Journal of Physics A: Mathematical and General

Dirac monopoles and the Hopf map S^3 to S^2

To cite this article: L H Ryder 1980 *J. Phys. A: Math. Gen.* **13** 437

J. Phys. A: Math. Gen., 13 (1980) 437–447. Printed in Great Britain

Dirac monopoles and the Hopf map $S^3 \to S^2$

L H Ryder
Physics Laboratory, University of Kent, Canterbury, UK

Received 7 November 1978, in final form 10 July 1979

Abstract. It is shown that the presence of point magnetic monopoles necessitates a fibre-bundle formulation of electrodynamics. The Hopf 'fibring' of S^3 over a base space S^2 with fibre S^1 yields the Wu–Yang potentials which describe the Dirac monopole. The Gauss–Bonnet–Chern theorem yields the Schwinger quantisation condition.

1. Introduction

Magnetic monopoles come in two varieties; Dirac monopoles and 't Hooft–Polyakov monopoles. The point magnetic charges of Dirac (1931) are singular, in the sense that they are singularities of the electromagnetic field. They may be introduced into, or left out of, the theory of electromagnetism, according to taste. If they are included in the theory, their mass is arbitrary. The only remarkable thing about them—but it is remarkable—is that the product of their magnetic charge with the electric charge of an electron (or other charged particle) is quantised. This may explain the quantisation of electric charge. The quantisation condition comes about by requiring that the wave-function of an electron in the field of a magnetic monopole be single-valued. The magnetic poles of 't Hooft (1974) and Polyakov (1976) could hardly be more different. They exist as particular solutions to the SU(2) gauge field equations in the presence of spontaneous symmetry breaking (Higgs field). If this gauge theory is right, 't Hooft–Polyakov monopoles *inevitably* exist. Moreover, their mass is not arbitrary and they are a finite size. The gauge field and the Higgs field, both carrying *electric charge only*, arrange themselves in a particular way so that when viewed from infinity they carry magnetic charge. It was quickly realised by Arafune *et al* (1975) that the origin of this charge is topological; that is to say that the boundary conditions on the fields are ones which cannot be changed continuously into constant values. The asymptotic field configuration is topologically non-trivial, and this gives rise to the quantised magnetic charge. The Dirac monopole, on the other hand, is usually considered to have no such topological structure associated with it. The quantisation condition is not regarded as topological in origin, but is the result of the famous 'Dirac veto'.

In this paper it is shown that the Dirac monopole *does* have a topological origin. In Dirac's original paper the vector potential A_μ has a line singularity, and the 'Dirac veto', which leads to the quantisation, is the requirement that the electron wavefunction vanish on this line. In recent years, Wu and Yang (1975) have reformulated Dirac's theory to avoid any singularities in A_μ. This is done by dividing the space surrounding a monopole into two overlapping regions, a and b, and defining A_μ^{a} and A_μ^{b} in each region. A_μ^{a} and A_μ^{b} are both finite in their own region, but are not identical in the

0305-4470/80/020437+11\$01.00

overlapping region; instead, they are related by a gauge transformation, and the condition for this to be single-valued is Dirac's quantisation condition. The mathematical structure of the Wu–Yang theory is that of fibre bundles.

We shall show below that the relevant fibre bundle is S^3, considered as an S^1 fibre over an S^2 base space. The S^1 fibre corresponds to electromagnetic gauge transformations (S^1 is the group space of the gauge group U(1)), and S^2 is the (unit) sphere surrounding the origin, where the monopole sits. Excising the origin from three-dimensional space gives $R^3 - \{0\} = S^2 \times R$, the cartesian product of the two-sphere and the real line (r coordinate). This space has a different topological structure from Euclidean space, and cannot be 'shrunk' to the origin. Locally, $S^3 \approx S^2 \times S^1$, but globally they are distinct, since their (co)homology groups are different.

The simplest mapping of S^3 into S^2 is the Hopf map. It is shown, using this map, that S^3 and $S^2 \times S^1$ are globally distinct, and that the Wu–Yang vector potentials result from projecting specific sections of S^3 onto S^2. Quantisation of the monopole charge is a consequence of the Gauss–Bonnet–Chern theorem in differential geometry, and yields the Schwinger condition $n = 2$. Results similar to these have been obtained by Trautman (1977), though his method differs from the one presented here.

