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

A *soliton* is a classical solution of nonlinear field equations which (1) is nonsingular, (2) has finite energy and (3) is localized in space. Further, more restrictive technical conditions may also be imposed, but will not concern us here. 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 this chapter, we will consider several examples of field theories that admit topological solitons.

In order to make the concept of topological soliton 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 the space $\mathcal{Q}$ defines the kinematics of the theory, but also encodes dynamical information via the condition of finiteness 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 $\mathcal{Q}$ 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 characterizes the 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 fields. 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 $\mathcal{Q}$. This is a formal procedure that would require much more technical work to be made precise, but holds great heuristic power. In the Schrödinger picture, the wave functions are complex functionals on $\mathcal{Q}$. If $\mathcal{Q}$ has several connected components, the Hilbert space $\mathcal{H}$ will split into subspaces called the topological sectors:

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

where $\mathcal{H}_i$ consists of wave functionals which are nonzero only on $\mathcal{Q}_i$. 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 $\mathcal{Q}_i$ to $\mathcal{Q}_j$ when $i \neq j$, so the transition amplitude between states in different sectors must vanish.

1.1. Scalar solitons in $1+1$ dimensions

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

$$S(\phi) = \int d^2x \left[-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right] \tag{1.1.1}$$

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, $i \in \mathcal{J}$, the minimum points. For definiteness one can think of the quartic potential

$$V = -\frac{1}{2}m^2\phi^2 + \frac{\lambda}{4}\phi^4 + \frac{m^4}{4\lambda} = \frac{\lambda}{4}\left(\phi^2 - f^2\right)^2, \tag{1.1.2}$$

with $f = \frac{m}{\sqrt{\lambda}}$ and m real and positive, with minima at points $y_\pm = \pm f$. With these assumptions, the energy:

$$E = \int_{-\infty}^{\infty} dx \left[\frac{1}{2}(\partial_0\phi)^2 + \frac{1}{2}(\partial_1\phi)^2 + V(\phi)\right] \tag{1.1.3}$$

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 (1.1.3) the first term represents the kinetic energy; the rest

$$E_S = \int_{-\infty}^{+\infty} dx \left[\frac{1}{2}(\partial_1\phi)^2 + V\right] \tag{1.1.4}$$

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 ϕ belongs to the space $\Gamma(\mathbf{R},\mathbf{R})$ 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 ϕ tends to one of the classical vacua, for otherwise the last two terms in E would diverge. We will call $\mathcal{Q}$ the subspace of $\Gamma(\mathbf{R},\mathbf{R})$ for which the static energy E_S is finite. If V has more that one minimum, then $\mathcal{Q}$ will not be connected. In fact, let

$$\mathcal{Q} = \bigcup_{i,j} \mathcal{Q}_{ij}\ , \qquad \mathcal{Q}_{ij} = \{\phi \in \mathcal{Q} \mid \phi \underset{x\to-\infty}{\longrightarrow} y_i, \phi \underset{x\to+\infty}{\longrightarrow} y_j\}\ .$$

Every path in $\Gamma(\mathbf{R},\mathbf{R})$ joining $\mathcal{Q}_{ij}$ to $\mathcal{Q}_{i'j'}$ (with $ij \neq i'j'$) must necessarily pass through the complement of $\mathcal{Q}$. In fact, to change the asymptotic behaviour of ϕ 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 $\mathcal{Q}_{ij}$ are separated by infinite energy barriers. For example in the case of the potential (1.1.2) there are four connected components of $\mathcal{Q}$, labelled $\mathcal{Q}_{++}$, $\mathcal{Q}_{+-}$, $\mathcal{Q}_{-+}$, $\mathcal{Q}_{--}$. In general, the set $\pi_0(\mathcal{Q})$ of connected components of $\mathcal{Q}$ is the cartesian product of two copies of the set indexing the minima: $\pi_0(\mathcal{Q}) = \mathcal{J} \times \mathcal{J}$.

Every $\phi \in \mathcal{Q}_{ij}$ can be written as the sum of an arbitrary given $\phi_0 \in \mathcal{Q}_{ij}$ (which we call the "basepoint" of $\mathcal{Q}_{ij}$) plus a function ψ which tends asymptotically to zero at $\pm\infty$. The function ψ can be regarded as a function $S^1 \to \mathbf{R}$, where $S^1 = \mathbf{R} \cup \{\infty\}$ is the one-point compactification of space. The space of such functions will be denoted $\Gamma_*(S^1,\mathbf{R})$. The subscript $*$ is there to remind us that we are dealing with functions which map a selected "basepoint" of S^1 (namely ∞) to the "basepoint" of $\mathbf{R}$ (namely 0). Therefore all connected components of $\mathcal{Q}$ are vectorspaces isomorphic to $\Gamma_*(S^1,\mathbf{R})$.

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 $\mathcal{Q}_{ii}$ the minima are the constant fields $\phi = y_i$. These are also the absolute minima of E on all $\mathcal{Q}$. In the case of the potential (1.1.2), 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 $\mathcal{Q}_{-+}$ and $\mathcal{Q}_{+-}$. Let us denote ℓ 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/ℓ, and hence decreases with ℓ, while the potential energy is of order $\lambda f^4 \ell$, and hence increases with ℓ. The static energy will have a minimum at some finite value of ℓ. The soliton will therefore be the result of a balance between elastic and potential energy.

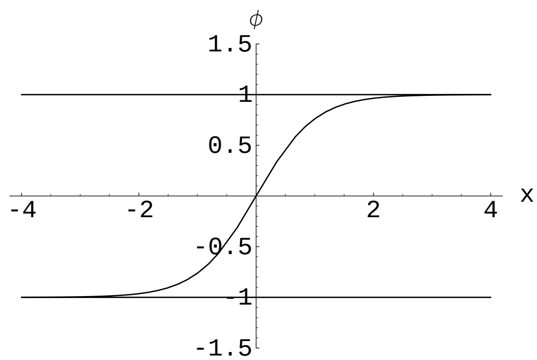
In order to find the explicit form of the soliton we have to solve the differential equation

$$\frac{d^2\phi}{dx^2} = \frac{\partial V}{\partial \phi} \tag{1.1.5}$$

with the appropriate boundary conditions. For the potential (1.1.2) the solutions of (1.1.5) in the sectors $\mathcal{Q}_{-+}$ and $\mathcal{Q}_{+-}$ are

$$\phi(x) = \pm \frac{m}{\sqrt{\lambda}} \tanh\left[\frac{m}{\sqrt{2}}(x - x_0)\right] , \tag{1.1.6}$$

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. The presence of the arbitrary point x_0, which marks the "center" of the kink is a reflection of the translational invariance of the action. The following figure shows a plot of ϕ/f as a function of $x\sqrt{2}/m$ for the kink at $x_0 = 0$. (The horizontal lines correspond to the minma of the potential.)



In the theory with potential (1.1.2), consider the current

$$J_T^\mu = \frac{1}{2f} \varepsilon^{\mu\nu} \partial_\nu \phi \; ; \tag{1.1.7}$$

clearly we have

$$\partial_\mu J_T^\mu = 0 \; . \tag{1.1.8}$$

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

$$Q_T = \int_{-\infty}^{\infty} dx \, J_T^0 = \frac{1}{2f} [\phi(+\infty) - \phi(-\infty)] \tag{1.1.9}$$

is known as the topological charge. It is clear that all fields in $\mathcal{Q}_{-+}$ have $Q_T = 1$, those in $\mathcal{Q}_{+-}$ have $Q_T = -1$ and those in $\mathcal{Q}_{++}$ and $\mathcal{Q}_{--}$ have $Q_T = 0$. Thus Q_T is a measure of the nontriviality of the boundary conditions of the fields.

