Monopoles and duality

Patrick Massot

April 2003

Abstract

This text reviews some aspects of monopoles and duality in gauge theories using a geometrical language. The reader is assumed to be quite familiar with connections on principal bundles, basic algebraic topology, electromagnetism and the Lagrangian formulation of classical mechanics but we don't assume any prior knowledge in quantum field theory.

Contents

Introduction

The subject of magnetic monopoles is a beautiful and large subject and there exist many good reviews. I think the best way to explore quickly this field may be to read the talks [1] and [2] then the reviews [3] then [4] or the lectures [5]. These references are very good but they are written for physicists and are quite difficult to read for a mathematician. So I have three good reasons to write this essay in a mathematical style : it is useless to recopy existing good reviews, it may be useful to write an original review aimed at mathematicians and actually it reflects an important part of my work during this project because I spent a lot of time trying to translate physicist style to mathematician style which I'm more familiar with.

A mathematical style allows us to avoid certain calculations at the expense of assuming some knowledge of differential forms and fiber bundles theory. This can be learnt from many places, e.g. [6], [7], [8]. As a concession to physical style, we use the words "connection 1-form" and "curvature 2-form" in their local, trivialization dependant meaning, except when we specify the contrary using the word "intrinsic".

The first section of this work is a crash course on gauge theories whose purpose is to give some physical intuition and to explain why fiber bundles with connections are relevant to particle physics. It also introduces a lot of vocabulary. This first section primary goal is to make the other parts understandable to a geometer but actually the other parts can also be seen as examples illustrating the first part and many things in this first part will hopefully become clearer with these examples. The second section presents the Dirac monopole which is a magnetic monopole in classical electromagnetism on a non trivial space-time discovered by Dirac in 1931. It is historically the first magnetic monopole and will be a reference for us. The third section deals with many general considerations and examples which are useful for the fourth section but have also their own interest and can also be viewed as examples and developments of the first section. The fourth section is devoted to the study of the 't Hooft-Polyakov monopole which is representative of monopoles in non-abelian gauge theories and had been discovered independently by 't Hooft and Polyakov in 1974. The fifth section explores the relation between magnetic and electric charges in the 't Hooft-Polyakov context by analogy with an example given in the third section. Every complicated calculations and notation specifications have been put in appendices to keep the stream of ideas flowing.

1 Physical background

In this section we review the ideas leading to the use of fiber bundles with gauge equivalence classes of connections in particle physics. Our description will be progressive with mathematical sophistication introduced in response to the physical discussion. From the physical point of view it will not be totally satisfactory, there are many simplifications, especially the fact that we totally overlook renormalization issues but it should be sufficient to understand the rest of this essay.

1.1 Particles and fiber bundles

The space-time is an open set $\mathcal{M}$ of the Minkowski space-time. So we assume that the space-time metric is as simple as possible but we don't assume that space-time is topologically trivial, it may be non-contractible.

We begin with the simplest possible description of the history of a particle as a worldline, *i.e.* a time-like curve in space-time. The second step is to assume that a particle has an internal state in addition to its space-time position. We assume that every interaction is local so that the internal state of a particle at a given point in $\mathcal{M}$ can affect another point only if it is carried over a linking path in space-time. This leads to the use of fiber bundles with connections. At each point $x \in \mathcal{M}$ we have an internal state space E_x and the connection specifies how to transport a state along a path. Note that we don't assume a priori that $E \to \mathcal{M}$ is a trivial bundle.

Each kind of interaction (electromagnetic, weak, strong...) is described by a principal bundle whose connection will be identified to a particle called the gauge boson which is said to mediate the interaction. This bundle is called the interaction bundle and its structure group is called the gauge group. In order to simplify the discussion in this section we assume that there is only one kind of interaction involved.

Experimentally, there are several kinds of particles reacting in their own way to the gauge boson so each kind of those particles is described by a vector-bundle associated to the interaction principal bundle. Each kind of particle is therefore associated with a representation of the structure group of the interaction. Such a representation simply describes how the connection (hence the gauge boson) acts on a specific kind of particle. These representations will be assumed to be irreducible since a reducible representation can be said to describe distinct particles. Also you can have different particles reacting in the same way to the gauge boson but with a different strength. This kind of difference is simply given by numerical factor in the induced Lie algebra representation. This numerical factor is called the charge of the particle (when several kind of interactions are involved you must precise which kind of charge you're talking about). We will return to this point later.

Because of locality, the laws of nature concerning the connection will naturally be given in terms of the covariant derivative rather than finite parallel transport. In particular we will only begin with the knowledge of the Lie algebra associated with the interaction instead of the Lie group. So we only begin with the universal cover of the gauge group and the precise group will be given by the exact spectrum of observed particles. This is a consequence of the fact that if you know the Lie algebra of a group and its representations you can determine the group. As an example of this problem let us consider electromagnetism. The Maxwell equations have been known for a long time and involve a 2-form F which really looks like a curvature form for a gauge theory. Because this form is real-valued we know that the gauge algebra is $\mathbb{R}$ but the gauge group could be $\mathbb{R}$ or $U(1)$. We will see that this is intimately related to the issue of charge quantization, *i.e.* the fact that all electric charges in nature seem to be a multiple of one fundamental charge. This will be discussed in the next section. See [9] for a quick discussion of the precise groups involved in the standard model.

It seems that at this stage we could simply specify some equations involving the connection and the other particles and try to do physics with them but we must take into account the fact that, experimentally, the internal state of a particle is not directly measurable. The only things which we can detect are differences of internal state. This leads to the idea that if $c_1, c_2 : \mathbb{R} \to E$ are two curves in the vector bundle E which describe the history of two particles and g is an element of the gauge group then $g \circ c_1$ and $g \circ c_2$ describe the same physical situation provided that their covariant derivatives are transformed in the same way, *i.e.* that D is transformed to D' such that for any section s we get $D'gs = gDs$. This can be summarized by saying that D and $g^{-1}Dg$ must describe the same physical situation for all g in G.

This symmetry can be explained from a physical point of view by examining the following special case : suppose we have a trivial interaction principal bundle and a p-dimensional associated vector bundle E. We can choose a trivialization $(X_1, \dots, X_p)$ of E and decide that each of these sections are different particles. These particles are said to form a multiplet. But our choice of trivialization is arbitrary and we can change of choice with an element $g \in G$ by replacing $(X_1, \dots, X_p)$ with $(gX_1, \dots, gX_p)$ and it is quite clear that it should not change the physical situation. For example, if we neglect electromagnetism and a small mass difference, the proton and the neutron are very similar and react in the same way to the weak force (gauge group $SU(2)$). So, if the weak interaction bundle is trivial, we can describe protons and neutrons by a trivial vector bundle with the fundamental representation of $SU(2)$ over $\mathbb{C}^2$ and we say that we have an $SU(2)$ doublet. Now we can say that the proton corresponds to the constant section $(1,0)$ and the neutron to $(0,1)$ but this is an arbitrary choice. So this symmetry tells us that the distinction we make between different particles of a multiplet is not important provided that we stick to our choice (because g is constant). Our equations must therefore be invariant under such transformations in the sense that one solution must be carried on another solution by such a transformation. In addition, this symmetry will lead, in QFT, to the "conservation of charge", in a sense that we will partly explain later.

All this seems quite natural but, if we look carefully at Maxwell's equations, we see that they are much more symmetric than that. Actually if we take a solution we get another solution by applying a transformation $D \mapsto g^{-1}Dg$ with g non-constant ! This kind of transformation is called a gauge transformation. Actually the fact that Maxwell equations are invariant under gauge transformations is rather trivial because they involve only the curvature of the electromagnetic connection and, because $\mathbb{R}$ is abelian, this curvature doesn't change under such a transformation. So there is no *a priori* reason to demand such a strong symmetry in the non-abelian case but Yang and Mills did it and it has proved to be a very good idea. Physically this seems to be very difficult to explain from a classical point of view and, indeed, it can be viewed as a trace of the quantum nature of things as we will show. There is a geometrical argument to promote this symmetry however. In our discussion of multiplets we have supposed that we have trivial bundles. In general a bundle is not trivial and the only way to generalize the argument about particle distinction to this case is to allow non-constant gauge transformations. More fundamentally, even with a trivial bundle $\mathcal{M} \times \mathbb{R}^n$, the choice of the sections $x \mapsto e_i$ where $(e_1, \ldots, e_n)$ is the canonical basis of $\mathbb{R}^n$ is not that natural from a geometrical point of view and no trivialization is natural so it is natural to allow non constant gauge transformations. However there is a more physical quantum argument.

1.2 Toward QFT

Now we turn to (semi)-quantum physics. All the preceding discussion seems to be ruined because, already in elementary quantum mechanics, particles can't be said to follow a definite path in space-time. The only thing we can hope in quantum physics is to calculate the probability that a certain set of particles will be detected at certain points in space-time given that a certain set of particles has been prepared at certain points in space-time. In this context, a natural idea if, say, we wanted the probability that a particle is detected at a point $y \in \mathcal{M}$ given that the same kind of particle has been detected at a point x, would be to transport the internal state of a classical particle from x to y along every possible path between these two points and add the results. Of course different paths would give different states and you'll get interferences. Actually this is really what is done only in the case of electromagnetism and for non-abelian theories it is a little bit different but the idea of a sum over all classical paths is still good. This sum over all paths is called the Feynman path integral and it is very difficult, if not impossible, to give it a precise mathematical meaning but it provides a very good ground to intuition. With this picture in mind we see that the relevant property of the interaction connection seems to be its holonomy. Actually even the holonomy is too much because what is really relevant is the trace of this holonomy. This trace is called a Wilson loop by physicists. This provides a very good reason to identify two connections related by a gauge transformation since Wilson loops are not affected by them. We are also forced to consider sections of our bundles rather than curves since particles are now "spread everywhere". However we need only one section for each associated bundle to describe a physical situation. In physical terms it means we need only one field for each kind of particle.

Earlier we talked about the charge as a numerical factor in the representation. We can now be a little bit more precise and see how it appears concretely in the theory. Two kinds of particle are said to be the same except for their charge if they are associated to the same interaction bundle and there exists two numbers g_1 and g_2 and a representation ρ such that the i-th particle has representation $g_i\rho$ of the gauge Lie algebra. To see how it appears concretely we look at the interaction bundle over a limited region U of $\mathcal{M}$ where it is a trivial bundle. Over this region the connection can be specified with a local connection 1-form $A \in \Gamma(\Lambda U \otimes \mathfrak{g})$ which is called a gauge potential. The covariant derivative of a section s representing a kind of particle of charge g is $Ds = ds + g\rho(A)s$. The basic representation ρ will often be implicit so the derivative will be written $Ds = ds + gAs$. In QFT each field will describe a non-constant number of particles of the given type and the gauge transformation symmetry will ensure via Noether theorem that $\sum n_i g_i$ where n_i is the number of particles of charge g_i is constant. Of course there is a lot of work to do to go from a classical field to this kind of description and we won't try to discuss it in details but we will discuss classical solutions and give hints of their relations to quantum fields. Note that the quantization of the associated fields is the easiest part and the quantization of gauge bosons is really tricky because you have to factor out the gauge freedom to really work with equivalence classes of connections.