In § 2 the 'conventional' treatment of the Dirac monopole is outlined, and in § 3 the Wu–Yang treatment. Section 4 is devoted to topology, differential geometry and the Hopf map. In § 5 it is shown how the Hopf map yields the Wu–Yang potentials, and the Gauss–Bonnet–Chern theorem the charge quantisation. Concluding remarks follow in § 6.

2. The Dirac monopole

We sketch the outline of the customary approach to magnetic monopoles, inaugurated by Dirac (1931). We have to derive two results; the quantisation condition, and an expression for the vector potential A_μ. We shall derive them in that order.

Consider a monopole of strength g at the origin. The magnetic field is

$$\boldsymbol{B} = (g/r^3)\boldsymbol{r} = -g\boldsymbol{\nabla}(1/r). \tag{2.1}$$

Since $\nabla^2(1/r) = -4\pi\delta^3 r$ we have

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 4\pi g\delta^3 r \tag{2.2}$$

corresponding to a point charge, as desired. Since $\boldsymbol{B}$ is radial, the total flux through a sphere surrounding the origin is

$$\Phi = 4\pi r^2 B = 4\pi g. \tag{2.3}$$

Consider a particle of electric charge e in the field of this monopole. Its wavefunction is

$$\psi = |\psi| \exp \frac{\mathrm{i}}{\hbar}(\boldsymbol{p} \cdot \boldsymbol{r} - Et).$$

In the presence of an electromagnetic field, $\boldsymbol{p} \to \boldsymbol{p} - (e/c)\boldsymbol{A}$, so

$$\psi \to \psi \exp[-(\mathrm{i}e/\hbar c)\boldsymbol{A} \cdot \boldsymbol{r}],$$

or the phase α changes by

$$\alpha \to \alpha - (e/\hbar c)\boldsymbol{A} \cdot \boldsymbol{r}.$$

Consider a closed path at fixed r, θ, with ϕ ranging from 0 to 2π. The total change in phase is

$$\Delta\alpha = \frac{e}{\hbar c}\oint \boldsymbol{A}\,.\,\mathrm{d}s = \frac{e}{\hbar c}\int \operatorname{curl}\boldsymbol{A}\,.\,\mathrm{d}\boldsymbol{\Sigma} = \frac{e}{\hbar c}\int \boldsymbol{B}\,.\,\mathrm{d}\boldsymbol{\Sigma}$$

$$= (e/\hbar c)(\text{Flux through cap}) = (e/\hbar c)\Phi(r,\theta). \tag{2.4}$$

$\Phi(r, \theta)$ is the flux through the cap defined by a particular r and θ, as shown by the shaded area in figure 1. As θ is varied, the flux through the cap varies. As $\theta \to 0$, the loop shrinks to a point and the flux passing through the cap approaches zero:

$$\Phi(r, 0) = 0.$$

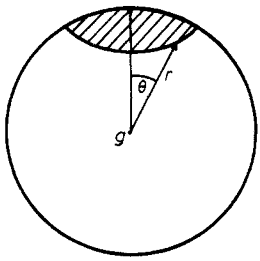

Figure 1. The shaded area of the sphere is the cap defined by a particular r and θ.

As the loop is lowered over the sphere, the cap encloses more and more flux, until, eventually, at $\theta = \pi$ we should have, from equation (2.3),

$$\Phi(r, \pi) = 4\pi g. \tag{2.5}$$

However, as $\theta \to \pi$, the loop has again shrunk to a point so the requirement that $\Phi(r, \pi)$ be finite entails, from equation (2.4), that $\boldsymbol{A}$ be singular at $\theta = \pi$. Since this argument holds for all spheres of all possible radii, $\boldsymbol{A}$ is then singular along the entire negative z axis. This is known as the Dirac string. It is clear that by a suitable choice of coordinates, the string may be chosen to be along any direction, and in fact need not be straight, but must be continuous.

The singularity in $\boldsymbol{A}$ gives rise to the Dirac veto—that the wavefunction should vanish along the negative z axis. Its phase is therefore indeterminate there and, referring to equation (2.4), there is *no* necessity that as $\theta \to \pi$, $\Delta\alpha \to 0$. We must have $\Delta\alpha = 2\pi n$, however, for ψ to be single-valued. Equations (2.4) and (2.5) then give

$$eg = \tfrac{1}{2} n\hbar c, \tag{2.6}$$

which is the Dirac quantisation condition.