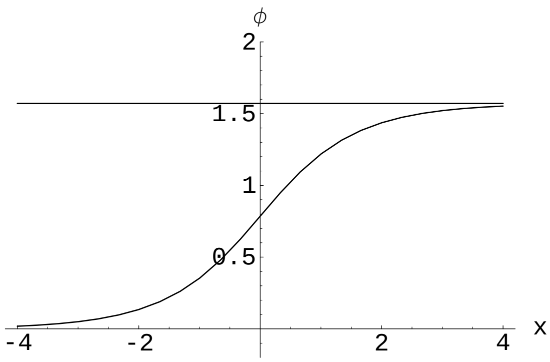
Another interesting potential is

$$V(\phi) = \frac{m^4}{\lambda} \left[1 - \cos\left(\frac{\sqrt{\lambda}}{m}\phi\right)\right] \; . \tag{1.1.10}$$

This corresponds to the so called "sine-Gordon" model. The indexing set of minima is the set of the integers $\mathcal{J} = \mathbf{Z}$, so there is a double infinity ($\mathbf{Z} \times \mathbf{Z}$) of connected components. The topological current and the topological charge are given again by (1.1.8) and (1.1.9), 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 $\mathcal{Q}_{0\,1}$ and $\mathcal{Q}_{0\,-1}$

$$\phi(x) = \pm \frac{4m}{\sqrt{\lambda}} \arctan\big\{\exp\big[(x - x_0)m\big]\big\} \tag{1.1.11}$$

This solution is plotted in the following figure (ϕ is measured in units of $\sqrt{\lambda}/4m$ and x is measured in units of m).



Just adding $2nf$ we get the soliton and antisoliton, still with $Q_T = \pm 1$, which minimize the energy in $\mathcal{Q}_{n\,n+1}$ and $\mathcal{Q}_{n\,n-1}$. Note that if in the field equation (1.1.5) with the potential (1.1.10) we reinterpret x as time and ϕ 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 (1.1.11) 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.

1.2. Linear and nonlinear scalar fields in higher dimensions

Next we consider the case of a single scalar field in $d > 1$ space dimensions. Finiteness of the static energy

$$E_S = \int d^d x \left[\frac{1}{2}\sum_{i=1}^{d}(\partial_i\phi)(\partial_i\phi) + V(\phi)\right] \tag{1.2.1}$$

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

$$\mathcal{Q} = \bigcup_{i\in\mathcal{J}} \mathcal{Q}_i\ , \qquad \mathcal{Q}_i = \{\phi \in \mathcal{Q} \mid \phi \underset{r\to\infty}{\longrightarrow} y_i\}$$

and $\mathcal{J}$ is the set of the minima of V. The absolute minimum of E_S in each $\mathcal{Q}_i$ 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_∞ 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^{d-1}_∞ is connected. By continuity the value of the field at infinity must be constant and there cannot be topological solitons in this theory.

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 $\Gamma(\mathbf{R}^d, \mathbf{R}^N)$. Assuming symmetry under $SO(N)$, the action is

$$S = \int d^{d+1}x \left[-\frac{1}{2}\partial_\mu\phi^a\partial^\mu\phi^a - V(|\phi|)\right]\ , \tag{1.2.2}$$

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

$$V=-\frac{1}{2}m^2|\phi|^2+\frac{\lambda}{4}|\phi|^4+\frac{m^4}{4\lambda}=\frac{\lambda}{4}\left(|\phi|^2-f^2\right)^2, \tag{1.2.3}$$

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

$$E_S=\int d^dx\left[\frac{1}{2}\partial_i\phi^a\partial_i\phi^a+V(|\phi|)\right]. \tag{1.2.4}$$

We are interested in the subspace $\mathcal{Q}\subset\Gamma(\mathbf{R}^d,\mathbf{R}^N)$ for which the static energy is finite. This demands again that as $r\to\infty$, ϕ tends to one of the minima of V.

One can ask whether it is necessary to allow ϕ 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 ϕ and ϕ' be two field configurations such that $\phi\underset{r\to\infty}{\longrightarrow}y$ and $\phi'\underset{r\to\infty}{\longrightarrow}y'$, where y and y' are two different points on S^{N-1}. Since all maps from $\mathbf{R}^d$ to $\mathbf{R}^N$ are homotopic, there exists a one-parameter family of maps $\phi_\tau(x)$, with $0\leq\tau\leq1$, 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 ∞ instead than from 0 to 1. For example, we can define

$$\tau=\frac{1}{2}+\frac{1}{\pi}\arctan t\,. \tag{1.2.5}$$

Writing $\phi_\tau(x)=\hat\phi(x,t)$, we can interpret t as time and $\hat\phi\in\Gamma(\mathbf{R}^{d+1},\mathbf{R}^N)$ as a *spacetime* field. The energy of this field is $E=E_K+E_S$ where $E_K=\int d^dx\,\frac{1}{2}\left(\frac{d\hat\phi}{dt}\right)^2$ 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 ϕ to ϕ' one must go through configurations with infinite kinetic energy, so the boundary value of ϕ cannot change in the course of the time evolution.

Using the $SO(N)$ invariance of the theory, we can assume without loss of generality that the value of ϕ 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}_∞; since the map ϕ must be constant on S^{d-1}_∞, all its points may be identified to a single point ∞. Then ϕ may be regarded as a map from the one-point compactification $\mathbf{R}^d\cup\{\infty\}=S^d$ into $\mathbf{R}^N$, mapping the "basepoint" ∞ of S^d to the "basepoint" y_0. Therefore $\mathcal{Q}=\Gamma_*(S^d,\mathbf{R}^N)$. All maps with these properties are homotopic to each other, so the space $\mathcal{Q}$ is connected.

These results imply that linear scalar field theories in dimensions $d\geq2$ cannot have topological solitons. There is an independent result, known as Derrick's theorem, saying that linear scalar field theories with action (1.2.2) do not admit nontrivial static solutions (whether topological or not) when $d\geq2$. The proof is based on a scaling argument.

Let us rewrite equation (1.2.4) 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 ϕ_λ be a one-parameter family of configurations defined by $\phi_\lambda(x)=\phi_1(\lambda x)$. We have

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

In order for ϕ_1 to be a stationary point of E_S it is necessary that

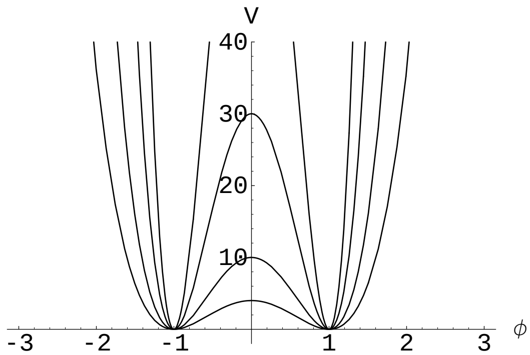
$$0=\frac{d}{d\lambda}E_S(\phi_\lambda)\Big|_{\lambda=1}=(2-d)E_1(\phi_1)-dE_2(\phi_1)\,. \tag{1.2.6}$$

Since E_1 and E_2 are positive semidefinite, for $d\geq3$ 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 $E_2(\phi_1)=0$ which implies $\frac{\partial V}{\partial\phi^a}=0$. Inserting in the equation of motion we obtain $\partial_\mu\partial^\mu\phi^a=0$, which, together with the given boundary conditions, implies again $\phi=$constant.

There are two very different ways in which this theorem can be evaded. One is to couple the scalars to gauge fields; this will be studied in Sections 1.6 and 1.7. The other way is to consider scalar fields that take values in a nonlinear space (e.g. a sphere). The dynamics will necessarily be modified, so there is a chance of escaping the scaling argument. Note that the topology of the configuration space will also be modified, so the existence of topological solitons is not excluded. We shall discuss such nonlinear scalar theories in the rest of this section, and in the next two sections.

