Gauge Theories in Loop Space

CHAN Hong-Mo and TSOU Sheung Tsun

July 1985
Gauge Theories in Loop Space

by

CHAN Hong-Mo
Kwétheld Appletov Laboratory
Chilton, Didcot, Oxon OX11 0QX, England

and

TSOU Sheung Tsun
Mathematical Institute, Oxford University
24-29 St Giles, Oxford, OX13LB, England

Abstract

Pure Yang-Mills theory is reformulated in terms of gauge-independent loop variables whose intrinsic redundancy is removed using a newly derived non-abelian generalization of the Poincaré lemma.
1. Why Loop Space?

Classical electromagnetism as conceived by Faraday and Maxwell can be formulated entirely in terms of the electromagnetic field tensor \( F_{\mu\nu}(x) \). For example, once \( F_{\mu\nu}(x) \) is given at a point \( x \), we know via the Lorentz equation exactly how a charged particle at \( x \) will move. This is no longer true, however, in the quantum theory; \( F_{\mu\nu}(x) \) there is inadequate to describe all electromagnetic effects on the wave function of an electron, as was beautifully demonstrated by the Bohm-Aharonov experiment.

Schematically, the experimental set up is shown in Fig. 1. Although the field tensor is identically zero throughout the region traversed by the electrons, diffraction patterns are observed which depend on the magnetic field through the core \( N \). Indeed these patterns correspond to the phase difference:

\[
\Phi = \frac{1}{\hbar} \oint_C A_\mu(x) \, dx^\mu
\]

along the two paths \( \Gamma_1 \) and \( \Gamma_2 \), given in terms of the gauge potential \( A_\mu(x) \), which depends on the magnetic field in \( N \), and can be nonzero in the region outside the core even when \( F_{\mu\nu}(x) \) there is identically zero. For this reason, the quantum theory of electromagnetism is formulated usually in terms of the gauge potential \( A_\mu(x) \) as field variables, not in terms of the field tensor \( F_{\mu\nu}(x) \).

The potential \( A_\mu(x) \), however, though adequate, overdescribes the system in the sense that different potentials can correspond to exactly the same physical conditions, being thus strictly not a physical observable. Indeed, if one replaces in (1.1) the potential \( A_\mu(x) \) by a gauge transform:

\[
A_\mu'(x) = A_\mu(x) + \partial_\mu \lambda(x),
\]

the phase difference will not change. Further, even if the phase difference were to change by an integral multiple of \( 2\pi \), there will be no observable change in the diffraction patterns. Therefore, what really matters in the Bohm-Aharonov experiment is not the value of \( A_\mu(x) \) itself, nor yet the phase difference (1.1), but just the phase factor:

\[
\Psi(C) = \exp \frac{i}{\hbar} \oint_C A_\mu(x) \, dx^\mu
\]

over closed loop \( C \). In contrast to \( F_{\mu\nu}(x) \) and \( A_\mu(x) \), \( \Psi(C) \) is both adequate and observable, neither underdescribing nor overdescribing the dynamical system of an electromagnetic field interacting with a (quantum) electron.

The situation is similar in nonabelian gauge theories, excepting that there the field tensor \( F_{\mu\nu}(x) \) is inadequate for a complete description already at the level of pure classical Yang-Mills fields. Again, what describes the theory exactly is not \( F_{\mu\nu}(x) \) or \( A_\mu(x) \), but the nonabelian generalisation of the phase factor (1.2), namely:

\[
\Psi(C) = \exp \left[ \frac{i}{\hbar} \oint_C A_\mu(x) \, dx^\mu \right]_{\text{ordered}}
\]

where now since \( A_\mu(x) \) do not in general commute, the expression has to be path ordered. (A careful analysis of this question for both the abelian and nonabelian theories can be found in ref. 2 and 3.)

In view of this, it is aesthetically appealing to attempt reformulating gauge theories entirely in terms of the phase factors \( \Psi(C) \) or some other equivalent loop variables, i.e. the so-called 'loop space formulation' of gauge theories. However, this is technically rather difficult for the following reason. The quantities \( \Psi(C) \) are labelled by loops \( C \) in space-time \( X \), which are much more numerous than the points \( x \) in \( X \) which label the gauge potential \( A_\mu(x) \). It follows therefore that as field variables \( \Psi(C) \) are highly redundant and have to be severely constrained. A priori, it is not even clear exactly what constraints are necessary on \( \Psi(C) \) to remove their redundancy, and how these constraints may best be imposed to reformulate the theory in terms of loop variables.

In any case, the necessity for such constraints makes the formulation rather cumbersome, and hence not particularly useful under normal circumstances in spite of its aesthetic appeal. Only in very special situations where the gauge dependence of \( A_\mu(x) \) has some reason to be strictly avoided in the loop space formulation likely to be of practical value.

