooth, finite energy configurations of the field at some instant of time. Note that $\mathcal{Q}$ defines the kinematics of the theory, but also knows about the form of the energy. The theories that we will consider in this chapter will have the common characteristic that their configuration space is not connected. Instead, it will be the disjoint union of several connected components, indexed by a set $\pi_0(\mathcal{Q})$ (the reason for this notation is explained in Appendix A):

$$\mathcal{Q} = \bigcup_{i \in \pi_0(\mathcal{Q})} \mathcal{Q}_i,$$

where $\mathcal{Q}_i$ are connected. Having determined the structure of the configuration space, the natural problem will be to find (if it exists) the absolute
minimum of the static energy in each connected component. Such minima will automatically be solutions of the classical equations of motion. The minimum of the energy in some connected components will be the classical vacuum configuration, but in others it will correspond to non trivial solutions; these will be our topological solitons.

The nonconnectedness of the configuration space $Q$ will manifest itself analytically in the existence of a conserved current known as the topological current. This current is not related to any symmetry of the theory and is identically conserved, i.e. it is conserved without making use of the equations of motion. (By contrast, Noether currents are conserved only upon using the equations of motion). Associated to the topological current is the topological charge, which is a functional on $Q$ that is locally constant. It is zero in the connected components containing the vacuum, and nonzero in those containing solitons.

The above definition of soliton is tailored to describe a classical extended particle. When the theory is quantized, the solitons behave like a new species of particles, in addition to the perturbative particle states of the field. This can be seen in various ways. In these lectures we will often find it convenient to think of a quantum field theory as the quantum mechanics of a system with configuration space $Q$. This is a formal definition that would require more technical work to be made precise, but is useful for heuristic considerations. In the Schrödinger picture, the wave functions are complex functionals on $Q$. If $Q$ has several connected components, the Hilbert space $\mathcal{H}$ will split into subspaces called the topological sectors:

$$\mathcal{H} = \bigoplus_{i \in \pi_0(Q)} \mathcal{H}_i,$$

where $\mathcal{H}_i$ consists of wave functionals which are nonzero only on $Q_i$. Each subspace $\mathcal{H}_i$ will be an eigenspace of the topological charge with eigenvalue $i$. It is clear that with any sensible definition of the measure the spaces $\mathcal{H}_i$ will be orthogonal to each other. The topological charge therefore defines a superselection rule: if the state vector belongs initially to the subspace $\mathcal{H}_i$, it will never leave it in the course of the time evolution. This fact can also be easily understood from the point of view of Feynman’s path integral, because there are no paths joining $Q_i$ to $Q_j$ when $i \neq j$, so the transition amplitude between states in different sectors must vanish.
1.1 Scalar solitons in 1 + 1 dimensions

1.1.1 Classical kinks

We begin by discussing the simplest case, that of a single scalar field in one space dimension, with action:

\[ S(\phi) = \int d^2x \left[ -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi) \right] \]  

(1.1)

with \( \partial_\mu \phi \partial^\mu \phi = -(\partial_0 \phi)^2 + (\partial_1 \phi)^2 \). We demand that the potential \( V \) be bounded from below, and we assume without loss of generality that the minimum value of \( V \) be zero. We call \( y_i, i \in J \), the minimum points. For definiteness one can think of the quartic potential

\[ V = -\frac{1}{2} m^2 \phi^2 + \frac{\lambda}{4} \phi^4 + \frac{m^4}{4\lambda} = \frac{\lambda}{4} (\phi^2 - f^2)^2, \]  

(1.2)

with \( f = \frac{m}{\sqrt{\lambda}} \) and \( m \) real and positive, with minima at points \( y_{\pm} = \pm f \).

With these assumptions, the energy:

\[ E = \int_{-\infty}^{\infty} dx \left[ \frac{1}{2} (\partial_0 \phi)^2 + \frac{1}{2} (\partial_1 \phi)^2 + V(\phi) \right] \]  

(1.3)

is positive semidefinite, and is zero only for the constant field configurations \( \phi(x, t) = y_i \). These are the absolute minima of \( E \); they are the classical vacua of the theory. Note that in (1.3) the first term represents the kinetic energy; the rest

\[ E_S = \int_{-\infty}^{+\infty} dx \left[ \frac{1}{2} (\partial_1 \phi)^2 + V(\phi) \right] \]  

(1.4)

will be called “static energy”. We will reserve the name “potential energy” for the second term in \( E_S \), while the first term could be appropriately called “elastic energy”.

The field \( \phi \) belongs to the space \( \Gamma(\mathbb{R}, \mathbb{R}) \) of smooth real functions of one variable. (In general we will use the notation \( \Gamma(X, Y) \) for the space of smooth maps from \( X \) to \( Y \), where \( X \) and \( Y \) are manifolds. This space is itself an infinite dimensional smooth manifold. See Appendix E) Finiteness of the energy demands that when \( |x| \) tends to infinity \( \phi \) tends to one of the
classical vacua, for otherwise the last two terms in $E$ would diverge. We will call $Q$ the subspace of $\Gamma(\mathbb{R}, \mathbb{R})$ for which the static energy $E_S$ is finite.

If $V$ has more than one minimum, $Q$ will not be connected. In fact, let

$$Q = \bigcup_{i,j} Q_{ij}, \quad Q_{ij} = \{ \phi \in Q \mid \phi \to y_i, \phi \to y_j \}.$$ 

Every path in $\Gamma(\mathbb{R}, \mathbb{R})$ joining $Q_{ij}$ to $Q_{i'j'}$ (with $ij \neq i'j'$) must necessarily pass through the complement of $Q$. In fact, to change the asymptotic behaviour of $\phi$ one has to go through fields which do not tend to one of the minima at infinity, and these have infinite energy. So, the spaces $Q_{ij}$ are separated by infinite energy barriers. For example in the case of the potential (1.2) there are four connected components of $Q$, labelled $Q_{++}, Q_{+-}, Q_{-+}, Q_{--}$. In general, the set $\pi_0(Q)$ of connected components of $Q$ is the cartesian product of two copies of the set indexing the minima: $\pi_0(Q) = \mathcal{J} \times \mathcal{J}$.

Every $\phi \in Q_{ij}$ can be written as the sum of an arbitrary given $\phi_0 \in Q_{ij}$ (which we call the “basepoint” of $Q_{ij}$) plus a function $\psi$ which tends asymptotically to zero at $\pm \infty$. The function $\psi$ can be regarded as a function $S^1 \to \mathbb{R}$, where $S^1 = \mathbb{R} \cup \{\infty\}$ is the one-point compactification of space.

The natural problem is then to find the minimum of the energy in each connected component, if it exists. It is clear that in the connected components $Q_{ii}$ the minima are the constant fields $\phi = y_i$. These are also the absolute minima of $E$ on all $Q$. In the case of the potential (1.2), one can easily convince oneself by means of the following qualitative argument that with the dynamics considered above there are going to be absolute minima of the static energy also in the sectors $Q_{-+}$ and $Q_{+-}$. Let us denote $\ell$ the “size of the soliton”, i.e. the length of the region where the field is significantly different from either vacua. It is clear that the elastic energy is of order $f^2/\ell$, and hence decreases with $\ell$, while the potential energy is of order $\lambda f^4 \ell$, and hence increases with $\ell$. The static energy will have a minimum at some finite value of order $\ell \approx 1/(\sqrt{\lambda} f)$. Inserting in the formula for the energy we also find that both elastic and potential energy of the soliton are of order $\sqrt{\lambda} f^3$. The soliton will therefore be the result of a balance between elastic and potential energy.
In order to find the explicit form of the soliton we have to solve the differential equation

$$\frac{d^2 \phi}{dx^2} = \frac{\partial V}{\partial \phi}$$

(1.5)

with the appropriate boundary conditions. For the potential (1.2) the solutions of (1.5) in the sectors $Q_{\pm}$ and $Q_{\pm}$ are

$$\phi(x) = \pm \frac{m}{\sqrt{\lambda}} \tanh \left[ \frac{m}{\sqrt{2}} (x - x_0) \right] ,$$

(1.6)

with the upper sign in the first case, the lower sign in the second. These solutions are known as the “kink” and the “antikink” respectively. Note that these solitons are not isolated solutions: they come in one-parameter families, parametrized by the “center of mass” coordinate $x_0$. This is a reflection of the translational invariance of the action. Figure (1.1) shows a plot of $\phi/f$ as a function of $x\sqrt{2}/m$ for the kink at $x_0 = 0$. (The horizontal lines correspond to the minima of the potential.)

Inserting (1.6) in (1.4) we obtain

$$E_S = \frac{2\sqrt{2}m^3}{3\lambda} = \frac{2\sqrt{2}}{3} f^3 \sqrt{\lambda} .$$

(1.7)

It is useful to note that there is equipartition between elastic and potential energy (i.e. each of the two terms in (1.4) contributes exactly $E_S/2$). To see this, multiply both sides of the equation of motion (1.5) by $d\phi/dx$. The resulting equation can be written

$$\frac{d}{dx} \left[ \frac{1}{2} \left( \frac{d\phi}{dx} \right)^2 - V \right] = 0 ,$$

implying that the quantity in square brackets is constant. We can evaluate the constant for $x \pm \infty$, and we find it must be zero. Thus, the density of elastic energy and the density of potential energy are equal. In particular, the total elastic and potential energies are equal.

In the theory with potential (1.2), consider the current

$$J_T^\mu = \frac{1}{2f} \varepsilon^{\mu\nu} \partial_\nu \phi ;$$

(1.8)

clearly we have

$$\partial_\mu J_T^\mu = 0 .$$

(1.9)
This current is conserved without recourse to the equations of motion, and it
is not related to any symmetry of the theory. It will be called the topological
current. The integral
\[ Q_T = \int_{-\infty}^{\infty} dx \; J_T^0 = \frac{1}{2f} [\phi(+\infty) - \phi(-\infty)] \]  
(1.10)
is known as the topological charge. It is clear that all fields in \( Q_{-} \) have
\( Q_T = 1 \), those in \( Q_{+} \) have \( Q_T = -1 \) and those in \( Q_{++} \) and \( Q_{--} \) have \( Q_T = 0 \).
Thus \( Q_T \) is a measure of the nontriviality of the boundary conditions of the
fields.

Another interesting potential is
\[ V(\phi) = \frac{m^4}{\lambda} \left[ 1 - \cos \left( \frac{\sqrt{\lambda}}{m} \phi \right) \right] . \]  
(1.11)
This corresponds to the so called “sine-Gordon” model. The indexing set of
minima is the set of the integers \( \mathcal{J} = \mathbb{Z} \), so there is a double infinity \((\mathbb{Z} \times \mathbb{Z})\) of
connected components. The topological current and the topological charge
are given again by (1.8) and (1.10), where \( f \), which is half the distance
between two successive minima of the potential, is now equal to \( \pi m/\sqrt{\lambda} \).
We give the form of the solitons with \( Q_T = \pm 1 \), which minimize the energy
in \( Q_{0\,1} \) and \( Q_{0\,-1} \)
\[ \phi(x) = \pm \frac{4m}{\sqrt{\lambda}} \arctan \left\{ \exp \left[ (x - x_0) m \right] \right\} \]  
(1.12)
1.1. SCALAR SOLITONS IN 1 + 1 DIMENSIONS

This solution is plotted in the Figure (1.2).

Just adding $2n f$ we get the soliton and antisoliton, still with $Q_T = \pm 1$, which minimize the energy in $Q_{nn+1}$ and $Q_{nn-1}$. Note that if in the field equation (1.5) with the potential (1.11) we reinterpret $x$ as time and $\phi$ as the coordinate of a particle on a line, then we can regard it as Newton’s equation of motion of the particle moving in the gravitational potential $-V$. Formula (1.12) represents a motion in which the particle rolls from one maximum of the gravitational potential to the next. Using this analogy it becomes intuitively clear that there cannot be any static soliton of the sine-Gordon model with $|Q_T| > 1$.

Note that this reinterpretation links a field theory in 1 + 1 dimensions to mechanics, regarded as a field theory in 0 + 1 dimensions. In chapter 2 we shall frequently use this trick of relating theories differing by one in dimension.

1.1.2 Quantum kinks

In this section we consider the quantum version of the $\phi^4$ theory with potential (1.2). This simple example already exhibits all the phenomena that characterize quantum solitons of more complicated systems.

We begin from the topologically trivial sectors. We can write a functional integral over fields $\phi \in Q_{++}$ as an integral over the shifted field $\varphi = \phi - f$, that has trivial boundary conditions $\varphi \to 0$ for $x \to \pm \infty$. 
The standard perturbative quantization procedure applied to the small fluctuations around the vacuum state $\phi = 0$ gives a Fock space of scalar particles, that we shall call “pions” with mass

$$m_\pi = \sqrt{V''(f)} = \sqrt{2m}.$$ 

Note that $\phi$ is dimensionless and $\lambda$ has dimensions of mass squared. In this theory weak coupling means $\lambda \ll m_\pi^2$.

The theory is superrenormalizable. The only divergence is logarithmic and renormalizes the pion mass, see figure (1.3). Evaluation of this diagram gives for the renormalized mass

$$m_{\pi R}^2 = m_\pi^2 - \frac{3\lambda}{2\pi} \log \left( \frac{\Lambda^2}{m_\pi^2} \right), \quad (1.13)$$

where we employed a simple momentum cutoff $\Lambda$.

The theory also contains kinks, that we can view as another type of particles. From (1.7) these particles have mass

$$m_k = \frac{m_\pi^3}{3\lambda}. \quad (1.14)$$

Taking the ratio to the pion mass we see that the solitons are much heavier than the pions at weak coupling.

Now one wonders what will become of (1.14) when the pion mass is renormalized. In order to answer this question we will now calculate the quantum corrections to the soliton mass. In the course of this calculation we will learn also several other interesting features of quantum solitons.

The path integral of the theory contains four distinct sectors, corresponding to paths that lie in each of the four connected components of configurations space $Q_{++}, Q_{+-}, Q_{-+}, Q_{--}$. Standard perturbation theory corresponds to the first or the last of these path integrals. We now consider the other two.
1.1. SCALAR SOLITONS IN 1 + 1 DIMENSIONS

The lowest energy state in $Q_{-+}$ is given by the kink, so the “vacuum-to-vacuum” amplitude is the sum over fields that are continuous deformations of the kink. We decompose

$$\phi(x) = \bar{\phi}(x) + \eta(x) ,$$

where $\bar{\phi}$ is the static solution (1.6), treated as a classical background, and $\eta$ is the quantum field.

Let us expand the action around the background:

$$S(\phi) = \int dt \left[ \frac{1}{2} \int dx \left( \frac{d\phi}{dt} \right)^2 - E_S(\phi) \right]$$

$$= S(\bar{\phi}) + \int dt \int dx \left[ \frac{1}{2} \left( \frac{d\eta}{dt} \right)^2 - \frac{1}{2} \eta L \eta - \lambda \left( \bar{\phi} \eta^3 + \frac{1}{4} \eta^4 \right) \right]$$

(1.15)

where

$$L = -\frac{d^2}{dx^2} + V''(\bar{\phi})$$

(1.16)

is essentially the second functional derivative of $E_S$ at $\bar{\phi}$.

Note that the terms on the r.h.s. of (1.15) are ordered in powers of $\lambda$: the term $S(\bar{\phi}) = -m_k \int dt$ is of order $\lambda^{-1}$ and hence non-perturbative; the first two terms in the square bracket are of order $\lambda^0$, the term cubic in $\eta$ is of order $\sqrt{\lambda}$ ($\bar{\phi}$ contains a factor $\lambda^{-1/2}$) and the term quartic in $\eta$ is of order $\lambda$. We are going to evaluate quantum corrections at order $\lambda^0$, which is equivalent to a standard saddle point (one-loop) evaluation of the path integral.