We come now to an expression for A_μ. As seen above, it is singular. This much is clear from equation (2.2), for if $\boldsymbol{B} = \operatorname{curl}\boldsymbol{A}$ and $\boldsymbol{A}$ is regular, div $\boldsymbol{B} = 0$, and no magnetic charges may exist. From the argument above, $\boldsymbol{A}$ is constructed by considering the pole as the end point of a string of magnetic dipoles (Dirac 1931) whose other end is at infinity. This gives (Wentzel 1966)

$$A_x = g\frac{-y}{r(r+z)}, \qquad A_y = g\frac{x}{r(r+z)}, \qquad A_z = 0 \tag{2.7}$$

or

$$A_r = A_\theta = 0, \qquad A_\phi = g(1-\cos\theta)/r\sin\theta. \tag{2.8}$$

$\boldsymbol{A}$ is clearly singular along $r=-z$. If, on the other hand, the Dirac string were chosen to be along $r=z$, we should have

$$A_r = A_\theta = 0, \qquad A_\phi = -g(1+\cos\theta)/r\sin\theta. \tag{2.9}$$

3. Wu-Yang treatment of the Dirac monopole

In recent years Wu and Yang (1975) have recast the theory of the Dirac monopole into a form which avoids the use of a singular vector potential. We here give a bare outline of their idea. The space surrounding the monopole—the sphere, essentially—is divided into two overlapping regions R_a and R_b. R_a excludes the negative z axis (S pole) and R_b excludes the positive z axis (N pole). In each region $\boldsymbol{A}$ is defined differently:

$$A_r^a = A_\theta^a = 0, \qquad A_\phi^a = g(1-\cos\theta)/r\sin\theta, \tag{3.1}$$

$$A_r^b = A_\theta^b = 0, \qquad A_\phi^b = -g(1+\cos\theta)/r\sin\theta. \tag{3.2}$$

By comparison with equations (2.8) and (2.9) it is clear that $\boldsymbol{A}^a$ and $\boldsymbol{A}^b$ are both finite in their own domain. In the region of overlap, however, they are not the same, but they are related by a gauge transformation

$$A_\mu^b = A_\mu^a - \frac{i\hbar c}{e} S \frac{\partial S^{-1}}{\partial x^\mu} \tag{3.3}$$

where $S = S_{ab}$ has to be a single-valued function. With A_ϕ^a and A_ϕ^b given as above, it is clear that

$$S = \exp \frac{2ige}{\hbar c}\phi \tag{3.4}$$

satisfies equation (3.3), and the requirement that S be single-valued yields the quantisation condition (2.6). To check that equations (3.1) and (3.2) do really represent a magnetic monopole, we calculate the total magnetic flux through a sphere surrounding the origin:

$$\Phi = \int F_{\mu\nu}\, dx^{\mu\nu} = \oint \operatorname{curl} \boldsymbol{A} \cdot d\boldsymbol{S} = \int_a \operatorname{curl} \boldsymbol{A} \cdot d\boldsymbol{S} + \int_b \operatorname{curl} \boldsymbol{A} \cdot d\boldsymbol{S}$$

Now, unlike the complete sphere, the regions a and b have boundaries which may be chosen to be the equator $\theta = \pi/2$. Stokes' theorem is then applicable, and since the equator bounds a in a positive orientation and b in a negative one, we have

$$\begin{aligned} \Phi &= \oint_{\theta=\pi/2} \boldsymbol{A}^a \cdot d\boldsymbol{l}^a - \oint_{\theta=\pi/2} \boldsymbol{A}^b \cdot d\boldsymbol{l}^b \\ &= \frac{i\hbar c}{e} \oint \frac{d}{d\phi}(\ln S^{-1})\, d\phi \\ &= 4\pi g \end{aligned} \tag{3.5}$$

in agreement with equation (2.5), where equations (3.3) and (3.4) have been invoked. The Wu-Yang potentials then define a magnetic monopole of strength g.