We ourselves are attracted to investigate loop space methods while attempting to develop a Lagrangian theory for monopoles, as reported in our other contribution to this school. In the presence of a monopole the gauge potential \( A_\mu(x) \) has to be patched, with the patches depending on the position of the monopole. Thus, when the position of the monopole is varied, as required in an action principle, the patches have to change also, leading to a variational problem we could not handle. We sought therefore to reformulate the theory in terms of field variables which need not be patched even in the presence of monopoles.

Now loop variables, such as \( \Psi(C) \), suit our purpose. Being gauge-independent, they are not altered by a patching transformation on the gauge potential, and hence also patch-independent. However, in order to reformulate the theory entirely in terms of such variables, we have to face
In full the attendant problem of constraints and redundancy which can in most other applications be avoided.

In what follows, we shall describe our solution to the problem 5). Although tailored to suit our specific purpose, we believe that this solution, as well as the analysis leading to it, should also be of use elsewhere.

2. Coordinates and Basic Operations

Our first task is to label the loops, or in other words to introduce co-ordinates for points in loop space. It will be seen that we need to consider only those loops passing through some fixed reference points, say \( \xi_0 = [\xi^{00}] \) in space-time \( X \). Such a loop can be parametrised as follows:

\[
C = \{ \xi^U(s) : s \in \mathbb{R}, \xi^U(0) = \xi^{00} \} \tag{2.1}
\]

where \( \xi^U(s) \) represent the co-ordinates in \( X \) of the points (2.1) on the loop. Thus,

\[
\Phi(C) = \text{exp} \left( \int_0^\infty ds \hspace{1em} \xi^U(s) \frac{d\xi^U(s)}{ds} \right), \tag{2.2}
\]

In (2.2) \( \Phi \) denotes ordering in \( s \), with \( s \) increasing in our convention from right to left, and the derivative \( d\xi^U(s)/ds \) is taken to be the derivative from below.

We may then label our loops just by the functions \( \{ \xi^U(s) \} \) or simply \( C \). However, such a labelling is incomplete in that the same loops in \( X \) can be parametrised differently. Indeed, the expression \( \Phi(C) \) in (2.2) itself depends only on the loop in \( X \), not in the manner it is parametrised. Hence, if we define one \( \Phi(C) \) for each parametrisation \( C = [\xi^U(s)] \), we have introduced a further redundancy into our variables.

The alternative is not to distinguish the different parametrisations, but to label the loops by the equivalent class \( \mathcal{F} \) of all \( C \)'s equivalent under reparametrisations. However, we have found it much easier to work in the space of all \( C \)'s than in the space of all \( \mathcal{F} \)'s; so much so, indeed, that it is worth the trouble of having later to remove the additional redundancy of variables so introduced.

We shall call the space of all \( C \)'s the 'parametrised loop space' \( \Omega^1X \), whose points are 'parametrised loops'. It has the virtue of being just a functional space with familiar properties. For example, we can define loop differentiation in \( \Omega^1X \) just as the ordinary functional derivative, thus, for any functional \( \Upsilon \) of \( C \):

\[
\frac{\delta}{\delta \xi^U(s)} \Upsilon(U) = \lim_{\delta \to 0} \frac{1}{\delta} \left[ \Upsilon(C') - \Upsilon(C) \right], \tag{2.3}
\]

with

\[
C' = \{ \xi^U(s) = \xi^U(s) + \delta \xi^U(s) \mid \delta(s-a) \}
\]

where, in case of ambiguity, \( \delta(s-a) \) may be regarded as the zero width limit of a smooth function. As usual, the derivative so defined when repeated gives symmetric second derivatives:

\[
\frac{\delta^2}{\delta \xi^U(s) \delta \xi^V(s')} = \frac{\delta^2}{\delta \xi^U(s) \delta \xi^V(s')}, \tag{2.5}
\]

Alternatively, had we decided to work in the space of 'unparametrised loops' \( \mathcal{F} \)'s, the loop derivative would not be so obviously defined. It has, for example, been suggested in the literature that the loop derivative in that case be defined as the continuum limit of the derivative on a discretised loop, as illustrated in Fig. 2(a). The derivative defined in this way, however, will in general give asymmetric second derivatives 6).

The reason for this difference with the preceding case is quite transparent. Varying \( C \) at \( s \) first in direction \( \delta \) and then in direction \( \delta' \) leads to Fig. 2(c), which differs from the result obtained from a variation in the opposite order, namely first \( \delta' \) then \( \delta \), giving Fig. 2(d). Whereas, for the previous definition in terms of functional derivatives, the variations in either order result in the same loop:

\[
C' = \{ \xi^U(s) = \xi^U(s) + \delta \xi^U(s-a) + \delta' \xi^U(s-a) \}
\]

as illustrated in Fig. 2(h) and (i), which is necessary for obtaining a symmetric second derivative.