Most of the complications of this problem derive from the fact that the “mass” term in the operator $L$ is actually a function of $x$: $V''(\bar{\phi}) = \lambda (3\bar{\phi}^2 - f^2)$

$$= m^2 \left( -1 + 3 \tanh^2 \left( \frac{mx}{\sqrt{2}} \right) \right) .$$

(1.17)

This function is shown in Figure (1.4). Away from the position of the kink it tends quickly to $m^2_\pi$, but near the kink it has a dip and becomes even negative.
The operator $L$ is a self-adjoint, second order differential operator and therefore its eigenfunctions $\eta_n$ form a basis for the space of square-integrable functions on the real line:

$$L\eta_n = \omega_n^2 \eta_n ; \quad \int dx \eta_n(x) \eta_m(x) = \delta_{nm} .$$

(1.18)

We have used a notation that is appropriate to a discrete spectrum, as would be obtained if the system was put in a box, but in infinite space the spectrum is actually mixed and consists of the following:

- an isolated eigenvalue $\omega_0^2 = 0$ with eigenfunction $\eta_0 = \frac{1}{\cosh^2 \left( \frac{mx}{\sqrt{2}} \right)}$;

- an isolated eigenvalue $\omega_1^2 = \frac{3}{2} m^2$ with eigenfunction $\eta_1 = \frac{\sinh \left( \frac{mx}{\sqrt{2}} \right)}{\cosh^2 \left( \frac{mx}{\sqrt{2}} \right)}$ describing an excited state of the kink;

- a continuous spectrum with eigenvalues $\omega_p^2 = m^2 + p^2$, with $-\infty < p < \infty$ describing scattering states of pions in the background of the kink.

We observe that in the absence of the kink there would just be the continuous spectrum with eigenvalues $\omega^2 = m^2 + p^2$, with $-\infty < p < \infty$. Each eigenvalue corresponds to a normal mode of the field with momentum $p$, which is $e^{ipx}$. The presence of the kink deforms the spectrum but in a rather simple way. Mathematically, solving the eigenvalue equation (1.18) is
equivalent to solving the Schrödinger equation for a particle moving in the potential (1.17). A “right-moving” mode with momentum $p > 0$ is given for large negative $x$ by $\bar{\eta}_p(x) \approx e^{ipx}$. Near the soliton the solution is more complicated, but it must have again a similar form for large positive $x$. It turns out that there is no reflected wave, and the transmitted wave, for large positive $x$ is simply

$$\bar{\eta}_p(x) \approx e^{ipx+\delta_p}$$  \hspace{1cm} (1.19)

where the phase shift is given by

$$e^{i\delta_p} = \left( \frac{1 + ip/m}{1 - ip/m} \right) \left( \frac{1 + 2ip/m}{1 - 2ip/m} \right).$$  \hspace{1cm} (1.20)

This is left as an exercise (see Exercise XXX). There is another eigenfunction with the same eigenvalue $\omega^2$, which is given by the “left-mover” $\eta_p(-x)$. The general solution with given $p$ is a linear combination of left- and right-moving waves:

$$A\bar{\eta}_p(x) + B\bar{\eta}_p(-x).$$  \hspace{1cm} (1.21)

At this point we put the system in a box of size $L \gg m^{-1}$ and impose boundary conditions on the pions, discretizing the continuous part of the spectrum. Imposing that (1.21) vanishes at $x = \pm L/2$ leads to $A = \pm B$ and $\bar{\eta}_p(L/2) = \pm \bar{\eta}_p(-L/2)$. Then, using the asymptotic behavior of the solutions, one obtains $\exp(ipL - i\delta_p) = \pm 1$, or

$$p = \bar{p}_n \equiv \frac{\pi n}{L} + \frac{\delta_{pn}}{L} \quad \text{with} \quad n = 0, 1, 2 \ldots$$

We denote $\tilde{\omega}_n^2 = m^2_\pi + \bar{p}_n^2$ the corresponding eigenvalues. We denote

$$p_n = \frac{\pi n}{L} \quad \text{with} \quad n = 0, 1, 2 \ldots$$

the momenta, and $\omega_n^2 = m^2_\pi + p_n^2$ the eigenvalues, in the absence of the kink.

It is natural to expand the quantum field $\eta$ on the basis of eigenfunctions of $L$, instead of ordinary Fourier modes:

$$\eta(t, x) = b_0(t)\eta_0(x) + b_1(t)\eta_1(x) + \sum_{n=0}^{\infty} a_n(t)\bar{\eta}_n(x),$$  \hspace{1cm} (1.22)

---

1See Morse and Feschbach, eq (12.3.22) and following.
where the first two terms correspond to the isolated modes and the sum to the “continuous” spectrum. Then, the quadratic part of the Hamiltonian of the fluctuation field becomes a sum of independent oscillators:

\[ H = \int dx \left[ \frac{1}{2} \dot{\eta}^2 + \frac{1}{2} \eta L \eta \right] \]

\[ = \frac{1}{2} b_0^2 + \frac{1}{2} \left( b_1^2 + \omega_1^2 a_1^2 \right) + \frac{1}{2} \sum_{n=1}^{\infty} \left( \bar{\omega}_n^2 a_n^2 \right). \] (1.23)

By choosing to work in the basis of eigenfunctions of \( L \) we have decomposed the system into infinitely many decoupled oscillators.

There is only one odd degree of freedom, namely the zero mode, which is not an oscillator. Since the potential for this mode is zero, its wave function will not remain localized near the center of the soliton. Recall that the semiclassical approximation rests on the assumption that the quadratic term in the Lagrangian is dominant with respect to the quartic one:

\[ \omega^2 \langle q^2 \rangle \gg \lambda \langle q^4 \rangle, \]

where \( q \) is to be identified with one of the normal modes. This is true for all the oscillator states, but not for the zero mode.

The physical origin of the zero mode can be understood by noting that \( \eta_0 \) proportional to the derivative of the classical solution (1.6). Among all possible deformations of the kink field, there is one that corresponds simply to an infinitesimal translation of the kink by \( \delta x \):

\[ \delta \phi(x) = \delta x d\phi \]

Such a deformation does not change the energy, because a translated kink is a solution of the field equations with the same energy as the original kink. This particular direction in the functional space of the fields corresponds to the bottom of a flat valley for the energy.

This suggests that instead of the zero mode \( b_0 \), which amounts to infinitesimal translations of a kink with a fixed center, we take the position of the center of the kink as a dynamical variable. To study the quantization of the center of the kink, let us consider a slowly moving kink, which can be described by the solution (1.6) with \( x_0 \) replaced by \( x_0(t) \): \( \phi(x, t) =  \phi(x - x_0(t)) \).
1.1. SCALAR SOLITONS IN 1 + 1 DIMENSIONS

2 Inserting in the action we find

\[ S = \int dt \, dx \left( \frac{1}{2} \dot{\phi}^2 - \frac{1}{2} \phi'^2 - V \right) \]

\[ = \int dt \left[ \frac{1}{2} \dot{x}_0^2 \frac{1}{2} \int dx \phi'^2 - \int dx \left( \frac{1}{2} \phi'^2 + V \right) \right], \]

where a prime denotes derivative with respect to \( x \). Now we recall that the energy of the kink is equally divided between elastic and potential energy. Thus the coefficient of \( \dot{x}_0^2 \) is \( m_k/2 \) and the second integral is \( m_k \):

\[ S = \int dt \left[ \frac{1}{2} m_k \dot{x}_0^2 - m_k \right]. \]

(1.24)

The corresponding Hamiltonian is therefore that of a free particle with mass \( m_k \):

\[ H = m_k + \frac{p^2}{2m_k}. \]

(1.25)

This collective degree of freedom can be quantized simply imposing the standard commutation relation \([x_0, p] = i\hbar\).

3 When the motion of the kink is taken into account in this way, we can remove the zero mode from the list of the degrees of freedom.

Then, the energy of the quantum state describing a kink at rest, with the pion field in the Fock vacuum, is given by

\[ H = m_k + \frac{\sqrt{3}}{4} m_\pi + \sum_{n=0}^{\infty} \frac{1}{2} \omega_n, \]

(1.26)

where the first term is energy of the classical solution, the second is the vacuum energy of the isolated non-zero mode and the sum extends on the vacuum energy of all the oscillators in the discretized continuous spectrum.

2 The condition of slow motion is necessary to ensure that the classical field remains at least approximately a solution of the equations of motion. Since the field equations are Lorentz-invariant, a kink in motion will be obtained by operating on the static kink with a boost, and not simply by giving a time dependence to its center. For sufficiently low velocity, however, the two coincide.

3 In the functional integral the transformation of the integration variable from \( a_0 \) to \( x_0 \) has to be accompanied by a Jacobian. We will not need to compute it here, but it will play a role later in other models.
For large \( n \), \( \bar{\omega}_n \sim n \), so the sum is quadratically divergent. This is the usual divergent contribution to the vacuum energy that one also encounters in any quantum field theory. It is also present in the vacuum sector \( \mathcal{Q}-- \). We are thus led to define the renormalization of the kink mass as the difference between the sum of the vacuum energies of all the oscillators in the presence of the kink and the sum of the vacuum energies of all the oscillators in the absence of the kink. Both sums are quadratically divergent, and in the difference this divergence is cancelled. The renormalization of the kink mass is therefore

\[
\delta m_k = \frac{\sqrt{3}}{4} m_\pi + \frac{1}{2} \sum_{n=1}^{\infty} (\bar{\omega}_n - \omega_n) = \frac{\sqrt{3}}{4} m_\pi + \frac{1}{2} \sum_{n=1}^{\infty} \frac{p_n \delta_p}{L \omega_n},
\]

where, in view of taking the limit \( L \to \infty \), in the second step we expanded:

\[
\bar{p}_n^2 = p_n^2 + 2 \frac{\delta p}{L} p_n + O(1/L^2).
\]

At this point we can take the limit \( L \to \infty \) and we return to continuous momenta:

\[
\delta m_k = \frac{\sqrt{3}}{4} m_\pi + \frac{1}{2\pi} \int dp \frac{p \delta_p}{\sqrt{m_\pi^2 + p^2}}
= \frac{\sqrt{3}}{4} m_\pi + \frac{1}{2\pi} \lim_{\Lambda \to \infty} \delta_p \sqrt{m_\pi^2 + p^2} \biggl|_0^\Lambda - \frac{1}{2\pi} \int dp \sqrt{m_\pi^2 + p^2} \frac{d \delta_p}{dp},
\]

where in the last line we have performed an integration by parts. Since we are only interested in a logarithmically divergent term, we neglect the first two terms, that are finite \((\delta_\Lambda \sim 1/\Lambda)\).

Using the explicit form of the phase shift given in (1.20), we find

\[
\frac{d \delta_p}{dp} = \frac{2}{m_\pi} \left( \frac{1}{1 + p^2/m_\pi^2} + \frac{2}{1 + 4p^2/m_\pi^2} \right).
\]

A direct calculation (see Exercise XXX) then yields for the renormalized kink mass, up to finite terms,

\[
m_{kR} = m_k + \delta m_k = m_k - \frac{3}{4\pi} m_\pi \log \left( \frac{\Lambda^2}{m_\pi^2} \right).
\]
For the unrenormalized mass on the r.h.s. we now use equation (1.14), which we can reexpress in terms of the renormalized pion mass, to first order in $\lambda/m_{\pi}^2$, as
\[ m_k = \frac{m_{\pi R}^3}{3\lambda} + \frac{3}{4\pi} m_{\pi R} \log \left( \frac{\Lambda^2}{m_{\pi}^2} \right). \]
We see that the logarithmic divergence cancels, so that the relation (1.14) is preserved under renormalization:
\[ m_{kR} = \frac{m_{\pi R}^3}{3\lambda}. \] (1.29)

1.1.3 Fermions and kinks *

We have considered the quantum properties of the scalar field fluctuating around a kink. Peculiar phenomena happen when fermions propagate in the background of a kink. In this section we consider the scalar theory with potential (1.2) and couple it to a Dirac fermion, a complex two-component field $\psi$ with Lagrangian
\[ \mathcal{L}_F = \bar{\psi} (\gamma^\mu \partial_\mu + g\phi) \psi \] (1.30)
The theory is invariant under global $U(1)$ transformations
\[ \psi \to e^{i\alpha} \psi ; \quad \bar{\psi} \to e^{-i\alpha} \bar{\psi} \]
as well as the discrete transformation
\[ \phi \to -\phi ; \quad \psi \to \gamma_A \psi ; \quad \bar{\psi} \to -\bar{\psi} \gamma_A \]
where $\gamma_A = \gamma^0 \gamma^3$ is the chirality operator. This $\mathbb{Z}_2$ symmetry is broken in the scalar vacuum $\phi = \pm f$, where the fermion acquires a mass $m_F = gf$.  

In the sectors $Q_{-\pm}$ and $Q_{+\pm}$, i.e. in scalar vacuum, the fermion field can be decomposed in plane waves
\[ \psi = \int \frac{dp}{2\pi \sqrt{2E}} \left[ b_p e^{-iEt} u_p(x) + d_p e^{iEt} v_p(x) \right]. \]

\[ ^4 \text{In general, the sign of the mass term in the fermionic Lagrangian is not physically significant because it can be changed by the field redefinition } \psi \to \gamma_A \psi, \quad \bar{\psi} \to -\bar{\psi} \gamma_A. \]
If we choose the representation $\gamma^0 = i\sigma_2$, $\gamma^1 = -\sigma_3$, $\gamma^A \equiv \gamma^0\gamma^1 = \sigma_1$, the elementary spinor solutions are

$$u_p(x) = e^{ipx} \left( \frac{\sqrt{E}}{\sqrt{p^2 + m^2}} \right); \quad v_p(x) = e^{-ipx} \left( \frac{\sqrt{E}}{\sqrt{p^2 + m^2}} \right).$$

(1.31)

The field is quantized by imposing the canonical anticommutation relations

$$\{b_p, b^\dagger_{p'}\} = \delta(p - p'); \quad \{d_p, d^\dagger_{p'}\} = \delta(p - p').$$

which are equivalent to canonical equal-time anticommutation relations for $\psi$ and $\psi^\dagger$.

For the fermion current it is best to use the definition

$$j^\mu = \frac{1}{2} \left( \bar{\psi} \gamma^\mu \psi - \bar{\psi}^c \gamma^\mu \psi^c \right),$$

(1.32)

where $\psi^c = \psi^*$ is the charge conjugate field, obeying the same equation as $\psi$. This expression has the advantage of avoiding the infinite charge of the Dirac sea that is present in the more familiar expression $j^\mu = \bar{\psi} \gamma^\mu \psi$. Indeed we have

$$Q = \int \frac{dp}{2\pi} \left( b^\dagger_p b_p - d^\dagger_p d_p \right)$$

(1.33)

whereas the Hamiltonian is given by

$$H = \int \frac{dp}{2\pi} E_p \left( b^\dagger_p b_p + d^\dagger_p d_p \right).$$

(1.34)

Let us now see what happens in the presence of a kink. In the chosen representation of the gamma matrices, the Dirac operator has the form

$$\begin{pmatrix} P^\dagger & \partial_t \\ -\partial_t & P \end{pmatrix} \quad \text{where} \quad P = \partial_x + g\bar{\phi}, \quad P^\dagger = -\partial_x + g\bar{\phi}.$$  

Normally squaring the Dirac operator (with a change of sign for the mass term) produces the Klein-Gordon operator times the unit matrix. This calculation requires commuting the mass with derivatives. Now, however, the mass has been replaced by the field $g\bar{\phi}$, which does not commute with the space derivative. We thus find:

$$\begin{pmatrix} -P & \partial_t \\ -\partial_t & -P^\dagger \end{pmatrix} \begin{pmatrix} P^\dagger & \partial_t \\ -\partial_t & P \end{pmatrix} = \begin{pmatrix} -\partial_t^2 - PP^\dagger & 0 \\ 0 & -\partial_t^2 - PP^\dagger \end{pmatrix}.$$
where
\[ P^\dagger P = -\partial_x^2 + g^2 \phi^2 - g\partial_x \phi, \quad PP^\dagger = -\partial_x^2 + g^2 \phi^2 + g\partial_x \phi. \]

The square of the Dirac operator therefore reads \(-\partial_t^2 + L\). where \(L\) is the self-adjoint operator
\[ L = \begin{pmatrix} PP^\dagger & 0 \\ 0 & P^\dagger P \end{pmatrix}. \]

Unlike the normal case, it is not proportional to the unit matrix.

As with the scalar field, it will prove convenient to decompose the spinor on the basis of eigenfunctions of this operator, instead of ordinary Fourier modes. We make the ansatz
\[ \psi = e^{-iEt} \begin{pmatrix} \tilde{u}_1(x) \\ \tilde{u}_2(x) \end{pmatrix} \]
and demand that these functions are annihilated by \(\partial_t^2 + L\). This implies that \(u_1\) must be an eigenfunction of \(PP^\dagger\) with eigenvalue \(E^2\) and \(u_2\) must be an eigenfunction of \(P^\dagger P\) with the same eigenvalue.