Consider first a linear scalar theory with action (1.2.2). 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 $\mathbf{R}^N$. 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 better way of studying the limit is to introduce a Lagrange multiplier field Λ and consider the action

$$S=\int d^{d+1}x\left[-\frac{1}{2}\partial_\mu\phi^a\partial^\mu\phi^a-\frac{2\Lambda}{\sqrt{\lambda}}\sqrt{V}+\frac{\Lambda^2}{\lambda}\right]. \tag{1.2.7}$$

The equation of motion for Λ is $\Lambda=\sqrt{\lambda V}$ and when this equation is used in (1.2.7) it gives back (1.2.2). Thus (1.2.7) is classically equivalent to (1.2.2). The advantage of the action (1.2.7) is that it remains well defined in the limit $\lambda\to\infty$. In fact, it reduces to

$$S=\int d^{d+1}x\left[-\frac{1}{2}\partial_\mu\phi^a\partial^\mu\phi^a-\Lambda(|\phi|^2-f^2)\right]. \tag{1.2.8}$$

The second term enforces the constraint $\phi^2=f^2$.

This argument can be generalized to scalar fields carrying a representation of any Lie group G. † If ϕ_0 is a minimum of the potential, every other point in the orbit of G through ϕ_0 is also a minimum. We assume that all the minima belong to a single orbit. As shown in Appendix C, if H is the stabilizer of ϕ_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.

All solutions of the nonlinear sigma model equations have $E_2=0$. Therefore (1.2.6) implies that if $d>2$ the only static solution of the nonlinear sigma model equations (coming from (1.2.8)) have $\phi=$ const, while in $d=2$ nontrivial solutions are possible. In the next section we will study this new class of solitons.

1.3. Nonlinear sigma model in $d=2$

We now consider the S^2-valued NSM in $d=2$. The action is given by (1.2.8), with $a=1,2,3$. It is convenient to solve the constraint $\phi^a\phi^a=f^2$ expressing the three fields ϕ^a in terms of only two independent

† See 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)

fields φ^α. There are infinitely many ways of doing this. For example we could choose φ^α to be the spherical coordinates ($\varphi^1 = \Theta$, $\varphi^2 = \Phi$):

$$\begin{aligned} \phi^1 &= f \sin\Theta\cos\Phi \\ \phi^2 &= f \sin\Theta\sin\Phi \\ \phi^3 &= f \cos\Theta \end{aligned} \tag{1.3.1}$$

Introducing into (1.2.8), we find the action

$$S = -\frac{f^2}{2}\int d^3x \left(\partial_\mu\Theta\partial^\mu\Theta + \sin^2\Theta\partial_\mu\Phi\partial^\mu\Phi\right) . \tag{1.3.2}$$

Another choice are the stereographic coordinates $\varphi^1 = \omega^1$, $\varphi^2 = \omega^2$:

$$\begin{aligned} \phi^1 &= f\frac{4\omega_1}{\omega_1^2+\omega_2^2+4} \\ \phi^2 &= f\frac{4\omega_2}{\omega_1^2+\omega_2^2+4} \\ \phi^3 &= f\frac{\omega_1^2+\omega_2^2-4}{\omega_1^2+\omega_2^2+4} \end{aligned} \tag{1.3.3}$$

Introducing in (1.2.8),

$$S = -\frac{f^2}{2}\int d^3x \,\frac{16}{(\omega_1^2+\omega_2^2+4)^2}\left(\partial_\mu\omega_1\partial^\mu\omega_1 + \partial_\mu\omega_2\partial^\mu\omega_2\right) . \tag{1.3.4}$$

In any case the action has the form

$$S = -\frac{f^2}{2}\int d^3x \,\partial_\mu\varphi^\alpha\partial^\mu\varphi^\beta h_{\alpha\beta}(\varphi) , \tag{1.3.5}$$

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 with the dimension of mass.

Let us now consider the configuration space of the NSM. So far φ is a map from $\mathbf{R^2}$ to S^2. Finiteness of the static energy

$$E_S = \frac{f^2}{2}\int d^2x \,\partial_i\varphi^\alpha\partial_i\varphi^\beta h_{\alpha\beta}(\varphi) \tag{1.3.6}$$

demands that $\partial_i\varphi \to 0$ as $r\to\infty$. Thus φ 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: $S^2 = \mathbf{R}^2\cup\{\infty\}$. 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 ∞, 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: $\pi_0(\mathcal{Q}) = \pi_2(S^2) = \mathbf{Z}$ (see Appendix F). So we can write $\mathcal{Q} = \bigcup_{i\in\mathbf{Z}}\mathcal{Q}_i$; the number i labelling the homotopy classes is known as the winding number. A general formula for this quantity is given in Eq.(A.3). In spherical and stereographic coordinates it has the expression

$$\begin{aligned} W(\Theta,\Phi) &= \frac{1}{4\pi}\int d^2x \,\sin\Theta\varepsilon^{ij}\partial_i\Theta\partial_j\Phi , \\ W(\omega^1,\omega^2) &= \frac{1}{4\pi}\int d^2x \,\frac{16}{(\omega_1^2+\omega_2^2+4)^2}\varepsilon^{ij}\partial_i\omega^1\partial_j\omega^2 , \end{aligned} \tag{1.3.7}$$

respectively. In any coordinate system

$$W(\varphi^\alpha) = \frac{1}{8\pi}\int d^2x\,\varepsilon^{ij}\partial_i\varphi^\alpha\partial_j\varphi^\beta\sqrt{\det h}\,\varepsilon_{\alpha\beta}\ . \tag{1.3.8}$$

One can think of W as a function on $\mathcal{Q}$: it is constant and equal to i on each connected component $\mathcal{Q}_i$. It is intuitively clear that since the time evolution is a continuous curve in $\mathcal{Q}$, the value of the winding number cannot change, so W must be a constant of motion of the theory. This conclusion can be substantiated by the following calculation. We define a topological current

$$J_T^\lambda = \frac{1}{8\pi}\varepsilon^{\lambda\mu\nu}\partial_\mu\varphi^\alpha\partial_\nu\varphi^\beta\sqrt{\det h}\,\varepsilon_{\alpha\beta}\ , \tag{1.3.9}$$

which is identically conserved:

$$\partial_\lambda J_T^\lambda = \frac{1}{8\pi}\varepsilon^{\lambda\mu\nu}\partial_\lambda\varphi^\gamma\partial_\mu\varphi^\alpha\partial_\nu\varphi^\beta\varepsilon_{\alpha\beta}\frac{\partial\sqrt{\det h}}{\partial\varphi^\gamma} = 0\ , \tag{1.3.10}$$

because the indices α, β, γ run from 1 to 2 and at least one index must be repeated. One sees immediately that the topological charge is equal to the winding number:

$$Q_T = \int d^2x\,J_T^0 = \frac{1}{8\pi}\int d^2x\,\varepsilon^{ij}\partial_i\varphi^\alpha\partial_j\varphi^\beta\varepsilon_{\alpha\beta}\sqrt{\det h} = W(\varphi)\ . \tag{1.3.11}$$

There follows that $Q_T = W$ is a constant of motion. (See also exercise 1.3.1).