4. The Hopf map

Since the spaces we are concerned with are spherical rather than Euclidean, theorems of integral calculus such as Stokes' theorem have to be generalised. This involves introducing a number of topological notions, which though hardly the stock-in-trade of physicists, are not so unfamiliar as they once were. We shall only give the briefest account here—for further information the reader should consult Choquet–Bruhat (1968), Flanders (1963), Pollard (1977) or von Westenholz (1978). A review for physicists appears in Misner and Wheeler (1957).

The first object of concern is the nth homology group $H_n(X)$ of a space X. It is the quotient group of the group of n-cycles Z_n divided by the group of n-boundaries B_n:

$$H_n(X) = Z_n(X)/B_n(X).$$

Loosely speaking, the number of generators of (say) the first homology group $H_1(X)$ of X is the number of inequivalent (non-homologous) closed curves in X which are not boundaries of pieces of area in X. For example, for the two-sphere S^2,

$$H_1(S^2) = 0,$$

since every closed curve on S^2 bounds a piece of area of S (see figure 2(a)). For the torus T^2, however,

$$H_1(T^2) = Z \times Z;$$

where Z is the group of integers under addition. In other words, $H_1(T^2)$ has two generators, since there are two types of closed curve on T^2 which do not bound areas on T^2, namely the major and minor circumferences c_2 and c_3 in figure 2(b). The general results of concern to us are

$$H_n(S^n) = Z, \qquad (4.1)$$

$$H_i(S^n) = 0 \qquad (0 < i < n), \qquad (4.2)$$

$$H_1(S^2 \times S^1) = Z. \qquad (4.3)$$

We next turn to differential forms. An n-form ω_n integrated over an n-chain c_n is a number

$$\int_{c_n} \omega_n \equiv \int_{c_n} f_{i_1 \ldots i_n}\, dx_{i_1} \wedge dx_{i_2} \wedge \ldots \wedge dx_{i_n} = \text{number}. \qquad (4.4)$$

A form ω_n is *closed* if $d\omega_n = 0$, and *exact* if $\omega_n = d\omega_{n-1}$. In R^n all closed forms are exact, but in a general space this is not true. The nth cohomology group $H^n(X)$ of a

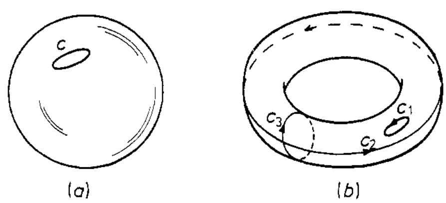

Figure 2. (a) Every closed curve c on the sphere S^2 encloses a piece of area of S^2. (b) The three closed curves c_1, c_2 and c_3 on the torus T^2 are non-homologous. c_1 encloses a piece of area of T^2, but c_2 and c_3 do not.

space X is the quotient group of closed n-forms divided by exact n-forms. The cohomology groups of a space are, by virtue of the duality expressed by equation (4.4), dual to the homology groups. It then follows from equation (4.1) that the second cohomology of the two-sphere, $H^2(S^2)$, is non-trivial, and hence that a closed 2-form ω_2 on S^2 is not necessarily exact.

$$\text{On } S^2, \qquad d\omega_2 = 0 \not\Rightarrow \omega_2 = d\omega_1. \tag{4.5}$$

In component language, this means that $\operatorname{div} \boldsymbol{B} = 0$ does not imply that $\boldsymbol{B} = \operatorname{curl} \boldsymbol{A}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are vector fields defined on S^2.

Finally, we have the generalised Stokes' theorem, according to which if ω is a p-form and c a $(p+1)$-chain with boundary ∂c,

$$\int_{\partial c} \omega = \int_c d\omega. \tag{4.6}$$

Putting $p = 1$ gives Stokes' theorem, and putting $p = 2$ gives the divergence theorem.

We come now to the Hopf map, which maps S^3 onto S^2. S^3 may be parametrised by x_1, x_2, x_3 and x_4 (coordinates in R^4) obeying

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1. \tag{4.7}$$

Putting $z_0 = x_1 + ix_2$, $z_1 = x_3 + ix_4$, this becomes

$$|z_0|^2 + |z_1|^2 = 1. \tag{4.8}$$

S^2 is parametrised by ξ_1, ξ_2 and ξ_3 (coordinates in R^3) obeying

$$\xi_1^2 + \xi_2^2 + \xi_3^2 = 1. \tag{4.9}$$

The Hopf map f is given by (Hopf 1931)

$$f: \quad \xi_1 = 2(x_1x_3 + x_2x_4), \qquad \xi_2 = 2(x_2x_3 - x_1x_4), \qquad \xi_3 = x_1^2 + x_2^2 - x_3^2 - x_4^2 \tag{4.10}$$

since, as may be verified, this implies that

$$\xi_1^2 + \xi_2^2 + \xi_3^2 = (x_1^2 + x_2^2 + x_3^2 - x_4^2)^2, \tag{4.11}$$

in agreement with equations (4.7) and (4.9).

