

Monopoles, gauge fields and de Rham's theorems

Gerald E Marsh

Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA

Received 21 November 1997

Abstract. The topology assumed by most authors for a spacelike hypersurface in a spacetime containing a monopole is generally $\mathbb{R}^3 - \{0\}$; save for the surface \mathbb{S}^2 isolating the monopole, this space is unbounded. For such a topology, a consistency relation of de Rham's theorems shows that a single isolated monopole cannot exist. Monopoles, with charge $\pm m$, if they exist at all, must occur in pairs having opposite magnetic charge. An extension of de Rham's theorems to non-Abelian monopoles which are generalizations of Dirac monopoles (those characterized by $\pi_1(G)$, the fundamental group of the gauge group G) is made using the definition of an ordered integral of a path-dependent curvature over a surface. This integral is similar to that found in the non-Abelian Stokes theorem. The implications of de Rham's theorems for non-Abelian monopoles are shown to be similar to the Abelian case.

1. Introduction

This paper deals with an area that dates back to Dirac's 1948 paper [1] on magnetic monopoles, and which introduced into the literature the concept of the Dirac 'string'. Even as late as 20 years ago, the reality of the Dirac string remained controversial[†]. Wu and Yang [3] sought to eliminate the string associated with a magnetic monopole by dividing spacetime into two or more overlapping regions, and defining a singularity-free vector potential in each region. In the region of overlap, the vector potentials are related by a gauge transformation. Yang [4] likened the problem of finding a singularity-free vector potential for the monopole to that of defining singularity-free coordinates on a globe. This 'patching' approach amounts to defining a non-singular vector potential on a covering of spacetime.

Because there is an equivalence between coverings and a special class of bundles, one may also use the language of fibre bundles to discuss monopoles. However, the topology must be more complex than \mathbb{R}^3 . If the space containing the monopole was \mathbb{R}^3 instead of $\mathbb{R}^3 - \{0\}$, the bundle would be trivial since \mathbb{R}^3 is contractible and a bundle over a contractible base space is trivial. One needs the topology of $\mathbb{R}^3 - \{0\}$ to define the non-trivial bundle needed for the presence of a monopole. The only boundary of $\mathbb{R}^3 - \{0\}$ is \mathbb{S}^2 , the surface isolating the monopole. Because $\mathbb{R}^3 - \{0\}$ has the same topology as \mathbb{S}^2 , Dirac monopoles correspond to non-trivial U1 bundles over \mathbb{S}^2 . This implies that it is not possible to choose a global cross section of the bundle, which is what leads to the necessity of patching the vector potential. The topology of $\mathbb{R}^3 - \{0\}$ also has implications for the physics. For example, the deletion of the origin from Minkowski space means that the resultant spacetime does not have a Cauchy surface [5].

[†] There is interesting literature that deals with the reality of the Dirac string. See, for example [2].

Defining a non-singular vector potential on a covering of spacetime is not very satisfactory because it does not alter the fact that a singularity-free vector potential does not exist on the underlying spacetime itself. Indeed, the existence of such a potential is precluded by the first of de Rham's theorems, a statement of which is contained in appendix A. These theorems are pertinent to the study of monopoles because of the presence of the boundary \mathbb{S}^2 in the space $\mathbb{R}^3 - \{0\}$ containing the monopole. de Rham's theorems imply that if monopoles of charge $\pm m$ exist, a single-valued vector potential can only be defined by treating the Dirac string as a real cut in spacetime. That the Dirac string has a definite physical and observable significance, and is not merely a mathematical device, has also been argued by Barut [2].

The discussion below begins by reformulating the Wu–Yang approach to Abelian magnetic monopoles and applying de Rham's theorems. Since the topological considerations given above also apply to non-Abelian monopoles, the question naturally arises as to whether there is a generalization of de Rham's theorems so as to make them applicable to non-Abelian fields. A possible generalization is given in the second part of the paper. Non-Abelian monopoles use the concept of a principal bundle where the fibre coincides with the structure group. If the structure group (or the base space) is contractible, the bundle will be trivial. The topology of $\mathbb{R}^3 - \{0\}$ is again necessary to have a non-trivial bundle.

What will be shown is that if, as must be done to obtain a non-trivial bundle, one isolates monopoles—Abelian or non-Abelian—by excising their location in spacetime, thereby isolating them by boundary components of the space, one cannot arbitrarily choose monopole types and charges. The implications of the non-Abelian extension of de Rham's theorems are similar to those for the Abelian version. There are some differences, however, which are discussed below.

2. A reformulation of Wu and Yang's approach

The connection between de Rham's theorems and monopoles is best illustrated by the Abelian case. This and the following section are concerned only with Dirac monopoles and serve to bring out the implications of the necessary choice of $\mathbb{R}^3 - \{0\}$ for the space containing a monopole.

Consider the following two *open* sets:

$$\begin{aligned} A &= \mathbb{R}^3 - \{(0, 0) \times (-\infty, 0]\} \\ B &= \mathbb{R}^3 - \{(0, 0) \times [0, \infty)\}. \end{aligned} \tag{1}$$

Set A is all of \mathbb{R}^3 with the negative z -axis and origin deleted, while set B is all of a second copy of \mathbb{R}^3 with the positive z -axis and origin deleted. The union of these two open sets is[†]

$$A \cup B = \mathbb{R}^3 - \{0\}. \tag{2}$$

The monopole is assumed to be located at the origin. It will be useful for what follows to note that

$$A \cap B = [\mathbb{R}^2 - \{0\}] \times \mathbb{R} \tag{3}$$

meaning all of \mathbb{R}^3 with the z -axis deleted, while $A - (A \cap B) \subset A$ and $B - (A \cap B) \subset B$ correspond respectively to the positive and negative z -axes without the origin.

[†] It is interesting to note that $\mathbb{R}^3 - \{0\}$ is also definable as the intersection of two sets A and B defined as follows. Let A be \mathbb{S}^3 with the north pole deleted and B be \mathbb{S}^3 with the south pole deleted. Then $\mathbb{S}^3 = A \cup B$ and $A \cap B$ is \mathbb{S}^3 with both the north and south poles deleted. Under a stereographic projection, $A \cap B \cong \mathbb{R}^3 - \{0\}$.

To obtain a sphere within which most authors usually isolate a monopole, one would delete a region about the origin $r \leq \varepsilon$, and with a slight change in the definition of sets A and B obtain

$$A \cup B = \mathbb{R}^3 - \{r \leq \varepsilon\}. \quad (4)$$

Note that since \mathbb{R}^3 is unbounded,

$$\partial[A \cup B] = -\partial\{r \leq \varepsilon\} = -\{r = \varepsilon\}. \quad (5)$$

This boundary is clearly equivalent to \mathbb{S}^2 , and 2-sphere, with the orientation being given by the inwardly directed normal.

Outside the region containing the monopole, the magnetic field B satisfies $dB = 0$; that is, B is a closed 2-form. There then exist 1-forms A_a and A_b such that

$$\begin{aligned} B &= dA_a & \text{on } A \\ B &= dA_b & \text{on } B. \end{aligned} \quad (6)$$

On $A \cap B$ one has $d(A_a - A_b) = 0$ so that $A_a - A_b = d\chi$, where χ is a 0-form (function) on $A \cap B$. This means that A_a and A_b are related by a gauge transformation.

The aim here is to define a 1-form on the cover $\{A, B\}$ of $\mathbb{R}^3 - \{0\}$ such that B is exact. This can be done by using what is known as a partition of unity: for any open cover of a manifold \mathbb{M} , there is a collection of functions $\varphi_i : \mathbb{M} \rightarrow [0, 1]$ such that

$$\begin{aligned} (1) \quad & \{\text{supp } \varphi_i\} \quad \text{is locally finite} \\ (2) \quad & \sum_i \varphi_i(p) = 1 \quad \forall p \in \mathbb{M}. \end{aligned}$$

The first condition means that each point of \mathbb{M} has a neighbourhood that intersects the support of φ_i for only finitely many values of i .