According to those principles, we must look for equations concerning a connection on a principal bundle and some sections of associated vector bundles such that a gauge transformation transforms a solution into a solution. The easiest way of getting such equations is to use variational principles for Lagrangian densities. A Lagrangian density is a function of the fields and their first derivatives. The integral of the Lagrangian density over space-time is called the action

$$\mathcal{A} = \int_{\mathcal{M}} \mathcal{L}$$

and the equations of the theory are given as stationarity conditions of the action, this is totally analogous to the Lagrangian formulation of classical mechanics. Actually from a mathematical point of view this construction is not perfect, notably because the action needs not to be finite in general but you can always do this formally and study the resulting equation (which makes sense) without considering its mythological least action principle origin.

Now we construct the simplest gauge invariant Lagrangian for a connection alone on a trivial G-bundle. Let $F \in \Gamma(\Lambda^2 T^*\mathcal{M} \otimes \mathfrak{g})$ be a curvature 2-form. The idea is to use a G-invariant quadratic form q on $\mathfrak{g}$, typically a multiple of the Killing form, to get the Lagrangian $\mathcal{L} = q(F \wedge \star F)$. For example, for $\mathfrak{g} = \mathfrak{su}(n)$, we will take the opposite of the Killing form as quadratic form. Note that the $F \wedge \star F$ part is just the scalar product on 2-forms coming from the Lorentzian structure on $\mathcal{M}$ multiplied by the Lorentzian volume form. The whole term is often be written as $||F||^2$. If we want to include other fields we use essentially two things : G-invariant functions of the non-gauge fields multiplied by the volume form and the norm of covariant derivatives of the fields. The first kind of term is called a potential, not to be confused with a gauge potential. The second kind is called minimal coupling because it involves both the connection and the fields in the simplest possible way. Let us be more precise concerning this kind of term. The covariant derivative of a section of a vector bundle E is a section of $\Lambda T^*\mathcal{M} \otimes E$. We assume that E has a G-invariant inner product and we use the Lorentzian inner product on $\Gamma(T^*\mathcal{M})$ and the Lorentzian volume form on $\mathcal{M}$ to get $||D\phi||^2$ which is a lovely G-invariant quantity that we can put in a Lagrangian.

Now we describe how to read the physical content of a Lagrangian. This physical content is called spectrum and consists of a list of particles with their mass and charges (and spin but we don't want to really talk about spin in this text). The first step is to identify the vacuum which is a configuration, *i.e.* a set of sections of bundles and connections, which minimizes the potential term. Quantum field theory describes particles as excitations of the fields near the vacuum. Actually the vacuum is not necessarily unique as we will see but suppose we have chosen a vacuum. We write the Lagrangian as a power series around the vacuum. More precisely, if a field ϕ takes the value

ϕ_0 in vacuum then we write each term where ϕ appears as a power series of the shifted field $\tilde{\phi} := \phi - \phi_0$. Then we use the following rules to find the particle spectrum :

- a term of the form $-\frac{1}{2}m^2|\tilde{\phi}|^2$ gives the mass m of the particle corresponding to $\tilde{\phi}$
- a minimal coupling term gives the charge of the particle for the interaction corresponding to the gauge covariant derivative appearing in the term (see our definition of charge)
- a term of the form $||F||^2$ is the kinetic term of a gauge boson
- a term of the form $||d\tilde{\phi}||^2$ is the kinetic term of the particle corresponding to $\tilde{\phi}$
- the other terms are interaction terms

Of course we have to be careful with the notion of multiplet. As we explained earlier, a multidimensional representation of the gauge group can be viewed as many closely related particles. Actually each field (including gauge potentials) must be decomposed so that each part can be consistently viewed as a particle with its (maybe zero) mass term, charge term etc. More details are given in section 3.2. Let us take an example. Suppose we have the following Lagrangian for a complex scalar field ϕ and a $U(1)$ connection with potential A :

$$\mathcal{L} = \frac{1}{2}||F||^2 + \frac{1}{2}||D\phi||^2 - \frac{m^2}{2}|\phi|^2 - \lambda|\phi|^4$$

with $D_\mu\phi = \partial_\mu\phi + eA_\mu\phi$. The vacuum is $\phi_0 \equiv 0$ and we have a particle of mass m and charge e and a massless gauge boson. Also we see a self interaction term for ϕ.

In QFT all the fields must be transformed in operators on some Hilbert space and we won't try to describe this procedure but the classical fields are supposed to be a kind of classical limit of the quantum fields and informations about particles can be seen from them. In the classical limit a particle of mass $m \neq 0$ is associated to a field with exponential decay $exp(-m|x|)$ at infinity so the mass gives a length scale. A zero mass particle is associated to a power decay at infinity. From a geometrical point of view it is clear that a conformal theory can only describe massless particles because there can not be a length scale. For example, vacuum electromagnetism is conformal invariant and describe the photon which is massless. Of course the link with our preceding definition of mass using the Lagrangian is not obvious at all but it has been proved in certain cases. Solitons are classical solutions which differ from vacuum only in small regions so, according to what we have just seen, they are interesting to try to understand the quantum fields. In this essay we will meet several examples and hints of this mysterious relationship between classical and quantum fields.

Before going further we digress to point out an important source of confusion. Concerning classical fields in general, physicists often use the words "scalar fields" and "vector fields". It is tempting to believe that a scalar field is just a section of a rank one bundle and that a vector field is a vector field but it's not true. Actually this vocabulary refers to the action of transformations of $\mathcal{M}$ on sections of fiber bundles over $\mathcal{M}$. When this action is simply by composition we have a scalar field, otherwise it is a vector field. So, for example, sections of $T\mathcal{M}$ or $\Lambda^p T\mathcal{M}$ are vector fields and sections of $\mathcal{M} \times R^p$ are scalar fields, even if $p > 1$.

2 Dirac monopole and charge quantization

First we review electromagnetism from a geometric point of view. The elementary historical way of describing electromagnetic phenomena uses two "vector fields" $\vec{E}$ and $\vec{B}$ constrained by the four Maxwell's equations. For a geometer it is puzzling that these fields are treated sometimes as 1-forms and sometimes as 2-forms. Indeed, when we use Gauss's law we integrate the electric field over surfaces but differences of electric potential are defined as integrals of the electric field along curves and we have the same kind of paradox with the magnetic field. This strongly suggests that the Hodge $\star$ operator is involved in electromagnetism. Actually the best way to build a geometrical version of electromagnetism is to see it as a gauge theory with gauge algebra $\mathbb{R}$ over space-time and we will recover the electric and magnetic fields in a special case.

Indeed, from elementary quantum mechanics we know that, for a spinless charged particle in the presence of an electromagnetic field F with 4-potential A, the relevant covariant derivative is $D = d - i\frac{q}{\hbar}A$ acting on complex-valued fields. This strongly suggests that we have a gauge theory with gauge algebra $\mathbb{R}$ on a rank one complex vector bundle. Of course the units used here and the electromagnetic field and potential were defined long before the birth of fiber-bundle theory so we have a kind of mismatch with our abstract definitions of charge and gauge potential such that $D = d + qA$ but this is not a big problem, we simply have to be careful to consider geometrical objects when using geometry theorems and then put the "traditional" constants to translate the result in terms of the "traditional" fields. Here the traditional field corresponding to the gauge curvature is the electromagnetic field $F =$

dA where A is the traditional potential so, comparing the abstract and the traditional definitions of the covariant derivative we see that the curvature 2-forms for the associated vector bundle are given in each trivialization by $-i\frac{q}{\hbar}F$.

Because the gauge Lie algebra $\mathbb{R}$ is abelian the electromagnetic 2-forms of different trivializations patch nicely to give a globally defined (real-valued) 2-form F which is also, for the same reason, invariant under gauge transformations. Because of this exceptional simplification we only need to study 2-forms on $\mathcal{M}$ and Maxwell's equation is now :

$$\star d \star F = J \tag{1}$$

where J is a 1-form describing the sources of electromagnetic field and, as foreseen, the Hodge operator is involved. Because F comes from the curvature forms we can use the Bianchi identity which gives

$$dF = 0 \tag{2}$$

When $J = 0$ we get

$$\begin{cases} dF &= 0 \\ d \star F &= 0 \end{cases}$$

These equations are clearly invariant under $F \mapsto \star F$, this is the famous vacuum electromagnetic duality, the mother of all dualities in physics. Note that when $F \mapsto \star F$ we have $\star F \mapsto -F$ because the signature is Lorentzian. This symmetry is so beautiful that we are very disappointed to see that it doesn't seem to hold when $J \neq 0$. It is tempting to fix this problem by adding a source term in (2) but this would not be very natural and, above all, it would ruin the geometrical interpretation of F as a constant times a curvature 2-form because Bianchi identity would be violated. Instead of loosing geometry we will see that more geometry comes to help. In order to explore this issue we now turn to a special case.

Everything we have said so far about electromagnetism applies in the case of any Lorentzian orientable 4-manifold $\mathcal{M}$. In the special case where $\mathcal{M}$ admits a (generally non unique) isometric decomposition $\mathcal{M} \simeq \mathcal{S} \times \mathbb{R}$ with $\mathcal{S}$ Riemannian, we have a time 1-form dt and we can decompose F and J in

$$\begin{aligned} F &= B + E \wedge dt \\ J &= j - \rho dt \end{aligned}$$

where B (resp E) is a time dependant 2-form (resp 1-form) on $\mathcal{S}$ which has been lifted to $\mathcal{M}$ in the obvious way. There is a slight abuse of notations because we don't distinguish between a form on $\mathcal{S}$ and its lifted version on $\mathcal{M}$ but there is no danger of confusion. We also define $\star_{\mathcal{S}}$ and $d_{\mathcal{S}}$ to be the Hodge operator and exterior derivative on $\mathcal{S}$ and they act in the obvious way on lifted forms. In this case, (1) and (2) become

$$\begin{aligned} \star_{\mathcal{S}} d_{\mathcal{S}} \star_{\mathcal{S}} E &= \rho \\ -\partial_t E + \star_{\mathcal{S}} d_{\mathcal{S}} \star_{\mathcal{S}} B &= j \\ d_{\mathcal{S}} B &= 0 \\ \partial_t B + d_{\mathcal{S}} E &= 0 \end{aligned}$$

We concentrate on the static case where $\partial_t = 0$ and $j = 0$ (no electric current). Because we consider a purely static case we work in $\mathcal{S}$ for the remaining of this section and drop the $\mathcal{S}$ index for notational convenience but it should still be there in d and $\star$. The preceding equations become

$$\begin{aligned} \star d \star E &= \rho \\ \star d \star B &= 0 \\ dB &= 0 \\ dE &= 0 \end{aligned}$$

and the vacuum electromagnetic duality becomes

$$\begin{cases} E & \mapsto & -\star B \\ B & \mapsto & \star E \end{cases}$$

and we see the obstacle to general electromagnetic duality is the absence of magnetic charge. As we have already said, we are not going to add a term in the equations because it wouldn't be geometrical but we will use *more* geometry.