The two-sphere may alternatively be parametrised by the coordinates on the equatorial plane by stereographic projection, as in figure 3. Denoting the two planar coordinates by the complex coordinate $z(= x + iy)$, we have, from the geometry of figure 3, and equation (4.10),

$$z = \frac{\xi_1 + i\xi_2}{1 - \xi_3} = \frac{x_1 + ix_2}{x_3 + ix_4} = \frac{z_0}{z_1}. \tag{4.12}$$

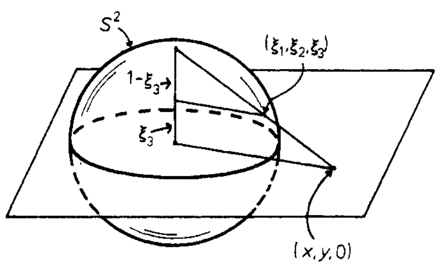

Figure 3. Stereographic projection of the two-sphere $\xi_1^2 + \xi_2^2 + \xi_3^2 = 1$ onto the plane (x, y).

Now z_0 and z_1 are not uniquely determined by z, for if (z_0, z_1) is replaced by $(\lambda z_0, \lambda z_1)$ with

$$|\lambda z_0|^2 + |\lambda z_1|^2 = |\lambda|^2 = 1, \tag{4.13}$$

z is unchanged. λ is then of the form $e^{i\alpha}$, and generates S^1, so *locally* $S^3 \approx S^2 \times S^1$. This is not true globally, however, since the first homology groups of the spaces are, from equations (4.2) and (4.3),

$$H_1(S^2 \times S^1) \approx Z, \qquad H_1(S^3) = 0. \tag{4.14}$$

From equation (4.13) we see that the Hopf map maps a 1-cycle (S^1) in S^3 onto a point in S^2; this is illustrated in figure 4. S^3 is a fibre bundle with base space S^2 and fibre S^1.

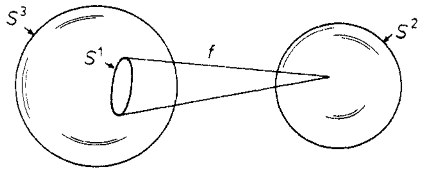

Figure 4. The Hopf map f maps $S^1 \subset S^3$ into a point in S^2.

We now make the observation that S^3 is the group space of the group SU(2), since this is the group of 2×2 matrices with the property

$$U = u_0 + i \sum_{i=1}^{3} u_i \sigma_i, \qquad u_0^2 + u_1^2 + u_2^2 + u_3^2 = 1.$$

S^3 may therefore be parameterised by the Euler angles $\psi\theta\phi$. The SU(2) element corresponding to an arbitrary rotation is

$$\begin{aligned} U &= \exp(i\sigma_3\phi/2)\exp(i\sigma_1\theta/2)\exp(i\sigma_3\psi/2) \\ &= \begin{pmatrix} \cos\theta/2 \exp i(\psi+\phi)/2 & i\sin\theta/2 \exp i(\phi-\psi)/2 \\ i\sin\theta/2 \exp i(\psi-\phi)/2 & \cos\theta/2 \exp -i(\phi+\psi)/2 \end{pmatrix} \\ &= \begin{pmatrix} z_0 & iz_1^* \\ iz_1 & z_0^* \end{pmatrix}. \end{aligned} \tag{4.15}$$

Putting $z_0 = x_1 + ix_2$, $z_1 = x_3 + ix_4$ then gives

$$\begin{aligned} x_1 &= \cos\left(\frac{\phi+\psi}{2}\right)\cos\frac{\theta}{2}, & x_2 &= \sin\left(\frac{\phi+\psi}{2}\right)\cos\frac{\theta}{2}, \\ x_3 &= \cos\left(\frac{\psi-\phi}{2}\right)\sin\frac{\theta}{2}, & x_4 &= \sin\left(\frac{\psi-\phi}{2}\right)\sin\frac{\theta}{2} \end{aligned} \tag{4.16}$$

The Hopf map (4.10) then gives

$$\xi_1 = \cos\phi\sin\theta, \qquad \xi_2 = \sin\phi\sin\theta, \qquad \xi_3 = \cos\theta,$$

so that (ϕ, θ) may be identified with the polar angles on the sphere S^2, and ψ is the angle of the S^1 fibre.