Hence, we can define in particular the derivative in \( \Omega^1X \) of the phase factor \( \Phi(C) \) according to (2.3). However, \( \Phi(C) \) being by construction an element of the gauge group, only its logarithmic derivative 7)
\[ F_\mu (\Phi(s)) = \frac{1}{\hbar} \delta^{-1}(\Phi(s)) \frac{\delta}{\delta \xi^\mu(s)} \Phi(s) \]  

(2.7)

is meaningful, which is an element of the Lie algebra.

The quantity \( F_\mu (\Phi(s)) \) is of paramount importance for our future discussion. From Fig. 3, it is readily seen that in terms of ordinary field variables, we can write:

\[ F_\mu (\Phi(s)) = \Phi^{-1}(s,0) F_{\mu \nu}(\ell(s)) \Phi(0,0) \frac{d\ell^\nu(s)}{ds} \]  

(2.8)

where

\[ F_{\mu \nu}(s) = 3 \frac{\partial A_\mu(s)}{\partial x_\nu} - 3 \frac{\partial A_\nu(s)}{\partial x_\mu} + \imath \epsilon_{\mu \nu \rho \delta} [A_\rho(s), A_\delta(s)] \]  

(2.9)

is the usual field tensor, and

\[ \Phi_{\nu}(s_2, s_1) = \frac{1}{\hbar} \exp \imath \int_{s_1}^{s_2} ds A_\nu(s) \frac{d\Phi(s)}{ds} \]  

(2.10)

is the parallel transport along \( \Phi \) from the point \( \ell^\mu(s_2) \) to the point \( \ell^\mu(s_1) \) in \( X \).

From the antisymmetry of \( F_{\mu \nu}(s) \) in (2.8) it follows that \( F_\mu (\Phi(s)) \) has only components transverse to the loop \( \Phi \), namely:

\[ F_\mu (\Phi(s)) \frac{d\ell^\nu}{ds} = 0. \]  

(2.11)

This statement is equivalent to the assertion that \( \Phi(s) \) in (2.2) is invariant under reparametrisation of \( \Phi \), since reparametrisation means just shifting points along the loop, and invariance must therefore lead to a zero derivative along the loop direction.

Further, one notes from (2.8) that \( F_\mu (\Phi(s)) \) depends only on the 'early' part of the loop \( \Phi \), i.e. it depends on \( \ell(s) \) only for \( s' < s \), not for \( s' > s \) (see note after (2.2)). This fact is reflected in the notation (C(s)) for its argument, which is henceforth to be so understood.

Geometrically, \( F_\mu (\Phi(s)) \) can be regarded as some sort of gauge potential or 'connection' in parametrised loop space \( \Omega^{1}X \). It represents the change in phase of \( \Phi(s) \) as one moves from a point \( \Phi \) in \( \Omega^{1}X \) to a neighbouring point \( \Phi + \delta \Phi \), in much the same way that the ordinary gauge potential \( A_\mu(s) \) represents the change in phase of a wave function in moving from a point \( x \) to a neighbouring point \( x + \delta x \) in usual 4-dimensional space-time \( X \).

Now, given the potential (connection) \( A_\mu(x) \) in \( X \), one can construct from it the field tensor (curvature) \( F_{\mu \nu}(x) \) via (2.9) and the phase factor or Wilson operator (holonomy) \( \Phi(s) \) via (2.10). So similarly, given now the connection \( \Phi(s) \), one can construct also the corresponding curvature and holonomy in \( \Omega^{1}X \).

Thus, the curvature is:

\[ C_{\mu \nu}(C; s, s') = \frac{\delta}{\delta \xi^\mu(s')} \Phi(s) - \frac{\delta}{\delta \xi^\mu(s)} \Phi(s') + \imath \epsilon_{\mu \nu \rho \delta} [F_{\rho \delta}(s), F_{\mu \nu}(s')]. \]  

(2.12)

As usual for the curvature, \( C_{\mu \nu}(C; s, s') \) represents the total change in phase as a point moves around an infinitesimal closed circuit. Here the circuit in \( \Omega^{1}X \) is obtained starting from the point \( C \) by first varying \( C \) at \( s \) in the direction \( \mu \), then at \( s' \) in direction \( \nu \), then back again at \( s \) in direction \( \mu \) and at \( s' \) in direction \( \nu \) in that order. In particular, for \( s=\ast \), this circuit in \( \Omega^{1}X \) will appear in ordinary space-time \( X \) like a skipping rope, sweeping out an infinitesimal 2-dimensional surface and enclosing a 3-dimensional element of volume, as illustrated in Fig. 2(j). On the other hand, for \( s \neq s' \), no such volume is enclosed. Hence, we may conclude that \( C_{\mu \nu}(C; s, s') \) is zero for \( s=\ast \), i.e. it is proportional to \( \delta(s=\ast) \). What its value is for \( s \neq s' \) will depend on what lies inside the enclosed volume, which will be the subject of the next section.