Let us look at the absolute minimum of the static energy (1.3.6) in each topological sector $\mathcal{Q}_i$. Consider the following inequality ¶

$$0 \le \int d^2x\,h_{\alpha\beta}\Big(\partial_i\varphi^\alpha \pm \varepsilon_{ik}\partial_k\varphi^\gamma\varepsilon_{\gamma\delta}h^{\delta\alpha}\sqrt{\det h}\Big)\Big(\partial_i\varphi^\beta \pm \varepsilon_{ij}\partial_j\varphi^\epsilon\varepsilon_{\epsilon\varphi}h^{\varphi\beta}\sqrt{\det h}\Big) =$$

$$= \int d^2x\Big[2h_{\alpha\beta}\partial_i\varphi^\alpha\partial_i\varphi^\beta \mp 2\varepsilon_{ij}\partial_i\varphi^\alpha\partial_j\varphi^\epsilon\varepsilon_{\alpha\epsilon}\sqrt{\det h}\Big] = \frac{4}{f^2}E_S \mp 16\pi W \tag{1.3.12}$$

where we used $h_{\gamma\epsilon} = (\det h)\varepsilon_{\gamma\delta}\varepsilon_{\epsilon\varphi}h^{\delta\varphi}$. If $W > 0$ (resp. $W < 0$) the inequality with the upper sign (resp. lower sign) is stronger. There follows that

$$E_S \ge 4\pi f^2|W|\ . \tag{1.3.13}$$

Furthermore, equality holds if and only if

$$\partial_i\varphi^\alpha = \mp\varepsilon_{ik}\partial_k\varphi^\gamma\varepsilon_{\gamma\delta}h^{\delta\alpha}\sqrt{\det h}\ . \tag{1.3.14}$$

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 (1.3.14) are first order equations, and therefore simpler than the second order Euler-Lagrange equations. It is convenient to specialize the discussion to stereographic coordinates ω^1 and ω^2. Equation (1.3.14) reduces to

$$\partial_i\omega_\alpha = \mp\varepsilon_{ik}\partial_k\omega^\gamma\varepsilon_{\gamma\alpha}\ , \tag{1.3.15}$$

and spelling these out

$$\begin{aligned}\partial_1\omega^1 &= \pm\partial_2\omega^2\ ,\\ \partial_2\omega^1 &= \mp\partial_1\omega^2\ .\end{aligned} \tag{1.3.16}$$

¶ A.M. Polyakov, A.A. Belavin, "Metastable States of Two-Dimensional Isotropic Ferromagnets", JETP Lett. **22** 245-248 (1975) Pisma Zh. Eksp. Teor. Fiz. **22** 503-506 (1975).

If we define $\omega = \omega^1 + i\,\omega^2$ and $z = x^1 + i\,x^2$ we recognize (1.3.16) 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 (1.3.16). For example $\omega(z) = z^n$ and $\omega(z) = (z^*)^n$, with $n \geq 0$, are solutions of (1.3.16). Note that for large $|z|$, ω 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 $\varphi \in \Gamma_*(S^2, S^2)$ with winding number $W = n$ and $W = -n$ respectively. They are absolute minima of the static energy in the sectors $\mathcal{Q}_n$ and $\mathcal{Q}_{-n}$ 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

$$\begin{aligned} \omega(z) &= \left(\frac{(z - z_0) e^{i\alpha}}{\lambda} \right)^n \\ \omega(z) &= \left(\frac{(z - z_0)^* e^{-i\alpha}}{\lambda} \right)^n \end{aligned} \tag{1.3.17}$$

where the complex number z_0 gives the position of the center of the soliton, the angle α its "internal orientation" and the positive real number λ its scale. The parameters z, α and λ will be called the "collective coordinates" of the soliton.

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 state describe massless particles (see below). The field φ 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.

We conclude this section with some remarks on the quantization of nonlinear sigma models. In n spacetime dimensions, the action can be written

$$S = -\frac{f^2}{2} \int d^n x\, \partial_\mu \varphi^\alpha \partial^\mu \varphi^\beta h_{\alpha\beta}(\varphi) \tag{1.3.18}$$

Since the metric is in general a nonpolynomial function of the fields, these have to be dimensionless. Therefore the constant f^2 multiplying the action must have dimension L^{2-n}; only in two spacetime dimensions one can choose $f^2 = 1$. In order to give the scalar fields their canonical dimension we absorb first the constant f^2 in the fields defining $\bar\varphi^\alpha = f\varphi^\alpha$. The dimension of $\bar\varphi$ is then $[\bar\varphi^\alpha] = L^{\frac{2-n}{2}}$. Now the action reads

$$S = -\frac{1}{2} \int d^n x\, \partial_\mu \bar\varphi^\alpha \partial^\mu \bar\varphi^\beta h_{\alpha\beta}\left(\frac{\bar\varphi}{f}\right) \tag{1.3.19}$$

The metric $h_{\alpha\beta}\left(\frac{\bar\varphi}{f}\right)$ 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 η:

$$h_{\alpha\beta}\left(\frac{\bar\varphi}{f}\right) = h_{\alpha\beta}\left(\frac{\bar\varphi_0}{f}\right) + \frac{1}{f}\partial_\gamma h_{\alpha\beta}\left(\frac{\bar\varphi_0}{f}\right)\eta^\gamma + \frac{1}{2f^2}\partial_\gamma\partial_\delta h_{\alpha\beta}\left(\frac{\bar\varphi_0}{f}\right)\eta^\gamma\eta^\delta + \cdots \tag{1.3.20}$$

where we write ∂_γ 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 $\mathbf{Z}_2$ invariance under the transformation $\eta \to -\eta$ forbids terms with an odd number of fields.) The dimension of the coupling constant in the m-th term, i.e. the coefficient of $\partial\eta\,\partial\eta\,\eta^m$ has dimension $\left[\left(\frac{1}{f}\right)^m\right] = L^{\frac{m}{2}(n-2)}$. In spite of the infinite number of couplings, this theory is renormalizable in a generalized sense for $n=2$. It is nonrenormalizable for $n>2$. Therefore, the nonlinear sigma model is not a good candidate for a fundamental theory in more than two dimensions. Instead, it is used in four dimensions as a low energy phenomenological approximation of QCD. This will be the topic of the next section.

1.4. Current Algebra and Skyrmions

Let us consider a nonlinear sigma model with values in some space N in $d>1$ space dimensions. 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 $\mathcal{Q} = \Gamma_*(S^d, N)$. 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 $\pi_3(G) = \mathbf{Z}$, 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.

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_μ, 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_α, where α is an index that runs over the six flavors. The quark part of the QCD action is

$$S_q = \sum_\alpha \int d^4x\, \bar\psi_\alpha \left(i\gamma^\mu D_\mu - m_\alpha\right)\psi_\alpha \ . \tag{1.4.1}$$

where D_μ 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

$$\delta_{V\alpha}\psi = i\alpha\psi \quad ; \quad \delta_{V\alpha}\bar\psi = -i\alpha\bar\psi (1.3.18)$$

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

$$\delta_{V\epsilon}\psi = \epsilon^a T_a\psi \quad ; \quad \delta_{V\epsilon}\bar\psi = -\bar\psi\epsilon^a T_a \ (1.3.18)$$

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:

$$\begin{aligned} \delta_{A\alpha}\psi &= i\alpha\gamma^A\psi \quad ; \quad \delta_{A\alpha}\bar\psi = i\alpha\bar\psi\gamma^A \qquad U(1)_A \ ; \\ \delta_{A\epsilon}\psi &= \epsilon^a T_a\gamma^A\psi \quad ; \quad \delta_{A\epsilon}\bar\psi = \bar\psi\epsilon^a T_a\gamma^A \quad SU(N)_A \ . \end{aligned} \tag{1.4.2}$$

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:

$$\begin{aligned} j^\mu_V &= \bar\psi\gamma^\mu\psi && \text{for } U(1)_V \\ j^\mu_A &= \bar\psi\gamma^\mu\gamma^A\psi && \text{for } U(1)_A \\ j^\mu_{V\epsilon} &= \bar\psi\gamma^\mu\epsilon^a T_a\psi && \text{for } SU(N)_V \\ j^\mu_{A\epsilon} &= \bar\psi\gamma^\mu\gamma^A\epsilon^a T_a\psi && \text{for } SU(N)_A \end{aligned} \tag{1.4.3}$$

where ϵ is an element of the Lie algebra of $SU(N)$. From the canonical equal-time anticommutation relations

$$\{\psi^{\alpha i}(\vec{x},t), \bar{\psi}_{\beta j}(\vec{y},t)\} = \delta^\alpha_\beta \delta^i_j \delta(\vec{x}-\vec{y}) \ , \tag{1.4.4}$$

where a, b are Dirac indices and i, j are $SU(N)$ indices, we obtain the following current algebra

$$\begin{aligned} [j^0_V, j^0_V] = [j^0_V, j^0_A] = [j^0_A, j^0_A] = 0 \ ; \\ [j^0_{V\epsilon_1}, j^0_{V\epsilon_2}] = j^0_{V[\epsilon_1,\epsilon_2]} \ ; \\ [j^0_{V\epsilon_1}, j^0_{A\epsilon_2}] = j^0_{A[\epsilon_1,\epsilon_2]} \ ; \\ [j^0_{A\epsilon_1}, j^0_{A\epsilon_2}] = j^0_{V[\epsilon_1,\epsilon_2]} \ ; \end{aligned} \tag{1.4.5}$$

One can verify that these are the algebras implied by (1.4.2).