Now let us take 'sections' of S^3 corresponding to the circles $\psi=\phi$ and $\psi=-\phi$. The above equations give

$$\psi=\phi:\quad \left.\begin{aligned} z_0&=e^{i\phi}\cos\theta/2\\ z_1&=\sin\theta/2\end{aligned}\right\}\; z_0=\frac{z}{(1+|z|^2)^{1/2}},\qquad z_1=\frac{1}{(1+|z|^2)^{1/2}} \tag{4.17}$$

$$\psi=-\phi:\quad \left.\begin{aligned} z_0&=\cos\theta/2\\ z_1&=e^{-i\phi}\sin\theta/2\end{aligned}\right\}\; z_0=\frac{1}{(1+|w|^2)^{1/2}},\qquad z_1=\frac{w}{(1+|w|^2)^{1/2}} \tag{4.18}$$

where $w=1/z$, and both are finite. Equation (4.17) maps $S^2-\{\infty\}$ into S^3, that is, referring to figure 3, the projective plane without the point at infinity; in other words, the sphere S^2 minus the N pole. Equation (4.18), on the other hand, maps $S^2-\{0\}$, or the sphere minus the S pole, into S^3. The fact that these maps do not agree in the overlap region is another indication that, globally, S^3 is distinct from $S^2\times S^1$.

Finally, the area of S^2 (of unit radius) is

$$\int_{S^2}\sin\theta\,d\theta\wedge d\phi=4\pi. \tag{4.19}$$

Defining the area 2-form σ_2 by

$$\sigma_2=\sin\theta\,d\theta\wedge d\phi \tag{4.20}$$

we then have

$$\int_{S^2}\sigma_2=4\pi. \tag{4.21}$$

The 2-form σ_2 is closed but not exact ($d\sigma_2=0$; $\sigma_2\neq d\sigma_1$); for if it were exact, we would have, using Stokes' theorem (4.6),

$$\int_{S^2}\sigma_2=\int_{S^2}d\sigma_1=\int_{\partial S^2}\sigma_1=0$$

since $\partial S^2=0$, and this conflicts with equation (4.20). When, however, σ_2 is expressed in S^3 coordinates, it *is* exact, since all closed forms on S^3 are exact, i.e. the second cohomology group of S is trivial $H^2(S^3)=0$, just as the second homology group is $H_2(S^3)=0$ (cf equation (4.2)):

$$\sigma_2=d\sigma_1\qquad(\text{on }S^3). \tag{4.22}$$

This equation plays a key role in determining the magnetic vector potentials, to which problem we now return.

5. Wu–Yang potentials and the Hopf map

We are now in a position to show the relation between monopoles and the Hopf map. The magnetic field of a monopole is described by a 2-form, since when integrated over an area (a 2-chain) it gives a number (we deal in units where flux is dimensionless). From equations (2.1) and (2.3) it is clear that the relevant 2-form is

$$B=g\sigma_2=g\sin\theta\,d\theta\wedge d\phi, \tag{5.1}$$

where σ_2 is given by equation (4.20). The flux is then

$$\Phi = \int_{S^2} B = 4\pi g. \tag{5.2}$$

Just as σ_2 is closed but not exact on S^2, so is B, so there is no global vector potential A on S^2 such that $B = \mathrm{d}A$, or, in components, $\boldsymbol{B} = \operatorname{curl} \boldsymbol{A}$. This is what we have already seen above.