As we had to turn to quantum mechanics to see precisely how F fits in the geometrical framework of gauge theories we now seek a gauge theoretical interpretation of B using quantum mechanics. The relevant covariant derivative is now $D = d - i\frac{q}{\hbar}A$ where A is now the three-dimensional "vector potential" (which is a real valued 1-form) and $B = dA$. This tells us that static magnetism is a gauge theory of a rank one complex vector bundle with connection over $\mathcal{S}$. The curvature of this connection can be expressed in terms of the traditional magnetic field as $-i\frac{q}{\hbar}B$ (we should say semi-traditional because the traditional magnetic field is a "pseudo-vector" rather than a 2-form). This interpretation forces B to be closed because of Bianchi identity and indeed this is required by Maxwell's equations. It is quite clear that this passage from F to B can be generalized to any gauge theory over space-time.

We can now examine the issue of the existence of magnetic charges using Chern classes (whose definitions and useful properties are recalled in appendix E) because we have a complex vector bundle whose curvature is proportional to B. It is very natural to interpret a non-zero flux of B through a 2-sphere as the indication of the presence of a magnetic monopole enclosed by the sphere. Such a flux is proportional to the first Chern number given by the integral of the curvature over the considered 2-sphere (see appendix E). This gives us two very interesting pieces of information. First we can have a magnetic monopole only if $\mathcal{S}$ is topologically non-trivial because if our 2-sphere Σ is the boundary of a 3-ball $\mathcal{B}$ then the magnetic charge is

$$g = \int_{\Sigma} B = \int_{\mathcal{B}} dB = 0$$

This simply corresponds to the fact that a magnetic monopole can not be inside space-time, it must be outside. Typically we can consider $\mathcal{S} \simeq R^3 \backslash \{0\}$. The second piece of information is that a magnetic monopole can exist only if the gauge group is not simply connected because the Chern number involved here can also be seen as a Chern number of the induced bundle over $\Sigma \simeq S^2$ for the cycle Σ and bundles over S^2 are classified by $\pi_1(G)$ for a structure group G (see [7] p 96) so if $\pi_1(G) = \{1\}$ then the induced bundle is trivial so all its first Chern numbers are zero and the magnetic flux is zero. So the existence of a magnetic monopole implies that the electromagnetic gauge group is $U(1)$ rather than $\mathbb{R}$.

This is especially interesting because the choice between $\mathbb{R}$ and $U(1)$ as the electromagnetic gauge group is equivalent to the question of charge quantization. Indeed we have said that different charges correspond to different multiples of a representation of the gauge algebra which come from irreducible representations of the gauge group and the irreducible representations of $U(1)$ are simply given by

$$\rho_n \left(\begin{array}{ccc} U(1) & \to & GL(\mathbb{C}) \\ e^{i\theta} & \mapsto & (z \mapsto e^{in\theta} z) \end{array} \right)$$

so if the gauge group is $U(1)$ then the possible charges are all multiples of a fixed charge whereas if the gauge group is $\mathbb{R}$ we have the representations

$$\rho_x \left(\begin{array}{ccc} \mathbb{R} & \to & GL(\mathbb{C}) \\ a & \mapsto & (z \mapsto e^{ixa} z) \end{array} \right)$$

so we have a continuum of charges. More generally, charge quantization is a consequence of the compactness of the gauge group, see [10].

That's the geometrical form of Dirac's argument leading to the statement that if a magnetic monopole exists in the universe then electric charge is quantized. The original Dirac argument involved the fact that if a magnetic monopole exists somewhere then each electrically charged particle can go near the monopole and quantum mechanics are consistent only if charge is quantized. Altough this is correct, it leaves the impression that electric charge need not be quantized if there are no monopole in the neighborhood. The discussion of the gauge group of electromagnetism is more abstract but avoids this difficulty : the existence of a magnetic monopole implies that $U(1)$ is the relevant group and this in turn implies the quantization of charge everywhere because the gauge group is a fundamental characteristic of the type of interaction which is not supposed to change from places to places.

Of course the link between the existence of magnetic monopoles and the gauge group is not an equivalence and we can postulate that $G = U(1)$ without postulating the existence of a magnetic monopole.

We now derive the precise Dirac quantization condition. Suppose we have a particle with electric charge q. The corresponding complex line bundle has covariant derivative $D = d - i\frac{q}{\hbar}A$ if we include every physical units and the curvature is $-i\frac{q}{\hbar}B$ so that the fact that the first Chern class of the induced bundle over Σ is integral gives

$$\frac{i}{2\pi}\int_{\Sigma} -i\frac{q}{\hbar}B \in \mathbb{Z}$$

so we have

$$\frac{qg}{2\pi\hbar} \in \mathbb{Z}$$

where $g = \int_{\Sigma} B$ is the magnetic charge enclosed by Σ. This is the celebrated Dirac quantization condition. Note that the magnetic charge is also quantized. In addition if we allow time variation of this configuration then the magnetic charge is conserved because $\mathbb{Z}$ is discrete so you can't continuously change the magnetic charge. The magnetic charge is therefore called a topological charge as opposed to the electric charge which is a Noether charge because its conservation is ensured even without monopole by Noether's theorem on continuous symmetries.

Let us construct an explicit example of such a magnetic field. We want to construct a "magnetically charged point" so we try to build a 2-form B over $R^3\backslash\{0\}$ such that $dB = 0$, $d \star B = 0$ and $\int_{S^2} B \neq 0$. Duality comes to help because we already know how to construct a "electrically charged point" outside space-time which gives an electric field $E \propto \frac{1}{4\pi r}dr$. Because of electromagnetic duality we know that $\star E$ will make a good magnetic field.

This is really electromagnetic duality at work ! We get

$$B \propto \star \frac{1}{4\pi r} dr = \frac{1}{4\pi r} \sin\phi \, d\theta \wedge d\phi$$

And of course the proportionality coefficient must satisfy the Dirac quantization condition.

Note that all this seems quite easy to understand but Dirac invented his monopole in 1931 before the birth of fiber bundles and characteristic classes and this is really an amazing intellectual achievement because at this time people didn't consider patched electromagnetic potentials. Actually Dirac did not really patch potentials, he considered a "singularity string" but this is equivalent to considering two patched potentials.

We won't go any further along this path in this essay but the relationship between space-time topology and magnetic fluxes has been investigated, see e.g. [11] and [12]. These papers also investigate the relation between topology and the existence of a spin structure on space time (spinors are useful to describe many particles) from a physical point of view and the link with electromagnetic charges.

The main problem of this approach is that such a magnetic monopole has necessarily an infinite energy and lives outside space-time. In the next section we show how to construct objects which look like a Dirac monopole when seen from far away but have a finite energy and an internal structure inside space-time.

3 Towards 't Hooft-Polyakov monopoles

Before considering those monopoles we must review some general ideas with examples.

3.1 Spontaneous symmetry breaking in physics

The concept of spontaneous symmetry breaking is very important in modern physics but, unfortunately, it is quite misnamed and this causes some confusion. We shall see that the name "hidden symmetry" which is sometimes used is better but we follow the tradition.

It is well-known that the set of solutions to a given problem has the symmetry of the problem but a particular solution can have less symmetry. The canonical example is an infinite 2-dimensional lattice of ferromagnets. Each ferromagnet interact with its nearest neighbors and the Hamiltonian is rotationally invariant. Figure 1.a shows a snapshot of such a system at high

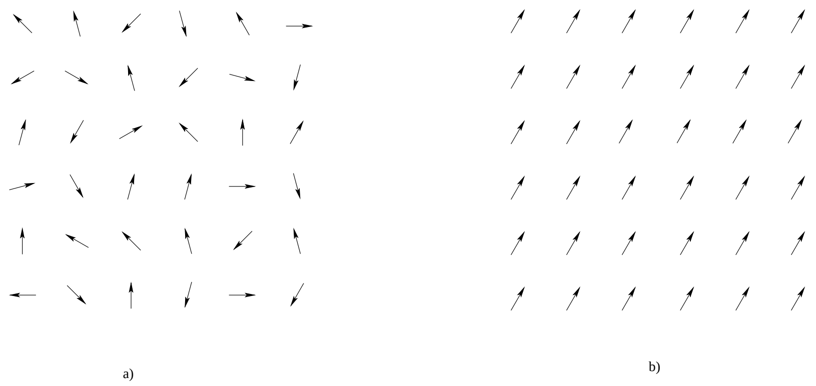

Figure 1: A typical example of SSB

temperature. We are interested only in fundamental states where energy is minimal. The interactions tends to align all ferromagnets so that a minimal energy state looks like figure 1.b. The striking feature of such a state is that is *not* rotationally symmetric. One direction is clearly privileged. The Hamiltonian is still symmetric and the set of all fundamental states is symmetric but each fundamental state is asymmetric. This situation is different from the situation where an external magnetic field forces a special direction. In the latter case the Hamiltonian is no more symmetric and we have a symmetry breaking. In the former case we say we have a spontaneous symmetry breaking (SSB). The name "hidden symmetry" is better because the symmetry is not really broken but any people living in such an environment will believe that there is no such symmetry. Also, because the system is supposed to be infinite you can't go from one fundamental state to another with a finite amount of energy. Note also that a SSB can be partial. Actually our example has more than a rotational invariance, it has also a translational invariance which is still present in any particular fundamental state.

In quantum field theory a fundamental state is called a vacuum and these states are very important since we have seen that (almost) everything in QFT is studied perturbatively around them.

3.2 SSB in gauge theories

In gauge theories we have a SSB when the vacuum is not invariant under the whole of the gauge group G but is only invariant under a subgroup H called the little group. This definition is consistent with what we have said about SSB in general and with the fact that in gauge theory the Lagrangian is invariant under the entire gauge group. Usually there is only one field which breaks the symmetry and this field is called a Higgs field.

We now investigate the consequences of SSB on the particle spectrum, the so called Higgs mechanism. Let G be a compact connected Lie group and H be a closed subgroup of G. Let $(T_1, \ldots, T_n)$ be a basis of the Lie algebra $\mathfrak{g}$ of G with $(T_1, \ldots, T_p)$ a basis of the Lie algebra $\mathfrak{h}$ of H. The elements of the basis of a gauge Lie algebra are called the generators of the gauge symmetry because the whole gauge action can be recovered from the derived action of the Lie algebra. Note that the infinitesimal counterpart of $g(s)\phi = \phi$ for a one parameter subgroup is $T\phi = 0$ where $T = \frac{dg}{ds}\big|_0$. Here we assume $\mathcal{M}$ is contractible and consider a generic Lagrangian with a connection and a field $\phi : \mathcal{M} \to E$ which we decompose on a given basis of the vector space E.

$$\mathcal{L} = \frac{1}{2}||F||^2 + \frac{1}{2}||D\phi||^2 - V(\phi)$$

where[1] $D\phi = d\phi + gA_aT^a\phi$ and V is G-invariant. Let ϕ_0 be a minimum of V with

$$\begin{aligned} T^a\phi &= 0 \quad 1 \leq a \leq p && (3)\\ T^a\phi &\neq 0 \quad p+1 \leq a \leq n && (4) \end{aligned}$$

So we have a SSB from G to H. The potential V is G invariant so for all $1 \leq a \leq n$:

$$\frac{\partial V}{\partial \phi^i} T^a_{ij} \phi^j = 0$$

and differentiation with respect to k-th component of ϕ gives

$$\frac{\partial^2 V}{\partial \phi^i \partial \phi^k} T^a_{ij} \phi^j + \frac{\partial V}{\partial \phi^i} T^a_{ik} = 0$$

Now we take the preceding equality at $\phi = \phi_0$ which is a minimum of V, the second term vanishes and we get

$$\frac{\partial^2 V}{\partial \phi^i \partial \phi^k}\bigg|_{\phi_0} T^a_{ij} \phi^j_0 = 0 \tag{5}$$

Furthermore, the power expansion of V near ϕ_0 is

$$V(\phi) = V(\phi_0) + \frac{1}{2} \frac{\partial^2 V}{\partial \phi^i \partial \phi^k}\bigg|_{\phi_0} (\phi^i - \phi^i_0)(\phi^k - \phi^k_0) + O(|\phi|^3)$$

[1] in this formula the charge is denoted by g. This is the usual convention in QFT, there is no risk of confusing it with an element of G

so, according to the principles of the first section, the masses of the particles are the square roots of the eigenvalues of the matrix $(\frac{\partial^2 V}{\partial\phi^i \partial\phi^k}\big|_{\phi_0})_{ik}$ and equations (4) and (5) tells us that there are $(n-p)$ massless particles.