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 γ^A satisfies $(\gamma^A)^2 = \mathbf{1}$, the operators

$$P_\pm = \frac{1 \pm \gamma^A}{2} \tag{1.4.6}$$

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

$$\begin{aligned} j^\mu_{L\epsilon} =& \frac{j^\mu_{V\epsilon} - j^\mu_{A\epsilon}}{2} = \bar{\psi}\gamma^\mu P_- \epsilon^a T_a \psi \ ; \\ j^\mu_{R\epsilon} =& \frac{j^\mu_{V\epsilon} + j^\mu_{A\epsilon}}{2} = \bar{\psi}\gamma^\mu P_+ \epsilon^a T_a \psi \ ; \end{aligned} \tag{1.4.7}$$

we can rewrite (1.4.5) as

$$\begin{aligned} [j^0_{L\epsilon_1}, j^0_{L\epsilon_2}] =& j^0_{L[\epsilon_1,\epsilon_2]} \ ; \\ [j^0_{L\epsilon_1}, j^0_{R\epsilon_2}] =& 0 \ ; \\ [j^0_{R\epsilon_1}, j^0_{R\epsilon_2}] =& j^0_{R[\epsilon_1,\epsilon_2]} \ ; \end{aligned} \tag{1.4.8}$$

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 eightfold 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 (1.4.8) 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 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 $U(x)\in SU(N)$.

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

$$S=\frac{f^2}{4}\int d^4x\,\mathrm{tr}\big(U^{-1}\partial_\mu UU^{-1}\partial^\mu U\big)\ . \tag{1.4.9}$$

(We are using an inner product in the Lie algebra such that $\mathrm{tr}(T_aT_b)=-\frac{1}{2}\delta_{ab}$. In the case $N=2$, $T_a=-\frac{i}{2}\sigma_a$, where σ^a are the Pauli matrices). If we choose a coordinate system on the group and call $\varphi^\alpha(x)$ the coordinates of the group element $U(x)$, the action (1.4.9) can be shown to be identical to the nonlinear sigma model action (1.3.5) with riemannian metric $h_g(v,w)=\mathrm{tr}(g^{-1}vg^{-1}w)$. (See Exercise 1.4.1). The advantage of the form (1.4.9) is that it makes the $SU(N)_L\times SU(N)_R$ invariance of the theory more transparent.

For the present purposes, the most useful coordinates on $SU(N)$ are the normal coordinates π^α, defined by:

$$U(x)=e^{2\pi^a(x)T_a/f} \tag{1.4.10}$$

where T_a is a basis in the Lie algebra of $SU(N)$, satisfying $\big[T_a,T_b\big]=f_{ab}{}^cT_c$, . Note that the coordinates have been scaled as in (1.3.19) so as to have the canonical dimension of mass.

Using (1.4.10), (1.4.9) can be expanded as

$$\int d^4x\left[-\frac{1}{2}\partial_\mu\pi^a\partial^\mu\pi^a+\frac{1}{f^2}\varepsilon^{abc}\pi^b\partial_\mu\pi^c\varepsilon^{ade}\pi^d\partial^\mu\pi^e+\ldots\right]\ . \tag{1.4.11}$$

This corresponds to the expansion (1.3.20) in normal coordinates in the neighborhhod of the identity. One observes that in this model the pions are massless. Furthermore, all interactions contain derivatives of the fields: this is as it should be, since a potential for π would certainly break the global invariance of the theory.

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 1.4.1):

$$W(U)=-\frac{1}{24\pi^2}\int d^3x\,\varepsilon^{\lambda\mu\nu}\mathrm{tr}\Big(U^{-1}\partial_\lambda UU^{-1}\partial_\mu UU^{-1}\partial_\nu U\Big)\ . \tag{1.4.12}$$

For other groups, the generator of $\pi_3(G)=\mathbf{Z}$ can be obtained by embedding $SU(2)$ in G and then considering the composition of this embedding with a map $S^3\rightarrow SU(2)$ 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 (1.4.12), that

$$W(U_1U_2)=W(U_1)+W(U_2)\ \ ;\qquad\qquad W(U^{-1})=-W(U)\ . \tag{1.4.13}$$

A field configuration of the form

$$U(\vec{x})=\exp\big[T_a\hat{x}^ag(r)\big] \tag{1.4.14}$$

where $\hat{x}^a = \frac{x^a}{r}$ and g is a function which is $-\pi$ in the origin and tends to zero as $r \to \infty$, has winding number one. From (1.4.13), configurations with arbitrary winding numbers can be constructed simply taking powers of (1.4.14). In the case of other semisimple Lie groups, if the basis in the Lie algebra of G is such that T_1, T_2, T_3 span the subalgebra of $SU(2)$, then the map (1.4.14) defines the generator of $\pi_3(G)$.

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 (1.4.9). In fact, from (1.2.6) we get

$$\frac{dE(\phi_\lambda)}{d\lambda}\Big|_{\lambda=1} = -E(\phi_1) < 0 ,$$

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. ‡ 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

$$S = \int d^4x \left(\frac{f^2}{4} \mathrm{tr}\big(U^{-1}\partial_\mu U U^{-1}\partial^\mu U\big) + \frac{1}{32e^2} \mathrm{tr}\big[U^{-1}\partial_\mu U, U^{-1}\partial_\nu U\big]\big[U^{-1}\partial^\mu U, U^{-1}\partial^\nu U\big] \right) , \qquad (1.4.15)$$

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 (1.4.14) in the equations of motion that come from (1.4.15), 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 ℓ 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 $1/fe$. Note that for $e \to \infty$, ℓ 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 (1.4.15) 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 theory

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 $\mathbf{g}$ of a group G: $A = A^a_\mu dx^\mu \otimes T_a$, where $\{T_a\}$ is a basis in $\mathbf{g}$. With A one can construct the nonabelian field strength

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + e f_{abc} A^b_\mu A^c_\nu , \qquad (1.5.1)$$

where $[T_a, T_b] = f_{ab}{}^c T_c$ and e is the coupling constant of the theory. The Yang–Mills action is

$$S_{YM} = -\frac{1}{4} \int d^{d+1}x\, F^a_{\mu\nu} F^{a\mu\nu} . \qquad (1.5.2)$$

‡ T.H.R. Skyrme, "A Nonlinear field theory" Proc. Roy. Soc. Lond. **A260** 127-138 (1961); "A Unified Field Theory of Mesons and Baryons" Nucl. Phys. **31** 556-569 (1962).

It is invariant under local gauge transformations

$$A_\mu \to g^{-1} A_\mu g + \frac{1}{e} g^{-1} \partial_\mu g \ , \qquad F_{\mu\nu} \to g^{-1} F_{\mu\nu} g \ , \tag{1.5.3}$$

where $g : \mathbf{R}^{d+1} \to G$ and $F_{\mu\nu} = F^a_{\mu\nu} T_a$.