Curvature, of course, is just the differential version of the more general concept of holonomy over finite loops. To construct the holonomy in \( \Omega^{1}X \) let us then parametrise a loop in space as follows. Let

\[ \Sigma = \left[ \ell^\mu(s) \right]; \quad s = 0 \rightarrow 2s, \quad t = U + 2t, \]  

(2.13)

with

\[ \ell^\mu(0) = \ell^\mu(2s), \quad t = U + 2t, \]  

(2.14)

\[ \ell^\mu(s) = \ell^\mu(0), \quad s = U + 2s. \]  

(2.15)
At each \( t \), \( \xi^u_t(s) \) represents for \( s = u + 2t \) a closed loop \( C_L \) in \( \mathbb{R}^3 \) passing through the fixed reference point \( P_v = \left[ \xi^u_0 \right] \):

\[
C_L = \left\{ \xi^u_t(s); s = u + 2t \right\}.
\]  
(2.16)

For \( t = 0, 2\pi \) according to (2.15), \( C_L \) shrinks to a point \( \left[ \xi^u_0 \right] \). As \( t \) varies, therefore, \( C_L \) traces out a closed loop in \( \mathbb{R}^3 \) beginning and ending at the zero loop at \( \left[ \xi^u_0 \right] \). For this parametrised loop \( C_L \) in \( \mathbb{R}^3 \), we can then construct the phase factor or Wilson operator in complete analogy to (2.2) as:

\[
\Theta(C_L) = \prod_o \exp i \int_0^{2\pi} ds \int_0^{2\pi} ds \prod_o \left( C_L \right)(s) \frac{\delta \xi^u_t(s)}{\delta \xi^u_t(s)}.
\]  
(2.17)

with \( \prod_o \left( C_L \right)(s) \), the connection in \( \mathbb{R}^3 \), taking the place of the connection \( A_{\mu}(s) \) in ordinary space-time \( X \). The only difference is that, loop space being infinite-dimensional, the connection \( \prod_o \left( C_L \right)(s) \) has many more components than \( A_{\mu}(s) \), being labelled by \( \mu \) as well as the continuous index \( s \), both of which have to be summed over in (2.17) — hence the additional integral over \( s \). Again, \( \prod_o \) denotes ordering on \( t \), with \( t \) increasing by one between diagrams. Hence, from below.

In ordinary space-time \( X \), the loop in \( \mathbb{R}^3 \) can be considered as a parametrised 2-dimensional surface swept out by the loops \( C_L \), as illustrated in Fig. 4. In general, it envelops a non-vanishing 3-dimensional volume in \( X \). As for \( \prod_o \left( C_L \right)(s, t) \), the value of \( \Theta(C_L) \) will depend on what is contained inside the volume enveloped by \( C_L \), which question we shall now examine.

3. Monopoles as Sources of Curvature

Let us begin by expressing the curvature \( C_{\mu \nu}(C_L; s, t) \) in \( \mathbb{R}^3 \) in terms of the usual field variables. This operation is a delicate and requires some care. We note first that by definition:

\[
\frac{\delta}{\delta \xi^u_t(s)} \left[ C_L(s, t) \right] = \lim_{\Delta \to 0} \frac{1}{\Delta} \left\{ \Phi^{-1}(C_L) \left[ \Phi(C_L) - \Phi^{-1}(C_L) \Phi(C_L) \right] \right\}.
\]  
(3.1)

where for \( C \) given as in (2.1), we obtain:

\[
C_1 = \left[ \xi^u_t(s) = \xi^u_t(s) + \delta \xi^u_t(s) \right],
\]  
(3.2)

\[
C_2 = \left[ \xi^u_2(s) = \xi^u_2(s) + \delta \xi^u_2(s) \right],
\]  
(3.3)

\[
C_3 = \left[ \xi^u_3(s) = \xi^u_3(s) + \delta \xi^u_3(s) \right],
\]  
(3.4)

through the variations:

\[
\delta \xi^u_t(s) = \Delta \xi^u_t \delta(s, s_1),
\]  
(3.5)

\[
\delta \xi^u_2(s) = \Delta \xi^u_2 \delta(s, s_2).
\]  
(3.6)