One easily sees that if \(u\) is an eigenfunction of \(PP^\dagger\) with a given eigenvalue, \(P^\dagger u\) is an eigenfunction of \(P^\dagger P\) with the same eigenvalue. The converse is also true, so these operators have the same eigenfunctions. If we choose the upper spinor component to be \(\tilde{u}_1(x)\), the corresponding lower spinor component must be \(\tilde{u}_2(x) = C_2 P^\dagger \tilde{u}_1(x)\), where \(C_2\) is some normalization constant. In the same way we find that if we choose the lower component \(\tilde{u}_2(x)\), the upper component must be \(\tilde{u}_1(x) = C_1 P \tilde{u}_2(x)\). For these two relations to be compatible we must have \(C_1 C_2 = 1/E^2\).  

The spectrum of \(L\) can be computed analytically, but we shall not need it in the following. Suffice it to say that it consists of a continuum of scattering states and a discrete spectrum with energies \(E^2 = 2rg - r^2\), where \(r = 0, 1, \ldots\) are integers less than \(g\). The continuum and the discrete states with \(r \geq 1\) come in pairs, as described above. The modes \(r = 0\), which have zero energy, behave in a drastically different way. The equation \(P\tilde{u}_0 = 0\) has solution
\[ \tilde{u}_0(x) \sim e^{-g \int_x dy \phi(y)}. \]

\[ ^5 \text{In the case } \phi = f \text{ these relations are satisfied by the solutions in (1.31), with } C_2 = -i/E. \]
CHAPTER 1. \( \pi_0(Q) \) AND SOLITONS

This is a normalizable zero-mode of \( P^\dagger P \), due to the asymptotic behavior of the function \( \bar{\phi} \). On the other hand the solution of the equation \( P^\dagger \tilde{u}_0 = 0 \) is

\[
\tilde{u}_0(x) \sim e^{g \int^x dy \bar{\phi}(y)}
\]

which is not normalizable, for the same reasons. Therefore \( PP^\dagger \) does not have a (normalizable) zero mode.

We can now decompose a spinor in the background of the kink as

\[
\psi = b_0 \left( \begin{array}{c} \tilde{0} \\ u_0(x) \end{array} \right) + \int \frac{dp}{2\pi \sqrt{2E}} \left[ b_p e^{-iEt} \tilde{u}_p(x) + d_p^\dagger e^{iEt} \tilde{v}_p(x) \right]
\]

where \( \tilde{u}_p \) and \( \tilde{v}_p \) are the eigenfunctions of \( L \) described above.

When this decomposition is used, the Hamiltonian still has the form (1.34), with the integral extending over all the non-zero modes. The zero mode is a discrete fermionic degree of freedom that can be in two quantum states: either free or occupied. The peculiar fact is that the occupied state has zero energy like the empty state. Therefore, the system has two degenerate vacua \( |0\rangle \) and \( |0'\rangle = b_0^\dagger |0\rangle \).

The surprise comes when we consider the charge of these states. When the decomposition (1.35) is inserted in the fermionic charge

\[
Q = \int dx \left( \psi^\dagger \psi - \psi^T \psi^* \right)
\]

due to the fact that they still come in degenerate pairs, the non-zero modes work out as in the absence of the kink and give back (1.33). However, the zero mode does not have a partner and its contribution is different:

\[
\frac{1}{2} \left( b_0^\dagger b_0 - b_0 b_0^\dagger \right) = b_0^\dagger b_0 - \frac{1}{2}
\]

In the vacuum state where the zero mode is empty

\[
Q|0\rangle = -\frac{1}{2} |0\rangle
\]

while in the vacuum state where the zero mode is occupied

\[
Q|0'\rangle = \frac{1}{2} |0'\rangle
\]

So we find that in the presence of the kink the fermionic field does not have a state of zero charge, and the charges are fractional. Creating fermions or antifermions will add integer charges to that of the vacuum, so all the states have a fractional charge. We could say that in the presence of the fermion field the kink itself carries a charge equal to \( \pm 1/2 \).
1.2 Scalar fields in other dimensions

1.2.1 Domain walls *

There is a way to use the preceding solution in higher dimensions. Consider the case of a single scalar field in $d > 1$ space dimensions. The equation of motion for a static solution is

$$\sum_i \partial_i^2 \phi = V', \quad (1.36)$$

We can make an ansatz for the field

$$\phi(x_1, \ldots, x_d) = \phi(x_1)$$

then the equation of motion reduces to that of a scalar in one dimension. We have already discussed solutions for this equation in section 1.1.1. Thus, inserting any of those solutions in the ansatz above gives a solution of the higher dimensional equations.

These kinks in higher dimensions are called domain walls. They separate two half-spaces where the scalar is in different vacua. The location of the wall is a linear subspace $W$ of codimension one where the scalar field vanishes. Domain walls are not solitons, because the energy of the solution is infinite:

$$E_S = \int_W d^{d-1}x \mathcal{E} \quad \text{where} \quad \mathcal{E} = \int dx_1 \left[ \frac{1}{2}(\partial_1 \phi)^2 + V(\phi) \right]$$

where $\mathcal{E}$ represents a surface density of energy. For example, for the potential (1.2), one has from (1.7)

$$\mathcal{E} = \frac{2\sqrt{2}}{3} f^3 \sqrt{\lambda}.$$  

(One has to bear in mind that the dimension of $f$ and $\lambda$ is now different from section 1.1, so that $\mathcal{E}$ has the correct dimension $d$ in mass.)

1.2.2 No go theorems

The existence of topological solitons requires that the configuration space has more than one connected components and that the equations of motion admit smooth, localized, finite energy solutions. These are separate conditions. In
this section we show that linear scalar theories with the usual two-derivative kinetic term and a potential, do not satisfy either of them.

We begin with a single scalar in higher dimensions. Finiteness of the static energy

$$E_S = \int d^d x \left[ \frac{1}{2} \sum_i (\partial_i \phi)^2 + V(\phi) \right]$$

demands that when $r = |\vec{x}| \to \infty$, $\phi$ tends to one of the minima of $V$. Thus the configuration space $Q$ will consist again of various connected components:

$$Q = \bigcup_{i \in \mathcal{J}} Q_i, \quad Q_i = \{ \phi \in Q \mid \phi \underset{r \to \infty}{\to} y_i \}$$

and $\mathcal{J}$ is the set of the minima of $V$. The absolute minimum of $E_S$ in each $Q_i$ is given by the constant $\phi = y_i$. These are just the classical vacua of the model. The essential difference with the case of the previous section is that in $d = 1$ the “sphere at infinity” $S_0^{d-1}$ defined by the limit $r \to \infty$ consists of two disconnected points, and the field can take different values at these two points, whereas in $d \geq 2$ the “sphere at infinity” $S_0^{d-1}$ is connected. By continuity the value of the field at infinity must be constant and there cannot be solutions with nontrivial boundary conditions.

Let us next consider the case of $N > 1$ scalar fields $\phi = \phi^a$ ($a = 1, \ldots, N$) in $d$ space dimensions. The space of all such fields is denoted $\Gamma(\mathbb{R}^d, \mathbb{R}^N)$. Assuming symmetry under $SO(N)$, the action is

$$S = \int d^{d+1} x \left[ -\frac{1}{2} \partial_\mu \phi^a \partial^\mu \phi^a - V(|\phi|) \right], \quad (1.37)$$

where $|\phi| = \sqrt{\phi^a \phi^a}$ and repeated indices are summed over. For definiteness we will consider only the case of a quartic potential

$$V = -\frac{1}{2} m^2 |\phi|^2 + \frac{\lambda}{4} |\phi|^4 + \frac{m^4}{4\lambda} = \frac{\lambda}{4} (|\phi|^2 - f^2)^2,$$

where $f = \sqrt{\frac{m^2}{\lambda}}$ and $m^2 > 0$. The locus of the minima is a sphere $S^{N-1}$. The static energy is now

$$E_S = \int d^d x \left[ \frac{1}{2} \partial_i \phi^a \partial_i \phi^a + V(|\phi|) \right]. \quad (1.38)$$
We are interested in the subspace $Q \subset \Gamma(\mathbb{R}^d, \mathbb{R}^N)$ for which the static energy is finite. This demands again that as $r \to \infty$, $\phi$ tends to one of the minima of $V$.

One can ask whether it is necessary to allow $\phi$ to go to an arbitrary point of $S^{N-1}$ when $r \to \infty$, or it suffices to consider fields that tend to a specific point of $S^{N-1}$. Let $\phi$ and $\phi'$ be two field configurations such that $\phi \xrightarrow{r \to \infty} y$ and $\phi' \xrightarrow{r \to \infty} y'$, where $y$ and $y'$ are two different points on $S^{N-1}$. Since all maps from $\mathbb{R}^d$ to $\mathbb{R}^N$ are homotopic, there exists a one-parameter family of maps $\phi_\tau(x)$, with $0 \leq \tau \leq 1$, such that $\phi_0 = \phi$ and $\phi_1 = \phi'$ (for more on homotopy theory see Appendix A). It is convenient to redefine the homotopy parameter to go from $-\infty$ to $\infty$ instead than from 0 to 1. For example, we can define

$$\tau = \frac{1}{2} + \frac{1}{\pi} \arctan t .$$

Writing $\phi_\tau(x) = \hat{\phi}(x,t)$, we can interpret $t$ as time and $\hat{\phi} \in \Gamma(\mathbb{R}^{d+1}, \mathbb{R}^N)$ as a spacetime field. The energy of this field is $E = E_K + E_S$ where $E_K = \int d^d x \frac{1}{2} (\frac{d\hat{\phi}}{dt})^2$ is the kinetic energy. Since $\frac{d\hat{\phi}}{dt}$ does not tend to zero as $r \to \infty$, it is clear that for finite $t$, $E_K$ is divergent. We conclude that to go from $\phi$ to $\phi'$ one must go through configurations with infinite kinetic energy, so the boundary value of $\phi$ cannot change in the course of the time evolution. For this reason, we will always assume that the configuration space consists of field with a fixed boundary condition at infinity. In particular, the only possible constant field is the field that is everywhere equal to the boundary value. It is worth noting that this restriction has a counterpart in homotopy theory, where one usually considers based maps, namely maps that have a predetermined value at a predetermined point.

Using the $SO(N)$ invariance of the theory, we can assume without loss of generality that the value of $\phi$ as $r \to \infty$ be $y_0 = (0,0,\ldots,0,f)$. The limit $r \to \infty$ defines a "sphere at infinity" $S^{d-1}_\infty$; since the map $\phi$ must be constant on $S^{d-1}_\infty$, all its points may be identified to a single point $\infty$. Then $\phi$ may be regarded as a map from the one-point compactification $\mathbb{R}^d \cup \{\infty\} = S^d$ into $\mathbb{R}^N$, mapping the "basepoint" $\infty$ of $S^d$ to the "basepoint" $y_0$. Therefore $Q = \Gamma_\ast(S^d, \mathbb{R}^N)$. All maps with these properties are homotopic to one another, so the space $Q$ is connected.

These results imply that linear scalar field theories in dimensions $d \geq 2$ cannot have topological solitons. There is an independent result, known as Derrick’s theorem, saying that linear scalar field theories with action 1.37 do
not admit nontrivial static solutions (whether topological or not) when $d \geq 2$. The proof is based on a scaling argument.

Let us rewrite equation (1.38) as $E_S = E_1 + E_2$, where $E_1$ and $E_2$ are the “elastic” and “potential” energy, in the terminology introduced in the previous section. Let $\phi_\lambda$ be a one-parameter family of configurations defined by $\phi_\lambda(x) = \phi_1(\lambda x)$. We have

$$E_1(\phi_\lambda) = \lambda^{2-d} E_1(\phi_1), \quad E_2(\phi_\lambda) = \lambda^{-d} E_2(\phi_1).$$

In order for $\phi_1$ to be a stationary point of $E_S$ it is necessary that

$$0 = \frac{d}{d\lambda} E_S(\phi_\lambda) \bigg|_{\lambda=1} = (2-d) E_1(\phi_1) - d E_2(\phi_1).$$

Since $E_1$ and $E_2$ are positive semidefinite, for $d \geq 3$ this implies $E_1(\phi_1) = 0$ and $E_2(\phi_1) = 0$, which is only satisfied by the trivial vacuum configuration.

For $d=2$ we get $E_2(\phi_1) = 0$. This means that the field must be everywhere in the minimum of $V$, which implies that $\frac{\partial V}{\partial \phi^a} = 0$. Inserting in the equation of motion we obtain $\partial_\mu \partial^\mu \phi^a = 0$, which, together with the given boundary conditions, implies again $\phi = \text{constant}$.

To escape the negative conclusions derived in this section, one has to modify either the kinematics or the dynamics of the theory, or both. One way is to couple the scalars to gauge fields. This will be discussed in section XXX. Another way is to consider nonlinear scalar theories.

### 1.2.3 Nonlinear sigma models

Let us start from a linear scalar theory with action (1.37). It is invariant under global internal rotations of the fields, forming the group $SO(N)$. In particular, the potential is constant on the orbits of $SO(N)$ in $\mathbb{R}^N$. The minima occur on a particular orbit $S^{N-1}=SO(N)/SO(N-1)$ (see Appendix C). If we take the limit $\lambda \to \infty$ with $f$ kept constant, the potential becomes unbounded everywhere except on the orbit of the minima, where it remains equal to zero. Thus in the strong coupling limit the potential constrains the field to lie on that particular orbit. This is illustrated by the following figure:

A mathematically more sensible way of studying the limit is to introduce a Lagrange multiplier field $\Lambda$ and consider the action

$$S = \int d^{d+1}x \left[ -\frac{1}{2} \partial_\mu \phi^a \partial^\mu \phi^a - \frac{2\Lambda}{\sqrt{\lambda}} \sqrt{V} + \frac{\Lambda^2}{\lambda} \right].$$

(1.41)
The equation of motion for $\Lambda$ is $\Lambda = \sqrt{\lambda V}$ and when this equation is used in (1.41) it gives back (1.37). Thus (1.41) is classically equivalent to (1.37). The advantage of the action (1.41) is that it remains well defined in the limit $\lambda \to \infty$. In fact, it reduces to

$$ S = \int d^{d+1}x \left[ -\frac{1}{2} \partial_\mu \phi^a \partial^\mu \phi^a - \Lambda (|\phi|^2 - f^2) \right]. $$

(1.42)

The second term enforces the constraint $\phi^2 = f^2$. This is called a “nonlinear sigma model with values in $S^{N-1}$,” or a “SO($N$)-nonlinear sigma model”.

It is usually quite inconvenient to work with constrained fields. This can be avoided by working directly with the coordinates of the target space. Let us illustrate how this works for the two-dimensional sphere. We can solve the constraint $\phi^a \phi^a = f^2$ expressing the three fields $\phi^a$ in terms of only two independent fields $\varphi^a$. There are infinitely many ways of doing this. For example we could choose $\varphi^a$ to be the spherical coordinates ($\varphi^1 = \Theta$, $\varphi^2 = \Phi$):

$$ \begin{align*}
\phi^1 &= f \sin \Theta \cos \Phi \\
\phi^2 &= f \sin \Theta \sin \Phi \\
\phi^3 &= f \cos \Theta
\end{align*} $$

(1.43)

(1.44)

(1.45)

Introducing into (1.42), we find the action

$$ S = -\frac{f^2}{2} \int d^{d+1}x \left( \partial_\mu \Theta \partial^\mu \Theta + \sin^2 \Theta \partial_\mu \Phi \partial^\mu \Phi \right). $$
Another choice are the stereographic coordinates $\varphi^1 = \omega^1$, $\varphi^2 = \omega^2$:

\[
\begin{align*}
\varphi^1 &= f \frac{4\omega_1}{\omega_1^2 + \omega_2^2 + 4} \\
\varphi^2 &= f \frac{4\omega_2}{\omega_1^2 + \omega_2^2 + 4} \\
\varphi^3 &= f \frac{\omega_1^2 + \omega_2^2 - 4}{\omega_1^2 + \omega_2^2 + 4}
\end{align*}
\] (1.46, 1.47, 1.48)

Introducing in (1.42),

\[
S = -\frac{f^2}{2} \int d^{d+1}x \frac{16}{(\omega_1^2 + \omega_2^2 + 4)^2} (\partial_\mu \omega_1 \partial^\mu \omega_1 + \partial_\mu \omega_2 \partial^\mu \omega_2).
\]

In any case the action has the form

\[
S = -\frac{f^2}{2} \int d^{d+1}x \partial_\mu \varphi^\alpha \partial^\mu \varphi^\beta h_{\alpha\beta}(\varphi),
\] (1.49)

where $h_{\alpha\beta}(\varphi)$ is the standard metric on the sphere $S^2$ of unit radius, written in the chosen coordinate system, and $f^2$ is a constant.