The cover $\{A, B\}$ of $A \cup B = \mathbb{R}^3 - \{0\}$ then has a partition of unity $\{\varphi_A, \varphi_B\}$ such that

$$\begin{aligned} \varphi_A + \varphi_B &= 1 \\ d\varphi_A + d\varphi_B &= 0 \\ \text{supp } \varphi_A &\subset A \\ \text{supp } \varphi_B &\subset B. \end{aligned} \quad (7)$$

This partition of unity allows the definition of the 0-forms $\varphi_A\chi$ and $\varphi_B\chi$. These have the values

$$\begin{aligned} \varphi_A\chi &= \begin{cases} \varphi_A\chi & \text{on } A \cap B \\ 0 & \text{on } B - (A \cap B) \end{cases} \\ \varphi_B\chi &= \begin{cases} \varphi_B\chi & \text{on } A \cap B \\ 0 & \text{on } A - (A \cap B). \end{cases} \end{aligned} \quad (8)$$

The 0-form $\varphi_A\chi$ is then well defined on set B since it vanishes in the region $B - (A \cap B)$, where χ is not defined; and similarly, the 0-form $\varphi_B\chi$ is well defined on set A .

Consider now the 1-form $A_a - d(\varphi_B\chi)$ defined on the set A . Using equations (7),

$$\begin{aligned} A_a - d(\varphi_B\chi) &= A_a - \varphi_B d\chi - \chi d\varphi_B \\ &= A_a + (\varphi_A - 1) d\chi + \chi d\varphi_A \\ &= A_a - d\chi + d(\varphi_A\chi) \\ &= A_b + d(\varphi_A\chi). \end{aligned} \quad (9)$$

The final result of this last equation is defined on set B . Moreover, again using equations (67), the difference of these two forms defined respectively on A and B is

$$\begin{aligned} A_a - A_b &= d(\varphi_B \chi) + d(\varphi_A \chi) \\ &= \chi(d\varphi_A + d\varphi_B) + d\chi(\varphi_A + \varphi_B) \\ &= d\chi \end{aligned} \tag{10}$$

showing that the relation between them is still a gauge transformation.

Thus, one can define a 1-form on $A \cup B = \mathbb{R}^3 - \{0\}$ by setting it equal to $A_a - d(\varphi_B \chi)$ on the set A , and $A_b + d(\varphi_A \chi)$ on the set B . Since, from equation (9),

$$\begin{aligned} B &= d(A_a - d[\varphi_B \chi]) = dA_a && \text{on } A \\ &= d(A_b + d[\varphi_A \chi]) = dA_b && \text{on } B \end{aligned} \tag{11}$$

B is not only a closed 2-form, it is also exact. This construction of a non-singular vector potential can clearly be generalized to coverings having more than two elements as was done by Wu and Yang.

While this reformulation of the technique used by Wu and Yang allows the construction of a single-valued vector potential for the covering $\{A, B\}$ of $A \cup B = \mathbb{R}^3 - \{0\}$, *this is not equivalent to the definition of a single-valued vector potential in the real spacetime outside the region isolating the monopole*. This possibility is precluded by the first of de Rham's theorems [6, 7].

3. de Rham's theorems and Abelian monopoles

From the discussion preceding equation (6), the boundary isolating the monopole can be considered to be a 2-cycle; since it is the only boundary, in the space outside the region containing the monopole

$$\Phi = \int_{\partial \mathbb{M}} B = \int_{\mathbb{M}} dB = 0 \tag{12}$$

because B is a closed 2-form. So, if $\partial \mathbb{M}$ is just the 2-cycle isolating the monopole, the flux Φ vanishes. On the other hand, if \mathbb{M} has additional boundary components such that $\sum_i a_i z_i = \partial \mathbb{M}$, the z_i being 2-cycles and the a_i constants, then

$$\sum_i a_i \int_{z_i} B = 0. \tag{13}$$

This relation between periods[†] constitutes a consistency relation for de Rham's theorems[‡] and gives an important result.

If the space containing a monopole is assumed to be unbounded, save for the boundary isolating the monopole, with spatial topology equivalent to $\mathbb{R}^3 - \{0\}$, then a single isolated magnetic monopole, with non-zero flux, cannot exist; if additional boundary components exist, and it is further assumed that the magnetic charge of a monopole is $\pm m$, then magnetic monopoles, if they exist, must occur in pairs having opposite magnetic charge.

[†] The integral $\int_{z_i} B$ for each 2-cycle z_i on \mathbb{M} is termed a period.

[‡] de Rham's existence theorems are often restricted to compact spaces (Goldberg [8]), but can be generalized to all dimensions and all differentiable manifolds. See, for example [7], Fenn [8] and for a proof, see Warner [8].

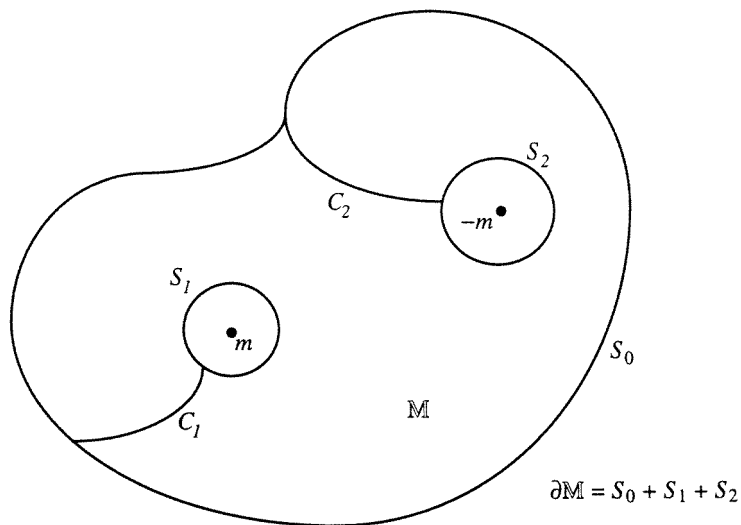


Figure 1. The domain \mathbb{M} is bounded by S_0 and the 2-cycles S_1 and S_2 isolating the monopoles. The open arcs C_1 and C_2 can be taken as cuts to allow the existence of a single-valued vector potential 1-form.

There is a possible caveat: if the charge of a monopole is allowed to take the values $0, \pm 1, \pm 2, \dots, \pm m$, then differently charged monopoles, each isolated by a boundary component, could exist provided consistency is maintained with the limitations imposed by equation (13).

It is the fact that $dB = 0$ which is responsible for the condition given by equation (12) or equation (13). Conversely, if one assumes the existence of a pair of monopoles with magnetic charge $+m$ and $-m$, then

$$\int_{\mathbb{M}} dB = \int_{\partial\mathbb{M}} B = \sum_i \int_{z_i} B = \text{per}(z_1) + \text{per}(z_2) = 0 \tag{14}$$

where $\text{per}(z_i) = 4\pi m_i$.

The condition required for the existence of a globally defined vector potential is given by de Rham's first theorem, which can be written in this case as follows.

I. A closed 2-form B is exact if and only if all its periods $\int_{z_i} B = \text{per}(z_i)$ vanish.

If $\partial\mathbb{M}$ has only one component, since $\Phi = \int_{\partial\mathbb{M}} B = 0$, there exists a vector potential A such that $B = dA$ globally. If $\partial\mathbb{M}$ has more than one component, and all of the periods do not vanish, the second of de Rham's theorems states the following.

II. There exists a closed 2-form B which has the assigned periods $\int_{z_i} B = \text{per}(z_i)$, subject to the consistency relation $\sum_i a_i \text{per}(z_i) = 0$ if $\sum_i a_i z_i$ is a boundary.

When this is the case, there is no globally defined vector potential. On the other hand, if the space is bounded by an additional boundary component S_0 as shown in figure 1, one could obtain a single-valued potential by defining the homology of open arcs modulo the boundary. Assume this is the case.