When considered as a 2-form on S^3, however, B is exact (cf equation (4.22)), so a 1-form A exists, with $B = \mathrm{d}A$. The expression for A is

$$A = 2g(x_2\,\mathrm{d}x_1 - x_1\,\mathrm{d}x_2 + x_4\,\mathrm{d}x_3 - x_3\,\mathrm{d}x_4) = -g(\mathrm{d}\psi + \cos\theta\,\mathrm{d}\phi), \tag{5.3}$$

where equation (4.16) has been used. Its exterior derivative is then

$$B = \mathrm{d}A = 4g(\mathrm{d}x_2 \wedge \mathrm{d}x_1 + \mathrm{d}x_4 \wedge \mathrm{d}x_3) = g\sin\theta\,\mathrm{d}\theta \wedge \mathrm{d}\phi, \tag{5.4}$$

as desired from equation (5.1). Now taking sections $\psi = \phi$ and $\psi = -\phi$, we obtain the following expressions for A:

$$\mathrm{a}: \psi = -\phi : A^{\mathrm{a}} = g(1-\cos\theta)\,\mathrm{d}\phi, \qquad \mathrm{b}: \psi = \phi : A^{\mathrm{b}} = -g(1+\cos\theta)\,\mathrm{d}\phi.$$

Putting the 1-form $A = A_\mu\,\mathrm{d}x^\mu$, this gives

$$A^{\mathrm{a}} = \frac{g}{r}\left(\frac{1-\cos\theta}{\sin\theta}\right) r\sin\theta\,\mathrm{d}\phi, \qquad \therefore \quad A^{\mathrm{a}}_\phi = \frac{g}{r}\left(\frac{1-\cos\theta}{\sin\theta}\right),$$

$$A^{\mathrm{b}} = -\frac{g}{\mathrm{r}}\left(\frac{1+\cos\theta}{\sin\theta}\right) r\sin\theta\,\mathrm{d}\phi, \qquad \therefore \quad A^{\mathrm{b}}_\phi = -\frac{g}{r}\left(\frac{1+\cos\theta}{\sin\theta}\right),$$

and $A^{\mathrm{a}}_\theta = A^{\mathrm{b}}_\theta = 0$. It is seen from equations (3.1) and (3.2) that these are precisely the Wu–Yang potentials. These results are summarised schematically in figure 5.

Finally, we come to the geometrical account of quantisation. A remarkable classical theorem in differential geometry, called the Gauss–Bonnet theorem, states that the integral of the Gaussian curvature K over a closed (two-dimensional) surface M is

$$\frac{1}{2\pi}\int_M K\,\mathrm{d}M = \chi(M),$$

where $\chi(M)$, the Euler number of M, is an integer. For the sphere, $\chi = 2$, and for a sphere with p handles, $\chi = 2 - p$. The remarkable property of the Gauss–Bonnet

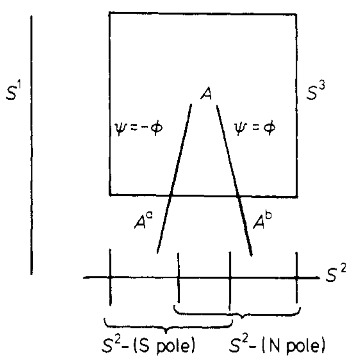

Figure 5. The Wu–Yang potentials derived by taking sections $\psi = \phi$ and $\psi = -\phi$ in S^3.

theorem is that it relates a purely local property of a surface (the Gaussian curvature) to a topological invariant of the surface. This theorem has been generalised to the case of vector fields defined on spheres—more particularly to the case when a 'connection form' is defined. A connection form is the geometric definition of a covariant derivative. The minimal prescription for an electromagnetic field, $\boldsymbol{p} \to \boldsymbol{p} - (e/c)\boldsymbol{A}$ implies that the derivative operator $\boldsymbol{\nabla}$ changes into $\boldsymbol{\nabla} - (\mathrm{i}e/\hbar c)\boldsymbol{A}$, which is called the covariant derivative (for an introduction to the geometrical treatment of gauge fields, see for example Drechsler and Mayer 1977). This identifies the connection form ω as

$$\omega = -(\mathrm{i}e/\hbar c)A.$$

The 'curvature form' Ω is now defined by

$$\Omega = \mathrm{d}\omega - \omega \wedge \omega.$$

It is clear from equation (5.3) that $\omega \wedge \omega = 0$, so

$$\Omega = -(\mathrm{i}e/\hbar c)\,\mathrm{d}A = -(\mathrm{i}e/\hbar c)g \sin\theta \,\mathrm{d}\theta \wedge \mathrm{d}\phi, \tag{5.5}$$

where equation (5.4) has been used.