We recall that the original linear model (1.37) is a standard example of the Goldstone theorem. Picking a vacuum state, for example $\phi = (0, \ldots, 0, f)$ breaks $SO(N)$ to $SO(N - 1)$, and gives rise to $N - 1$ massless Goldstone bosons. The model contains an additional "radial" scalar degree of freedom with mass $\sqrt{2\lambda}f$. If we consider phenomena at energies much lower than this mass, the radial mode cannot be excited and we remain just with the Goldstone bosons, whose dynamics is described by the action (1.49).

This discussion can be generalized to scalar fields carrying a representation of any Lie group $G$. If $\phi_0$ is a minimum of the potential, every other point in the orbit of $G$ through $\phi_0$ is also a minimum. We assume that all the minima belong to a single orbit. If $H$ is the stabilizer of $\phi_0$, the orbit of the minima is diffeomorphic to the coset space $G/H$. Then, the procedure described above gives a nonlinear sigma model with values in $G/H$.

In fact, equation (1.49) is valid for any target space $G/H$, provided we interpret $h_{\alpha\beta}$ as the components of a $G$-invariant metric. The $G$-invariance of

---

6See for example L. Michel, "Minima Of Higgs-Landau Polynomials", Contribution to Colloq. on Fundamental Interactions, in honor of Antoine Visconti, Marseille, France, Jul 5-6, 1979. Published in Marseille Collog. 157 (1979) (CERN-TH-2716)
the action can be proven as follows. Let us first consider a general variation of the field. We have
\[
\delta S = -\frac{f^2}{2} \int d^{d+1}x \left[ 2\partial_\mu \delta \varphi^\alpha \partial^\mu \varphi^\beta h_{\alpha\beta} + \partial_\mu \varphi^\alpha \partial^\mu \varphi^\beta \partial_\gamma h_{\alpha\beta} \delta \varphi^\gamma \right].
\] (1.50)

Assume that \(\delta \varphi^\gamma = \epsilon^a K_a^\gamma(\varphi)\), where \(\epsilon^a\) are constant infinitesimal parameters (which can be thought of as an element of the Lie algebra of \(G\)) and \(K_a\) are vectorfields, satisfying the Killing equation
\[
K_a^\gamma \partial_\gamma h_{\alpha\beta} + h_{\alpha\gamma} \partial_\beta K_a^\gamma + h_{\beta\gamma} \partial_\alpha K_a^\gamma = 0.
\]

Then it is easy to check that \(\delta S = 0\). On the other hand, if we keep the variation arbitrary, but going to zero at infinity so that integrations by parts do not leave any boundary term, then one obtains the field equation
\[
\partial_\mu \partial^\mu \varphi^\gamma + \partial_\mu \varphi^\alpha \partial^\mu \varphi^\beta \Gamma^\gamma_{\alpha\beta}(\varphi) = 0.
\] (1.51)

where \(\Gamma^\gamma_{\alpha\beta}\) are the Christoffel symbols of \(h_{\alpha\beta}\).

One can further generalize this discussion by considering nonlinear sigma models with values in a completely arbitrary target manifold, as long as it is endowed with a metric \(h_{\alpha\beta}\). This is relevant for example in string theory. However, we will not need to consider such models here. For us a nonlinear sigma model will always be a theory of Goldstone bosons.

### 1.2.4 Power counting

We conclude this section with some remarks on the quantization of nonlinear sigma models, with the action (1.49). For this discussion it is convenient to define \(f = 1/g\). Since the metric is in general a nonpolynomial function, the fields have to be dimensionless. Therefore the constant \(g^2\) must have dimension \(L^{n-2}\), where \(n = d + 1\) is the dimension of spacetime. In two spacetime dimensions, and only in two, we can choose \(g^2 = 1\). In order to give the scalar fields their canonical dimension, we absorb first the constant \(g^2\) in the fields, defining \(\varphi^\alpha = \varphi^\alpha / g\). The dimension of \(\varphi^\alpha\) is then \([\varphi^\alpha] = L^{\frac{d-2}{2}}\).

Now the action reads
\[
S = -\frac{1}{2} \int d^n x \partial_\mu \varphi^\alpha \partial^\mu \varphi^\beta h_{\alpha\beta}(g \varphi).
\] (1.52)
The metric $h_{\alpha\beta}(g\bar{\varphi})$ is still dimensionless. In order to separate the kinetic term from the interaction terms we have to fix some constant background $\bar{\varphi}_0^\alpha$, write $\bar{\varphi}^\alpha = \bar{\varphi}_0^\alpha + \eta^\alpha$, and expand the metric in Taylor series in $\eta$:

$$h_{\alpha\beta}(g\bar{\varphi}) = h_{\alpha\beta}(g\bar{\varphi}_0) + g\partial_\gamma h_{\alpha\beta}(g\bar{\varphi}_0)\eta^\gamma + g^2\partial_\gamma \partial_\delta h_{\alpha\beta}(g\bar{\varphi}_0)\eta^\gamma \eta^\delta + \cdots \quad (1.53)$$

where we write $\partial_\gamma$ for $\frac{\partial}{\partial \varphi_\gamma}$. The coefficients of this expansion are now field-independent and represent the coupling constant of the theory. Note that there is in general an infinite number of couplings and all couplings involve derivatives of the fields. (In most models of interest, a $\mathbb{Z}_2$ invariance under the transformation $\eta \rightarrow -\eta$ forbids terms with an odd number of fields.)

The dimension of the coupling constant in the $m$-th term, i.e. the coefficient of $\partial\eta \partial\eta \eta^m$, is $[g^m] = L^{\frac{m}{n} - 2}$. In spite of the infinite number of couplings, this theory is renormalizable in a generalized sense for $n = 2$. It is nonrenormalizable for $n > 2$. We note that in some cases, such as the sphere, the metric is entirely determined up to an overall scale by symmetry requirements. In these cases there is really only one independent coupling constant $g$: All the coefficients of the expansion of the metric are determined by the requirement of $G$-invariance.

In conclusion, the nonlinear sigma models are not good candidates for fundamental theories in more than two dimensions. Instead, they are widely used in condensed matter physics and also in particle physics, as low energy phenomenological models.

### 1.3 Nonlinear sigma model in $d=2$

Let us ask whether the nonlinear sigma models could have nontrivial solutions in $d > 1$. All solutions of the nonlinear sigma model equations have $E_2 = 0$, so (1.40) implies that if $d > 2$ the only static solution of the field equations is constant, while in $d = 2$ nontrivial solutions are possible.

For the existence of topological solitons one also needs a suitable target space. The simplest example is $S^2$, so we now turn to the $S^2$-nonlinear sigma model in $d = 2$. 
1.3. NONLINEAR SIGMA MODEL IN D=2

1.3.1 Topology

We start by discussing the configuration space. We work with unconstrained fields \( \varphi \) representing a map from \( \mathbb{R}^2 \) to \( S^2 \). Finiteness of the static energy

\[
E_S = \frac{f^2}{2} \int d^2x \partial_i \varphi^\alpha \partial_i \varphi^\beta h_{\alpha\beta}(\varphi)
\]

(1.54)
demands that \( \partial_i \varphi \to 0 \) as \( r \to \infty \). Thus \( \varphi \) must tend to a constant at infinity. Without loss of generality we can take this constant value to be the north pole. In spherical coordinates it is given by \( \Theta = 0 \); in stereographic coordinates it is given by

\[
\sqrt{\omega_1^2 + \omega_2^2} \to \infty.
\]

Since from now on we will restrict our attention to this particular class of maps, we can compactify space to a sphere by adding a point at infinity: \( S^2 = \mathbb{R}^2 \cup \{\infty\} \). In homotopy theory it is often very convenient to pick a special point in each space, called the “basepoint”. In the present context it is natural to choose the basepoint of the spatial \( S^2 \) to be the point \( \infty \), and the basepoint of the internal \( S^2 \) to be the north pole. There follows that any finite energy configuration can be regarded as a map from \( S^2 \) to \( S^2 \) preserving basepoints. The space of such maps is denoted \( \mathcal{Q} = \Gamma_*(S^2, S^2) \). This space consists of infinitely many connected components: \( \pi_0(\mathcal{Q}) = \pi_2(S^2) = \mathbb{Z} \) (see Appendix XXX). So we can write

\[
\mathcal{Q} = \bigcup_{n \in \mathbb{Z}} \mathcal{Q}_n.
\]

The integer \( n \) labelling the homotopy classes is known as the winding number. In any coordinate system, it can be written as

\[
W(\varphi^\alpha) = \frac{1}{8\pi} \int d^2x \varepsilon^{ij} \partial_i \varphi^\alpha \partial_j \varphi^\beta \sqrt{\text{det} h} \varepsilon_{\alpha\beta}.
\]

(1.55)

For example, in spherical and stereographic coordinates it has the expression

\[
W(\Theta, \Phi) = \frac{1}{4\pi} \int d^2x \sin \Theta \varepsilon^{ij} \partial_i \Theta \partial_j \Phi,
\]

(1.56)

\[
W(\omega_1, \omega_2) = \frac{1}{4\pi} \int d^2x \frac{16}{(\omega_1^2 + \omega_2^2 + 4)^2} \varepsilon^{ij} \partial_i \omega^1 \partial_j \omega^2,
\]

(1.57)

respectively. It is not obvious from these formulae that it is an integer. However, we can easily prove that \( W \) is locally constant. To this effect, let us write

\[
\sqrt{\text{det} h(\varphi)} \varepsilon_{\alpha\beta} = \omega_{\alpha\beta}(\varphi)
\]

(1.58)
for the components of the volume form of $S^2$. Varying infinitesimally we get
\[
\delta W = \frac{1}{8\pi} \int d^2 x \varepsilon^{ij} \left[ 2 \partial_i \delta \varphi^\alpha \partial_j \varphi^\beta \omega_{\alpha\beta} + \partial_i \varphi^\alpha \partial_j \varphi^\beta \delta \varphi^\gamma \partial_{\gamma} \omega_{\alpha\beta} \right].
\]
Since the variation is supposed to preserve the boundary conditions, it must vanish at infinity. Thus we can integrate the first term by parts. Factoring $\delta \varphi^\gamma$ and antisymmetrizing the first term, we arrive at
\[
\delta W = \frac{1}{8\pi} \int d^2 x \varepsilon^{ij} \partial_i \varphi^\alpha \partial_j \varphi^\beta \delta \varphi^\gamma \left( \partial_{\alpha} \omega_{\beta\gamma} + \partial_{\beta} \omega_{\gamma\alpha} + \partial_{\gamma} \omega_{\alpha\beta} \right) = 0,
\]
since the exterior derivative of the form $\omega$ vanishes. Thus $W$ is a functional on $Q$ that is constant on each connected component $Q_n$. We shall encounter in the next section explicit solutions of the field equations for which one can check, by explicit calculation, that $W = n$ is an integer. Then, $W$ is constant and equal to $n$ for all fields belonging to the same connected component $Q_n$. A more general definition of the winding number and a theorem proving its integrality are discussed in Appendix XXX.

Since the time evolution is a continuous curve in $Q$, the value of the winding number cannot change; the winding number must be a constant of motion of the theory. This can be confirmed by the following argument. We define a topological current
\[
J_T^\lambda = \frac{1}{8\pi} \varepsilon^{\lambda\mu\nu} \partial_\mu \varphi^\alpha \partial_\nu \varphi^\beta \omega_{\alpha\beta} ,
\]
which is identically conserved:
\[
\partial_\lambda J_T^\lambda = \frac{1}{8\pi} \varepsilon^{\lambda\mu\nu} \partial_\lambda \varphi^\gamma \partial_\mu \varphi^\alpha \partial_\nu \varphi^\beta \partial_{\gamma} \omega_{\alpha\beta} = 0,
\]
again because the form $\omega$ is closed. One sees immediately that the topological charge is equal to the winding number:
\[
Q_T = \int d^2 x J_T^0 = \frac{1}{8\pi} \int d^2 x \varepsilon^{ij} \partial_i \varphi^\alpha \partial_j \varphi^\beta \omega_{\alpha\beta} = W(\varphi) .
\]
(There is no contribution to the boundary integral coming from spatial infinity, because $J^0$ is proportional to spatial derivatives, that are required to vanish at infinity.) There follows that $Q_T = W$ is a constant of motion.
1.3. NONLINEAR SIGMA MODEL IN $D=2$

1.3.2 Dynamics

Let us look at the absolute minimum of the static energy (1.54) in each topological sector $Q_i$. Consider the following inequality\(^7\)

$$0 \leq \int d^2 x \, h_{\alpha\beta} \left( \partial_i \varphi^\alpha \pm \varepsilon_{ik} \partial_k \varphi^\gamma \omega_{\gamma}^\alpha \right) \left( \partial_i \varphi^\beta \pm \varepsilon_{ij} \partial_j \varphi^\gamma \omega_{\gamma}^\beta \right) =$$

$$= \int d^2 x \left[ 2h_{\alpha\beta} \partial_i \varphi^\alpha \partial_i \varphi^\beta \mp 2\varepsilon_{ij} \partial_i \varphi^\alpha \partial_j \varphi^\gamma \omega_{\gamma}^\alpha \right] = \frac{4}{f^2} E_S \mp 16\pi W$$

where in the product of the last two terms we used

$$\omega_{\gamma}^\alpha \omega_{\epsilon}^\alpha = \varepsilon_{\gamma\alpha} \varepsilon_{\epsilon\delta} h^{\alpha\delta} \det h = h_{\gamma\epsilon}.$$  

If $W > 0$ (resp. $W < 0$) the inequality with the upper sign (resp. lower sign) is stronger. There follows that

$$E_S \geq 4\pi f^2 |W|.$$  \hspace{1cm} (1.60)

Furthermore, equality holds if and only if

$$\partial_i \varphi^\alpha = \mp \varepsilon_{ik} \partial_k \varphi^\gamma \varepsilon_{\gamma\delta} h^{\delta\alpha} \sqrt{\det h}.$$  \hspace{1cm} (1.61)

The fields for which this equation is satisfied are the absolute minima of the static energy and are also static solutions of the Euler-Lagrange equations of the theory. Note that (1.61) are first order equations, and therefore simpler than the second order Euler-Lagrange equations.

It is convenient to specialize the discussion to stereographic coordinates $\omega^1$ and $\omega^2$. Equation (1.61) reduces to

$$\partial_i \omega_\alpha = \mp \varepsilon_{ik} \partial_k \omega^\gamma \varepsilon_{\gamma\alpha},$$  \hspace{1cm} (1.62)

and spelling these out

$$\partial_1 \omega^1 = \pm \partial_2 \omega^2,$$

$$\partial_2 \omega^1 = \mp \partial_1 \omega^2.$$  \hspace{1cm} (1.63)

If we define \( \omega = \omega^1 + i \omega^2 \) and \( z = x^1 + i x^2 \) we recognize (1.63) as the Cauchy-Riemann equations for the function \( \omega = \omega(z) \). The solutions are the functions which are analytic or antianalytic depending on the sign in (1.63). For example \( \omega(z) = z^n \) and \( \omega(z) = (z^*)^n \), with \( n \geq 0 \), are solutions of (1.63). Note that for large \( |z| \), \( \omega \) does not tend to an angle-independent limit, but since \( |\omega| \to \infty \) it does not matter since all these points represent the north pole of \( S^2 \). These functions describe smooth maps \( \varphi \in \Gamma_\ast(S^2, S^2) \) with winding number \( W = n \) and \( W = -n \) respectively. They are absolute minima of the static energy in the sectors \( Q_n \) and \( Q_{-n} \) respectively \((n \geq 0)\).