Two points are said to be homologous if they bound an arc. If these points are on the boundary of \mathbb{M} , then two arcs C and C' , contained in \mathbb{M} , are said to be homologous modulo the boundary, $C \sim C' [\text{mod } \partial\mathbb{M}]$, if their zero-dimensional boundaries are homologous in

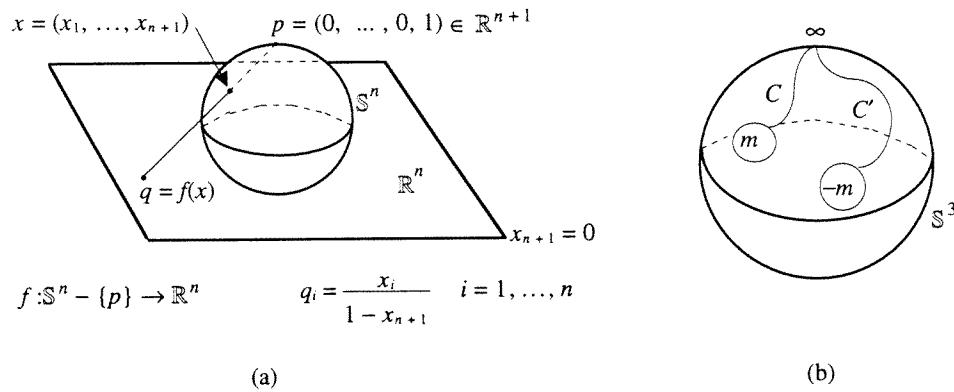


Figure 2. (a) The stereographic projection $f : \mathbb{S}^n - \{p\} \rightarrow \mathbb{R}^n$. (b) A pair of monopoles with Dirac strings C and C' .

$\bar{\mathbb{M}}$. Thus, when $C \sim C' [\text{mod } \partial\mathbb{M}]$ it is possible to complete $C - C'$ to a closed, bounding curve in \mathbb{M} by adding arcs on the boundary.

Consider the situation shown in figure 1 for two monopoles having magnetic charge m and $-m$. The choice of C_1 and C_2 is arbitrary since different choices are homologous modulo the boundary. Taking C_1 and C_2 as cuts makes the domain \mathbb{M} simply connected, so that it is possible to construct a single-valued vector potential in \mathbb{M} . As S_0 recedes to infinity, the cuts C_1 and C_2 correspond to the Dirac strings associated with each monopole.

If the space is without boundary (except for the 2-cycles isolating the monopoles) this procedure, and the Dirac strings, may not be well defined. This can be seen by using a stereographic projection, as shown in figure 2(a). Note that the point p , corresponding to infinity, is excluded since \mathbb{R}^n is homeomorphic to $\mathbb{S}^n - \{p\}$. It is $\mathbb{R}^n \cup \{\infty\}$ that is homeomorphic to \mathbb{S}^n . In figure 2(b) the Dirac strings C and C' are shown going towards infinity. Because $\mathbb{S}^n - \{p\}$ is homeomorphic to \mathbb{R}^n , it is an open (unbounded) space. Consequently, C and C' do not have endpoints, a circumstance that does not necessarily rule out the presence of a monopole. By adding the point at infinity, the Dirac strings can be given endpoints and the space will consequently have the topology of \mathbb{S}^n .

An interesting possibility, for the case where the space is assumed to be unbounded, is shown in figure 3. The 2-form B will not be exact unless $\int_{\partial\mathbb{M}} B = 0$. But this is indeed the case here because the total magnetic charge contained within the boundary is zero, since the cut connects the 2-cycles isolating the monopoles. A single-valued vector potential is then possible for a pair of monopoles provided a cut is made along an arc joining them; i.e. the Dirac string is held in common, a possibility raised by Barut [2].

The above formalism can also be used to compute the total flux of the monopole as well as the Dirac quantization condition. For the sake of completeness, these are given in appendix B.

4. Non-Abelian gauge fields

In non-Abelian gauge theories, one generalizes the 1-form A to a 1-form matrix such that

$$A = A_\mu^a T_a dx^\mu = A_\mu dx^\mu \quad (15)$$

where the T_a are the generators of a given group. In Abelian electrodynamics, the group is $U(1)$.

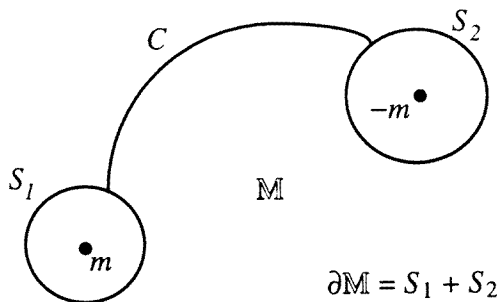


Figure 3. The domain \mathbb{M} is bounded only by the 2-cycles isolating the monopoles. A single-valued vector potential will exist if the open arc C is taken as a cut: the Dirac string is shared by the monopoles.

$U(1)$ is interesting because its group manifold S^1 is multiply connected, implying that the physical charges are some multiple of e . The group $SU(n)$, on the other hand, is simply connected and is the universal covering group of its Lie algebra. Now the connected Lie groups (not necessarily simply connected) which are locally isomorphic to $SU(n)$ are obtained from a covering homomorphism whose kernel is the centre of $SU(n)$ or a subgroup of the centre. Recall that the centre of a continuous semisimple group is discrete and cannot contain more elements than its rank. The centre, being the set of elements that commute with all other group elements must, by Schur's lemma, have the general form ωI_n , ω being a complex number. The condition that elements of $SU(n)$ have unit determinant then implies that $\omega^n = 1$. Thus, the centre of $SU(n)$ is isomorphic to the cyclic group \mathbb{Z}_n , whose elements may be conveniently labelled by the n th roots of unity [9]. The discussion here will involve the multiply connected groups locally isomorphic to $SU(n)$ obtained via the above homomorphism; i.e. the groups of interest are $SU(n)/\mathbb{Z}_n$. Note that if n is prime, one has only $SU(n)/\mathbb{Z}_n$, while if $n = pq$, there are two additional groups $SU(n)/\mathbb{Z}_p$ and $SU(n)/\mathbb{Z}_q$.

The basis for generalizing de Rham's theorems is contained in the literature dealing with the non-Abelian Stokes theorem [10]. There one defines the non-Abelian group element U in terms of the differential equation for parallel transport, with connection A , along the path $x(s)$. Assuming the boundary condition $U(0) = I$, one may rewrite the equation of parallel transport as a Volterra integral equation and use Picard's process of successive approximations to solve the integral equation in terms of the path ordered exponential [11]

$$U = P \exp \int_{x(s)} A. \tag{16}$$

The condition $U(0) = I$ is concerned with what Gantmacher [11] calls the 'normalization' of U . Such a normalization allows the infinite series of matrix integrals representing the solution to the differential equation satisfied by U to be written in the symbolic form $P \exp \int$.

Using an adaptation of the notation of Fishbane *et al* [10] the non-Abelian Stokes theorem, for a surface S with boundary ∂S , and fixed reference point a on ∂S , can be written as

$$P \exp \int_{\partial S} A = \mathcal{P} \exp \int_S U_{ay} G(y) U_{ya} \tag{17}$$

where $G = \frac{1}{2} G_{\mu\nu} dx^\mu dx^\nu$, $G_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu} + [A_\mu, A_\nu]$ or $G = dA + A \wedge A = DA$, and \mathcal{P} is some surface ordering.