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

$$S_{YM} = -\frac{1}{4e^2} \int d^{d+1}x \, F^a_{\mu\nu} F^{a\mu\nu} \ , \tag{1.5.4}$$

where the curvature is now defined by

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f_{abc} A^b_\mu A^c_\nu \ , \tag{1.5.5}$$

The nonabelian gauge transformations then read

$$A_\mu \to g^{-1} A_\mu g + g^{-1} \partial_\mu g \ , \qquad F_{\mu\nu} \to g^{-1} F_{\mu\nu} g \ . \tag{1.5.6}$$

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 (1.5.2). In later chapters we will use the rescaled fields, with action (1.5.4).

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

$$\mathcal{L}_{YM} = \frac{1}{2} E^a_i E^a_i - \frac{1}{4} F^a_{ij} F^a_{ij} \ , \tag{1.5.7}$$

where $E^a_i = F^a_{0i} = \partial_0 A^a_i - D_i A^a_0$ is the nonabelian "electric" field (we have used the notation $D_i A^a_0 = \partial_i A^a_0 + e f_{abc} A^b_i A^c_0$; 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

$$P^0_a = \frac{\partial \mathcal{L}_{YM}}{\partial_0 A^a_0} = 0 \ , \qquad P^i_a = \frac{\partial \mathcal{L}_{YM}}{\partial_0 A^a_i} = E^a_i \ .$$

The phase space of the theory is initially coordinatized by the variables A^a_μ and P^μ_a. However, the relation between velocities and momenta is not invertible, so this is a theory with constraints, in the terminology of Dirac. We will not go through the whole canonical analysis here. The canonical Hamiltonian can be written

$$H = \int d^d x \left[\frac{1}{2} E^a_i E^a_i + \frac{1}{4} F^a_{ij} F^a_{ij} - A^a_0 G_a \right] \ , \tag{1.5.8}$$

where $G_a = D_i P^i_a = D_i E^a_i$. The fields A^a_0 play the role of Lagrange multipliers enforcing the Gauss law $G_a = 0$. Note that the last term of the Hamiltonian vanishes upon using the equations of motion. 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 $g(x,t) = \mathrm{P} \exp\left(-e \int^t dt' \, A_0(x,t') \right)$, where P stands for path ordering). This leaves the freedom of performing time-independent gauge transformations. In this gauge $E^a_i = \dot{A}^a_i$, so the first term in (1.5.8) 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 the gauge field A_μ 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 YM theory does not admit static solitons if $d\neq 4$. ∗ 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 $E_1=\frac{1}{2}\int d^dx\,(D_iA_0)^2>0$ and $E_2=\frac{1}{4}\int d^dx\,(F^a_{ij})^2>0$. Consider the two-parameter family of configurations $A_{(\sigma,\lambda)}$ defined by

$$
\begin{aligned}
A_{(\sigma,\lambda)}{}^a_0(x) &= \sigma\lambda A^a_0(\lambda x)\ ,\\
A_{(\sigma,\lambda)}{}^a_i(x) &= \lambda A^a_i(\lambda x)\ .
\end{aligned}
\tag{1.5.9}
$$

We have $E_1\big(A_{(\sigma,\lambda)}\big)=\sigma^2\lambda^{4-d}E_1\big(A_{(1,1)}\big)$ and $E_2\big(A_{(\sigma,\lambda)}\big)=\lambda^{4-d}E_2\big(A_{(1,1)}\big)$. For $A_{(1,1)}$ to be a solution of the field equations we must have

$$
\begin{aligned}
0 &= \frac{d}{d\lambda}L\Big|_{\lambda=\sigma=1} = (4-d)L\big(A_{(1,1)}\big)\ ,\\
0 &= \frac{d}{d\sigma}L\Big|_{\lambda=\sigma=1} = 2E_1\big(A_{(1,1)}\big)\ ,
\end{aligned}
\tag{1.5.10}
$$

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

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 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. The Nielsen-Olesen vortex

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

$$
S=\int d^3x\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}-\frac{1}{2}|D_\mu\phi|^2-\frac{\lambda}{4}\Big(|\phi|^2-f^2\Big)^2\right]\ ,
\tag{1.6.1}
$$

where $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$, $D_\mu\phi=\partial_\mu\phi-ieA_\mu\phi$. (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 $A=A^1T=-iA^1$. The field A_μ used in this section is A^1_μ stripped of the index 1. The gauge transformations can then be obtained from (1.5.3) by putting $g=e^{i\alpha}$.) The theory is invariant under the local gauge transformations

$$
A_\mu\to A'_\mu=A_\mu+\frac{i}{e}g^{-1}\partial_\mu g=A_\mu-\frac{1}{e}\partial_\mu\alpha\ ,\qquad \phi\to\phi'=g^{-1}\phi=e^{-i\alpha(x)}\phi\ .
\tag{1.6.2}
$$

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

$$
E_K=\int d^2x\left[\frac{1}{2}E_iE_i+\frac{1}{2}|D_0\phi|^2\right]\ .
\tag{1.6.3}
$$

and E_S is the static energy

$$
E_S=\int d^2x\left[\frac{1}{2}B^2+\frac{1}{2}|D_i\phi|^2+\frac{\lambda}{4}\Big(|\phi|^2-f^2\Big)^2\right]\ .
\tag{1.6.4}
$$

∗ S. Coleman, Comm. Math. Phys. **31** 259 (1973).

♣ H.B. Nielsen and P. Olesen, Nucl. Phys. **B61** 45 (1973).

where $B = F_{12}$. We will look for static solitons of this theory, assuming that the gauge in which the field is time-independent is the gauge $A_0 = 0$. The absolute minimum of E_S occurs for

$$B = 0 \ , \qquad D_i\phi = 0 \ , \qquad |\phi| = f \ . \tag{1.6.5}$$

This statement is a invariant under the residual time-independent gauge transformations. A particular solution of these conditions is $A_i = 0$, $\phi = f$; any gauge transformation of this solution, $A_i = \frac{i}{e} g^{-1}\partial_i g$, $\phi = g^{-1} f$ is also a solution (here $g = e^{i\alpha}$ is a smooth map $\mathbf{R}^2 \to U(1)$).

The classical configuration space of this theory consists of regular fields with finite static energy. Clearly (A, ϕ) will have finite energy only if the conditions (1.6.5) are satisfied asymptotically as $r \to \infty$. This requires that

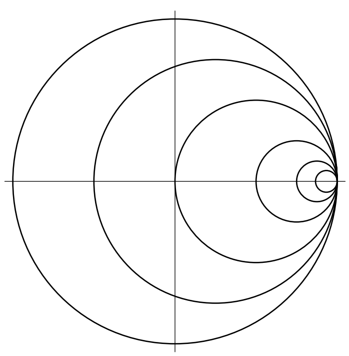
$$\begin{aligned} \phi(r,\theta) \underset{r\to\infty}{\longrightarrow} \phi_\infty = f\, e^{-i\alpha_\infty} \ , \\ A_i(r,\theta) \underset{r\to\infty}{\longrightarrow} -\frac{1}{e}\partial_i \alpha_\infty \ , \end{aligned} \tag{1.6.6}$$

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

The asymptotic behaviour of the field ϕ 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

$$W(\phi_\infty) = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, \frac{d\alpha_\infty}{d\theta} = -\frac{i}{2\pi}\int_0^{2\pi} d\theta\, \frac{1}{\phi_\infty}\frac{d\phi_\infty}{d\theta} \ . \tag{1.6.7}$$

The field ϕ 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_∞.