The theory is invariant under rotations, translations and dilatations, so applying these transformations to the solutions we get other solutions. This means that the solitons are not isolated, but rather come in four-parameter families. Applying these transformations to the solutions mentioned above we find

\[
\omega(z) = \left( \frac{(z - z_0)e^{i\alpha}}{\lambda} \right)^n \quad (1.64)
\]

\[
\omega(z) = \left( \frac{(z - z_0)^*e^{-i\alpha}}{\lambda} \right)^n \quad (1.65)
\]

where the complex number \( z_0 \) gives the position of the center of the soliton, the angle \( \alpha \) its “internal orientation” and the positive real number \( \lambda \) its scale. The parameters \( z, \alpha \) and \( \lambda \) are the collective coordinates, or moduli, of the soliton. \(^8\)

1.3.3 No ferromagnetic transition in \( d = 2 \)

This model can be regarded as the continuum limit of a planar ferromagnetic crystal, with unit spins allowed to point in any direction in a three-dimensional embedding space. Classically, the state of lowest energy of the system is a perfect ferromagnet with all spins aligned in a fixed direction. It has \( W = 0 \). The direction of the spins breaks the rotational invariance of the system and from Goldstone’s theorem one expects to find massless excitations in the spectrum. In fact, small perturbations of the field around this

\(^8\)Since the theory has internal \( SO(3) \) invariance, any rotation of the solution is also a solution. However, as discussed in Section 1.2.2, the boundary condition at infinity breaks \( SO(3) \) to \( SO(2) \approx U(1) \). A space rotation, which in the chosen coordinates is represented by \( z \to e^{i\alpha}z \), can be undone by an internal \( U(1) \) transformation, so of these two abelian groups only one remains as a modulus.
state describe massless particles. The field $\varphi$ is itself the Goldstone boson and its quanta are the fundamental excitations of the system.

However, it is also possible to excite states with $W \neq 0$, namely solitons. Since a soliton with $|W| = 1$ has mass $4\pi f^2$, at a fixed temperature $T$ there will be a density of solitons of order $e^{-f^2/kT}$. If the solitons had fixed size (as the kinks of section 1.1), for very small $T$ this would describe an ordered state with a few localized defects. But in this theory solitons can be arbitrarily large without paying any price in energy. Thus in a given box of finite size there will be solitons/antisolitons that occupy much of the (two dimensional) volume and since a soliton has spins pointing in any direction, the ferromagnetic order will be destroyed.

This is a special case of the so-called Mermin-Wagner theorem, stating that in two (or less) space dimensions at temperature $T > 0$, there cannot be a phase where a continuous symmetry is spontaneously broken.

1.4 Current algebra and solitons in $d = 3$

Let us consider a nonlinear sigma model with values in some target space $N$. The scaling argument rules out static solitons for the action (1.49) in dimensions other than two. Nevertheless let us see for what choices of dimension and target space the configuration space would have more than one connected component. Then we shall look for some alternative action functional that could have stationary points in the nontrivial topological sectors.

Following the same reasoning as in the case of the $S^2$ sigma model, the space of smooth finite energy configurations of the field is $\mathcal{Q} = \Gamma^*_s(S^d, N)$. Therefore, there is room for the existence of topological solitons whenever $\pi_0(\mathcal{Q}) = \pi_d(N) \neq 0$. One important case is when $N = G$, a Lie group. This is called a principal sigma model. If $G$ is semisimple one has $\pi_3(G) = \mathbb{Z}$, the fundamental class being realized by a homomorphism $SU(2) \equiv S^3 \to G$. These models appear in the description of strong interactions at low energies. To motivate this we will give first a brief review of current algebra.

1.4.1 The chiral models

The strong interactions are described by QCD, a gauge theory with gauge group $SU(3)$. The fields entering the QCD action are the gauge fields $A_{\mu}$, describing particles called gluons, and spinor fields describing the quarks. There
are six known types (or flavors) of quarks: \( u \) (up), \( d \) (down), \( s \) (strange), \( c \) (charm), \( b \) (bottom or beauty) and \( t \) (top), in order of increasing mass. Each of them is described by a Dirac spinor. We can collect these quark fields into a column vector \( q_\alpha \), where \( \alpha \) is an index that runs over the six flavors. The quark part of the QCD action is

\[
S_q = \sum_\alpha \int d^4x \, \bar{\psi}_\alpha (i\gamma^\mu D_\mu - m_\alpha) \psi_\alpha .
\]  

(1.66)

where \( D_\mu \) denotes the covariant derivative with respect to the gluon fields. For arbitrary masses, the only invariance of this action are the constant phase transformations. Infinitesimally, these are given by

\[
\delta V_\alpha \psi = i\alpha \psi ; \quad \delta V_\alpha \bar{\psi} = -i\alpha \bar{\psi}
\]

(1.67)

The corresponding group is called the vector \( U(1) \), or \( U(1)_V \). Assuming that \( N \) masses are equal, then also the transformations

\[
\delta V_\epsilon \psi = \epsilon^a T_a \psi ; \quad \delta V_\epsilon \bar{\psi} = -\bar{\psi} \epsilon^a T_a
\]

(1.68)

with \( T_a \) a basis in the Lie algebra of \( SU(N) \), are symmetries. This group is called \( SU(N)_V \).

The masses of the quarks are distributed over a large range, so it is sometimes possible to pretend that some of them are massless. This is a good approximation for the \( u \) and \( d \) quarks and, to a lesser extent, also for the \( s \) quark. Let us suppose that the masses of the \( N \) lightest quarks can be neglected (this is usually called the chiral limit of QCD). Then, in addition to the above, the QCD action is invariant also under axial \( U(1) \) and \( SU(N) \) transformations:

\[
\delta A_\alpha \psi = i\alpha \gamma^A \psi ; \quad \delta A_\alpha \bar{\psi} = i\alpha \bar{\psi} \gamma^A \quad U(1)_A ;
\]

\[
\delta A_\epsilon \psi = \epsilon^a T_a \gamma^A \psi ; \quad \delta A_\epsilon \bar{\psi} = -\bar{\psi} \epsilon^a T_a \gamma^A \quad SU(N)_A .
\]

(1.69)

Here \( \gamma^A = \gamma^5 \) is the chirality operator, which anticommutes with the gamma matrices. We shall now forget about the heavy quarks and have a closer look at the symmetries of massless QCD with \( N \) flavors. The generators of the transformations written above are the charges constructed with the following currents:

\[
\begin{align*}
\hat{j}_V^\mu &= \bar{\psi} \gamma^\mu \psi & \text{for } U(1)_V \\
\hat{j}_A^\mu &= \bar{\psi} \gamma^\mu \gamma^A \psi & \text{for } U(1)_A \\
\hat{j}_{V_\epsilon}^\mu &= \bar{\psi} \gamma^\mu \epsilon^a T_a \psi & \text{for } SU(N)_V \\
\hat{j}_{A_\epsilon}^\mu &= \bar{\psi} \gamma^\mu \gamma^A \epsilon^a T_a \psi & \text{for } SU(N)_A
\end{align*}
\]

(1.70)
where $\epsilon$ is an element of the Lie algebra of $SU(N)$. From the canonical equal-time anticommutation relations
\begin{equation}
\{\psi^{\alpha i}(\vec{x}, t), \bar{\psi}_{\beta j}(\vec{y}, t)\} = \delta^{\alpha}_{\beta} \delta^i_j \delta(\vec{x} - \vec{y}) ,
\end{equation}
where $a, b$ are Dirac indices and $i, j$ are $SU(N)$ indices, we obtain the following current algebra
\begin{align}
[j_0^V, j_0^V] &= [j_0^A, j_0^A] = 0 ; \\
[j_{V[\epsilon_1, \epsilon_2]}^0, j_{V[\epsilon_1, \epsilon_2]}^0] &= j_{V[\epsilon_1, \epsilon_2]}^0 ; \\
[j_{A[\epsilon_1, \epsilon_2]}^0, j_{A[\epsilon_1, \epsilon_2]}^0] &= j_{A[\epsilon_1, \epsilon_2]}^0 ; \\
[j_{A[\epsilon_1, \epsilon_2]}^0, j_{V[\epsilon_1, \epsilon_2]}^0] &= j_{V[\epsilon_1, \epsilon_2]}^0 ;
\end{align}
(1.72)
One can verify that these are the algebras implied by (1.69).

The vector and axial transformations are entangled; in particular, the axial transformations do not form a subalgebra. It is convenient to reshuffle the $SU(N)_V$ and $SU(N)_A$ transformations in a different way. Since the chirality operator $\gamma^A$ satisfies $(\gamma^A)^2 = 1$, the operators
\begin{equation}
P_\pm = \frac{1 \pm \gamma^A}{2}
\end{equation}
are projectors and can be used to decompose the Dirac spinors (for each flavor) as the sum of a left handed (negative chirality) and right handed (positive chirality) part: $\psi = \psi_+ + \psi_-$, where $\psi_\pm = P_\pm \psi$. Defining
\begin{align}
\bar{j}_L^\mu &= \frac{j_{V[\epsilon_1, \epsilon_2]}^\mu - j_{A[\epsilon_1, \epsilon_2]}^\mu}{2} = \bar{\psi} \gamma^\mu P_- \epsilon^a T_a \psi ; \\
\bar{j}_R^\mu &= \frac{j_{V[\epsilon_1, \epsilon_2]}^\mu + j_{A[\epsilon_1, \epsilon_2]}^\mu}{2} = \bar{\psi} \gamma^\mu P_+ \epsilon^a T_a \psi ;
\end{align}
(1.74)
we can rewrite (1.72) as
\begin{align}
[j_0^L_{\epsilon_1}, j_0^L_{\epsilon_2}] &= j_{L[\epsilon_1, \epsilon_2]}^0 ; \\
[j_0^L_{\epsilon_1}, j_0^R_{\epsilon_2}] &= 0 ; \\
[j_0^R_{\epsilon_1}, j_0^R_{\epsilon_2}] &= j_{R[\epsilon_1, \epsilon_2]}^0 ;
\end{align}
(1.75)
showing that the global symmetry group is $SU(N)_L \times SU(N)_R$.

The generator of the the group $U(1)_V$ is baryon number, and is therefore an observed symmetry of nature. The group $U(1)_A$ is not realized in nature,
because if it was, for every hadron there would be another hadron with the same mass but opposite parity. We shall defer a discussion of the fate of this group to Section 5.3.

In the case $N = 2$ the group $SU(2)_V$ corresponds to isospin (this can be deduced for example by looking at the transformation of the proton and neutron, which are composites of quarks). In the case $N = 3$ the group $SU(3)_V$ corresponds to the $SU(3)$ of the octet way (again this follows for example from the action on the octet of baryons). These are not strictly speaking symmetry groups of the real world, because if they were the masses of the proton and neutron (in the case $N = 2$) or of all baryons of the octet (in the case $N = 3$) would be equal. However, to the extent that the mass differences between these particles can be neglected, they are an unbroken symmetry.

The “axial $SU(N)$” transformations cannot be a symmetry of nature, however, not even approximately, for if it was then for each multiplet of baryons and mesons there would exist another multiplet with the same masses but opposite parity. On the other hand, the phenomenology of hadrons shows that the current algebra (1.75) is realized in nature to good approximation for $N = 2$ and to a slightly lesser extent also for $N = 3$. One concludes that $SU(N)_L \times SU(N)_R$ is a symmetry of the Lagrangian but not of the vacuum, or in other words it is a spontaneously broken symmetry. From Goldstone’s theorem, then, there should exist $N^2 - 1$ massless scalar particles (Goldstone bosons). There do indeed exist scalar particles whose masses are small compared to those of the other hadrons: these are the pions and, to a lesser extent, all the mesons in the pion/kaon octet. In the case $N = 2$, it is therefore possible to interpret the pions as the Goldstone bosons that come from the spontaneous breaking of $SU(2)_A$. In the case $N = 3$, it is also possible to interpret the pions and kaons as the Goldstone bosons that come from the spontaneous breaking of $SU(3)_A$.

The upshot of this discussion is that in the chiral limit in which $N$ quarks are massless, the vacuum state of QCD breaks $SU(N)_L \times SU(N)_R$, leaving $SU(N)_V$ unbroken, and therefore defines a point $U$ in the coset space $SU(N)_L \times SU(N)_R / SU(N)_V$. This coset space can be geometrically identified with the group $SU(N)$ itself. Suppose now that we want to study low momentum/low energy phenomena. The state of the system is no longer the vacuum state, but in a sufficiently small spacetime region it can still be described as the vacuum. We can describe such a state by giving the vacuum vector a weak dependence on the spacetime point, so at low energy strong in-
1.4. CURRENT ALGEBRA AND SOLITONS IN $D = 3$

Interactions can be described by a map from spacetime into $SU(N)$. It is quite convenient to represent this map by a matrix-valued field $U(x) \in SU(N)$ ($U$ is in the fundamental representation).

This is a phenomenological description of low energy QCD, so the action can in principle contain all terms consistent with the symmetries of the theory. However, at low momenta the terms with the lowest number of derivatives will dominate. There cannot be any potential term, and the term with the lowest number of derivatives is

$$S = f^2 \int d^4x \text{tr}(U^{-1}\partial_\mu UU^{-1}\partial^\mu U) \ . \quad (1.76)$$

To relate this to previous formulas, given a coordinate system on $SU(N)$, we call $\varphi^\alpha(x)$ the coordinates of the group element $U(x)$ and we decompose

$$U^{-1}\partial_\mu U = \partial_\mu \varphi^\alpha L^\alpha_\alpha(\varphi)T_a \ , \quad (1.77)$$

where $L^\alpha_\alpha$ are the components of the Maurer-Cartan form on $SU(N)$. In the case of $SU(2)$, these are given explicitly in Appendix XXX. The basis in the Lie algebra, in the fundamental representation, is chosen such that $\text{tr} T_a T_b = -\frac{1}{2}\delta_{ab}$. (In the case $N = 2$, $T_a = -\frac{i}{2}\sigma_a$, where $\sigma^a$ are the Pauli matrices). Then we choose the $Ad$-invariant inner product in the Lie algebra

$$-2\text{tr}(T_a T_b) = \delta_{ab} \quad (1.78)$$

and we define a Riemannian metric on $SU(N)$ by declaring the Maurer-Cartan forms to be an orthonormal field of co-frames:

$$h_{\alpha\beta} = L^a_\alpha L^b_\beta \delta_{ab} \ . \quad (1.79)$$

In this way we see that the action (1.76) can be is identical to the nonlinear sigma model action (1.49). The advantage of the form (1.76) is that it makes the $SU(N)_L \times SU(N)_R$ invariance of the theory very transparent: if we transform

$$U \rightarrow g_L^{-1}U g_R \ ,$$

the (constant) group elements $g_L$ and $g_R$ cancel (the latter using ciclicity of the trace). \footnote{When the action is written in the form (1.49), its invariance is less evident. It follows from the fact that the metric $h_{\alpha\beta}$ is both left- and right-invariant, which can be proven by showing that the vectorfields with components $L^a_\alpha$ and $R^a_\beta$, that generate right- and left-multiplications, respectively, are Killing vectors for $h$.} On the other hand, choosing a particular $U$ breaks this
invariance, leaving a residual unbroken group. For the choice \( U = 1 \) the
unbroken group is the diagonal subgroup with \( g_L = g_R \).

For phenomenological purposes, the most useful coordinates on \( SU(N) \)
are the normal coordinates \( \pi^\alpha \), defined by:
\[
U(x) = e^{2\pi^\alpha(x)T_\alpha/f}
\]
where \( T_\alpha \) is a basis in the Lie algebra of \( SU(N) \), satisfying
\[
[T_\alpha, T_\beta] = f^{\alpha\beta\gamma}T_\gamma.
\]
Note that the coordinates have been scaled as in (1.52) so as to have the
canonical dimension of mass.

Using (1.80), (1.76) can be expanded as
\[
\int d^4x \left[ -\frac{1}{2} \partial_\mu \pi^\alpha \partial^\mu \pi^\alpha + \frac{1}{f^2} \epsilon^{abc} \pi^b \partial_\mu \pi^c \epsilon^{ade} \pi^d \partial_\mu \pi^e + \ldots \right].
\]
(1.81)
This corresponds to the expansion (1.53) in normal coordinates in the neigh-
borhood of the identity. One observes that in this model the pions are mass-
less. Furthermore, all interactions contain derivatives of the fields: this is as
it should be, since a potential for \( \pi \) would certainly break the global invari-
ance of the theory.

1.4.2 The Skyrmion

We have mentioned in the beginning of this section, that principal models
with values in semisimple groups have topological sectors. To describe these
sectors in the present formalism let us consider first the case \( G = SU(2) = S^3 \).
The topological sectors in this case are classified by the winding number,
which in terms of the fields \( U \) can be written (see Exercise XXX):
\[
W(U) = -\frac{1}{24\pi^2} \int d^3x \epsilon^{\lambda\mu\nu} \text{tr} \left( U^{-1} \partial_\lambda UU^{-1} \partial_\mu UUU^{-1} \partial_\nu U \right).
\]
(1.82)
For other groups, the generator of \( \pi_3(G) = \mathbb{Z} \) can be obtained by embedding
\( SU(2) \) in \( G \) and then considering the composition of this embedding with a
map \( S^3 \to SU(2) \) of winding number one.