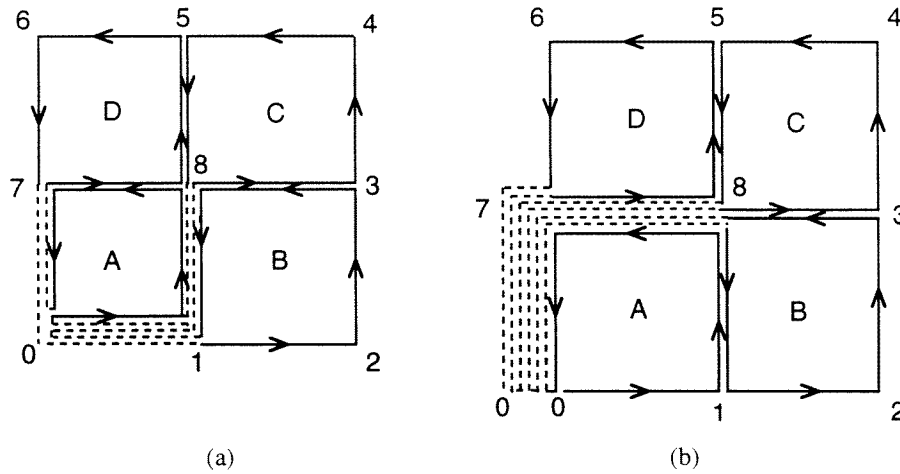


Figure 4. The surface ordering used by (a) Aref'eva and (b) Fishbane *et al.* The boundary of the surface is distorted so as to divide the surface into elementary areas without the boundary crossing itself.

Various approaches to surface ordering appear in the literature, each consistent with distorting the boundary of S so as to cover S with oriented elementary areas that can, in the limit, be made infinitesimal. In forming the elementary areas, one must also include the paths that lead from the reference point to the elementary area and return to the reference point in such a way that none of the paths cross; i.e. the distortion of the boundary must be such as to break up the surface into elementary areas without the boundary crossing itself. The ordering is therefore important.

Figures 4(a) and (b) show, respectively, the surface ordering used by Aref'eva [10] and Fishbane *et al* [10]. In each of these examples, $P \exp \int_{\partial S} A$ can be written as

$$P \exp \int_{\partial S} A = U_{01} U_{12} U_{23} U_{34} U_{45} U_{56} U_{67} U_{70}. \quad (18)$$

For the surface ordering used by Fishbane *et al* the right-hand side of equation (18) can be expressed as

$$[U_{01} U_{18} U_{87} U_{70}]_A | U_{07} U_{78} [U_{81} U_{12} U_{23} U_{38}]_B U_{87} U_{70} | U_{07} U_{78} \\ \times [U_{83} U_{34} U_{45} U_{58}]_C U_{87} U_{70} | U_{07} [U_{78} U_{85} U_{56} U_{67}]_D U_{70}$$

where the terms in square brackets correspond to the loops around the elementary squares A , B , C and D of figure 4 (full curves); the vertical bars separate the four loops and their associated U terms that lead from the reference point 0 to and from the loop (broken curves); and the terms appearing in the right-hand side of equation (18) are in bold. Note that the terms between the bold terms are equivalent to the identity, and that the terms between the vertical bars can be reduced to the form $U_{0n}[\dots]U_{n0}$. Similarly, for the example used by Aref'eva, one has

$$U_{01} [U_{12} U_{23} U_{38} U_{81}]_B U_{10} | U_{01} U_{18} [U_{83} U_{34} U_{45} U_{58}]_C U_{81} U_{10} | \\ \times [U_{01} U_{18} U_{87} U_{70}]_A | U_{07} [U_{78} U_{85} U_{56} U_{67}]_D U_{70}.$$

Keeping in mind that the $G(y)$ of equation (17) has an infinitesimal area—corresponding to the elementary squares of figure 4—associated with it, each bracketed term may be

evaluated as $[1 + G(y)]_X$, where $X = A, B, C$ or D . An elementary example of this derivation is given by Moriyasu [12]. Looking only at first-order terms, the product becomes a sum of terms of the form $U_{0y}[1 + G(y)]U_{y0}$, and in the limit, one obtains the right-hand side of equation (17).

In what follows, whether a path or surface ordering is appropriate should be clear from the context and distinguishing ordering symbols will not be used. The integrals depend on ∂S but not on the parametrization of S , i.e. not on the partition of S into infinitesimal components and their ordering. The U factors on the right-hand side of equation (17) are needed in the integrand to ensure that the expression is gauge invariant, and that the surface integral is independent of the choice of surface (see Bralić [10] and Simonov [13]).

That equation (17) is indeed gauge covariant can be seen as follows. Under a gauge transformation, given the definition of U in equation (16), the left-hand side of equation (17) becomes $U' = g^{-1}(a)Ug(a)$; and the right-hand side becomes

$$\begin{aligned} P \exp \int_S U'_{ay} G'(y) U'_{ya} &= P \exp \int_S g^{-1}(a) U_{ay} G(y) U_{ya} g(a) \\ &= g^{-1}(a) \left[P \exp \int_S U_{ay} G(y) U_{ya} \right] g(a) \end{aligned}$$

where $U'_{ay} = g^{-1}(a)U_{ay}g(y)$, $U'_{ya} = g^{-1}(y)U_{ya}g(a)$, $G(y)' = g^{-1}(y)G(y)g(y)$, and the last equality is due to $g(a)$ being a constant matrix. The constant matrices can now be eliminated from both sides of the equation, showing that equation (17) is indeed gauge invariant and independent of the reference point a . The U_{ay} and U_{ya} in equation (17) satisfy the two relations

$$dU_{ay} = -U_{ay}A(y) \quad \text{and} \quad dU_{ya} = A(y)U_{ya} \tag{19}$$

as well as $U_{ay}U_{ya} = I$.

The integrand $U_{ay}G(y)U_{ya}$ in equation (17) can be thought of as a 2-form matrix derived from the 1-form matrix $U_{ay}A(y)U_{ya}$, that is,

$$d(U_{ay}A(y)U_{ya}) = U_{ay}DA(y)U_{ya} = U_{ay}G(y)U_{ya}. \tag{20}$$

Similarly,

$$dd(U_{ay}A(y)U_{ya}) = d(U_{ay}G(y)U_{ya}) = U_{ay}DG(y)U_{ya} = 0 \tag{21}$$

where $DG(y) = 0$ because of the Bianchi identity. It is interesting to note that here the Bianchi identity guarantees the validity of the Poincaré lemma†.

One may then write

$$dA = \mathcal{G} \quad d\mathcal{G} = ddA = 0 \tag{22}$$

where $A(y) = U_{ay}A(y)U_{ya}$ and $\mathcal{G}(y) = U_{ay}G(y)U_{ya}$. Under a gauge transformation g , where $A' = g^{-1}Ag + g^{-1}dg$ and $U'_{ya} = g^{-1}(y)U_{ya}g(a)$, it is readily shown that

$$\begin{aligned} (dA)' &= g^{-1}(a) dA g(a) \\ \mathcal{G}(y)' &= g^{-1}(a)\mathcal{G}(y)g(a) \\ (d\mathcal{G}(y))' &= g^{-1}(a) d\mathcal{G}(y)g(a). \end{aligned} \tag{23}$$

† Some authors take the Poincaré lemma to be the statement that in a contractible space all closed forms are exact. Others take it to simply mean that two applications of the exterior derivative always yield zero. It is the latter meaning that is intended with regard to equation (21). The Bianchi identity fails at the location of a monopole, but will be valid so long as any monopoles present are excluded from the space by being isolated by boundary components.

These quantities are then gauge invariant except for a position-independent gauge rotation at the reference point a . The expression $d\mathcal{A} = \mathcal{G}$ is itself fully gauge invariant and independent of the reference point.

From equation (22) it can be seen that \mathcal{G} is a closed 2-form matrix which is exact if there exists a 1-form matrix \mathcal{A} such that $d\mathcal{A} = \mathcal{G}$. A non-Abelian de Rham's theorem would give the conditions needed for the *global* existence of \mathcal{A} . In order to state such a theorem, however, one must first define the concept of a non-Abelian period for the 2-form matrix \mathcal{G} . For this one needs an expression representing the non-Abelian flux through a closed surface. In addition, a statement of a non-Abelian form of Gauss' theorem, really a generalized form of Stokes' theorem, is needed. The latter can be written symbolically, for a region \mathbb{M} , as

$$P \exp \int_{\partial\mathbb{M}} \mathcal{G} = P \exp \int_{\mathbb{M}} d\mathcal{G} \quad (24)$$

where the exact meaning of these integrals has been discussed in detail by both Fishbane *et al* [10] and Simonov [13]. Since \mathcal{G} is a closed 2-form matrix, the integrand of the right-hand side of equation (24) vanishes unless the Bianchi identity is violated due to the presence of a magnetic monopole in \mathbb{M} . Note that, by equations (23), equation (24) is gauge invariant.