To see what happens with the remaining components of ϕ we want to eliminate the massless part of ϕ with a change of gauge so we use the following parametrization for ϕ near ϕ_0 :

$$\phi = \exp\left(\sum_{c=p+1}^{n} \frac{\xi_c T^c}{|\phi_0|}\right)(\phi_0 + \eta)$$

where we sum over the broken generators. We are now ready to gauge out the massless part with the following gauge transformation :

$$\begin{aligned} g &= \exp\left(-\sum_{c=p+1}^{n} \frac{\xi_c T^c}{|\phi_0|}\right) \\ \phi &\mapsto \phi' = g\phi = \phi_0 + \eta \\ A &\mapsto A' = gAg^{-1} + gdg^{-1} \end{aligned}$$

The Lagrangian has now a quadratic term in A'

$$-\frac{g^2}{2}(T^c\phi_0, T^d\phi_0)(A'_c, A'_d)$$

and the gauge bosons mass matrix has p zero eigenvalues corresponding to the p unbroken generators so there are p massless gauge bosons and the other components of the gauge potential acquire mass. This mechanism is very important because it fixes the problem explained in the first section that pure Yang-Mills theories only describe massless particles. This problem is the reason why Yang-Mills theories have been originally considered as unphysical. Now the Higgs mechanism is at the heart of the unification between electromagnetism and weak interactions which, together with quantum chromodynamics constitute the standard model of elementary particles.

3.3 Solitons and topological charges

We now make a digression to describe the Sine-Gordon model which is a simple, very interesting and inspiring example of a model with solitons and duality. It will then be our reference point. It is a field theory in $1+1$ dimensions[2] with one real-valued field and no gauge connection whose Lagrangian is

$$\mathcal{L} = \frac{1}{2}((\partial_t)^2 - (\partial_x)^2)\phi - \frac{\alpha}{\beta}(1 - \cos(\beta\phi))$$

[2] $d+1$ dimensions means d space-like dimensions and 1 time-like dimension

This Lagrangian describes the continuous limit of the following discrete mechanical system : consider a infinite clothesline with pegs attached at equal distance along it, each peg can rotate in a plane perpendicular to the clothesline and is acted on by a constant gravitational field orthogonal to the line and by its neighbors. The angle between a peg located at x and the gravitational vector field is given by $\beta\phi(x)$ (suitable units are chosen). The three terms in the Lagrangian are the kinetic energy, the torsion potential of the line and the gravitational potential, in this order. Obviously, the ground state is when all the pegs are hanging down (the constant is the Lagrangian is here so that the corresponding Hamiltonian is zero in this state). The states we are really interested in are those when all the pegs hang (almost) down except in a small region where they make one or several turns around the line as in figure (2). Solutions to the continuous model corresponding to

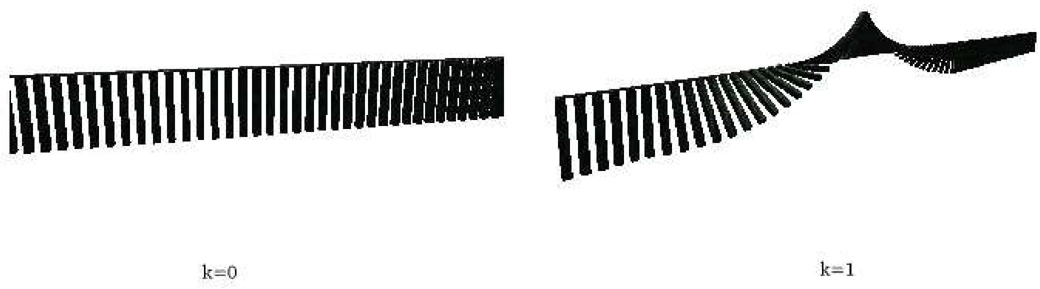

Figure 2: Solitons in the Sine-Gordon model

these states of the discrete model are called solitons (or kinks). The number of turns k is called the soliton (or kink) number. We concentrate on the case $k = 1$. This number is conserved in time evolution, heuristically because it would need an infinite amount of energy to flip one half of the infinite number of pegs to go from k to $k \pm 1$. It is tempting to think that this conserved quantity comes from the symmetry $\phi \mapsto \phi + 2n\pi$ of the Lagrangian but this symmetry is not continuous so Noether theorem does *not* apply. This conserved quantity is a topological charge because it comes from the fact that the map

$$\phi_\infty \left(\begin{array}{lcl} \{-1,+1\} & \to & 2\pi\mathbb{Z} \\ \epsilon & \mapsto & \lim_{x\to\epsilon\infty}\phi(x) \end{array} \right)$$

for a given k is not homotopic to a ϕ_∞ map for another k because of a topological obstruction. This kind of "field at infinity" will play an important role in 't Hooft-Polyakov monopoles.

It is very tempting to identify the $k = 1$ soliton to a particle because it has a finite energy, can be located in a small region of space and can move with any velocity up to the speed of light because the equations of motion are Lorentz invariant. In addition for $k > 1$ we can study solitons scattering.

All this is totally classical and we haven't quantized anything yet. Actually we can see that the relationship between solitons and quantum particles in this model is not so simple because we don't get the same mass. On the classical side we get the mass according to special relativity as the energy of the soliton in its rest frame : $M = \frac{8\sqrt{\alpha\beta}}{\beta^2}$. On the quantum side the mass is seen on the perturbative development of the Lagrangian around the ground state :

$$\mathcal{L} = \frac{1}{2}((\partial_t)^2 - (\partial_x)^2))\phi - \frac{1}{2}\alpha\beta\phi^2 + O(\phi^4)$$

so $m = \sqrt{\alpha\beta} \neq M$. So the soliton is a classical object distinct from quantum particles.

What is really very interesting is that there exists another field theory in $1 + 1$ dimensions which describes the dynamics of the solitons of the Sine-Gordon model using standard quantum particles with a Noether charge replacing the topological charge. This is the massive Thirring model whose Lagrangian is

$$\mathcal{L} = \frac{1}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - \frac{1}{2}g\bar{\psi}\gamma^\mu\psi\bar{\psi}\gamma_\mu\psi$$

The precise understanding of this Lagrangian is not necessary here, the important thing is that it is symmetric under $\psi \mapsto e^{i\alpha}\psi$ (α being constant, this is not a gauge theory) and this symmetry gives a Noether conserved quantity and when everything is quantized in both models with

$$\frac{\beta^2\hbar}{4\pi} = \frac{1}{1 + g\hbar/\pi}$$

the quantum theory is the same with an exchange between the topological charge of Sine-Gordon and the Noether charge of Thirring, see [3] for more details. This is very intriguing because topological charges and Noether charges seem to be very different so we have an hint of the relation between classical solitons and quantum fields. This relation will be investigated further in the case of the BPS monopoles in section 5

3.4 Boundary conditions and homotopy groups

In the context of SSB in gauge theories the "field at infinity" concept of the Sine-Gordon model can be generalized in the following way. Suppose we have a gauge theory on a trivial bundle with a potential term V which is a G-invariant function of one field ϕ. Since the bundle is trivial we can see ϕ as a map from $\mathcal{M}$ to a finite dimensional linear space F and V as a function from F to $\mathbb{R}$. Put

$$X := \{\zeta \in F \mid V(\zeta) \,\text{is minimal}\}$$

Suppose that G acts transitively on X. Let's call H the stabilizer of an element of X (every stabilizers are isomorphic so it doesn't matter which element we choose) so that we have a SSB from G to H and X is diffeomorphic to the homogeneous space G/H (as usual we assume that H is closed in G so H is a Lie group and G/H is a nice homogeneous space).

We now examine the consequences of requiring that $\phi(x)$ tends to an element of X at infinity. More precisely we define

$$\phi_\infty \begin{pmatrix} S^2 & \to & X \\ \hat{x} & \mapsto & \lim_{r\to\infty} \phi(r\hat{x}) \end{pmatrix}$$

and suppose that appropriate decay conditions are imposed to ensure that it is a well defined continuous map, see [13] for a precise statement and proof when $G = SU(2)$ and $H = U(1)$. This map defines an element of $\pi_2(G/H)$ because you can always gauge transform ϕ such that a fixed point of S^2 goes to a fixed point in X (remember that G acts transitively on X). We have a fiber bundle $G \to G/H$ with fiber H so we can compute this homotopy group using the exact sequence of homotopy groups for a fiber bundle (see [7]) :

$$\cdots \to \pi_2(H) \to \pi_2(G) \to \pi_2(G/H) \to \pi_1(H) \to \pi_1(G) \to \pi_1(G/H) \to \cdots$$

If we assume that G is compact then a theorem by Cartan says that $\pi_2(G) = \{0\}$ and if G is simply connected then the preceding exact sequence gives

$$\pi_2(G/H) \simeq \pi_1(H)$$

This result gives a rigorous ground to the idea that when we have a SSB and we require that all fields tend to vacuum at infinity then the asymptotic properties of the field are governed by the topology of the little group instead of the full gauge group (note however that there is an assumption on G so this intuitive idea can be misleading).