Now define

$$\det[1 + (\mathrm{i}/2\pi)\Omega] = 1 + \omega_1 + \omega_2 + \ldots + \omega_n.$$

In our case, where Ω is not a matrix, this simply gives

$$\omega_1 = (\mathrm{i}/2\pi)\Omega = (eg/2\pi\hbar c) \sin\theta \,\mathrm{d}\theta \wedge \mathrm{d}\phi.$$

The Gauss–Bonnet–Chern theorem now states that each form ω_i defines a cohomology class whose integral is given by (in our case) (Chern 1972, Baum 1970, Drechsler and Mayer 1977)

$$\int_{S^2} \omega_1 = \chi(S^2) = 2,$$

which gives

$$eg = \hbar c,$$

which is the Dirac condition with $n = 2$. The Gauss–Bonnet–Chern theorem has yielded quantisation but allowed only one value of n. The case $n = 2$ is called the Schwinger condition.

6. Concluding remarks

We have shown that the magnetic monopole is described by the fibre bundle S^3 over the base space S^2. The electromagnetic vector potential is not globally defined over the sphere S^2, and appropriately chosen sections in S^3 yield the Wu–Yang vector potentials which describe the Dirac monopole. Quantisation of magnetic charge follows from the Gauss–Bonnet–Chern theorem, and gives the Schwinger condition. The origin of this quantisation is different from that of solitons, for example the 't Hooft–Polyakov monopole.

Finally, we may expect the instanton (Belavin *et al* 1975) to be described by the analogous Hopf map $S^7 \to S^4$, with fibre S^3. S^4, the base space, is compactified

Euclidean space E^4, and S^3 the group space of SU(2), corresponds to Yang–Mills gauge transformations. This has already been anticipated by Trautman (1977).

Very closely related work, including a treatment of non-abelian gauge fields and instantons, has been done by S Minami (1979a,b).

Acknowledgments

I should like to acknowledge the stimulus for this investigation provided by a seminar given by Professor A Trautman in 1977. I am also particularly grateful to Dr J Merriman for taking a lively interest in this work and for giving me a lot of instruction in differential topology. In addition I should like to thank Miss R Farwell, Dr P W Higgs and Dr B J Vowden for helpful conversations, and Professor G Rickayzen for his comments on the manuscript. I thank Dr S Minami for informing me of his work.

References

Arafune J, Freund P G O and Goebel C J 1975 *J. Math. Phys.* **16** 433
Baum P F 1970 *Bull. Am. Math. Soc.* **76** 1202
Belavin A A, Polyakov A M, Schwartz A S and Tyupkin Yu S 1975 *Phys. Lett.* B **59** 85
Chern S S 1972 *Proc. 13th bienn. Sem. of Can. Math. Congr.* vol I (Montreal: Canadian Mathematical Congress) pp 1–40
Choquet-Bruhat Y 1968 *Géométrie Différentielle et Systèmes Extérieurs* (Dunod)
Dirac P A M 1931 *Proc. R. Soc.* A **133** 60
Dreschsler W and Mayer M E 1977 *Lecture Notes in Physics no. 67* (Berlin: Springer)
Flanders H 1963 *Differential forms with Applications to the Physical Sciences* (New York: Academic)
Hopf H 1931 *Math. Ann.* **104** 637
—— 1964 *Selecta Heinz Hopf* 38 (Berlin: Springer)
Minami S 1979a *Prog. Theor. Phys.* **62**
—— 1979b *Preprint* RIMS-299 Kyoto University
Misner C W and Wheeler J A 1957 *Ann. Phys., NY* **2** 525
Pollard B R 1977 Bristol University lecture notes
Polyakov A M 1976 *Sov. Phys.-JETP* **41** 988
't Hooft G 1974 *Nucl. Phys.* B **79** 276
Trautman A 1977 *Int. J. Theor. Phys.* **16** 561
Wentzel G 1966 *Supp. Prog. Theor. Phys.* **37** and **38** 163
von Westenholz C 1978 *Differential Forms in Mathematical Physics* (Amsterdam: North-Holland)
Wu T T and Yang C N 1975 *Phys. Rev.* D **12** 3845