In their seminal paper [3], Wu and Yang made the point that for non-Abelian fields it is not possible to define the magnetic flux (in the usual sense) through a loop since the divergence of \mathbf{H} does not usually vanish for such fields. This does not, however, mean that Stokes' theorem or Gauss' theorem are ill defined. While it is true that $\mathcal{G}(y) = U_{ay}G(y)U_{ya}$ depends on the choice of path used in calculating U_{ay} , so that $\mathcal{G}(y)$ does not have a unique value at a point, the integrals of this quantity over a surface are well defined in the following sense. In the case of Stokes' theorem, equation (17), the integral over a surface bears a fixed relationship to the integral of the gauge potential over the boundary of the surface[†]. As noted before, this relationship is gauge invariant and independent of the choice of surface or its parametrization. In the case of Gauss' theorem, if the integrand of the right-hand side of equation (24) vanishes because of the Bianchi identity, then the left-hand side will have the gauge invariant value of I . Integrals of the form $P \exp \int_{\Sigma} \mathcal{G}$ represent the total non-Abelian flux through the closed surface Σ . However, it can be seen from equations (23) that the value of such integrals is indeterminate up to a position-independent gauge rotation at the reference point a .

It is important, for what follows, to note that the term 'non-Abelian monopole' is used in the sense of Tse and Ezawa [14], Wu and Yang [3], and Chan and Tsou [15], as distinct from the SO(3) solitons carrying magnetic charge of 't Hooft [16] and Polyakov [17]. That is, closed curves on the manifold of the gauge group G fall into homotopy classes $\pi_1(G)$, or gauge types, corresponding to a topological charge. If the gauge-group manifold is simply connected (this will be the case if $\pi_1(G) = 0$), there are no monopoles; if G is compact and semisimple, there are a finite number of monopole types; and if $G = U(n)$, there are infinitely many monopole types. The group manifold of SU(2), for example, is simply connected and has no monopoles, while $SU(2)/\mathbb{Z}_2 \cong SO(3)$ is not simply connected and has two gauge types, one corresponding to the presence of a monopole, and the other to the vacuum.

[†] Note added in proof. The term 'integral' is used here to mean $P \exp \int_S$ rather than just the integral in the exponential. Hirayama and Matsubara (Hirayama M and Matsubara S 1998 *Prog. Theor. Phys.* **99** 691) have shown that the value of the integral in the exponential *does* vary under deformations of S with ∂S fixed, but the quantity $P \exp \int_S$ nevertheless remains constant under such deformations provided the Bianchi identity is satisfied.

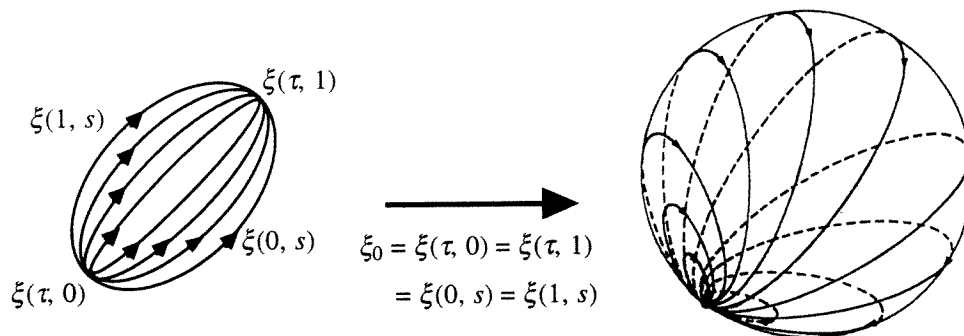


Figure 5. The family of paths (a) used by Bralić in the non-Abelian Stokes theorem; and (b) the corresponding family of loops enveloping a closed two-dimensional surface in spacetime obtained by identifying the points $\xi(\tau, 0)$ and $\xi(\tau, 1)$ and requiring $\xi(\tau, s)$ to smoothly decrease in length as τ approaches 0 or 1 so that the loops $\xi(0, s)$ and $\xi(1, s)$ are of zero length, and the points $\xi(\tau, 0) = \xi(\tau, 1) = \xi(0, s) = \xi(1, s)$ correspond to the base point ξ_0^μ .

When a monopole is present, an expression for the non-Abelian flux through a closed surface is most easily defined by using the parametrized loop space introduced by Polyakov [18] and used by Chan and Tsou [15]. They define a one-parameter family of loops which envelope a closed surface Σ . The loops, each of which is parametrized along its length by s , are defined as†

$$\Sigma : \xi_t^\mu(s) \quad 0 \leq s \leq 2\pi, 0 \leq t \leq 2\pi \tag{25}$$

with

$$\begin{aligned} \xi_t^\mu(0) = \xi_t^\mu(2\pi) = \xi_0^\mu & \quad 0 \leq t \leq 2\pi \\ \xi_0^\mu(s) = \xi_{2\pi}^\mu(s) = \xi_0^\mu & \quad 0 \leq s \leq 2\pi \end{aligned} \tag{26}$$

where ξ_0^μ corresponds to the base point a above. When the parameter t takes the value 0 or 2π , the loop is the base point ξ_0^μ , and as t varies continuously from 0 to 2π , the loops cover a closed two-dimensional spacelike surface in spacetime. Since the base point also corresponds to the identity element in group space, as t varies continuously from 0 to 2π , the phase factors U_{ξ_t} associated with each loop trace out a path in the group manifold beginning and ending with the identity. This is what Yang and Wu called the ‘total circuit’ for the gauge field, and was used by them to replace the electromagnetic concept of the total magnetic flux through a closed surface surrounding a given point.

The loop parametrization of the surface Σ is similar to that used by Bralić for a sheet (see figure 5). There one has the family of paths $\xi(\tau, s)$ with τ being the path label and s the parameter along the path. The loop parametrization is obtained from the sheet parametrization used by Bralić by identifying the points $\xi(\tau, 0)$ and $\xi(\tau, 1)$ and requiring $\xi(\tau, s)$ to smoothly decrease in length as τ approaches 0 or 1 so that $\xi(0, s) = \xi(1, s)$ are loops of zero length corresponding to the base point $\xi_0 = \xi(\tau, 0) = \xi(\tau, 1)$. This process is equivalent to the mapping used by Goddard *et al* [20] where the unit square in the (s, t)

† At an arbitrary base point x_0 in a space Y , one can define the loop space ΩY and show that $\pi_n(Y) \cong \pi_{n-1}(\Omega Y)$. The parametrized set of loops defined in equation (26) then corresponds to the mapping used to define $\pi_1(\Omega Y)$. The isomorphism with $\pi_2(Y)$ implies that $\pi_1(\Omega Y)$ is topologically equivalent to the sphere \mathbb{S}^2 since $\pi_n(Y, x_0)$ is defined as the homotopy classes of the map that takes the boundary of the n -cube to x_0 and the interior into Y . See, for example [19].

plane is mapped onto the sphere, with the boundary of the square corresponding to a fixed point on the sphere.

If one chooses an arbitrary point O , a family of closed parametrized curves with O as origin, and a set of maps that associate with each curve an affine transformation of the tangent space at O , these transformations form a group known as the holonomy group at O . An expression for the loop space holonomy characterizing the non-Abelian flux through a closed surface may then be written, in terms of the parametrized loop space, as†

$$\Theta(\Sigma) = P_t \exp \int_0^{2\pi} dt \int_0^{2\pi} ds U_{\xi_t}^{-1}(s, 0) G_{\mu\nu}[\xi_t(s)] U_{\xi_t}(s, 0) \frac{d\xi_t^\mu(s)}{dt} \frac{d\xi_t^\nu(s)}{ds} \quad (27)$$

where

$$U_{\xi_t}(s, 0) = P_s \exp \int_0^s ds A_\mu[\xi_t(s)] \frac{d\xi_t^\mu(s)}{ds}. \quad (28)$$

In the more compact symbolic notation used earlier, $\Theta(\Sigma)$ may be written as

$$\Theta(\Sigma) = P_t \exp \int_\Sigma \mathcal{G}. \quad (29)$$

Note that the left-hand side of equation (24) is identical to the right-hand side of equation (29) with $\partial\mathbb{M}$ identified with Σ . The value of the integral in equation (27) does not depend on the parametrization of the loops. Bralić used $0 \leq s, \tau \leq 1$ whereas τ is replaced here by t and the range is changed to $0 \leq s, t \leq 2\pi$ to be consistent with Chan and Tsou.