According to (2.6), we may write:

\[
\Phi(C_L) = \Phi(C) - \Phi(C_1) \left[ \Phi(C_2) \right] F_{\mu \nu}(5(s)) \frac{\delta \xi^u_t(s)}{\delta \xi^u_t(s)} \delta \xi^u_2(s) \Phi(C_2(s, t), \)  
(3.7)

and an analogous formula for \( \Phi(C_L) \). For \( \Phi(C_2) \), however, we have:

\[
\Phi(C_2) = \Phi(C_1) - \Phi(C_1) \left[ \Phi(C_2) \right] F_{\mu \nu}(5(s)) \frac{\delta \xi^u_t(s)}{\delta \xi^u_t(s)} \delta \xi^u_2(s) \Phi(C_2(s, t)).
\]  
(3.8)

Now in (3.8):

\[
\Phi(C_1) - \Phi(C_1) \left[ \Phi(C_2) \right] \left[ \Phi(C_2) \right] F_{\mu \nu}(5(s)) \frac{\delta \xi^u_t(s)}{\delta \xi^u_2(s)} \delta \xi^u_2(s) \Phi(C_2(s, t)),
\]  
(3.9)

where the last term in (3.9) is due to variation of the end-point in the integral (2.10) for \( \Phi(C(s, t)) \). A similar formula holds for \( \Phi(C_2(s, t)) \). Furthermore, we have:

\[
F_{\mu \nu}(5(s)) = F_{\mu \nu}(5(s)) + \delta \Phi_{\mu \nu}(5(s)) \delta \xi^u_t(s),
\]  
(3.10)
\[
\frac{dL^\nu(s)}{ds} = \frac{dL^\nu(s)}{ds} + \frac{d}{ds} \left( \delta L^\nu(s) \right).
\] (3.11)

Hence, collecting all variations to first order in \(\delta L^\nu(s)\) and \(\delta \nu^\nu(s)\), we obtain:

\[
\frac{\delta}{\delta \nu^\nu(s)} G_\nu^\nu(\mathbf{s}, s') \bigg|_{s_1}^{s_2}
= \chi \left[ \nu_\lambda \left( \mathbf{C}(s_1) \right), \nu_\lambda \left( \mathbf{C}(s_2) \right) \right] \bigg|_{s_1}^{s_2}
+ \frac{\delta L^\nu(s)}{\delta \nu^\nu(s)} \bigg|_{s_1}^{s_2} \left( \frac{\partial L^\nu(s)}{\partial s} \right) \delta(s - s')
+ \frac{\delta L^\nu(s)}{\delta \nu^\nu(s)} \bigg|_{s_1}^{s_2} \left( \frac{\partial L^\nu(s)}{\partial s} \right) \delta(s - s'),
\] (3.12)

where

\[
U_\lambda F_{\nu \lambda}(s) = \partial_\lambda F_{\nu \lambda}(s) - i \epsilon \left[ A_\lambda(s), F_{\nu \lambda}(s) \right] \quad \text{(3.13)}
\]

is the usual covariant derivative of the field tensor. Substituting (3.12) into (2.12), then gives:

\[
G_{\nu \mu}(\mathbf{s}, s')
= \frac{\delta L^\nu(s)}{\delta \nu^\nu(s)} \bigg|_{s_1}^{s_2} \left( \frac{\partial L^\nu(s)}{\partial s} \right) \delta(s - s'),
\] (3.14)

or equivalently:

\[
G_{\nu \mu}(\mathbf{s}, s')
= -\frac{1}{2} \epsilon_{\nu \mu \rho \sigma} U_\alpha \frac{\partial F_{\rho \alpha}(s)}{\partial s} \bigg|_{s_1}^{s_2} \epsilon^\rho_\sigma(s(s)) \frac{\partial s(s)}{\partial s} \delta(s - s').
\] (3.15)

in terms of the dual field tensor:

\[
G_{\nu \mu}(\mathbf{s}, s') = \frac{1}{2} \epsilon_{\nu \mu \rho \sigma} U_\alpha \frac{\partial F_{\rho \alpha}(s)}{\partial s} \bigg|_{s_1}^{s_2} \epsilon^\rho_\sigma(s(s)) \frac{\partial s(s)}{\partial s} \delta(s - s').
\] (3.16)

We note that this nice and symmetric expression for \(G_{\nu \mu}(\mathbf{s}, s')\) depends on our defining the loop derivative as the functional derivative in parameterised loop space \(\mathcal{S}_\lambda\). Had we operated with the other definition mentioned above of the loop derivative as the continuum limit of a discretised version, we would have missed the second term on the right of (3.11), and hence also the last term inside the bracket in (3.14). This term in (3.14) may be interpreted as a sort of flux flowing along the direction of the loop through the elemental volume depicted in Fig. 2(c). For the discretised version, it corresponds to the flux through the two shaded end plaquettes of the box in Fig. 2(e), and is exactly what will be missed in taking the difference between Fig. 2(c) and (d), thus reflecting the asymmetry in this approach of the second derivative as we have discussed.