4 Georgi-Glashow model and 't Hooft-Polyakov monopole

The Georgi-Glashow model is an early attempt to model electro-weak interactions. It has still an interest as a toy model because it is simple and exhibits many very interesting features as we shall see. We are interested in the bosonic part of the Georgi-Glashow model which is a gauge theory on a trivial $SU(2)$-bundle with one scalar field in the adjoint representation. We choose once and for all a trivialization and write this Higgs field as $\phi : \mathcal{M} \to \mathcal{M} \times \mathfrak{su}_2$. The connection acts on ϕ as $D\phi = d\phi + q[W, \phi]$ where W is a potential which can be defined over all $\mathcal{M}$ because the bundle is trivial and is a $\mathfrak{su}_2$-valued 1-form. The gauge field is $F = dW + qW \wedge W$. The wedge product in this equation is defined in appendix A and the precise relationship between F and the curvature of the gauge connection on the adjoint bundle is explained in appendix C. The Lagrangian is

$$\mathcal{L} = \frac{1}{2}\tilde{K}(F, F)vol + \frac{1}{2}\tilde{K}(D\phi, D\phi)vol - \frac{\lambda}{4}(|\phi|^2 - a^2)^2 vol$$

See appendices B and C for the notations used. The equations of motion are the Yang-Mills-Higgs equations derived in appendix C :

$$\begin{cases} d_{\bar{D}}F & = & 0 \qquad \text{(Bianchi identity)} \\ \star d_{\bar{D}} \star F & = & -q[\phi, D\phi] \\ \star d_D \star D\phi & = & \lambda\phi(|\phi|^2 - a^2) \end{cases}$$

The potential is minimal when $|\phi| = a$. We don't have a unique vacuum and every vacuum is invariant under a $U(1)$ subgroup of $SU(2)$ so we have a SSB from $SU(2)$ to $U(1)$. The particle spectrum of this Lagrangian can be seen by expanding it around a vacuum state ϕ_0 as explained in subsection 3.2. The unbroken $U(1)$ subgroup leads to a massless gauge particle with potential $A = \frac{1}{a}(\phi_0, W)$ which we call the photon because it is related to a $U(1)$ symmetry which can be identified to electromagnetism as we will see. The broken generators absorb the part of ϕ which is orthogonal to ϕ_0 to make two massive electrically charged gauge particles $W^{\pm}$ and the remaining part of ϕ becomes a massive neutral Higgs boson (cf. section 3.2).

particle	mass	electric charge
photon	0	0
Higgs	$a\sqrt{2\lambda}$	0
$W^{\pm}$	aq	$\pm q$

We are interested only in finite energy solutions. For such solutions every fields must approach a vacuum state at spatial infinity, *cf.* section 3.4. Here we must have $|\phi| \to a$ and $D\phi \to 0$ at spatial infinity. We can use the result of the end of section 3.4 because $SU(2)$ is compact and simply connected. So each solution is associated with an homotopy class in $\pi_1(U(1)) \simeq \mathbb{Z}$. This homotopy class is called the charge of the solution. Here G/H is topologically a 2-sphere and the homotopy class is indeed the degree of ϕ_∞, see *e.g.* [14] and [15] for the definition and needed properties the degree of a map. We will see later that it can also be identified with a Chern number. Because this number is a integer, it is preserved under continuous deformations, notably under time evolution so we have a topological charge.

4.1 Static solutions and the 't Hooft-Polyakov Ansatz

We are interested in static solutions, i.e. everything is time independent and $W_0 = 0$. Now we seek a spherically symmetric solution using the 't Hooft-Polyakov Ansatz (Ansatz is the German word physicists like to use to say that they look for a special form of solution), we decompose W as $W^\mu = W^{\mu,\alpha} T_\alpha$ and ϕ as $\phi = \phi^\alpha T_\alpha$ using the basis given in appendix B and specify :

$$\begin{cases} \phi^\alpha(x) &= \frac{x^\alpha}{qr^2} H(aqr) \\ W^{i,\alpha}(x) &= -\epsilon_{\alpha ij} \frac{x^j}{q||r||^2}(1 - K(aqr)) \\ W^{0,\alpha}(x) &= 0 \end{cases}$$

Latin indices run over space-like dimensions only, H and K are two functions from $\mathbb{R}$ to $\mathbb{R}$ and $r = ||x||$ for $x \in \mathbb{R}^3$. A more general Ansatz (including non-static solutions) and a discussion of its origin can be found in [3]. We put $\xi = aqr$. The finiteness of energy requires $K \to 0$ and $H/\xi \to 1$ as $\xi \to \infty$ and $K - 1 \leq O(\xi)$ and $H \leq O(\xi)$ as $\xi \to 0$. The first consequence of these boundary conditions is that ϕ_∞ has degree 1. The equations of motion give the following equations for H and K :

$$\begin{aligned} \xi^2 \frac{d^2 K}{d\xi^2} &= KH^2 + K(K^2 - 1) \\ \xi^2 \frac{d^2 H}{d\xi^2} &= 2K^2 H + \frac{\lambda}{q^2} H(H^2 - \xi^2) \end{aligned}$$

Of course this system can not be solved exactly but the existence of a solution can be proved. In addition the asymptotic behavior when $\xi \to \infty$ (hence when $r \to \infty$) is very easy to study because the preceding equations become

$$\begin{aligned} \frac{d^2K}{d\xi^2} &= K \\ \frac{d^2h}{d\xi^2} &= 2\frac{\lambda}{q^2}h \end{aligned}$$

where $h = H - \xi$. These equations are easy to solve and using the boundary conditions we get

$$\begin{aligned} K &\sim \exp(-\frac{M_W r}{\hbar}) \\ h &\sim \exp(-\frac{M_H r}{\hbar}) \end{aligned}$$

and this is very interesting because it means that this solution differs from a Higgs vacuum only in a region whose size is of order $\max(\frac{\hbar}{M_W}, \frac{\hbar}{M_H})$ so we have in this case the relationship between mass and length scale which was alluded in section 1.2. This region is called the core of the magnetic monopole and we now explain why the whole configuration is called a magnetic monopole by examining what happens outside the core.

Here we will use the very compact but sometimes confusing convention used implicitly in [16] and [17] that (A, B) means whatever is natural and meaningful depending of what A and B are. This convention is very convenient to expose quickly the results but we use a different one when we derive things carefully in appendices.

Suppose we are very far away from the core so $|\phi| = a$ and $D\phi = 0$. Then $A = \frac{1}{a}(\phi, W)$ and in the same way we define the electromagnetic field to be $f = \frac{1}{a}(\phi, F)$. Indeed it deserves this name because

$$df = \frac{1}{a}d(\phi, F) = \frac{1}{a}(D\phi, F) + \frac{1}{a}(\phi, d_{\bar{D}}F) = 0$$

because $D\phi = 0$ (vacuum condition) and $d_{\bar{D}}F = 0$ (Bianchi identity). And

$$d \star f = \frac{1}{a}d \star (\phi, F) = \frac{1}{a}d(\phi, \star F) = \frac{1}{a}(D\phi, \star F) + \frac{1}{a}(\phi, d_{\bar{D}} \star F) = 0$$

because $D\phi = 0$ and $\star d_{\bar{D}} \star F = -q[\phi, D\phi]$ (equation of motion) so $d_{\bar{D}} \star F = 0$. So f satisfies Maxwell's equation very far away from the core. Those calculations of df and $d \star f$ are not carried out in details in the appendices because there is not enough room for it but it is very similar to the calculation of appendix D.

As in section 2 we write $f = B + E \wedge dt$ and because everything is static $E = 0$ and we can use the asymptotic form of H and K to see that, far away from the core, B is exactly the same that for a Dirac monopole with charge 1. The 't Hooft-Polyakov Ansatz can be slightly modified to get a charge -1. In the next section these two monopoles will be denoted by $M^{\pm}$. As a first approximation, the solutions corresponding to higher charges can be seen as a dilute gas of $M^{\pm}$ (note the analogy with the kinks of Sine-Gordon). The exact situation is actually much richer and I've spent a lot of time studying it but there is not enough room in this essay to include my work on these issues.

When $\lambda \to 0$ we have an interesting limiting case called the Prasad-Sommerfield limit. In this case the equations for the 't Hooft-Polyakov Ansatz can be solved in closed form and we get

$$\begin{aligned} H(\xi) &= \xi \coth \xi - 1 \\ K(\xi) &= \frac{\xi}{\sinh \xi} \end{aligned}$$

We still have a finite total energy but this monopole can be distinguished from the Dirac monopole even from a very large distance because the Higgs field is now a massless long range field. From now we will consider only this case.

4.2 The Bogomolny bound

We now turn to a more geometrical approach to static 't Hooft-Polyakov monopoles in order to calculate the mass of a monopole and find new insights of the topological issues. In the same way that we went from a four dimensional theory to a static three dimensional theory in section 2, we now study the $SU(2)$ gauge theory with an adjoint Higgs field on $\mathbb{R}^3$. We still denote by F the gauge field because B is already the magnetic part of f.

The energy density is given by the opposite of what remains of the Georgi-Glashow Lagrangian in the static case :

$$\mathcal{H} = -\frac{1}{2}\tilde{K}(F, F)vol - \frac{1}{2}\tilde{K}(D\phi, D\phi)vol$$

(remember that $\lambda = 0$, our space time metric has signature $(-, +, +, +)$ and K is negative definite). We now calculate the total energy of a static solution to Yang-Mills-Higgs equations which tends to vacuum at infinity. Because we consider static solutions, this energy is also the mass of the solution according to the theory of special relativity. Let $\mathcal{B}_R$ be a ball of radius R in $\mathbb{R}^3$ with boundary S_R. In appendix D we show that

$$\int_{\mathcal{B}_R} \mathcal{H} = -\frac{1}{2}\int_{\mathcal{B}_R} \tilde{K}(F - \star D\phi, F - \star D\phi)vol - \int_{S_R} (\phi, F) \tag{6}$$

Now consider the rank 2 complex vector bundle E associated with our principal $SU(2)$-bundle. Since $|\phi| \to a > 0$ as $R \to \infty$ then, for R sufficiently large, ϕ has two one dimensional complex eigenspaces with eigenvalues $\pm i|\phi|^2/4$ which split E in two complex line bundles L and L^* over S_R . We equip L with the connection obtained by orthogonal projection of the original connection of E onto it. We assume that ϕ decays to vacuum such that $\tilde{K}(D\phi, D\phi) = O(R^{-2})$ so the curvature of the connection on L differs from the projection of the curvature of E onto it by a term of order $O(R^{-4})$ so it is given by $2i(qF,\phi)/|\phi| + O(R^{-3})$ (remember that there is a q in the relation between F and $\hat{F}$) and so the integral over S_R in (6) differs from 2π times the first Chern number of L by a term of order $O(R^{-1})$ and for $R \to \infty$ we get

$$\int_{\mathbb{R}^3} \mathcal{H} = -\frac{1}{2}\int_{\mathbb{R}^3} \tilde{K}(F - \star D\phi, F - \star D\phi)vol + \frac{4\pi ak}{q}$$

where k is the degree of ϕ_∞. If $k/q \geq 0$ then the absolute minimum of E is $4\pi ak/q$ so we have a bound on the mass of a monopole in this model

$$M \geq \frac{4\pi ak}{q}$$

This is the celebrated Bogomolny bound which is a further advantage of 't Hooft-Polyakov monopoles compared to Dirac monopoles. This bound is saturated when $F = \star D\phi$. This condition is called the Bogomolny equation. When it holds, the equations of motion automatically hold. This is rather analogous to the self-duality condition which ensures that the Yang-Mills equation is satisfied. Indeed the Bogomolny equation which deals with fields on $\mathbb{R}^3$ can be viewed as a special case of self-dual Yang-Mills equation on $\mathbb{R}^4$. These finite energy solutions to the Georgi-Glashow model in the Prasad-Sommerfield limit verifying the Bogomolny equation are called BPS monopoles (Bogomolny-Prasad-Sommerfield monopoles). I have spent quite a lot of time studying BPS monopoles but I don't have enough room to develop this subject here, see [4] for a review and [18], [5], [17],[16],[8], [19] for more details.