With the identification given above between Polyakov's loop parametrization and the sheet parametrization used by Bralić, Chan and Tsou's $\Theta(\Sigma)$ is formally equivalent to the sheet variable $S(\xi)$ defined by Bralić. This can be seen by using the multiplicative derivative introduced by Gantmacher,

$$D_t X = \frac{dX}{dt} X^{-1}. \quad (30)$$

Applied to $\Theta(\Sigma)$ this gives

$$\begin{aligned} D_t \Theta(\Sigma) &= \frac{d}{dt} (\Theta(\Sigma)) \Theta(\Sigma)^{-1} \\ &= \int_0^{2\pi} ds U_{\xi_t}^{-1}(s, 0) G_{\mu\nu}[\xi_t(s)] U_{\xi_t}(s, 0) \frac{d\xi_t^\mu(s)}{dt} \frac{d\xi_t^\nu(s)}{ds} =: \mathcal{A}. \end{aligned} \quad (31)$$

A similar expression appears in the review by Goddard and Olive [22] where it is credited to Christ [23] and Goldstone (unpublished 1976 lectures).

Thus,

$$\frac{d}{dt} (\Theta_\Sigma(t)) = \mathcal{A} \Theta_\Sigma(t) \quad (32)$$

so that $\Theta(\Sigma)$ is equivalent to $S(\xi)$ defined by Bralić.

The value of $\Theta(\Sigma)$ depends on the gauge group and whether or not a monopole is contained within Σ . If the gauge group is $SU(n)/\mathbb{Z}_n$, the n different gauge types can be characterized by $\pi_1(SU(n)/\mathbb{Z}_n) = \mathbb{Z}_n$. As discussed earlier, \mathbb{Z}_n is the cyclic group of order

† The expression $U_{\xi_t}^{-1}(s, 0) G_{\mu\nu}[\xi_t(s)] U_{\xi_t}(s, 0) \frac{d\xi_t^\mu(s)}{ds}$ is viewed by Chan *et al* as a connection in loop space. They have used this to define a corresponding curvature, the vanishing of which implies that there are no monopoles present. The vanishing of the loop space curvature then has the same physical content as the Bianchi identity of Abelian electromagnetism (see [21]). Note that the 'connection' as defined by Chan *et al* differs from that of Bralić.

n that is the centre of the gauge group, and its elements may be conveniently labelled by the n th roots of unity. The charge[†] enclosed by Σ may then be labelled by the loop space holonomy; i.e. the value of $\Theta(\Sigma)$ corresponding to the different gauge types can then be labelled (\mathcal{L}) by

$$\mathcal{L}[\Theta(\Sigma)] = \exp(i2\pi r/n)I_n \quad r = 0, 1, 2, \dots, (n - 1) \tag{33}$$

where I_n is the $n \times n$ unit matrix. $\mathcal{L}[\Theta(\Sigma)] = I_n$, corresponding to $r = 0$, indicates the absence of a monopole.

The value of $\Theta(\Sigma)$ given by equation (29) can be taken as a definition of the ‘period’ needed to express a non-Abelian form of de Rham’s theorem, at least for the 2-form matrices $\mathcal{G}(y)$ discussed above. For this to make sense, it is explicitly assumed that if the space \mathbb{M} contains any monopoles, they are isolated by boundary components of the space. Thus, if \mathbb{M} has only one boundary component isolating a single monopole, then $P_t \exp \int_{\Sigma} \mathcal{G}$ must, by the generalized Stokes theorem of equation (24), be the identity matrix since $\mathcal{G}(y)$ is a closed 2-form matrix in \mathbb{M} . The implications of this are similar to those for Abelian monopoles where the topology assumed for a spacelike hypersurface in a spacetime containing a monopole is also $\mathbb{R}^3 - \{0\}$:

If the space containing a monopole is assumed to be unbounded, save for the boundary isolating the monopole, with spatial topology equivalent to $\mathbb{R}^3 - \{0\}$, then a single isolated non-Abelian monopole, with $\Theta(\Sigma) \neq I_n$, cannot exist.

If, on the other hand, \mathbb{M} has m boundary components such that $\sum_{i=1}^m a_i z_i = \partial\mathbb{M}$, the z_i being the boundary component 2-cycles, and the a_i constants, then by assuming a parametrized set of loops for each of the z_i , one must have by the generalized Stokes theorem of equation (24)

$$\sum_{i=1}^m a_i P_i \exp \int_{z_i} \mathcal{G} = P_t \exp \int_{\partial\mathbb{M}} \mathbb{G} = P_t \exp \int_{\mathbb{M}} d\mathcal{G} = I_n \tag{34}$$

where $d\mathcal{G}$ vanishes because \mathcal{G} is a closed 2-form matrix in \mathbb{M} . All of the periods are group elements, and the physical meaning of this equation is that the sum of the non-Abelian flux over all boundary components must yield the identity matrix. Differently charged monopoles, each isolated by a boundary component, could exist—provided one maintains consistency with the limitations imposed by equation (34).

Given that the periods of equation (34) are group elements, some discussion of the meaning of the addition operation in this expression may be of value. The law of addition is the group law of combination in the fundamental group of the homotopy classes that correspond to the different monopole types. Since the fundamental group $\pi_1(\text{SU}(n)/\mathbb{Z}_n) = \mathbb{Z}_n$ is an Abelian subgroup of $\text{SU}(n)$, it is not only independent of the base point, but independent of the path connecting different base points. Similarly, the value of $\Theta(\Sigma)$ in equation (27) does not depend on the location of the base point used for the parametrized set of loops. This means, for example, that if \mathbb{M} has two boundary components, for the purposes of computation one may identify the two base points of the parametrized set of loops associated with each boundary component. If one then reparametrizes with respect to t , as t ranges from 0 to 2π , the concatenation of the two sets of loops allows

[†] Corrigan and Olive [9] and Chan *et al* [24] have shown that the relationship between the non-Abelian electric and magnetic charge for each of the gauge types is $g = r/2ne$, where $r = 0, 1, \dots, (n - 1)$. For $\text{U}(1) \times G$, where G is any compact, semisimple group, the number of possible intersections of $\text{U}(1)$ and G are finite and lie in a cyclic subgroup of the centre of G . If p is the number of elements in this subgroup, the generalization of the Dirac quantization condition $eg = n/2$, $n = \pm 1, \pm 2, \dots$ is $eg = n/2p$, n having the range just stated.

the loops to cover both boundary components with the law of combination being that of combination in the fundamental group of homotopy classes. This is the same addition law as that used by Lubkin [17] when considering the addition or combination of two ‘bags’.

Consider $SU(2)/\mathbb{Z}_2 \cong SO(3)$, which is obtained from $SU(2)$ by identifying pairs of elements with opposite signs; since the group $SU(2)$ has the topology of the hyperspherical surface \mathbb{S}^3 , one must identify antipodal points on \mathbb{S}^3 to obtain the group manifold of $SO(3)$. All closed curves in $SO(3)$ thus fall into two homotopy classes corresponding to the fundamental group \mathbb{Z}_2 : those that wind around $SO(3)$ once, and cannot be smoothly contracted to a point, and those which wind around twice and can be so contracted. If both boundary components of \mathbb{M} are now assumed to contain a monopole, as the loops range over the first boundary component a path is traced out halfway around the surface \mathbb{S}^3 ; this is a closed curve since antipodal points are identified, and it cannot be smoothly contracted to a point. As the loops range over the second boundary component, a second such path is traced out in \mathbb{S}^3 which, when combined with the first yields a closed curve on \mathbb{S}^3 which now can be continuously contracted to a point. Thus, for $SU(2)/\mathbb{Z}_2$ two monopoles combine to give a ‘charge’ of zero so that a monopole is its own antimonopole [15].