From (3.14), one concludes that so long as \(F_{\nu \mu}(s)\) is a gauge field derivable via (2.9) from a gauge potential, then \(G_{\nu \mu}(\mathbf{s}, s')\) is zero by virtue of the Bianchi identity. However, it need not vanish at \(s = s'\) if \(\nu^\nu(s)\) happens to sit on the world-line of a monopole, since at a monopole \(F_{\nu \mu}(s)\) is not a gauge field and a potential need not exist there. Indeed, if we are considering the abelian theory, then \(G_{\nu \mu}(\mathbf{s}, s')\) is manifestly nonvanishing in the presence of a monopole since \(\nabla^\nu F_{\mu \nu}(s) - \frac{1}{2} \epsilon_{\nu \mu \rho \sigma} F_{\rho \sigma}(s)\) in (3.15) is just the magnetic' current due to the monopole charge \(g:\)

\[
\epsilon_{\nu \mu \rho \sigma} F_{\rho \sigma}(s) = -4\pi g \int \frac{d^4(x)}{v^4} \partial_{\nu}(\chi) \delta(x - Y(t)).
\] (3.17)

We may therefore assert in that case that monopoles appear in loop space as sources of the curvature' \(\nu^\nu\). For nonabelian theories on the other hand, the situation is not yet transparent since at the monopole position the right-hand-side of (3.14) becomes meaningless, the potential \(A_\lambda(x)\) being undefined, hence also the covariant derivative \(\partial_\lambda\). To evaluate \(G_{\nu \mu}(\mathbf{s}, s')\) in that case we shall have to analyse the problem further in what follows.
Initially, to avoid unnecessary confusion about infinitesimals, let us first, as it were, enlarge the region around a monopole under a microscope, and consider the holonomy $\theta(\Sigma)$ over a finite loop in $\Sigma X$. As explained in the last section, $\Sigma$ is a parametrisation of a closed 2-dimensional surface $\Sigma$ in space-time $X$. What we wish to know now in the non-abelian theory is the value that $\theta(\Sigma)$ will take if $\Sigma$ encloses a monopole.

Consider for concreteness a pure $su(2)$ Yang-Mills theory. As explained in our other contribution to this school, a monopole there is characterised by a patched potential, say $A_\mu^{(N)}(x)$ and $A_\mu^{(S)}(x)$, defined respectively in

\begin{align}
(N) : & \quad 0 < t < \pi , \quad 0 < \varphi < 2\pi \\
(S) : & \quad 0 < t < \pi , \quad 0 < \varphi < 2\pi \tag{3.18}
\end{align}

and related to each other in the overlapping region by the patching condition:

\[ A_\mu^{(S)}(x) = \theta^{-1}(x) A_\mu^{(N)}(x) \theta(x) \cdot \frac{1}{i} \theta^{-1}(x) A_\mu^{-1}(x) . \tag{2.19} \]

The potential $A(x)$ is an element of the gauge Lie algebra $su(2)$, while the patching function $\theta(x)$ may be regarded as representing a closed curve in the gauge group $SO(3)$ belonging to the nontrivial homotopy class $\xi$. We may as usual represent the potential by:

\[ A_\mu(x) = A_\mu^{(S)}(x) \cdot (\Sigma/2) \tag{3.20} \]

where $\Sigma_i$, $i = 1, 2, 3$ are Pauli matrices. On substituting this into (2.10), one obtains on the right-hand side a $2 \times 2$ unitary matrix $\tilde{\Phi}(C)$, say, that is an element of the group $SU(2)$, which covers, but is not the same as the gauge group $SU(3)$. To obtain $\Phi(C)$, which is by definition an element of the gauge group, all we need to do is to identify matrices differing only by a sign, and regard the pair as corresponding to the same group element. It is in this manner that (2.10) is originally meant to be understood.

Take now a $\xi$ enclosing a monopole. As $t$ varies for $C_\xi$ in (2.16), $\Phi(C_\xi)$ traces a closed curve in the gauge group $SO(3)$. However, suppose we choose instead to consider $\tilde{\Phi}(C_\xi)$, then for $t$ in the range $[0, 2\pi]$ $\tilde{\Phi}(C_\xi)$ will trace a curve in the covering group $SU(2)$ which is still continuous but need no longer be closed. Indeed, starting at $t = 0$ where $C_\xi$ shrinks to the point $P_0$, the matrix $\tilde{\Phi}(C_\xi)$ is by definition $I$. As $t$ increases, $C_\xi$ shrinks to the point $P_0$, but at still may be considered as lying entirely in the patch (N) so that the matrix $\tilde{\Phi}(C_\xi)$ as calculated with $A_\mu^{(N)}(x)$ varies continuously with $t$. However, in order to envelop completely the surface $\Sigma$ as depicted in Fig. 4, $C_\xi$ must at some value of $t$ pass through the south pole and can no longer remain entirely in the northern patch (N). One must then at some value $0 < t < 2\pi$, perform a patching transformation from $A_\mu^{(N)}$ to $A_\mu^{(S)}$ at which point, as can be seen from (3.19) for $\theta$ of homotopy class $\xi$, $\Phi(C_\xi)$ suffers a discontinuous change of sign. For $t < \xi$, the loop $C_\xi$ can now be considered as lying entirely in the patch (S) and $\Phi(C_\xi)$ again varies continuously with $t$, returning to the value I at $t = 2\pi$ when $C_\xi$ shrinks again to a point at $P_0$. One sees therefore that the curve traced by $\tilde{\Phi}(C_\xi)$ for $t$ in the range $[0, 2\pi]$ does in fact form a continuous piece stretching from the value of $\tilde{\Phi}(C_\xi)$ at $t = \xi$ through I to its value at $t = \xi$.