5 Duality and Montonen-Olive conjecture

We have seen an example of two theories, the Sine-Gordon and the Thirring models, describing the same thing with an exchange of solitons/usual particles and topological/Noether charges. In addition, in the Prasad-Sommerfield limit of the Georgi-Glashow model, we have magnetic monopoles which are solitons with a topological charge and we have electrically charged particles $W^{\pm}$ with a Noether charge. So, in 1977, C. Montonen and D. Olive made in [20] the bold conjecture that the fully quantized version of this theory could have two dual formulations which exchange solitonic and non-solitonic particles, topological and Noether charges, magnetic and electric charges and that, in addition, they would both have the same Georgi-Glashow Lagrangian, up to numerical factors. In this section we review the evidences favoring conjecture as well as the possible obstacles.

If this conjecture is true then the massive vector particles $W^{\pm}$ and the magnetic monopoles $M^{\pm}$ must have many common points and indeed we have the following spectrum :

particle	mass	electric charge	magnetic charge
photon	0	0	0
Higgs	0	0	0
$W^{\pm}$	aq	$\pm q$	0
$M^{\pm}$	ag	0	$\pm g$

which tells us that everything is OK and that we don't even have to change the constant a to go from one theory to the other.

We also have to check that the forces between particles look the same. The force between two W particles is given by a QFT calculation. At the first order we only have two Feynman diagrams, a photon exchange and a Higgs exchange. These two interactions have the same magnitude, the Higgs exchange is always attractive and the photon exchange is attractive for opposite charges and repulsive for the same charge. Combining the two, we find that two W with opposite charges are attracted to each other and that for two W with the same charge there is a force cancellation. In [21], Manton showed that, at least to order $O(1/r^2)$, the results are the same for two widely separated monopoles. I have studied this calculation in details but I don't have enough room to explain it. Note that the result in the case of two monopoles of the same charge has been confirmed by the discovery of static solutions with many separated monopoles.

Despite these evidences the conjecture may fail for at least two reasons. First we don't know precisely how to quantize completely all this and it may be that the equality of masses get lost because of renormalization. Also there may be a problem with spin because the W's have spin 1 and the M's seem to have spin 0 because they are spherically symmetric. These two problems may be solved in the framework of a supersymmetric theory which could "protect" the masses from renormalization and introduce supersymmetric partners with the right spin.

There have been recent progress in this direction but unfortunately I lacked time to study this, see [5] and [22].

Conclusion

As we have seen, the subject of magnetic monopoles leads to very beautiful applications of topology and geometry together with exciting physical discussions. As alluded at the end of section 4, the theory of BPS monopoles is very rich and uses a lot more mathematics, especially complex geometry, than what we have discussed. For more applications of algebraic topology to field theories see [23]. Also there are several directions that we have not explored during this project. The most obvious one is that of generalization to arbitrary semi-simple compact Lie group. It is clear that the calculations of appendix C extent without any difficulty but the topological classification of solutions and the corresponding topological charges are more complicated and I have only looked superficially at [24]. The Montonen-Olive duality can also be extended in this general case, see [25] and [26]. There are also papers like [27] which study the dynamics of Dirac monopoles and there are a lot of papers studying dualities in supersymmetric theories related to what we examined in this essay.

Now I would like to thank my supervisor Niall McKay, especially for having let me free to explore issues he hadn't quite expected. I would also like to thank Ed Corrigan, Jose Figueroa-O'Farrill and Ian McIntosh for useful discussions about this project and Vincent Minerbe for reading an early version of this essay (before I added the appendices).

Appendices

A Connections on a vector bundle

Here we recall some definitions and results about connections on a vector bundle. Let $E \to M$ be a vector bundle with a connection D. We work with a local trivialization so that D can be expressed as $D = D^0 + A$ where D^0 is the flat connection coming from the trivialization and A is a $End(E)$-valued 1-form called the gauge potential which can be seen as the pull-back by the trivialization of the intrinsic connection 1-form. We now recall how the covariant derivative can be extended to various objects.

E-valued forms D is extended to $\Gamma(\Lambda^\bullet T^*M \otimes E)$ by

$$d_D(s \otimes \omega) := Ds \wedge \omega + s \otimes d\omega$$

where $\omega \in \Gamma(\Lambda^\bullet T^*M)$ and $s \in \Gamma(E)$ so $d_D(s_I \otimes dx^I) = D_\mu s_I \otimes dx^\mu \wedge dx^I$. Note that for a 0-form we get $d_D = D$. Also we put $d := d_{D^0}$ on E-valued forms. This notation is reasonable since D^0 is flat so $d_{D^0}^2 = 0$.

$End(E)$-valued forms First we extend D to the sections of $End(E)$ by using $End(E) \simeq E \otimes E^*$ and the usual constructions of connections on a dual and a tensor product.

$$(\bar{D}T)(s) := D(T(s)) - T(Ds) \tag{7}$$

where $T \in \Gamma(End(E))$, $s \in \Gamma(E)$ and $T(Ds) := T(D_\mu s) \otimes dx^\mu$. Then D can be extended to $End(E)$-valued forms by the same construction that in the preceding paragraph but using $\bar{D}$ instead of D. We also define the wedge product of a $End(E)$-valued form and a E-valued form :

$$(T \otimes \omega) \wedge (s \otimes \mu) := T(s) \otimes (\omega \wedge \mu) \tag{8}$$

which is a E-valued form. The wedge product of two $End(E)$-valued forms is

$$(S \otimes \omega) \wedge (T \otimes \mu) := ST \otimes (\omega \wedge \mu)$$

and the graded commutator of two $End(E)$-valued forms α and β of degrees p and q respectively is

$$[\alpha, \beta] := \alpha \wedge \beta - (-1)^{pq} \beta \wedge \alpha$$

In particular when $\alpha = S \otimes \omega$ and $\beta = T \otimes \mu$ with $deg(\omega)deg(\mu) \in 2\mathbb{Z}$ then

$$[\alpha, \beta] = [S, T]\omega \wedge \mu \tag{9}$$

where the bracket on the left hand side is the endomorphism commutator. More generally we have the following properties :

- $d_{\bar{D}}(\alpha \wedge \beta) = (d_{\bar{D}}\alpha) \wedge \beta + (-1)^p \alpha \wedge (d_{\bar{D}}\beta)$
- $[\alpha, \beta] = (-1)^{pq+1}[\beta, \alpha]$

- $[\alpha, \alpha] = 0$
- $[\alpha, \alpha \wedge \alpha] = 0$

Note that $\alpha \wedge \alpha \neq 0$ in general, even if $deg(\alpha)$ is odd.

The derivatives d_D and $d_{\bar{D}}$ can be rewritten using those operations and the gauge potential A. Let ω be a E-valued form, we have

$$d_D\omega = d\omega + A \wedge \omega$$

Let η be a $End(E)$-valued form, we have

$$d_{\bar{D}}\eta = d\eta + [A, \eta] \tag{10}$$

Note also that, in a semi-Riemannian context, the Hodge $\star$ operator can be extended to E-valued and $End(E)$ valued forms simply by acting on the $\Lambda^\bullet T^*M$ part. Concerning the Hodge operator on differential form we use the following conventions : for any p-form

$$\omega = \frac{1}{p!}\omega_{i_1 \dots i_p} e^{i_1} \wedge \dots \wedge e^{i_p}$$

where (e^i) is an orthonormal local basis of T^*M we take $(j_1 \dots j_{n-p})$ the complement of $(i_1 \dots i_p)$ in $(1 \dots n)$ and we have

$$(\star\omega)_{j_1 \dots j_{n-p}} = \frac{1}{p!}\epsilon^{i_1 \dots i_p}{}_{j_1 \dots j_{n-p}}\omega_{i_1 \dots i_p}$$

We denote by Tr the trace of a $End(E)$ valued form which is an ordinary form. We have the following properties for two $End(E)$-valued forms α and β of degrees p and q respectively :

- $Tr(\alpha \wedge \beta) = (-1)^{pq} Tr(\beta \wedge \alpha)$
- $Tr[\alpha, \beta] = 0$ (note that this is not as obvious as it seems)
- $Tr(d_{\bar{D}}\alpha) = dTr\alpha$

And if α or β is compactly supported

- $\int_M Tr(d_{\bar{D}}\alpha \wedge \beta) = (-1)^p \int_M Tr(\alpha \wedge d_{\bar{D}}\beta)$ with $p + q = dimM - 1$
- $\int_M Tr(\alpha \wedge \star\beta) = \int_M Tr(\beta \wedge \star\alpha)$

Curvature The curvature is a $End(E)$-valued 2-form $F := dA + A \wedge A$. In a coordinate chart, $F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ with $F_{\mu\nu} := [D_\mu, D_\nu]$. We have the following properties :

- $d_{\bar{D}} F = 0$ (Bianchi identity)
- the squared derivative of a E-valued form η is $d_D^2 \eta = F \wedge \eta$

B The Lie algebra $\mathfrak{su}_2$

We see $\mathfrak{su}_2$ as the vector space spanned by $\{T_a = i\sigma_a; a \in \{1, 2, 3\}\}$ where the σ's are the Pauli matrices so that

$$T_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad T_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad T_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

with the Lie bracket

$$[T_a, T_b] = -2\epsilon_{abc} T_c$$

This explicit basis will be needed only for the 't Hooft-Polyakov Ansatz. We denote by K the Killing form on $\mathfrak{su}_2$. This unusual notation is used because B is already used for the magnetic field. As usual the Killing form is defined by

$$K(X, Y) = tr(adX \circ adY : \mathfrak{su}_2 \to \mathfrak{su}_2)$$

and we have the useful property that, for all X, Y, Z in $\mathfrak{su}_2$,

$$K([X, Y], Z) = K(X, [Y, Z]) \tag{11}$$

The definition of the Killing form and this property are very general, see [28, p. 207]. Also, because $\mathfrak{su}_2$ is semi-simple, K is non degenerate.

The Killing form is a natural choice of norm in gauge theories because it is gauge invariant.