A possible statement of a non-Abelian version of de Rham’s theorems is now straightforward. If \mathcal{G} is a closed 2-form matrix, for each 2-cycle z one can define a period of \mathcal{G} by $P_z \exp \int_z \mathcal{G}$. de Rham’s theorems can then be expressed for the non-Abelian gauge potential \mathcal{A} and the corresponding field strength \mathcal{G} as follows.

I. The closed 2-form matrix \mathcal{G} is exact if and only if all its periods yield the identity matrix.

Note that this does not imply the existence of a unique \mathcal{A} , consistent with the results of Wu and Yang [3] that the field strength does not determine a unique gauge potential.

If $\partial\mathbb{M}$ has more than one component, and all of the periods do not yield the identity matrix, then the non-Abelian version of the second of de Rham’s theorems would be as follows.

II. If each 2-cycle z is assigned an element in the centre of the gauge group, $\text{per}(z)$, there is closed 2-form matrix \mathcal{G} which has the assigned periods $P_z \exp \int_z \mathcal{G} = \text{per}(z)$ for each 2-cycle z (subject to the consistency relation $\sum_m a_i \text{per}(z_i) = I_n$ if $\sum_m a_i z_i$ is a boundary).

Here a *global* non-Abelian gauge potential does not exist. The consistency relation, which comes from equation (34), continues to hold since $d\mathcal{G}$ still vanishes because \mathcal{G} remains a closed 2-form matrix in \mathbb{M} (the Bianchi identity guarantees this as shown by equation (21)).

It should also be noted that, unlike the Abelian case, if one constructs a cover for the space $\mathbb{R}^3 - \{0\}$ containing a single isolated monopole (despite the fact that this possibility has been ruled out above), where the gauge potential on each element of the cover is related by $A' = g^{-1}Ag + g^{-1}dg$, the field (curvature) on the different elements of the cover is not identical. In fact, the gauge transformation $A' = g^{-1}Ag + g^{-1}dg$ implies that $G' = g^{-1}Gg$ or $\mathcal{G}(y)' = g^{-1}(a)\mathcal{G}(y)g(a)$, so that the field on different elements of the cover depends on a gauge rotation, which is position independent in the case of \mathcal{G} . As mentioned above, the field strength does not uniquely determine the gauge potential, although one may choose a specialized gauge where this is the case [25].

Acknowledgment

This work was supported in part by the US Department of Energy under contract no W-31-109-ENG-38.

Appendix A

de Rham's theorems applied to electromagnetics specify the conditions needed for the global existence of a vector potential \mathbf{A} . To facilitate contact with de Rham's theorems, the vector potential is replaced by a 1-form A and the magnetic field by a 2-form B . If A is a 1-form such that $dA = B$, then the Poincaré lemma states that B is a closed form; that is, $dB = 0$. Locally, the converse of the Poincaré lemma is also true: if B is a closed 2-form, then there is a 1-form A such that $B = dA$, which means that B is an exact form[†]. In general, no such 1-form exists *globally*.

de Rham's theorems, which specify the conditions needed for the global existence of A , are perhaps most easily understood in the form given by Flanders [26]. They need not be restricted to 2-forms but hold for p -forms. In general then, if ω is a closed p -form, for each p -cycle[‡] z one can define a period of ω by $\int_z \omega$. The period only depends on the homology class of z . If z is a boundary, then by Stokes' theorem the period vanishes. Therefore, if $\sum a_i z_i = \text{boundary}$, then $\sum a_i \int_{z_i} \omega = 0$. de Rham's first and second theorems can then be expressed as follows.

I. A closed p -form is exact if and only if all its periods vanish.

II. If each p -cycle z is assigned a number, $\text{per}(z)$, there is a closed p -form ω which has the assigned periods $\int_z \omega = \text{per}(z)$ for each p -cycle z (subject to the consistency relation $\sum a_i \text{per}(z_i) = 0$ if $\sum a_i z_i$ is a boundary).

If ω is set equal to the 2-form B , the substance of de Rham's first theorem is as follows. Since the magnetic field B is solenoidal ($dB = 0$), a vector potential A exists locally such that $B = dA$; i.e. B is locally exact. For A to exist globally, the closed surface periods $\int_z B$ must vanish (z being a 2-cycle). Another way of stating the first theorem uses the cohomology group $H^p(\mathbb{M}, \mathbb{R})$, where \mathbb{M} is an n -dimensional differential manifold. The group $H^p(\mathbb{M}, \mathbb{R})$ is isomorphic to the factor group of the group of closed p -forms on \mathbb{M} by the subgroup of exact p -forms on \mathbb{M} . Thus, the statement that every closed p -form on \mathbb{M} is exact is equivalent to $H^p(\mathbb{M}, \mathbb{R}) = 0$. The second of de Rham's theorems deals with the case where a global vector potential does not exist.

[†] In general, the domain within which a closed p -form will be exact depends on p : if $p = 0$, the domain must be connected; if $p = 1$, the domain must be simply connected; and if $p = 2$, one must be able to shrink any spherical surface to a point.

[‡] If one generalizes the concept of the triangulation of a surface to higher-dimensional spaces where the basic elements of the triangulation are n -simplices, recall that a cycle is defined as a chain whose boundary vanishes, while an n -chain is a formal sum, with constant coefficients, of n -simplices. The collection of all n -cycles forms the n -dimensional cycle group Z_n , while the collection of all n -boundaries forms the n -dimensional boundary group B_n . The factor or quotient group $H_n = Z_n/B_n$ is known as the n -dimensional homology group. The elements of H_n are called homology classes, and cycles in the same homology class are said to be homologous. The H_n can be given a geometrical interpretation: H_0 is a measure of the number of connected components of the space; H_1 of the number of closed curves that are not boundaries of a two-dimensional part of the space; and H_2 measures the number of closed two-dimensional surfaces that are not boundaries of some three-dimensional part of the space.

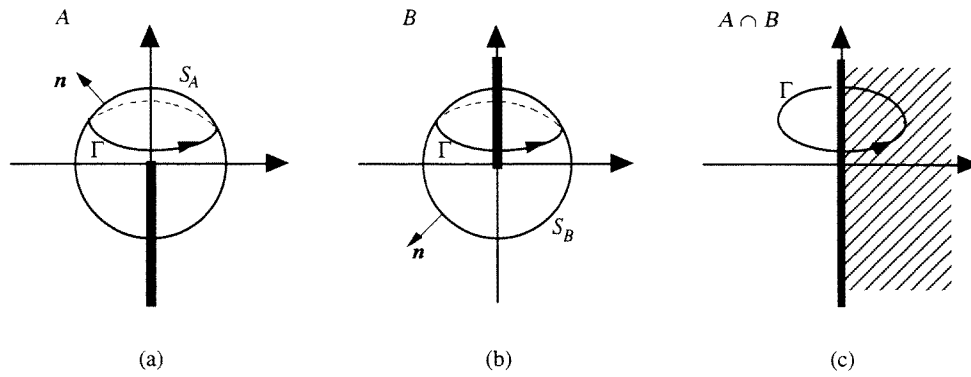


Figure 6. (a) The contour Γ bounding the cap S_A is shown on the open set A (the origin and negative z -axis has been deleted); (b) the contour Γ bounding the cap S_B is shown on the open set B (the origin and positive z -axis has been deleted); and (c), the contour Γ is shown on the open set $A \cap B$ where the entire z -axis has been deleted. A possible cut needed to make the scalar function χ single-valued is indicated by the cross-hatching.