The total change in the matrix $\Phi(C_\xi)$ over the whole range $[0, 2\pi]$ is simply given by the jump $\Phi^{-1}(C_\xi) \Phi(C_\xi)$ at $\xi$, which is always $-I$ independently of exactly where one chooses $\xi$ to perform the patching transformation. But this change is exactly what is given by $\theta(\Sigma)$ when regarded as an element of the gauge group $SU(3)$ but of its covering group $SU(2)$. One can see this as follows. Remembering the definition (2.7) of $F_\mu(\Sigma)$ as the logarithmic loop derivative of $\Phi(C)$, one sees that the factor $i \Theta_\mu \left[ \int d\varphi F_\mu(\Sigma) \frac{d}{dt} \right] \frac{1}{\Theta_\mu} \frac{d\Theta_{\mu}}{dt} \left[ \Phi^{-1}(C_\xi) \Phi(C_\xi) \right]$ in (2.17) may be interpreted as representing the quantity $\Phi^{-1}(C_\xi) \Phi(C_\xi)$ itself. Thus $\theta(\Sigma)$, being an ordered product of such factors, is just the total change in $\Phi(C_\xi)$ over the whole range.

Hence we conclude that for $\xi$ enclosing the monopole, $\theta(\Sigma) = -I$. Whereas, if $\Sigma$ does not enclose the monopole, $A_\mu(x)$ need no longer be patched, and one easily deduces that $\theta(\Sigma) = I$. In other words, we have shown that, for the pure $su(2)$ Yang-Mills theory:

\[ \theta(\Sigma) = \xi \tag{3.21} \]
where $\gamma$ and $\beta$ are the same as in the previous section. The following expression for the monopole charge density is derived:

$$\rho_m(r) = \frac{1}{4\pi r^2} \int d^3k \left( \frac{1}{m_{\gamma}^2} \right) \frac{\rho_{\gamma}^2(k) \cdot \rho_{\beta}(k)}{2m_{\gamma}^2 + m_{\beta}^2}.$$

This expression is evaluated numerically to obtain the monopole charge distribution. The resulting charge distribution is then used to calculate the induced magnetic field around the monopole. The calculations are carried out using numerical methods, and the results are consistent with the expected behavior of the monopole charge.

In the last section, we have obtained a good understanding of the monopole charge distribution. This is due to the careful analysis of the monopole wavefunction, which is a solution to the Schrödinger equation in the presence of a magnetic field. The wavefunction is found to be a non-trivial function of the magnetic field, and this leads to the formation of the monopole charge distribution.

Finally, we note that the monopole charge distribution is a fundamental aspect of the theory of monopoles. It is through the study of monopole charge distribution that we gain insight into the behavior of monopoles in various physical situations. The results obtained in this work are expected to have applications in various fields of physics, including condensed matter physics, particle physics, and astrophysics.

In conclusion, the study of monopole charge distribution is a crucial aspect of the theory of monopoles. The results obtained in this work provide a solid foundation for future studies on the behavior of monopoles in various physical situations.
to the whole loop space formulation of gauge theories.

Again, this situation is not entirely new. A parallel already exists in the abelian theory. Although classical electromagnetism can be described in terms of the antisymmetric field tensor $f_{\mu\nu}(x)$ as variables, they are not independent variables but must satisfy the Bianchi Identity:

$$\delta_{\nu} f_{\mu\nu}(x) = 0,$$  \hspace{1cm} (4.1)

or, in other words, be expressible in terms of some potential $a_{\mu}(x)$, thus:

$$f_{\mu\nu}(x) = \partial_{\nu} a_{\mu}(x) - \partial_{\mu} a_{\nu}(x).$$ \hspace{1cm} (4.2)

The fact that these two statements (4.1) and (4.2) are equivalent is sometimes known as the Poincaré lemma. If we think of $a_{\mu}(x)$ as the independent variables of the theory, then $f_{\mu\nu}(x)$ are redundant, being more numerous than the original variables $a_{\mu}(x)$, and must therefore be constrained. The constraints necessary and sufficient for removing their redundancy are exactly the Bianchi identity (4.1). Now, the physical content of (4.1) is that there should be no monopole (or magnetic source) at $x$. Hence we arrive at the conclusion that the absence of monopole is the essence of the constraint required for removing the redundancy of $f_{\mu\nu}(x)$ as field variables.