A peculiar feature of principal sigma models is that their configuration
space is itself a group. The product of two field configurations is defined
by pointwise multiplication: \( (U_1U_2)(x) = U_1(x)U_2(x) \). One can then verify
directly from (1.82), that
\[
W(U_1U_2) = W(U_1) + W(U_2) \ ; \ \ \ \ W(U^{-1}) = -W(U) .
\]
(1.83)
A field configuration of the form

\[ U(\vec{x}) = \exp[T_0 \hat{x}^a g(r)] \]  

(1.84)

where \( \hat{x}^a = \frac{x^a}{r} \) and \( g \) is a function which is \(-2\pi\) in the origin and tends to zero as \( r \to \infty \), has winding number one. From (1.83), configurations with arbitrary winding numbers can be constructed simply taking powers of (1.84).

Unfortunately, it follows from the discussion in the end of Section 2 that such fields cannot be solutions of the field equations obtained from the action (1.76). In fact, from (1.40) we get

\[ \left. \frac{dE(\phi_\lambda)}{d\lambda} \right|_{\lambda=1} = -E(\phi_1) < 0 , \]

so they are unstable against deformations that shrink the size of the soliton to zero. The way of stabilizing the solitons is to add higher order terms to the action. \(^{10}\) This may seem a bit artificial, but one has to bear in mind that this theory is to be thought of as an effective low energy theory and hence in principle one should consider all terms in the action consistent with the desired symmetry properties. The total action considered by Skyrme was

\[
S = \int d^4x \left[ \frac{f^2}{4} \text{tr}(U^{-1} \partial_\mu UU^{-1} \partial^\mu U) + \frac{1}{32e^2} \text{tr}\left[U^{-1} \partial_\mu U, U^{-1} \partial_\nu U\right]\left[U^{-1} \partial^\mu U, U^{-1} \partial^\nu U\right] \right] \quad (1.85)
\]

where \( e \) is a new coupling constant. Out of all possible terms containing four derivatives of the fields, only the one with the commutators was chosen, because it contains only two time derivatives of the fields and is therefore better amenable to canonical analysis. This is not essential for what follows, however.

In order to find the soliton with unit winding number, we have to insert the Ansatz (1.84) in the equations of motion that come from (1.85), and solve for the radial function \( g \). Unfortunately the dynamics is sufficiently complicated to prevent an explicit solution. Numerical approximations are

necessary. However, as in section (1.1), we can apply qualitative arguments to infer the existence of a solution and derive some of its properties. Suppose that the function \( g \) goes from \(-\pi\) to zero within a distance \( \ell \) of the origin, corresponding to the size of the soliton. Then the static energy is of the order

\[
E_S(\ell) \approx \ell^3 \left[ \frac{f^2}{\ell^2} + \frac{1}{e^2 \ell^4} \right].
\]

The size of the soliton results from a balance between these two terms, and turns out to be of order \( 1/f e \). Note that for \( e \to \infty \), \( \ell \) tends to zero, in accordance with the argument in the end of section (1.2). The mass of the soliton is of the order \( f/e \).

These solitons are known as skyrmions. Skyrme suggested that the solitons of the theory (1.85) be interpreted as the baryons. In order to understand this claim, we have to study the quantum numbers of the skyrmions. This we shall do much later, in section 4.3.

### 1.5 Yang-Mills theories

In this section we will consider the question whether a pure Yang–Mills theory can have static solitons. Before doing this, however, it will be useful to review some generalities about these theories, and to establish the notation. The dynamical variable is a one-form with values in the Lie algebra \( g \) of a group \( G \): \( A = A^a_{\mu} dx^\mu \otimes T_a \), where \( \{T_a\} \) is a basis in \( g \). With \( A \) one can construct the nonabelian field strength

\[
F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + e f_{abc} A^b_\mu A^c_\nu ,
\]

where \([T_a, T_b] = f_{abc} T_c\) and \( e \) is the coupling constant of the theory. The Yang–Mills action is

\[
S_{YM} = -\frac{1}{4} \int d^{d+1}x \, F^a_{\mu\nu} F^{a\mu\nu} .
\]

It is invariant under local gauge transformations

\[
A_\mu \to g^{-1} A_\mu g + \frac{1}{e} g^{-1} \partial_\mu g , \quad F_{\mu\nu} \to g^{-1} F_{\mu\nu} g ,
\]

where \( g : \mathbb{R}^{d+1} \to G \) and \( F_{\mu\nu} = F^a_{\mu\nu} T_a \).
1.5. YANG-MILLS THEORIES

This formulation of the theory is best suited for the perturbative expansion. In many cases it is more convenient to rescale the field \( A \) by a factor \( 1/e \). In this case the Yang–Mills action reads

\[
S_{YM} = -\frac{1}{4e^2} \int d^{d+1}x F^a_{\mu\nu} F^{a\mu\nu},
\]

where the curvature is now defined by

\[
F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\nu,
\]

The nonabelian gauge transformations then read

\[
A_\mu \to g^{-1}A_\mu g + g^{-1}\partial_\mu g, \quad F_{\mu\nu} \to g^{-1}F_{\mu\nu} g.
\]

This formulation is better suited for the discussion of geometrical properties of the Yang–Mills fields. In this context one often refers to \( A \) as a connection and \( F \) as its curvature. In the present chapter dealing with solitons we will use the former definition of the theory, with action (1.87). In later chapters we will use the rescaled fields, with action (1.88).

Let us define the Yang–Mills Lagrangian density by

\[
S_{YM} = \int d^{d+1}x \mathcal{L}_{YM}.
\]

Separating the space and time components of the curvature we have

\[
\mathcal{L}_{YM} = \frac{1}{2} E^a_i E^a_i - \frac{1}{4} F^a_{ij} F^a_{ij},
\]

where \( E^a_i = F^a_{0i} = \partial_0 A^a_i + \partial_i A^a_0 \) is the nonabelian “electric” field (we have used the notation \( D^a_i A^a_0 = \partial_i A^a_0 + e f^{abc} A^b_0 A^c_0 \); this quantity is a covariant derivative with respect to time independent gauge transformations). The space components of the field strength \( F_{ij} \) are related to the nonabelian “magnetic” field: in \( d = 3 \) we define \( F_{ij} = \varepsilon_{ijk} B_k \), while in \( d = 2 \), \( F_{ij} = \varepsilon_{ij} B \).

The momenta canonically conjugate to the the fields are

\[
P^0_a \equiv \frac{\partial \mathcal{L}_{YM}}{\partial_0 A^a_0} = 0, \quad P^i_a \equiv \frac{\partial \mathcal{L}_{YM}}{\partial_0 A^a_i} = E^a_i.
\]

The relation between velocities and momenta is not invertible, so the proper way to formulate the Hamiltonian dynamics is via Dirac’s theory of constrained systems. In the present case the equation \( P^0_a = 0 \) is known as a “primary constraint”. The canonical Hamiltonian can be written

\[
H_c = \int d^dx \left[ \frac{1}{2} P^a_i P^a_i + \frac{1}{4} F^a_{ij} F^a_{ij} - A^a_0 G_a \right],
\]

(1.91)
where \( G_a = D_i P_a^i = D_i E_a^a \). We have to impose that the primary constraint holds for all time. This means that \( \{P_a^0(x), H\} = 0 \), which results in the “secondary constraint” \( G_a = 0 \). In the Hamiltonian formalism the fields \( A_0^a \) play the role of Lagrange multipliers enforcing the Gauss law \( G_a = 0 \).

When studying the canonical formulation of a YM theory it is often very convenient to choose the gauge \( A_0 = 0 \) (this can be done by performing the gauge transformation \( g(x,t) = P \exp \left( -e \int^t dt' A_0(x,t') \right) \), where \( P \) stands for path ordering). This leaves the freedom of performing time-independent gauge transformations. In this gauge \( E_i^a = A_i^a \), so the first term in (1.91) is seen as a kinetic term, the second as a potential term. We will mostly use this gauge in later sections.

Let us now come to the question whether a pure Yang–Mills theory can have static solitons. There is here a slight complication: if a gauge field configuration is time-independent, it can acquire a time dependence after a gauge transformation. In a gauge theory one calls a field “static” if there is a gauge in which \( A_\mu \) is time-independent. This implies that all gauge invariant quantities constructed with the field (such as, for example, the energy density) are time-independent. Note that for a static configuration, the gauge \( A_0 = 0 \) may not be the gauge in which \( \partial_0 A_\mu = 0 \), so we do not make this gauge choice here.

We shall now prove that pure YM theory does not admit static solitons if \( d \neq 4 \) (i.e. in five-dimensional spacetime). 11

For a static field in a gauge in which \( \partial_0 A_\mu = 0 \), the lagrangian is given by \( L = E_1 - E_2 \), where

\[
E_1 = \frac{1}{2} \int d^d x \left( D_i A_0^a \right)^2 > 0 \quad \text{and} \quad E_2 = \frac{1}{4} \int d^d x \left( F_{ij}^a \right)^2 > 0 .
\]

Consider the two-parameter family of configurations \( A_{(\sigma,\lambda)} \) defined by

\[
A_{(\sigma,\lambda)}^a_0(x) = \sigma \lambda A_0^a(\lambda x) , \quad A_{(\sigma,\lambda)}^a_1(x) = \lambda A_1^a(\lambda x) .
\]

We have \( E_1(A_{(\sigma,\lambda)}) = \sigma^2 \lambda^{4-d} E_1(A_{(1,1)}) \) and \( E_2(A_{(\sigma,\lambda)}) = \lambda^{4-d} E_2(A_{(1,1)}) \). For

---

1.6. Vortices

$A_{(1,1)}$ to be a solution of the field equations we must have

$$0 = \frac{d}{d\lambda} L \bigg|_{\lambda=\sigma=1} = (4 - d)L(A_{(1,1)}) \ , \quad (1.94)$$

$$0 = \frac{d}{d\sigma} L \bigg|_{\lambda=\sigma=1} = 2E_1(A_{(1,1)}) \ , \quad (1.95)$$

which implies that for $d \neq 4$, $E_1 = E_2 = 0$, which in turn implies $F_{\mu\nu} = 0$.

This argument rules out nontrivial static solitons for pure YM theories except in five spacetime dimensions, which are not physically interesting. Static solitons do indeed exist in five spacetime dimensions, but we will discuss them later in a different context, where they have a different physical interpretation and are known as instantons.

1.6 Vortices

1.6.1 The Nielsen-Olesen vortex

We now consider scalar electrodynamics in two space dimensions. The dynamical variables are a $U(1)$ gauge field $A_\mu$ coupled to a complex scalar field $\phi$, with action

$$S = \int d^3x \left[ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} |D_\mu \phi|^2 - \frac{\lambda}{4} (|\phi|^2 - f^2)^2 \right] , \quad (1.96)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $D_\mu \phi = \partial_\mu \phi - i e A_\mu \phi$. (The Lie algebra of $U(1)$ consists of the purely imaginary numbers and one can take as a basis element $T = -i$. The Lie algebra valued gauge potential is therefore an imaginary one-form $A = A^1 T = -i A^1$. The field $A_\mu$ used in this section is $A^1_\mu$ stripped of the index 1. The gauge transformations can then be obtained from (1.5) by putting $g = e^{i\alpha}$.) The theory is invariant under the local gauge transformations

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{i}{e} g^{-1} \partial_\mu g = A_\mu - \frac{1}{e} \partial_\mu \alpha \ , \quad \phi \rightarrow \phi' = g^{-1} \phi = e^{-i\alpha(x)} \phi . \quad (1.97)$$

---

In the gauge $A_0 = 0$, $E_i = F_{0i} = \dot{A}_i$ and $D_0\phi = \dot{\phi}$; in this gauge the energy reads $E = E_K + E_S$, where

$$E_K = \int d^2 x \left[ \frac{1}{2} \dot{A}_i \dot{A}_i + \frac{1}{2} |\phi|^2 \right].$$

and $E_S$ is the static energy

$$E_S = \int d^2 x \left[ \frac{1}{2} B^2 + \frac{1}{2} |D_i\phi|^2 + \frac{\lambda}{4} (|\phi|^2 - f^2)^2 \right]. \quad (1.98)$$

where $B = F_{12}$. The absolute minimum of $E_S$, the classical vacuum, occurs for

$$B = 0 \ , \ D_i\phi = 0 \ , \ |\phi| = f \ . \quad (1.99)$$

A particular solution of these conditions is

$$A_i = 0 \ , \ \phi = f \ . \quad (1.100)$$

This is the starting point for the usual perturbative discussion of the Higgs phenomenon, showing that the small fluctuations around this vacuum comprise a vector field with mass $m_A = ef$ and a scalar field with mass $m_S = \sqrt{2\lambda}f$. Any gauge transformation of (1.100) $A_i = \frac{i}{e} g^{-1} \partial_i g$, $\phi = g^{-1} f$ is obviously still a solution (here $g = e^{i\alpha}$ is a smooth map $\mathbb{R}^2 \to U(1)$). However, there are other interesting states.

We will now look for static solitons, assuming that the gauge in which the field is time-independent is the gauge $A_0 = 0$. The classical configuration space of this theory consists of regular fields with finite static energy. Clearly $(A,\phi)$ will have finite energy only if the conditions (1.99) are satisfied asymptotically as $r \to \infty$. This requires that

$$\phi(r, \theta) \xrightarrow{r \to \infty} \phi_\infty = f e^{-i\alpha_\infty} \ , \quad (1.101)$$

$$A_i(r, \theta) \xrightarrow{r \to \infty} -\frac{1}{e} \partial_i \alpha_\infty \ , \quad (1.102)$$

where $\alpha_\infty$ depends only on the angular variable $\theta$ parameterizing the “circle at infinity” $S_\infty^1$. We see that unlike the case of the sigma model, the condition $D_i\phi \to 0$ does not imply that $\phi$ tends to a constant at infinity: as long as $|\phi| \to f$, any dependence of $\phi$ on the angle $\theta$ is permitted, because one can always compensate for this dependence by choosing $A_i = \frac{1}{ie} \partial_i \phi$. 


The asymptotic behaviour of the field $\phi$ as $r \to \infty$ defines a map $\phi_\infty : S^1_\infty \to U(1)$. We have seen that such maps fall into homotopy classes, labelled by the winding number

$$W(\phi_\infty) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \frac{d\alpha_\infty}{d\theta} = \frac{i}{2\pi} \int_0^{2\pi} d\theta \frac{1}{\phi_\infty} \frac{d\phi_\infty}{d\theta}.$$  \hspace{1cm} (1.103)$$

The field $\phi$ has values in a linear space and therefore any field configuration can be smoothly deformed into any other. The following figure shows a homotopy between a field with $W = 1$ and a constant field $\phi = f$ (having $W = 0$). The circles represent the images in field space of $S^1_\infty$.

Figure 1.6: A homotopy in fields space. The circle of unit radius is the locus of the minima of the potential.

It is clear that in the intermediate steps of the deformation the field $|\phi|$ does not tend to $f$ as $r \to \infty$. Such fields have infinite static energy, so there is an infinite energy barrier between configurations with different winding numbers of $\phi_\infty$, or in other words the configuration space consists of infinitely many connected components, labelled by $W(\phi_\infty)$.

The time evolution cannot change the winding number of $\phi_\infty$, so there must be in the theory a topological conservation law. In fact, consider the topological current

$$J^\lambda_T = \frac{1}{2\pi i} \varepsilon^{\lambda\mu\nu} \partial_\mu \hat{\phi}^* \partial_\nu \hat{\phi},$$
where \( \hat{\phi} = \phi/|\phi| \). This current is identically conserved and the corresponding topological charge is

\[
Q_T = \int d^2x J^0_T = W(\phi_\infty) .
\]

The physical meaning of the winding number can be understood by using (1.102) in (1.103) and then applying Stokes’ theorem:

\[
W(\phi_\infty) = \frac{e}{2\pi} \oint_{S^1_\infty} A_i dx^i = \frac{e}{2\pi} \int_{\mathbb{R}^2} d^2x B = \frac{e}{2\pi} \Phi ,
\]

where \( \Phi \) is the magnetic flux through \( \mathbb{R}^2 \) (thinking of \( B \) as a magnetic field orthogonal to \( \mathbb{R}^2 \)).