Appendix B

B.1. The total flux

With reference to figure 6, the equation

$$A_a - A_b = d\chi \quad (\text{B.1})$$

can be integrated around a closed contour, Γ , on a sphere isolating the monopole to obtain

$$\int_{\Gamma} A_a - \int_{\Gamma} A_b = \int_{\Gamma} d\chi. \quad (\text{B.2})$$

Note that the first integral is defined on the set A , the second on the set B , and the third on $A \cap B$.

The integral, $\int_{\Gamma} A_a$, is the flux through S_A . The second integral, $\int_{\Gamma} A_b$, is the flux through S_B , but a minus sign must be introduced because of the relative orientation of the normal to this surface and the contour Γ . If one introduces the usual spherical coordinates, the total flux is explicitly given by

$$\Phi = \frac{m}{r^2} \int_0^{2\pi} d\phi \int_0^{\theta} r^2 \sin \theta d\theta + \frac{m}{r^2} \int_0^{2\pi} d\phi \int_{\theta}^{\pi} r^2 \sin \theta d\theta = 4\pi m. \quad (\text{B.3})$$

As seen in figure 6(c), the integral, $\int_{\Gamma} d\chi$, is not trivially zero by Stokes' theorem because the contour Γ cannot be smoothly contracted to a point on $A \cap B$ if it circles the z -axis. To evaluate this integral one must make a cut that consists, for example, of the half-plane $y = 0$ bounded by the z -axis. The value of the integral is then the jump in crossing this plane and is equal to the total monopole flux (see [7]).

Let us briefly return to figure 6(a) where the contour Γ is shown bounding the cap S_A . Topologically, the cap S_A bounded by Γ is homeomorphic to the closed disk \mathbb{D}^2 . It is often stated in the literature that as $\theta \rightarrow \pi$ the contour Γ shrinks to zero yielding the closed surface of the sphere. Such a transformation takes a surface with a boundary into a surface without a boundary, and cannot be a topological mapping since it does not preserve interior and boundary points, which have different local homology groups. The sphere \mathbb{S}^2 is homeomorphic to \mathbb{R}^2 and \mathbb{D}^2 , the open disk, not \mathbb{D}^2 . It is the factor space $\mathbb{D}^2/\mathbb{S}^1$ that is homeomorphic to \mathbb{S}^2 .

B.2. The Dirac quantization condition

For a classical charged particle, the equation of motion may be written as

$$m_e \frac{d^2 x_\mu}{ds^2} - e F_{\mu\nu} \frac{dx^\nu}{ds} = 0 \quad (\text{B.4})$$

where m_e is the mass of a particle with charge e , and $F_{\mu\nu}$ is the electromagnetic field tensor. The corresponding action is

$$S[x(\tau)] = -m_e \int ds - e \int A_\mu dx^\mu. \quad (\text{B.5})$$

In the space $A \cup B = \mathbb{R}^3 - \{0\}$ there is no globally defined vector potential, and consequently equation (B.5) is not in a usable form. However, if one chooses the path $x(\tau)$ to be a closed spacetime path[†], the action can be written, using Stokes' theorem as

$$S[x(\tau)] = -m_e \int ds - e \int_\Sigma F \quad (\text{B.6})$$

where Σ is the surface spanned by the closed path $x(\tau)$, and F is the electromagnetic 2-form. Consider only the last term. In the context of the above formalism, this term becomes

$$S[\Gamma] = -e \int_\Gamma d\chi. \quad (\text{B.7})$$

In quantum mechanics, for $\exp(iS/\hbar)$ to be well defined (single-valued), one must have

$$\int_\Gamma d\chi = \frac{n\hbar}{e}. \quad (\text{B.8})$$

However the integral, $\int_\Gamma d\chi$ is the total flux of the monopole which is $4\pi m$, so that $m = n\hbar/2e$, which is Dirac's quantization condition. It should be noted that there is nothing in the topology that mandates the quantization of the monopole field. It is simply the requirement that $\exp(iS/\hbar)$ is single-valued that is responsible for this condition.

References

- [1] Dirac P A M 1931 *Proc. R. Soc. A* **133** 60
Dirac P A M 1948 *Phys. Rev.* **74** 817 (the term 'string' was coined by Dirac in this paper)
Dirac P A M 1978 *Int. J. Theor. Phys.* **17** 235
- [2] Barut A O and Bornzin G L 1974 *Nucl. Phys. B* **81** 477
Usachev Yu D 1973 *Sov. J. Part. Nucl.* **4** 94
Barut A O 1979 *Sov. J. Part. Nucl.* **10** 209
- [3] Wu T S and Yang C N 1975 *Phys. Rev. D* **12** 3845
- [4] Yang C N 1978 *Proc. 19th Int. Conf. High Energy Physics (Tokyo)* p 497
- [5] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Space-time* (Cambridge: Cambridge University Press)
- [6] Flanders H 1963 *Differential Forms* (New York: Academic) section 5.9
- [7] Marsh G E 1996 *Force-free Magnetic Fields: Solutions, Topology and Applications* (Singapore: World Scientific) ch 5
- [8] Goldberg S I 1982 *Curvature and Homology* (New York: Dover)
Fenn R A 1983 *Techniques of Geometric Topology* (Cambridge: Cambridge University Press) p 228, comment 18
Warner F W 1971 *Foundations of Differentiable Manifolds and Lie Groups* (Glenview, IL: Scott, Foresman and Co.)

[†] This is not an essential restriction and the action can be defined for general spacetime paths; see, for example [27].

- [9] Corrigan E and Olive D 1976 *Nucl. Phys. B* **110** 237 (see also Tze and Ezawa [14] who used the same labelling scheme for non-Abelian $SU(n)/\mathbb{Z}_n$ vortices)
- [10] Aref'eva Ya 1980 *Theor. Math. Phys.* **43** 353
Bralić N E 1980 *Phys. Rev. D* **22** 3090
Fishbane P M, Gasiorowicz S and Kaus P 1981 *Phys. Rev. D* **24** 2324
Diósi L 1983 *Phys. Rev. D* **27** 2552
Broda B 1992 *J. Math. Phys.* **33** 1511
Mensky M B 1979 *Lett. Math. Phys.* **3** 513
- [11] Gantmacher F R 1959 *Applications of the Theory of Matrices* (New York: Interscience) section 5
- [12] Moriyasu K 1983 *An Elementary Primer for Gauge Theory* (Singapore: World Scientific) section 3.6
- [13] Simonov Yu A 1989 *Sov. J. Nucl. Phys.* **50** 134
- [14] Tze H C and Ezawa Z F 1976 *Phys. Rev. D* **14** 1006
- [15] Chan Hong-Mo and Tsou Sheung Tsun 1993 *Some Elementary Gauge Theory Concepts* (Singapore: World Scientific) section 2.2
Tsou Sheung Tsun 1981 *Nucl. Phys. B* **189** 381
- [16] 't Hooft G 1974 *Nucl. Phys. B* **79** 276
- [17] Polyakov A M 1976 *Sov. Phys.-JETP* **41** 988
Lubkin E 1963 *Ann. Phys., NY* **23** 233
- [18] Polyakov A M 1979 *Nucl. Phys. B* **164** 171
- [19] Hocking J G and Young G S 1961 *Topology* (New York: Dover) section 4.7
- [20] Goddard P, Nuyts J and Olive D 1977 *Nucl. Phys. B* **125** 1
- [21] Chan Hong-Mo, Faridani J and Tsou Sheung Tsun 1995 *Phys. Rev. D* **52** 6134
- [22] Goddard P and Olive D I 1978 *Rep. Prog. Phys.* **41** 91 (section 5.6)
- [23] Christ N H 1975 *Phys. Rev. Lett.* **34** 355
- [24] Chan Hong-Mo, Scharbach P and Tsou Sheung Tsun 1986 *Ann. Phys.* **166** 396
- [25] Gu Chaohao 1981 *Phys. Rep.* **80** 251
- [26] Flanders H 1963 *Differential Forms* (New York: Academic)
- [27] Schwarz A S 1993 *Quantum Field Theory and Topology* (Berlin: Springer) ch 22