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 ϕ_∞, 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 ϕ_∞, so there must be in the theory a topological conservation law. In fact, consider the topological current

$$J^\lambda_T = \frac{1}{2\pi i}\varepsilon^{\lambda\mu\nu}\partial_\mu\hat\phi^*\partial_\nu\hat\phi \ , \tag{1.6.8}$$

where $\hat\phi = \phi/|\phi|$. This current is identically conserved and the corresponding topological charge is

$$Q_T = \int d^2x\, J^0_T = W(\phi_\infty) \ . \tag{1.6.9}$$

The physical meaning of the winding number can be understood by using the second equation in (1.6.6) in (1.6.7) and the applying Stokes' theorem:

$$W(\phi_\infty) = \frac{e}{2\pi}\oint_{S^1_\infty} A_i dx^i = \frac{e}{2\pi}\int_{\mathbf{R}^2} d^2x\, B = \frac{e}{2\pi}\Phi \ , \tag{1.6.10}$$

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

Since W is an integer, we get flux quantization:

$$\Phi = \frac{2\pi}{e} n \ . \tag{1.6.11}$$

Field configurations of this type occur in type II superconductors. Due to the Meissner effect, the magnetic field is zero in the bulk of the material, but if the external magnetic field exceeds a critical value, it penetrates the superconductor in the form of thin tubes, each carrying a quantized unit of magnetic flux. These tubes are called Abrikosov vortices. The name vortex comes from the form of the magnetic potential around the tube. In condensed matter physics the abelian Higgs model is known as the Landau–Ginzburg theory and is regarded as a mean field approximation of a more fundamental microscopic theory. In the case of superconductivity, this microscopic theory is the BCS theory, and in this theory one finds a relation similar to (1.6.11):

$$\Phi = \frac{2\pi\hbar}{2q} n \ , \tag{1.6.12}$$

where q is the electron charge ($2q$ is the charge of a Cooper pair). Note the different meaning of these equations. Equation (1.6.11) is purely classical; the constant e is a coupling constant of a classical field theory. On the other hand (1.6.12) is of quantum mechanical origin and q is the charge of the electron. 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.

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

$$\begin{aligned} A_0 &= 0 \ , \\ A_i &= -\varepsilon_{ij}\hat{x}^j A(r) \ , \\ \phi &= F(r)e^{i\varphi} \ , \end{aligned} \tag{1.6.13}$$

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.

1.7. The 't Hooft-Polyakov monopole

Maxwell's equations can be written in the form

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= 4\pi J^\nu_{(E)} \ , \\ \partial_\mu {}^*F^{\mu\nu} &= 0 \ , \end{aligned} \tag{1.7.1}$$

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 $F \to {}^*F$, ${}^*F \to {}^{**}F = -F$. Writing

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & +B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \qquad {}^*F_{\mu\nu} = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ B_1 & 0 & E_3 & -E_2 \\ B_2 & -E_3 & 0 & E_1 \\ B_3 & E_2 & -E_1 & 0 \end{pmatrix} \tag{1.7.2}$$

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

$$\begin{aligned} F &\to \quad \cos\theta F + \sin\theta\, {}^*F \\ {}^*F &\to -\sin\theta F + \cos\theta\, {}^*F \ . \end{aligned} \tag{1.7.3}$$

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 (1.7.1) is identically zero. They could be made symmetric under duality transformations by introducing a magnetic current J^ν_M such that

$$\partial_\mu\, {}^*F^{\mu\nu} = 4\pi J^\nu_M \tag{1.7.4}$$

and postulating the transformation

$$\begin{aligned} J_E &\to \quad \cos\theta J_E + \sin\theta J_M \ , \\ J_M &\to -\sin\theta J_E + \cos\theta J_M \ . \end{aligned} \tag{1.7.5}$$

That J^ν_M is a magnetic current is seen by observing for example that the time component of (1.7.4) would read $\mathrm{div}B = 4\pi\rho_M$, 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_μ 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

$$B_i = \frac{Q_M}{r^2}\hat{x}^i \ , \tag{1.7.6}$$

describing a static pointlike magnetic monopole in the origin, solves the equation $\mathrm{div}B = 4\pi Q_M\delta(r)$. Since the field is singular in the origin, one can remove this point from space and regard (1.7.6) as a smooth field on $\mathbf{R}^3\backslash\{0\}$. Since the field B given in (1.7.6) is divergence free on $\mathbf{R}^3\backslash\{0\}$, 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 paricular 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.

We consider the Georgi-Glashow model, consisting of an $SU(2)$ gauge field $A_\mu = A^a_\mu T_a$ coupled to a Higgs field ϕ^a in the adjoint (triplet) representation. We use the unscaled gauge fields, with curvature (1.5.1) and action (1.5.2). The total Lagrangian density is

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} - \frac{1}{2}D_\mu\phi^a D^\mu\phi^a - \frac{\lambda}{4}\left(\phi^a\phi^a - f^2\right)^2 \tag{1.7.7}$$

where $D_\mu\phi^a = \partial_\mu\phi^a + e\varepsilon_{abc}A^b_\mu\phi^c$. The structure constants of the Lie algebra of $SU(2)$ are $f_{abc} = \varepsilon_{abc}$ (in the adjoint representation the generators are $(T_a)_{bc} = -\varepsilon_{abc}$). The action is invariant under the local gauge transformations (1.5.3), 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

$$E_S = \int d^3x\left[\frac{1}{4}\left(F^a_{ij}\right)^2 + \frac{1}{2}\left(D_i\phi^a\right)^2 + \frac{\lambda}{4}\left(\phi^a\phi^a - f^2\right)^2\right] \ . \tag{1.7.8}$$

Its absolute minimum is obtained for

$$\begin{aligned} &F_{ij} = 0 \ , \\ &D_i\phi^a = 0 \ , \\ &\phi^a\phi^a = f^2 \ , \end{aligned} \tag{1.7.9}$$

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 $A_i^a=0$, $\phi^a=\bar{\phi}^a=(0,0,f)$ 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 (1.7.9) 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 ϕ defines a map $\phi_\infty: S^2_\infty\to S^2_{\rm int}$, where S^2_∞ denotes the "sphere at infinity" in $\mathbf{R}^3$ and $S^2_{\rm int}$ 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 (1.7.9) does not restrict the map ϕ itself. The asymptotic field ϕ^a_∞ can depend on the angles in an arbitrary way; the condition $D_i\phi\to 0$ can then be solved by

$$A_i^a=\frac{1}{f^2e}\varepsilon^{abc}\partial_i\phi^b\phi^c+\alpha_i\phi^a+O(1/r^2)\ , \tag{1.7.10}$$

for an arbitrary constant α_i.

The scalar fields ϕ fall into classes, labelled by the winding number of the map ϕ_∞. 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 ϕ_∞. 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

$$J_T^\mu=\frac{1}{8\pi}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{abc}\partial_\nu\hat{\phi}^a\partial_\rho\hat{\phi}^b\partial_\sigma\hat{\phi}^c\ , \tag{1.7.11}$$

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

$$\begin{aligned} Q_T=\int d^3x J_T^0=\frac{1}{8\pi}\int d^3x\varepsilon^{ijk}\varepsilon_{abc}\partial_i\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c= \\ =\frac{1}{8\pi}\int_{S^2_\infty} d^2x\varepsilon^{ij}\varepsilon_{abc}\hat{\phi}^a\partial_i\hat{\phi}^b\partial_j\hat{\phi}^c=W(\phi_\infty)\ . \end{aligned} \tag{1.7.12}$$

The last equality can be proven by choosing a particular coordinate system on S^2, for example the spherical coordinates (1.3.1), and comparing with (1.3.7).

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 $\mathcal{F}_{\mu\nu}=\partial_\mu A^3_\nu-\partial_\nu A^3_\mu$.