C Derivation of the Yang-Mills-Higgs equations

In this appendix we derive carefully the Yang-Mills-Higgs equations. First we must be precise with notations. There are several inner products involving elements of $\mathfrak{su}_2$, the Lorentzian metric and combinations of these two ones. We adopt the convention suggested by Ian McIntosh that all symmetric bilinear functions give scalar results. In particular we resist the temptation to include the Lorentzian volume form in the inner product on differential forms coming from the space-time metric. We also adopt the convention that all bilinear functions take two objects of the same nature. For example we don't use $K(\phi, D\phi)$ to mean $K(\phi, D_\mu\phi)dx^\mu$. The gain in clarity given by these conventions is worth the little loss of compactness of notations and the apparition of some indices. We denote by $\tilde{K}$ the inner product defined on $\mathfrak{su}_2$-valued 1-forms by

$$\tilde{K}(\alpha_\mu dx^\mu, \beta_\nu dx^\nu) := K(\alpha_\mu, \beta^\mu) \tag{12}$$

and on $\mathfrak{su}_2$-valued 2-forms by :

$$\tilde{K}(\frac{1}{2}\alpha_{\mu\nu}dx^\mu \wedge dx^\nu, \frac{1}{2}\beta_{\sigma\rho}dx^\sigma \wedge dx^\rho) := \frac{1}{2}K(\alpha_{\mu\nu}, \beta^{\mu\nu})$$

We now turn to the Georgi-Glashow setting. We start with a trivial $SU(2)$-bundle $P \to \mathcal{M}$ with a $SU(2)$-connection. We use the adjoint representation of $SU(2)$ on $\mathfrak{su}_2$ (multiplied by a charge q) to build an associated vector bundle E. The connection on P induces a connection D on E. We choose once and for all a trivialization of E. Let $\hat{F}$ be the corresponding curvature 2-form for D. Because E is trivial, $\hat{F}$ is defined globally. Because D is a $SU(2)$-connection and because we use the adjoint representation, $\hat{F}$ (which, *a priori*, is a $End(E)$-valued 2-form) can be written $\hat{F} = adF$ for a $\mathfrak{su}_2$-valued 2-form F. This fact is useful but we must be careful not to confuse F with a E-valued form. Indeed, if ϕ is a section of E then $D\phi$ is a E-valued form, so it can be seen as a $\mathfrak{su}_2$-valued form thanks to our trivialization, so it seems to be like F but the derivative of $D\phi$ is $d_D D\phi$ whereas the derivative of F is $d_{\bar{D}}F$. This is why we make a notational difference between D and $\bar{D}$ altough it is not usual.

Now the pure Yang-Mills Lagrangian is

$$\begin{aligned}
\mathcal{L}_{YM} &= \frac{1}{4}tr(\hat{F}_{\mu\nu} \circ \hat{F}^{\mu\nu})vol = \frac{1}{4}tr(adF_{\mu\nu} \circ adF^{\mu\nu})vol \\
&= \frac{1}{4}K(F_{\mu\nu}, F^{\mu\nu})vol = \frac{1}{2}\tilde{K}(F, F)vol
\end{aligned}$$

And the Georgi-Glashow Lagrangian is

$$\mathcal{L} = \frac{1}{2}\tilde{K}(F, F)vol + \frac{1}{2}\tilde{K}(D\phi, D\phi)vol - \frac{\lambda}{4}(|\phi|^2 - a^2)^2 vol$$

where the Higgs field ϕ is a section of E, a and λ are positive real numbers and $|\phi|^2 = -K(\phi, \phi)$. We now derive the equations of motion.

Variation of the connection Let W be a $\mathfrak{su}_2$-valued 1-form such that $D\phi = D^0\phi + q[W, \phi]$. We derive the stationarity condition of the action under $W \mapsto W + \delta W$ with δW compactly supported so that all the following integrals are well defined except in the left hand side of the following equation which is a formal expression whose definition is the right hand side. Unless otherwise specified all the integrals are over $\mathcal{M}$.

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) := \int \delta(\tilde{K}(D\phi, D\phi))vol = 2\int \tilde{K}(\delta D\phi, D\phi)vol$$

with $\delta D\phi = \delta(D^0\phi + q[W, \phi]) = q[\delta W, \phi]$ so

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) = 2q\int \tilde{K}([\delta W, \phi], D\phi)vol$$

and by (12) and (11) we get

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) = 2q\int \tilde{K}(\delta W, [\phi, D\phi])vol \tag{13}$$

The calculation for the pure Yang-Mills term can be found in [6] and easily translated into our notations if we are careful with the q in the definition of F and W, we get

$$\delta\left(\int \tilde{K}(F, F)vol\right) = 2\int \tilde{K}(\delta W, \star d_{\bar{D}} \star F)vol$$

So the variation of the total action with respect to W is

$$\begin{aligned} \delta\mathcal{A} &= \int \tilde{K}(\delta W, \star d_{\bar{D}} \star F)vol + q\int \tilde{K}(\delta W, [\phi, D\phi])vol \\ &= \int \tilde{K}(\delta W, \star d_{\bar{D}} \star F + q[\phi, D\phi])vol \end{aligned}$$

And using the facts that the Killing form on $\mathfrak{su}_2$ and the metric on $\mathcal{M}$ are non-degenerate and that δW is arbitrary we get the first equation of motion

$$\star d_{\bar{D}} \star F = -q[\phi, D\phi]$$

Variation of the Higgs field We now derive the stationarity condition of the action under $\phi \mapsto \phi + \delta\phi$ with $\delta\phi$ compactly supported.

We being with the variation of the potential term of the Lagrangian when the Higgs field varies.

$$\begin{aligned} \delta((|\phi|^2 - a^2)^2) &= 2(|\phi|^2 - a^2)\delta(|\phi|^2 - a^2) \\ &= -2(|\phi|^2 - a^2)\delta(K(\phi, \phi)) \end{aligned}$$

so

$$\delta((|\phi|^2 - a^2)^2) = -4(|\phi|^2 - a^2)K(\delta\phi, \phi) \tag{14}$$

This was the easy part. Now we calculate the variation of the $\tilde{K}(D\phi, D\phi)$ term.

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) := \int \delta(\tilde{K}(D\phi, D\phi))vol = 2\int \tilde{K}(\delta D\phi, D\phi)vol$$

with $\delta D\phi = \delta(D^0\phi + q[W, \phi]) = D^0\delta\phi + q[W, \delta\phi]$ so

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) = 2\int \tilde{K}(D^0\delta\phi, D\phi)vol + 2q\int \tilde{K}([W, \delta\phi], D\phi)vol \tag{15}$$

We treat these two terms separately. Ultimately we want to express this variation as something like $\int K(\delta\phi, ?)vol$ and conclude as we did when varying W so the D^0 in $\int \tilde{K}(D^0\delta\phi, D\phi)vol$ is annoying and the natural idea to get rid of it is to use Stokes theorem.

We can calculate :

$$\begin{aligned} d(K(\delta\phi, D_\nu\phi) \star dx^\nu) &= d(K(\delta\phi, D_\nu\phi)) \wedge \star dx^\nu \\ &= K(\partial_\mu\delta\phi, D_\nu\phi)dx^\mu \wedge \star dx^\nu + K(\delta\phi, \partial_\mu D_\nu\phi)dx^\mu \wedge \star dx^\nu \end{aligned}$$

The first term is easy to calculate :

$$\begin{aligned} K(\partial_\mu\delta\phi, D_\nu\phi)dx^\mu \wedge \star dx^\nu &= K(\partial_\mu\delta\phi, D_\nu\phi)g^{\mu\nu}vol \\ &= K(\partial_\mu\delta\phi, g^{\mu\nu}D_\nu\phi)vol \\ &= K(\partial_\mu\delta\phi, D^\mu\phi)vol \\ &= \tilde{K}(D^0\delta\phi, D\phi)vol \end{aligned}$$

To calculate the second one we use the following intermediate result :

$$\begin{aligned} d\star D\phi &= d(\star D_\nu\phi dx^\nu) \\ &= d(D_\nu\phi) \wedge \star dx^\nu + D_\nu\phi\, d\star dx^\nu \\ &= \partial_\mu D_\nu\phi\, dx^\mu \wedge \star dx^\nu \\ &= \partial_\mu D_\nu\phi\, g^{\mu\nu}\, vol \end{aligned}$$

So $\partial_\mu D_\nu\phi\, g^{\mu\nu} = \star d\star D\phi$ and our second term was

$$K(\delta\phi, \partial_\mu D_\nu\phi)dx^\mu \wedge \star dx^\nu = K(\delta\phi, \partial_\mu D_\nu\phi)\, g^{\mu\nu}\, vol = K(\delta\phi, \star d\star D\phi)vol$$

Putting the two terms together we get

$$d(K(\delta\phi, D_\nu\phi) \star dx^\nu) = \tilde{K}(D^0\delta\phi, D\phi)vol + K(\delta\phi, \star d \star D\phi)vol$$

We integrate over a ball $\mathcal{B}_R$ of radius R and boundary S_R

$$\int_{\mathcal{B}_R} d(K(\delta\phi, D_\nu\phi) \star dx^\nu) = \int_{\mathcal{B}_R} \tilde{K}(D^0\delta\phi, D\phi)vol + \int_{\mathcal{B}_R} K(\delta\phi, \star d \star D\phi)vol$$

and, as originally planned, use Stokes theorem to get

$$\int_{S_R} K(\delta\phi, D_\nu\phi) \star dx^\nu = \int_{\mathcal{B}_R} \tilde{K}(D^0\delta\phi, D\phi)vol + \int_{\mathcal{B}_R} K(\delta\phi, \star d \star D\phi)vol$$

In addition, $\delta\phi$ is compactly supported and, because D^0 is a local operator, $D^0\delta\phi$ is also compactly supported. So we can choose R sufficiently large so that the preceding equation becomes

$$0 = \int_{\mathcal{M}} \tilde{K}(D^0\delta\phi, D\phi)vol + \int_{\mathcal{M}} K(\delta\phi, \star d \star D\phi)vol$$

Note that there is no convergence issue because we haven't taken any limit, we have chosen *one* large R.

We can now return to (15) and get

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) = -2\int K(\delta\phi, \star d \star D\phi)vol + 2q\int \tilde{K}([W, \delta\phi], D\phi)vol$$

and calculate on the second term

$$\begin{aligned}
\tilde{K}([W, \delta\phi], D\phi) &= K([W_\mu, \delta\phi], D^\mu\phi) \\
&= -K(\delta\phi, [W_\mu, D^\mu\phi]) \\
&= -K(\delta\phi, [W_\mu, D_\nu\phi]g^{\mu\nu}) \\
&= -K(\delta\phi, [W_\mu, D_\nu\phi] \star (dx^\mu \wedge \star dx^\nu)) \\
&= -K(\delta\phi, \star([W_\mu, D_\nu\phi] dx^\mu \wedge \star dx^\nu)) \\
&= -K(\delta\phi, \star(W \wedge \star D\phi))
\end{aligned}$$

See equation (8) in appendix A for the definition of $W \wedge \star D\phi$. So we have

$$\begin{aligned}
\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) &= -2\int K(\delta\phi, \star d \star D\phi)vol - 2qK(\delta\phi, \star(W \wedge \star D\phi))vol \\
&= -2\int K(\delta\phi, \star(d \star D\phi + qW \wedge \star D\phi)vol
\end{aligned}$$

So, using equation (10)

$$\delta\left(\int \tilde{K}(D\phi, D\phi)vol\right) = -2\int K(\delta\phi, \star d_D \star D\phi)vol$$

Putting together this and equation (14) we get the variation of the total action with respect to ϕ :

$$\begin{aligned}\delta\mathcal{A} &= -\int K(\delta\phi, \star d_D \star D\phi)vol + \int \lambda(|\phi|^2 - a^2)K(\delta\phi, \phi)vol \\ &= \int K(\delta\phi, -\star d_D \star D\phi + \lambda(|\phi|^2 - a^2)\phi)vol\end{aligned}$$