Since \( W \) is an integer, we get flux quantization:

\[
\Phi = \frac{2\pi}{e} n .
\]

Finally, we would like to find explicit “vortex” solutions in each topological sector. For the solitons with unit flux we make the ansatz

\[
\begin{align*}
A_0 &= 0 , \\
A_i &= -\varepsilon_{ij}\hat{x}^j A(r) , \\
\phi &= F(r)e^{i\varphi} ,
\end{align*}
\]

where \( A \) and \( F \) are functions of the radius such that \( A(r) \to \frac{1}{er} \) and \( F(r) \to f + O(r^{-1}) \) when \( r \to \infty \). Clearly the asymptotic conditions are satisfied and \( W(\phi_\infty) = 1 \). However, it has so far proved impossible to solve explicitly the equations of motions (proofs of existence have been given, though). One has to resort to numerical calculations.

### 1.6.2 Superconductivity *

Now consider scalar electrodynamics in \( d = 3 \). If we assume that \( A_3 = 0 \) and that all the fields are independent of \( x_3 \), then the equations of the theory reduce to those of scalar electrodynamics in \( d = 2 \). Thus, the vortex soliton of \( d = 2 \) becomes as an infinite vortex line in \( d = 3 \). It now has infinite energy on account of its infinite length, so it is not a soliton, but it has important physical application that we review briefly here.
1.6. VORTICES

What in relativistic quantum field theory would be called scalar QED, is called in condensed matter physics a Landau-Ginzburg theory. It is an approximation of Bardeen-Cooper-Schrieffer (BCS) theory, which is itself an (approximate) microscopic model. The important properties of superconductors are independent of the approximate nature of these models and follow simply from the assumption that in the bulk of the material, electromagnetic gauge invariance is in the Higgs phase.

In the BCS theory the charge-carriers are weakly-bound pairs of electrons. Such pairs can be described by a field transforming under $U(1)$ as

$$\phi(x) \rightarrow \exp(\ii 2e\alpha(x)/\hbar)\phi(x),$$

where $-e$ is the electron charge and $\alpha$ is identified mod$2\pi$. The field is invariant under transformations with $\alpha = \pi\hbar/e$, so a nontrivial VEV for this field would break $U(1)$ to $\mathbb{Z}_2$. In the ungauged case, there would then be a Goldstone boson with values in $U(1)/\mathbb{Z}_2$. It is a real field identified mod$\pi\hbar/e$ and transforming under $U(1)$ by

$$\varphi \rightarrow \varphi + \alpha \quad (1.106)$$

Given any other field $\psi$, transforming linearly under gauge transformation, with charge $q$, we can construct a gauge invariant field

$$\tilde{\psi}(x) = \psi(x) \exp(\ii q\varphi(x)/\hbar), \quad (1.107)$$

The Lagrangian has to have the form

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \mathcal{L}_m(\varphi, \tilde{\psi}),$$

where the matter Lagrangian depends only on the gauge invariant fields and on the Goldstone boson, the latter entering only through the covariant derivative

$$D_\mu \varphi = \partial_\mu \varphi - A_\mu.$$

We can define the current

$$J^\mu = \frac{\delta \mathcal{L}_m}{\delta A_\mu} \quad (1.108)$$

The equation of motion of the field $\varphi$ is then seen to be equivalent to the statement of current conservation:

$$0 = \frac{\delta \mathcal{L}_m}{\delta \varphi} - \partial_\mu \frac{\delta \mathcal{L}_m}{\delta \partial_\mu \varphi} = \partial_\mu \frac{\delta \mathcal{L}_m}{\delta A_\mu} = \partial_\mu J^\mu.$$
CHAPTER 1. $\pi_0(\mathcal{Q})$ AND SOLITONS

The main assumption that is needed to describe superconductivity is that the Goldstone boson is covariantly constant:

$$D_\mu \varphi = 0 .$$  \hspace{1cm} (1.109)

This condition will follow if we make the rather reasonable assumption that in the Lagrangian for $\varphi$, the lowest term is quadratic in $D_\mu \varphi$. This is indeed what one has in the Landau-Ginzburg theory: if we consider the action (1.96) and take the strong coupling limit, following the procedure of section 1.2.3, we will arrive at

$$\mathcal{L} = -\frac{f^2}{2}(D \varphi)^2 .$$

Consider a static situation with $\partial_0 \varphi = 0$, $A_0 = 0$. The space component of (1.109)

$$A_i = \partial_i \varphi$$  \hspace{1cm} (1.110)

implies that the magnetic field must be zero:

$$B_i = 0 .$$

This is known as Meissner effect and holds in the bulk of a piece of superconductor. If there is an external magnetic field, the field lines will be deformed so as to avoid going through the semiconductor.

Absence of electrical resistance can be gleaned from the following argument. In any simply connected piece of superconductor, $\varphi$ can be set to any fixed constant by a transformation (1.106). Now consider a thick torus made of superconductor and let $\ell$ be a closed loop deep in the material. Integrating (1.110) on this loop and using Stokes' theorem we find that

$$\Delta \varphi = \int_\ell A = \int_S B = \Phi ,$$

where $\Phi$ is the magnetic flux through a surface $S$ bounded by the loop $\ell$. Since the Goldstone field is periodically identified, it can jump by integral multiples of $\pi h/e$. We thus find that flux must be quantized:

$$\Phi = \frac{\pi h}{e} n .$$  \hspace{1cm} (1.111)

Because of this, the current in the superconductor cannot decay continuously. The fact that the current is not affected by ordinary electrical resistance can
be proven more generally by considering the time dependence of the current. We will not discuss this here.

We have seen that the crucial field in the description of the superconducting state is a real Goldstone boson. On the other hand, in the ordinary state of matter, electromagnetic $U(1)$ is not Higgsed, and the fields carry ordinary linear representations of $U(1)$. By continuity, near the transition also the Goldstone boson must be accompanied by a dynamical modulus field $\rho$ that acts as an order parameter: it is zero in the normal state and nonzero in the superconducting state.

If the superconductor is exposed to an external magnetic field, the field lines will penetrate the material but only for a depth of order

$$\lambda = \frac{1}{ef} = \frac{1}{m_V},$$

(1.112)

which is called the penetration depth and is the inverse of the mass of the photons in the bulk. In addition to $\lambda$, superconductors are characterized by another length scale called the superconducting coherence length

$$\xi = \frac{1}{\sqrt{2}m_S}.$$  

(1.113)

In the Landau-Ginzburg description, these lengths are just the inverse masses of the gauge fields and the scalar. The ratio of these lengths

$$\kappa = \frac{\lambda}{\xi} = \frac{\sqrt{2}m_S}{m_A}$$

characterizes the behavior of the material in a strong magnetic field: a superconductor is said to be of type I if $0 < \kappa < 1/\sqrt{2}$ and of type II if $\kappa > 1/\sqrt{2}$.

In a type I superconductor, when the magnetic field exceeds a critical value, the material undergoes a phase transition to a normal state. In a type II superconductor, when the external magnetic field exceeds a critical value, it is energetically favorable for the magnetic field to penetrates the superconductor in the form of thin tubes, called Abrikosov vortices. In Landau-Ginzburg theory they correspond to the vortex solution discussed in the preceding section. In the core of each tube the modulus field is zero, but elsewhere the material remains superconducting. This is called the vortex phase. The density of vortices increases with the external magnetic field, up to a second, higher, critical field, where superconductivity is lost.
Since the core of the flux tube is not superconducting, the topology of a piece of superconductor that is pierced by a vortex line is the same as that of the thick torus discussed earlier. Therefore, the flux through the tube must be quantized as in (1.111). Note the similarity between the classical quantization conditions (1.104) of Landau-Ginzburg theory and the quantum condition (1.111). In the former $e$ is a classical parameter in the Lagrangian that could have any value, in the latter it is identified with the electron charge (the factor of two is due to the charge of the Cooper pairs). Yet somehow we see that the topological information is preserved in the approximate phenomenological theory. This is a rather general phenomenon of which we shall encounter other examples later on.

### 1.7 Monopoles

Maxwell’s equations can be written in the form

$$\partial_\mu F^{\mu\nu} = 4\pi J_\nu^{(E)},$$

$$\partial_\mu *F^{\mu\nu} = 0,$$

(1.114)

where $*F^{\mu\nu} = \frac{1}{2} g_{\mu\rho} g_{\nu\sigma} \varepsilon^{\rho\sigma\alpha\beta} F_{\alpha\beta}$ is the dual of the field strength. (Recall that $g_{\mu\nu} = (-+++)$ and $\varepsilon^{0123} = 1$. In Minkowski space $**F = -F$, whereas in Euclidean space one would have $**F = F$). In vacuum ($J_\nu^{E} = 0$) these equations are invariant under the duality transformation $F \rightarrow *F$, $*F \rightarrow **F = -F$. Writing

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & +B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \quad \quad *F_{\mu\nu} = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ B_1 & 0 & E_3 & -E_2 \\ B_2 & -E_3 & 0 & E_1 \\ B_3 & E_2 & -E_1 & 0 \end{pmatrix}$$

we see that duality transformations amount to the replacements $E \rightarrow -B$, $B \rightarrow E$. In fact the vacuum Maxwell equations are invariant under a whole $U(1)$ group of transformations of the form

$$F \rightarrow \cos \theta F + \sin \theta *F$$

$$*F \rightarrow -\sin \theta F + \cos \theta *F.$$  \hspace{1cm} (1.115)

In the presence of sources an asymmetry is seen to arise, due to the empirical fact that the r.h.s. of the second equation in (1.114) is identically zero. They
1.7. MONOPOLES

could be made symmetric under duality transformations by introducing a
\[ \partial_\mu *F^{\mu\nu} = 4\pi J^\nu_M \]  
(1.116)
and postulating the transformation
\[ J_E \rightarrow \cos \theta J_E + \sin \theta J_M, \]
\[ J_M \rightarrow -\sin \theta J_E + \cos \theta J_M. \]  
(1.117)

That \( J^\nu_M \) is a magnetic current is seen by observing for example that the
time component of (1.116) would read \( \text{div} B = 4\pi \rho_M \), and therefore acts
as the source of the magnetic potential, i.e. has to be interpreted as the
magnetic charge density. Such a modification would introduce essential new
features in the theory. Most important, if \( J_M \neq 0 \) it would become impossible
to introduce a magnetic potential \( A_\mu \) such that \( F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \). This
complication does not arise if we limit ourselves to the study of pointlike
magnetic sources. The Coulomb–like field
\[ B_i = \frac{Q_M}{r^2} \hat{x}^i, \]  
(1.118)
describing a static pointlike magnetic monopole in the origin, solves the equa-
tion \( \text{div} B = 4\pi Q_M \delta(r) \). Since the field is singular in the origin, one can re-
move this point from space and regard (1.118) as a smooth field on \( \mathbb{R}^3 \setminus \{0\} \).
Since the field \( B \) given in (1.118) is divergence free on \( \mathbb{R}^3 \setminus \{0\} \), it is possible
to introduce the magnetic potential there.

This solution of Maxwell’s equations has interesting properties that we
shall study in detail in Section 3.1. In particular we will find that the magnetic
monopole can be regarded as a \( U(1) \) gauge field only if \( Q_M \) is quantized in
certain units. For the time being we merely observe that it is a singular field
and has infinite energy, so it does not satisfy the requirements for a soliton.
The remarkable fact is that certain nonabelian gauge theories with Higgs
fields admit solitons whose behaviour at large \( r \) approaches that of a Dirac
monopole. We will now discuss this type of solutions.

1.7.1 The ’t Hooft-Polyakov monopole

We consider the Georgi-Glashow model, consisting of an \( SU(2) \) gauge field
\( A_\mu = A_\mu^a T_a \) coupled to a Higgs field \( \phi^a \) in the adjoint (triplet) representation.
We use the unscaled gauge fields, with curvature (1.86) and action (1.87). The total Lagrangian density is
\[ L = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu a} - \frac{1}{2} D_\mu \phi^a D^\mu \phi^a - \frac{\lambda}{4} (\phi^a \phi^a - f^2)^2 \] (1.119)
where \( D_\mu \phi^a = \partial_\mu \phi^a + \epsilon_{abc} A^b_\mu \phi^c \). The structure constants of the Lie algebra of \( SU(2) \) are \( f_{abc} = \epsilon_{abc} \) (in the adjoint representation the generators are \((T^a)_{bc} = -\epsilon_{abc}\)). The action is invariant under the local gauge transformations (1.5), acting on the scalar as \( \phi \to g^{-1} \phi \) (here \( g \) is in the adjoint representation). It is convenient to choose the gauge so that \( A^a_0 = 0 \). Then \( F^a_{0i} = \partial_0 A^a_i \), \( D_0 \phi^a = \partial_0 \phi^a \).

The static energy is
\[ E_S = \int d^3 x \left[ \frac{1}{4} (F^a_{ij})^2 + \frac{1}{2} (D_i \phi^a)^2 + \frac{\lambda}{4} (\phi^a \phi^a - f^2)^2 \right] \] (1.120)
Its absolute minimum is obtained for
\[
F_{ij} = 0 , \\
D_i \phi^a = 0 , \\
\phi^a \phi^a = f^2 ,
\]
in which case \( E_S = 0 \). This is the classical vacuum of the theory. Due to the shape of the potential, the Higgs phenomenon occurs. This can be seen by choosing a gauge in which \( A^a_i = 0 \), \( \phi^a = \bar{\phi}^a = (0, 0, f) \) and expanding the action to second order in \( A \) and in the shifted field \( \phi - \bar{\phi} \). Invariance under local \( SU(2) \) transformations is not broken, however, and any gauge transform of this solution is also a solution.

Finiteness of \( E_S \) demands that the conditions (1.122) be satisfied asymptotically when \( r \to \infty \). In particular for large \( r \) we must have \( \phi^2 = f^2 + O(1/r^2) \), so the asymptotic behaviour of \( \phi \) defines a map \( \phi_\infty : S^2_\infty \to S^2_{\text{int}} \), where \( S^2_\infty \) denotes the “sphere at infinity” in \( \mathbb{R}^3 \) and \( S^2_{\text{int}} \) is the locus of the minima of the potential in the field space. The covariant derivative and the magnetic field have to go to zero like \( 1/r^2 \). As in the abelian case, discussed in the previous section, the second condition in (1.122) does not restrict the map \( \phi \) itself. The asymptotic field \( \phi^a_\infty \) can depend on the angles in an arbitrary way; the condition \( D_i \phi \to 0 \) can then be solved by
\[ A^a_i = \frac{1}{f^2} \epsilon^{abc} \partial_i \phi^b \phi^c + \alpha_i \phi^a + O(1/r^2) \] (1.122)
for an arbitrary constant $\alpha_i$.

The scalar fields $\phi$ fall into classes, labelled by the winding number of the map $\phi_\infty$. Fields with different winding numbers at infinity are separated by an infinite energy barrier. There follows that the configuration space of smooth finite energy configurations for this model consists of infinitely many connected components, labelled by the winding number of $\phi_\infty$. The configuration with $W=0$ is the vacuum, the other one is called a “hedgehog”. The winding number cannot be altered in the course of the time evolution, so there will be a topological conservation law. We define the topological current

$$J_T^\mu = \frac{1}{8\pi} \varepsilon^{\mu\rho\sigma} \varepsilon_{abc} \hat{\phi}^a \partial_\rho \hat{\phi}^b \partial_\sigma \hat{\phi}^c ,$$

(1.123)

where $\hat{\phi}^a = \frac{\phi^a}{\sqrt{\phi^b \phi^b}}$. This current is identically conserved and the corresponding charge is

$$Q_T = \int d^3x J_T^0 = \frac{1}{8\pi} \int d^3x \varepsilon^{ijk} \varepsilon_{abc} \hat{\phi}^a \partial_i \hat{\phi}^b \partial_j \hat{\phi}^c$$

$$= \frac{1}{8\pi} \int_{S^2} d^2x \varepsilon^{ijk} \varepsilon_{abc} \partial_i \hat{\phi}^a \partial_j \hat{\phi}^b \partial_k \hat{\phi}^c = W(\phi_\infty) .$$

(1.124)

The last equality can be proven by choosing a particular coordinate system on $S^2$, for example the spherical coordinates (1.45), and comparing with (1.57).

We are now in a position to explain why configurations with $W \neq 0$ can be interpreted as monopoles. When the Higgs phenomenon occurs, we can interpret the projection of the gauge field along the Higgs VEV as an abelian gauge field. If $\hat{\phi}^a = (0, 0, 1)$, the corresponding field strength is $F_{\mu\nu} = \partial_\mu A_3^\nu - \partial_\nu A_3^\mu$.