Following 't Hooft, we can generalize this to position-dependent Higgs fields. § Let $\mathcal{A}_\mu=A^a_\mu\hat{\phi}^a$. We define an abelian electromagnetic field $\mathcal{F}_{\mu\nu}$ by

$$\mathcal{F}_{\mu\nu}=\partial_\mu\mathcal{A}_\nu-\partial_\nu\mathcal{A}_\mu-\frac{1}{e}\varepsilon_{abc}\hat{\phi}^a\partial_\mu\hat{\phi}^b\partial_\nu\hat{\phi}^c\ .$$

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

$$\mathcal{F}_{\mu\nu}=\hat{\phi}^aF^a_{\mu\nu}-\frac{1}{e}\varepsilon_{abc}\hat{\phi}^aD_\mu\hat{\phi}^bD_\nu\hat{\phi}^c\ , \tag{1.7.13}$$

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

$$\partial_\nu\,{}^*\mathcal{F}^{\mu\nu}=-\frac{4\pi}{e}J_T^\mu\ . \tag{1.7.14}$$

§ G. 't Hooft, Nucl. Phys. **B79** 276 (1974); A.M. Polyakov, Pisma v. Zh. E.T.F. **20** 430 (1974), JETP Lett. **20** 194 (1974).

Comparing with (1.7.4), we see that we can interpret $\frac{1}{e}J^\mu_T$ as a magnetic current. The corresponding magnetic charge is

$$Q_M = \frac{1}{e}Q_T = \frac{1}{e}W \ . \tag{1.7.15}$$

Since W is an integer, we get a quantization condition for the magnetic charge, analogous to the flux quantization condition (1.6.11). 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 (1.6.11) and (1.6.12).

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

$$\begin{cases} \phi^a = \frac{x^a}{r}F(r) \ , \\ A^a_i(x) = \varepsilon_{aij}\frac{x^j}{r}A(r) \ , \\ A^a_0 = 0 \ , \end{cases} \tag{1.7.16}$$

where $F(r)\to f$ and $A(r)\to\frac{1}{er}$ for $r\to\infty$. 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

$$\mathcal{B}_i = \frac{1}{2}\varepsilon_{ijk}\mathcal{F}_{jk}\to\frac{1}{e}\frac{\hat{x}^i}{r^2} \tag{1.7.17}$$

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

When the ansatz (1.7.16) 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.

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 λ 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

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

$$= \frac{1}{4}\int d^3x \left(F^a_{ij}\mp\varepsilon_{ijk}D_k\phi^a\right)^2 \pm \frac{1}{2}\int d^3x\, \varepsilon_{ijk}F^a_{ij}D_k\phi^a \ . \tag{1.7.18}$$

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

$$\frac{1}{2}\int d^3x\,\partial_k\left(\varepsilon_{ijk}F^a_{ij}\phi^a\right) = f\int_{S^2_\infty} d\sigma^k\mathcal{B}_k = 4\pi f\, Q_M = \frac{4\pi f}{e}W \ , \tag{1.7.19}$$

where we have used (1.7.17). Using this in (1.7.18) we get the so-called Bogomol'nyi bound ⋄

$$E \geq \frac{4\pi f}{e}|W| \ , \tag{1.7.20}$$

with equality holding if and only if

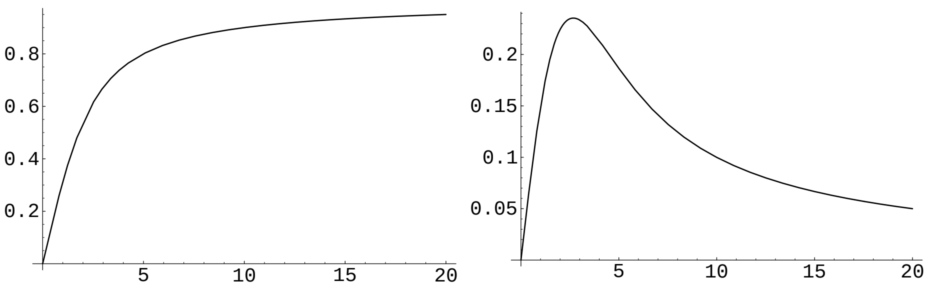
$$F^a_{ij} = \pm\varepsilon_{ijk}D_k\phi^a \ . \tag{1.7.21}$$

⋄ E.B. Bogomol'nyi, Sov. J. Nucl. Phys. **24** 449 (1976)

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 (1.7.21). In the Prasad-Sommerfield limit, ‖ the explicit form of the functions appearing in (1.7.16) is

$$\begin{aligned} F(r) &= \frac{f}{\tanh(efr)} - \frac{1}{er} \ , \\ A(r) &= \frac{1}{er} - \frac{f}{\sinh(efr)} \ . \end{aligned} \tag{1.7.22}$$

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



Exercises

Exercise 1.3.1: check by direct calculation that the winding number is invariant under infinitesimal variations of the fields.

Exercise 1.4.1. In the case of a principal sigma model one often employs a different formalism from the one given so far. Choose any faithful matrix representation of the group (for example in the case $G = SU(N)$ one can consider the fundamental representation of $SU(N)$ as unitary, unimodular $N \times N$ matrices). In the case $d = 3$, the field is a matrix-valued map U. We define an invariant metric $h_{\alpha\beta}$ on G as follows: we identify the tangent space to G at the identity e with the Lie algebra $\mathbf{g}$ and introduce in this space a positive definite invariant inner product $h_{(e)}(v, w) = c\,\mathrm{tr}(vw)$, where c is some real constant. This inner product in $\mathbf{g}$ is invariant under the adjoint action: $h\big(\mathrm{Ad}(g)v, \mathrm{Ad}(g)w\big) = \mathrm{tr}(g^{-1}vg \cdot g^{-1}wg) = h(v, w)$. The inner product at some other point g is then defined by translating the vectors v and w to the origin with g^{-1}: $h_g(v, w) = c\,\mathrm{tr}(g^{-1}vg^{-1}w)$. In this way the action can be rewritten as in (1.4.9) .

Exercise 1.4.2. Similarly we can define a left-invariant volume form on $SU(2)$ starting from the Ad-invariant three-form $\omega_{(e)}(v, w, z) = \frac{1}{3}\left(\mathrm{tr}(v, [w, z]) + \mathrm{cyclic}\right)$ in the origin and then translating it to other points. Then the general formula (A.3) for the winding number can be rewritten as in (1.4.12).

Exercise 1.4.3. Using the definition (1.4.12) *, check the product rule* (1.4.13).

Exercise 1.4.4. Using the formula

$$\exp(gn^aT_a) = \cos(\frac{g}{2}) + n^aT_a\sin(\frac{g}{2})$$

show that the field given in (1.4.14) *has winding number one.*

Exercise 1.7.1. The proper gauge group of the theory consists of functions that tend to the identity at infinity. Such transformations obviously do not affect the winding number of ϕ_∞. However, there are transformations that do not behave like this. Perform a singular gauge transformation that aligns the Higgs field in the 3rd direction (the resulting gauge is called the unitary gauge). Under this transformation the gauge potential becomes singular. (See J. Arafune, P.G.O. Freund, C.J. Goebel, J. Math. Phys. **16** *433 (1975).)*

Exercise 1.7.2. Generalize the preceding discussion by replacing $SU(2)$ and $U(1)$ with arbitrary groups G and H. The topological condition for the existence of monopoles is then $\pi_2(G/H) \neq 0$. For example in the Weinberg-Salam model, $G = SU(2) \times U(1)_Y$, $H = U(1)_Q$ and $G/H \approx S^3$, so there are no monopoles. In order to discuss more general cases, write the homotopy exact sequence of the bundle $G \to G/H$. If G is compact, connected and semisimple, $\pi_2(G) = 0$; use the homotopy exact sequence to deduce that $\pi_2(G/H) = \pi_1(H)$. So if H contains a factor $U(1)$, $\pi_2(G/H) = \mathbf{Z}$ and we are in the same situation considered in the text. This is the case for example in the minimal $SU(5)$ model.

‖ M.K. Prasad, C.H. Sommerfield, Phys. Rev. Lett. **35** 760 (1975).