And using the facts that the Killing form on $\mathfrak{su}_2$ is non-degenerate and that $\delta\phi$ is arbitrary we get the second equation of motion

$$\star d_D \star D\phi = \lambda(|\phi|^2 - a^2)\phi$$

D Derivation of the Bogomolny bound

Here we derive in details the Bogomolny bound for the mass of a t' Hooft-Polyakov monopole using the notational conventions of appendix C. The link between the compact notation convention and the explicit conventions is $(\phi, F) = \frac{1}{2}K(\phi, F_{\mu\nu})dx^\mu \wedge dx^\nu$. We also introduce the notation $dx^{i_1 \cdots i_l} := dx^{i_1} \wedge \cdots \wedge dx^{i_l}$. We begin with a calculation which will be useful in the energy calculation :

$$\begin{aligned}d(K(\phi, F_{\mu\nu}) \wedge dx^{\mu\nu}) &= (K(\partial_\rho\phi, F_{\mu\nu})dx^\rho + K(\phi, \partial_\rho F_{\mu\nu})dx^\rho) \wedge dx^{\mu\nu} \\ &= K(D_\rho\phi, F_{\mu\nu})dx^{\rho\mu\nu} \\ &\quad + (K(\phi, \partial_\rho F_{\mu\nu}) - qK([A_\rho, \phi], F_{\mu\nu}))\, dx^{\rho\mu\nu}\end{aligned}$$

By the usual trick with the bracket and the Killing form we can rewrite the second term as $K(\phi, \partial_\rho F_{\mu\nu} + qK([A_\rho, F_{\mu\nu}])dx^{\rho\mu\nu}$ and this is zero because of the Bianchi identity. Indeed, the Bianchi identity says that $d_{\bar{D}}\hat{F} = 0$ and the relation between $\hat{F}$ and F is a Lie algebra homomorphism between the space of endomorphisms of our vector bundle with commutator bracket and $\mathcal{C}^\infty(\mathbb{R}^3, \mathfrak{su}_2)$ with the obvious bracket so we conclude using the expression of $d_{\bar{D}}$ in terms of the gauge potential and (9). So we have proved that

$$K(D_\rho\phi, F_{\mu\nu})dx^{\rho\mu\nu} = d(K(\phi, F_{\mu\nu}) \wedge dx^{\mu\nu})$$

and we can now calculate as announced

$$\begin{aligned}\int_{\mathcal{B}_R}\mathcal{H} &= -\frac{1}{2}\int_{\mathcal{B}_R}\tilde{K}(F,F)vol - \frac{1}{2}\int_{\mathcal{B}_R}\tilde{K}(D\phi,D\phi)vol \\ &= -\frac{1}{2}\int_{\mathcal{B}_R}\tilde{K}(F-\star D\phi,F-\star D\phi)vol - \int_{\mathcal{B}_R}\tilde{K}(\star D\phi,F)vol\end{aligned}$$

with

$$\begin{aligned}2\tilde{K}(\star D\phi,F)vol &= K((\star D\phi)^{\mu\nu},F_{\mu\nu})vol \\ &= K(\epsilon^{\rho\mu\nu}D_\rho\phi,F_{\mu\nu})vol \\ &= K(D_\rho\phi,F_{\mu\nu})\epsilon^{\rho\mu\nu}vol \\ &= K(D_\rho\phi,F_{\mu\nu})dx^{\rho\mu\nu}\end{aligned}$$

(see our conventions for the $\star$ operator in appendix C). So

$$\begin{aligned}\int_{\mathcal{B}_R}\mathcal{H} &= -\frac{1}{2}\int_{\mathcal{B}_R}\tilde{K}(F-\star D\phi,F-\star D\phi)vol - \frac{1}{2}\int_{\mathcal{B}_R}d(K(\phi,F_{\mu\nu})\wedge dx^{\mu\nu}) \\ &= -\frac{1}{2}\int_{\mathcal{B}_R}\tilde{K}(F-\star D\phi,F-\star D\phi)vol - \frac{1}{2}\int_{S_R}K(\phi,F_{\mu\nu})\wedge dx^{\mu\nu}\end{aligned}$$

using Stokes theorem for the second equality.

E The first Chern class

Here we review very briefly what is needed about Chern classes in this essay. Chern classes are topological invariants of smooth complex vector bundles living in the de Rham cohomology spaces of the base space.

Let $E \to M$ be a smooth (as opposed to holomorphic) complex vector bundle with connection D. We seek a way of relating the topology of E as a fiber bundle to the topology of M via the differential geometry provided by D. On one hand we have the intrinsic connection 1-form and curvature 2-form but they are forms on E and we want to use the topology of M. On the other hand we have the connection 1-forms and curvature 2-forms defined on M but there are trivialization dependent and not globally defined. However, this latter approach can be fixed up by applying to the curvature forms some operation that suppress the trivialization dependence and so, at the same time, gives a global object. The first Chern form comes from this idea.

Let $\mathcal{U} = (U_\alpha)_{\alpha\in I}$ be an open covering of M such that for all α we have a trivialization of E over U_α and a corresponding curvature form F_α. We define :

$$c_1(E,D)_\alpha := \frac{i}{2\pi}tr(F_\alpha)$$

so $c_1(E,D)_\alpha$ is a (real valued) 2-form on U_α. Let $g_{\alpha\beta}$ be the transition function from U_α to U_β. On $U_\alpha \cap U_\beta$ we have $F_\beta = g_{\alpha\beta} F_\alpha g_{\alpha\beta}^{-1}$ so

$$c_1(E,D)_\beta = \frac{i}{2\pi} tr(F_\beta) = \frac{i}{2\pi} tr(g_{\alpha\beta} F_\alpha g_{\alpha\beta}^{-1}) = \frac{i}{2\pi} tr(F_\alpha) = c_1(E,D)_\alpha$$

so we have a globally defined object $c_1(E,D)$ because our locally defined objects agree on intersections. The same calculation shows that it doesn't depend on the choice of trivializations. The same calculation (again) shows that it is gauge invariant (it doesn't change under gauge transformations). This latter point is a serious hint of the physical significance of this object called the first Chern form of (E,D).

An even more interesting point is that the integral of $c_1(E,D)$ over any 2-cycle of M (*i.e.* any surface without boundary in M) is an integer called a Chern number. This and the gauge invariance make the first Chern form a very good candidate for being an explanation of the mysterious classical quantum numbers, these physical numbers which are quantized before quantization ! Indeed in this essay we show that both the Dirac and 't Hooft-Polyakov magnetic charges can be seen as Chern numbers. Actually the status of these numbers is not so clear because they arise in what is called classical field theory but these fields are introduced because of quantum arguments as explained in the first section and in the second section we see that the precise link between physical field and geometrical fields is given by quantum mechanics. This is developed in the main stream of the essay.

There is another mathematical property of the first Chern form which is useful in this essay. The first Chern form is closed so it defines an element of the de Rham cohomology of M. It turns out that this cohomology class doesn't depend on D, it only depends on E. It means that the projection on $H^2_{DR}(M)$ looses track of exactly the right pieces of information to retain only the link between E and the topology of M. This cohomology class is called the first Chern class of E. Besides being mathematically beautiful and unexpected, this property has physical consequences because a trivial bundle has a zero first Chern class and so all its Chern numbers are zero. As explained in the main stream of the text, this is the central point of the link between the existence of a Dirac magnetic monopole and electric charge quantization.

The theory of Chern classes is rich and beautiful and, as suggested by its name, the first Chern class is not the only kind of Chern class. More details about them can be found in [8] and all the proofs that have been omitted here are in [29].

References

[1] D. I. Olive, "Exact electromagnetic duality,". Prepared for NATO Advanced Study Institute on Strings, Branes and Dualities, Cargese, France, 26 May - 14 Jun 1997.

[2] D. I. Olive, "The quantisation of charges," `hep-th/0104063`.

[3] P. Goddard and D. I. Olive, "Magnetic monopoles in gauge field theories," *Rept. Prog. Phys.* **41** (1978) 1357.

[4] P. M. Sutcliffe, "BPS monopoles," *Int. J. Mod. Phys.* **A12** (1997) 4663–4706, `hep-th/9707009`.

[5] J. Figueroa-O'Farrill, "Electromagnetic duality for children,". http://www.maths.ed.ac.uk/ jmf/English/EDC.html.

[6] J. Baez and J. P. Munian, *Gauge fields, knots and gravity.* World scientific, 1994.

[7] N. Steenrod, *The topology of fiber bundles.* Princeton University Press, 1951.

[8] R. S. Ward and R. O. Wells, *Twistor geometry and field theory.* Cambridge University Press, 1990.

[9] S. Tsou, "Concepts in gauge theory leading to electric-magnetic duality," *Geometric methods for quantum field theory* `hep-th/0006178`.

[10] C.-N. Yang, "Charge quantization, compactness of the gauge group, and flux quantization," *Phys. Rev.* **D1** (1970) 2360.

[11] D. I. Olive, "Spin and electromagnetic duality: An outline," `hep-th/0104062`.

[12] M. Alvarez and D. I. Olive, "The dirac quantization condition for fluxes on four-manifolds," *Commun. Math. Phys.* **210** (2000) 13–28, `hep-th/9906093`.

[13] A. Jaffe and C. Taubes, *Vortices and monopoles. Structure of static gauge theories.* Birkhäuser, 1980.

[14] W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature and Cohomology Vol 1.* Academic Press, 1972.

[15] G. E. Bredon, *Topology and geometry.* Springer, 1993.

[16] M. Atiyah and N. Hitchin, *The geometry and dynamics of magnetic monopoles.* Princeton University Press, 1988.

[17] N. Hitchin, *Monopoles, minimal surfaces and algebraic curves.* Presses de l'université de Montréal, 1987.

[18] N. J. Hitchin, "Monopoles and geodesics," *Commun. Math. Phys.* **83** (1982) 579–602.

[19] C. H. Taubes, "Stability in Yang-Mills theories," *Commun. Math. Phys.* **91** (1983) 235.

[20] C. Montonen and D. I. Olive, "Magnetic monopoles as gauge particles?," *Phys. Lett.* **B72** (1977) 117.

[21] N. S. Manton, "The force between 't Hooft-Polyakov monopoles," *Nucl. Phys.* **B126** (1977) 525.

[22] A. Sen, "Dyon - monopole bound states, selfdual harmonic forms on the multi - monopole moduli space, and $SL(2, Z)$ invariance in string theory," *Phys. Lett.* **B329** (1994) 217–221, `hep-th/9402032`.

[23] P. Goddard and P. Mansfield, "Topological structures in field theories," *Rept. Prog. Phys.* **49** (1986) 725–781.

[24] P. A. Horvathy and J. H. Rawnsley, "Topological charges in monopole theories," *Commun. Math. Phys.* **96** (1984) 497.

[25] E. Corrigan and D. I. Olive, "Color and magnetic monopoles," *Nucl. Phys.* **B110** (1976) 237.

[26] P. Goddard, J. Nuyts, and D. I. Olive, "Gauge theories and magnetic charge," *Nucl. Phys.* **B125** (1977) 1.

[27] H.-M. Chan and S. T. Tsou, "Nonabelian generalization of electric-magnetic duality: A brief review," *Int. J. Mod. Phys.* **A14** (1999) 2139–2172, `hep-th/9904102`.

[28] W. Fulton and J. Harris, *Representation theory.* Springer, 1991.

[29] P. Griffiths and J. Harris, *Principles of algebraic topology.* Wiley-Interscience, 1978.