Following ‘t Hooft, we can generalize this to position-dependent Higgs fields.\textsuperscript{13} Let $A_\mu = A_\mu^a \hat{\phi}^a$. We define an abelian electromagnetic field $F_{\mu\nu}$ by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \frac{1}{e} \varepsilon_{abc} \hat{\phi}^a \partial_\mu \hat{\phi}^b \partial_\nu \hat{\phi}^c .$$

(1.125)

The last term has been added to compensate the $SU(2)$ non-invariance of $A$. In fact, this can also be written as

$$F_{\mu\nu} = \hat{\phi}^a F_{\mu\nu}^a - \frac{1}{e} \varepsilon_{abc} \hat{\phi}^a D_\mu \hat{\phi}^b D_\nu \hat{\phi}^c ,$$

(1.126)

which is manifestly invariant under $SU(2)$ gauge transformations. This tensor does not obey the Bianchi identities. Instead,

$$\partial_\nu \ast F^{\nu\mu} = \frac{4\pi}{e} J^\mu_T .$$

(1.127)

as one can check most easily using (1.125). Comparing with (1.116), we see that we can interpret $\frac{1}{e} J^\mu_T$ as a magnetic current. The corresponding magnetic charge is

$$Q_M = \frac{1}{e} Q_T = \frac{1}{e} W .$$

(1.128)

Since $W$ is an integer, we get a quantization condition for the magnetic charge, analogous to the flux quantization condition (1.104). We shall see in section 3.1 that quantum mechanics requires the magnetic charge to be quantized in units of $\frac{\hbar}{2e}$, where $e$ is the charge of the electron. The relation between these two conditions is the same as that between (1.104) and (1.111).

We would like to get an explicit solution to the Euler-Lagrange equations realizing these nontrivial boundary conditions. Consider the ansatz

$$\begin{cases}
\phi^a = \frac{x^a}{r} F(r) , \\
A^a_i(x) = \varepsilon_{aij} \frac{x^j}{r} A(r) , \\
A^a_0 = 0 .
\end{cases}$$

(1.129)

In order for the potential energy to be finite, $F(r) - f$ must go to zero faster than $r^{-3/2}$. Then we calculate

$$D_i \phi^a = (\delta_{ia} - \hat{x}_i \hat{x}_a) \left( \frac{1}{r} - eA \right) F + \hat{x}_i \hat{x}_a F' ,$$

(1.130)

where a prime stands for derivative with respect to $r$. The contribution to the energy coming from the covariant derivatives will be finite provided $A(r) \to \frac{1}{er}$ for $r \to \infty$. For the non-abelian magnetic field we have

$$B^a_i = - (\delta_{ia} - \hat{x}_i \hat{x}_a) A' - \frac{1}{r} \delta_{ia} A + \hat{x}_i \hat{x}_a \left( eA^2 - \frac{1}{r} A \right) .$$

(1.131)

It behaves at large $r$ like $1/r^2$, so the magnetic field energy will be automatically finite.
Clearly, the conditions for finiteness of the energy are satisfied and this configuration belongs to the sector $W = 1$. Since $D\phi \to 0$ for $r \to \infty$, the abelian magnetic field

\[ B_i = \frac{1}{2} \varepsilon_{ijk} F_{jk} \to \hat{\phi}^a B_i^a = -\frac{1}{e} \frac{\hat{\phi}}{r^2} \]  

(1.132)

while $\mathcal{E}_i = F_{0i} = 0$. Therefore, for large $r$, the abelian field strength becomes identical to the one of the Dirac monopole.

When the ansatz (1.129) is inserted into the Euler-Lagrange equations, these become coupled second order differential equations for the functions $F$ and $A$. The exact solution to these equations has not been found; only numerical solutions have been given.

### 1.7.2 The Prasad-Sommerfield limit

There is one particular limit, known as the Prasad-Sommerfield limit, in which the functions $F$ and $A$ can be solved exactly: it is the limit in which $\lambda$ and $m^2$ tend to zero with $f = \sqrt{m^2/\lambda}$ constant. In this limit one can derive a useful bound on the energy. We have

\[ E = \int d^3x \left[ \frac{1}{4} F_{ij}^a F_{ij}^a + \frac{1}{2} D_i \phi^a D_i \phi^a \right] \]

\[ = \frac{1}{4} \int d^3x \left( F_{ij}^a \mp \varepsilon_{ijk} D_k \phi^a \right)^2 \pm \frac{1}{2} \int d^3x \varepsilon_{ijk} F_{ij}^a D_k \phi^a. \]

(1.133)

In the second term on the r.h.s. the covariant derivative can be integrated by parts, and using the Bianchi identities for $F_{ij}^a$ it becomes

\[ \frac{1}{2} \int d^3x \partial_k \left( \varepsilon_{ijk} F_{ij}^a \phi^a \right) = f \int_{S^2_{\infty}} d\sigma^k B_k = 4\pi f Q_M = \frac{4\pi f}{e} |W|, \]

(1.134)

where we have used (1.132). Using this in (1.133) we get the so-called Bogomol’nyi bound \(^{14}\)

\[ E \geq \frac{4\pi f}{e} |W|, \]

(1.135)

with equality holding if and only if

\[ F_{ij}^a = \pm \varepsilon_{ijk} D_k \phi^a. \]

(1.136)

The solutions of these equations are the absolute minima of the static energy and therefore automatically satisfy the Euler-Lagrange equations of the theory. In this way we have been able to replace the second-order Euler-Lagrange equations with the first-order equations (1.136). In the Prasad-Sommerfield limit, the explicit form of the functions appearing in (1.129), for the lower sign in (1.136), is

\[ F(r) = \frac{f}{\tanh(efr)} - \frac{1}{er}, \]
\[ A(r) = \frac{1}{er} - \frac{f}{\sinh(efr)}. \]  

The profiles of these functions is shown in the following figures.

Figure 1.7: Monopole profiles in the Prasad-Sommerfield limit.

1.7.3 Symmetries and moduli

The symmetries of the Georgi-Glashow model are the Poincaré group and internal $SO(3)$ transformations with constant parameters (usually called global gauge transformations). These are transformations that correspond to observable transformations on the fields, and they do not include local gauge transformations, that correspond to unobservable transformations of the fields.

We now ask which of these transformations are also symmetries of the monopole solution. Time translation invariance is preserved, because the

\[ ^{15} \text{M.K. Prasad, C.H. Sommerfield, Phys. Rev. Lett. 35 760 (1975).} \]
solution is static, but space translations are broken, because we can distinguish a monopole from a translated monopole. Boosts are also broken: acting with a boost generates another solution that describes a monopole in motion. There remain to discuss internal rotations and space rotations.

Let us consider the effect these transformations have on the scalar field $\phi^a = F(r)\dot{x}^a$. An internal transformation with constant parameter $\epsilon_I^a$ transforms

$$\delta_I \phi^a = \varepsilon_{abc} \epsilon^b_I \phi^c.$$  \hspace{1cm} (1.138)

Under the rotation group a scalar transforms by

$$\delta_R \phi^a = \delta x^k \partial_k \phi^a,$$

where

$$\partial_k \phi^a = \frac{1}{r} \left[ \delta_{ka} - \frac{x_k x_a}{r^2} (r F' - F) \right].$$

A space rotation corresponds to $\delta_R x^i = \epsilon^a_R \varepsilon_{aij} x^j$, so

$$\delta_R \phi^a = \varepsilon_{abc} \epsilon^b_R \phi^c.$$  \hspace{1cm} (1.139)

Combining (1.138) and (1.139) we see that $\phi$ is invariant under the combined transformation

$$(\delta_R - \delta_I) \phi^a = 0$$

where the infinitesimal parameters are the same in the two cases. One can show that also $A^a_i$ is invariant under the same transformations, so the monopole has a symmetry $SO(3)$ consisting of simultaneous internal and space rotations.

This subgroup is unbroken and does not give rise to moduli. There remains one $SO(3)$ subgroup that we can choose to correspond to the internal transformations and could give rise to moduli. However, recall that we work in a functional space with fixed boundary conditions. In this space the field $\phi$ at infinity is fixed and we do not allow transformations that change it. Since the field $\phi^a$ at infinity is direction-dependent, the transformations that leave it invariant must also be direction-dependent and are not strictly speaking a subgroup of the rigid internal $SO(3)$ rotations. They can be described in the following way. As is always the case, a field configuration in the Higgs phase can be brought to the unitary gauge, where the Higgs is aligned along
CHAPTER 1. \( \pi_0(\mathcal{Q}) \) AND SOLITONS

the third direction. The transformation that does this for the monopole is

\[
T = \begin{bmatrix}
\frac{n_2^2 + n_1^2 n_3}{1 - n_3^2} & -\frac{n_1 n_2}{1 + n_3} & -n_1 \\
-\frac{n_1 n_2}{1 + n_3} & \frac{n_1^2 + n_2^2 n_3}{1 - n_3^2} & -n_2 \\
n_1 & n_2 & n_3
\end{bmatrix}
\] (1.140)

This is clearly a singular transformation, since it changes the winding number of the Higgs field at infinity, but it defines a valid gauge locally. In this gauge the last remaining modulus consists just of the group of internal rotations around the third axis. Alternatively, in the regular gauge the modulus parameterizes the rotations of the form

\[ T^{-1} e^{\alpha t^3} T \]

In conclusion, the monopoles come in a four-parameter family, characterized by the coordinates of the center of mass and an internal angle.

1.7.4 Monopoles in GUTs

The preceding discussion of the Georgi-Glashow model can be generalized to arbitrary groups \( G \) and \( H \). The condition for the existence of monopoles is that the map \( \phi_\infty \), mapping the sphere at infinity to the minima of the potential, can be topologically nontrivial. Since the orbit where the potential is minimized is diffeomorphic to the coset space \( G/H \), the condition is that \( \pi_2(G/H) \) be nontrivial.

The homotopy groups of this space are related to the homotopy groups of \( G \) and \( H \) by the so-called homotopy exact sequence. This is discussed in general in Appendix XXX. The part of the sequence that is relevant to us is

\[
\cdots \xrightarrow{\partial} \pi_2(H) \xrightarrow{\iota_*} \pi_2(G) \xrightarrow{\mu_*} \pi_2(G/H) \xrightarrow{\partial} \pi_1(H) \xrightarrow{\iota_*} \pi_1(G) \xrightarrow{\mu_*} \cdots
\] (1.141)

Here \( \iota_* \) are the homomorphisms of homotopy groups induced by the embedding \( \iota : H \to G \) and \( \mu_* \) are the homomorphisms induced by the projection \( \mu : G \to G/H \). The basepoints in the groups are the identity elements \( e \) and the basepoint in the coset space is the coset of the identity, \( eH \).

The following facts are known about Lie groups. The homotopy groups of \( U(1) \) are given in the table of Appendix XXX. Only the fundamental group is nonzero. If \( G \) is a compact, connected, simple Lie group \( G \), \( \pi_2(G) = 0 \) and \( \pi_3(G) = \mathbb{Z} \).
1.7. MONOPOLES

We can use these properties to deduce the second homotopy group of the coset space. One has to use the fact that the maps in the exact sequence are such that the image of each map is the kernel of the next. For example, in the case of GUTs, the group $G$ is compact, simple and simply connected and the subgroup $H$ contains an abelian factor (the unbroken electromagnetic $U(1)_Q$). Thus

$$
\ldots \rightarrow 0 \rightarrow 0 \rightarrow \pi_2(G/H) \rightarrow \mathbb{Z} \rightarrow 0 \rightarrow \ldots
$$

(1.142)

The fact that $\pi_2(G) = 0$ implies that the map $\mu_*$ is injective and the fact that $\pi_1(G) = 0$ implies that $\mu_*$ is surjective. Thus $\pi_2(G/H)$ is isomorphic to $\mathbb{Z}$, and the theory will have monopoles.

This argument does not apply to the Standard Model, because the group $G$ contains an abelian factor $U(1)_Y$. Even though this subgroup is not the same as the electromagnetic, unbroken group $U(1)_Q$, one can continuously deform one into the other by

$$
Q_t = tT_3 + Y , \quad 0 \leq t \leq 1
$$

and therefore $\iota_* : \pi_1(U(1)) \rightarrow \pi_1(G)$ is still an isomorphism. This implies that the image of $\partial$ is zero. On the other hand, since $\pi_2(SU(2) \times U(1)) = 0$, the map $\partial$ is still injective. Therefore we must have $\pi_2(SU(2) \times U(1)/U(1)) = 0$, and we conclude that the Standard Model does not admit monopole solutions. In this case one could have come to the same conclusion more easily by noting that the orbit of the minima is the locus where the norm of the complex Higgs doublet vanishes:

$$
|\phi_1|^2 + |\phi_2|^2 = v^2 ,
$$

which is a three-sphere.

Instead of appealing to the existence of the homotopy exact sequence, one can give an ad hoc proof of the above results, that goes as follows. \(^{16}\) We start from the map $\phi_\infty : S^2_\infty \rightarrow G/H$. As usual in homotopy, we think of it as a map $I \times I \rightarrow G/H$, such that $\phi_\infty(t_1, t_2) = eH$ whenever $t_1$ or $t_2$ is equal to 0 or 1. Using the gauge field $A$ we construct a map $g_\infty : S^2_\infty \rightarrow G$ as follows:

$$
g_\infty(t_1, t_2) = P \exp \int_0^{t_1} dt A(t, t_2) .
$$

\(^{16}\)See Coleman’s lectures on “Classical lumps and their quantum descendants”.

CHAPTER 1. \(\pi_0(\mathcal{Q})\) AND SOLITONS

The integral is along the line \((t, t_2)\) with constant \(t_2\). Since \(D\phi_\infty = 0\),

\[
g_\infty(t_1, t_2)eH = \phi_\infty ,
\]

or in other words \(\phi_\infty = \mu \circ g_\infty\). Clearly \(g(t_1, 0) = g(t_1, 1) = g(0, t_2) = e\). Since \(\mu(g_\infty(1, t_2)) = \phi_\infty(1, t_2) = eH, g_\infty(1, t_2) \in H\). We define \(h(t) = g_\infty(1, t_2)\). In this way we have constructed a map \(\partial : \pi_2(G/H) \to \pi_1(H)\) that maps \([\phi_\infty]\) to \([h]\). It can be checked that this map is a homomorphism.

Next we observe that the map \(g_\infty\) defines a homotopy of \(\iota \circ h\) (for \(t_1 = 1\)) to a constant (for \(t_1 = 0\)). Thus \(\text{im} \partial \subset \ker \iota^*\). Running the above argument backwards, given \(h : S^1 \to H\) such that \(\iota \circ h\) is homotopic to a constant, we construct a map \(\phi_\infty\) such that \(\partial([\phi_\infty]) = [h]\). Thus \(\partial\) is surjective. \(^{17}\)

To complete the proof, we must show that \(\partial\) is also injective, \(i.e.\) that if \(h\) is homotopic to a constant, also \([\phi_\infty]\) is homotopic to a constant. To this end, let us define \(\gamma : I \times I \to G\) by

\[
\gamma(t_1, t_2) = \begin{cases} 
g_\infty(2t_1, t_2) & \text{for } 0 \leq t_1 \leq \frac{1}{2} \\
g_\infty(1, t_2) & \text{for } \frac{1}{2} \leq t_1 \leq 1 .
\end{cases}
\]

and \(\varphi : S^2 \to G/H\) by

\[
\varphi(t_1, t_2) = \begin{cases} 
\phi_\infty(2t_1, t_2) & \text{for } 0 \leq t_1 \leq \frac{1}{2} \\
eH & \text{for } \frac{1}{2} \leq t_1 \leq 1 .
\end{cases}
\]

These maps are such that \(\varphi = \mu \circ \gamma\). Then, let \(h_t\) be a homotopy between \(h_0 = h\) and \(h_1 = eH\). If we replace \(\gamma\) by the map

\[
\gamma'(t_1, t_2) = \begin{cases} 
g_\infty(2t_1, t_2) & \text{for } 0 \leq t_1 \leq \frac{1}{2} \\
2t_1-1(t_2) & \text{for } \frac{1}{2} \leq t_1 \leq 1 .
\end{cases}
\]

we have again \(\varphi = \mu \circ \gamma'\), but now the map \(\gamma'\) is equal to \(e\) on the boundary of \(I \times I\), and therefore can be viewed as a map \(S^2 \to G\). Since \(\pi_2(G) = 0\), \(\gamma'\) is homotopic to a constant, and therefore also \(\varphi\) is homotopic to a constant, which is equivalent to saying that \(\phi_\infty\) is homotopic to a constant, QED.

\(^{17}\)This also proves the exactness of the homotopy sequence at \(\pi_1(H)\).