Notice that all the statements in the preceding paragraph are local in $x$ and are not affected by what happens elsewhere. Suppose, for example, that there is a monopole moving along the world-line $Y(t)$, say, then the Bianchi identity no longer holds everywhere but is replaced by (3.17). In that case, $a_{\mu}(x)$ no longer exists on $Y(t)$, but since (3.17) still implies that (4.1) is satisfied for $x \neq Y(t)$, $f_{\mu\nu}(x)$ is still necessarily expressible in terms of some gauge potential $a_{\mu}(x)$ at all these points.

Consider now the situation in nonabelian theories. Again we started with the potential $A_{\mu}(x)$ as independent variables but wish to replace them by another gauge independent set. The field tensor $F_{\mu\nu}(x)$ is no longer adequate, being also not gauge invariant, only covariant. We chose to express the theory instead in terms of the loop variables $F_{\mu}(C)(s)$. These, being much more numerous than the original variables $A_{\mu}(x)$ are therefore highly redundant and have to be constrained. A comparison with the parallel in the abelian theory would make it seem at least plausible that the constraint will take the form of a sort of Bianchi identity ensuring the absence of monopole charges wherever the potential $A_{\mu}(x)$ exists.

Such a conjecture turns out to be very near the truth. Indeed, we were able to prove for $F_{\mu}(C)(s)$ the following assertion, which may be considered as the generalisation of the Poincaré lemma to nonabelian theories:

(i) Except on monopole world-lines $F_{\mu}(C)(s)$ is expressible via (2.7) and (2.2) in terms of some potential $A_{\mu}(x)$ which is unique up to gauge equivalence.

If and only if:

(1) $F_{\mu}(C)(s)$ satisfies:

(A) $0 (C) = \zeta_{C}$ for all $C$, or equivalently (3.22)

(B) $F_{\mu}(C)(s) \frac{df_{\mu}(s)}{ds}$ = 0, i.e. transversality.

(C) $F_{\mu}(C)(s)$ is independent of $\xi(s')$ for $s' > s$ as implicit already in our notation (C)(s) for its argument.

The condition (A) is the analogue of (3.17) in loop space for nonabelian theories, and guarantees that there are no further monopole charges in $X$ other than those already specified on the world-lines $Y(t)$. A reference to the expression (3.8) for $F_{\mu}(C)(s)$ in terms of external field variables easily shows that the condition (B) for transversality is equivalent to $F_{\mu}(x)$ being antisymmetric, which is understood also in the usual Poincaré lemma. Only the condition (C), relevant for the ordering of noncommuting quantities, has apparently no abelian analogue.

That (i) implies (ii) is easy to prove and has been shown already in the last two sections. The verification of the converse is much less straightforward and cannot be given here in detail. In outline, our proof consists essentially of three steps. First, from the condition (A) being satisfied for all $C$, one deduces that the parallel transport in loop space:

$$\partial_{x}(t,0) = F_{\mu}(C), \exp \int_{0}^{t} \frac{2\pi}{|\mu|} d\delta F_{\mu}(C) \frac{d\phi_{\mu}}{dt},$$ \hspace{1cm} (4.3)

apart from a possible factor $\zeta_{C}$ belonging to $\zeta_{n}$, can depend on the path along $C$ from 0 to $t$ in $U_{\mu}$ only through its end-point $C_{t}$. Since matrices in $SU(n)$ differing only by a factor $\zeta_{C}$ correspond to the same element in the gauge group $SU(n)/U_{\mu}$, we may then define a $\phi(C)$ such that $\partial_{x}(C)$ is the image of $\partial_{x}(t,0)$ in the gauge group, which by construction satisfies the required
relation (2.7) with the given \( \Phi_\mu(C|s) \). Second, this \( \Phi(C) \) so defined is seen (i) to be parametrisation independent, and (ii) to satisfy the composition law:

\[
\Phi(C_2 \ast C_1) = \Phi(C_2) \circ \Phi(C_1),
\]

for any \( C_1, C_2 \), where \( C_2 \ast C_1 \) is the loop obtained by composing \( C_1 \) with \( C_2 \) in that order. That (i) is satisfied is obvious from condition (b). The result (ii) is proved by constructing a particular path \( \Sigma \) in \( G^1X \) along which the parallel transports defined in (4.4) satisfy a composition law corresponding to (4.4). This construction requires condition (C) to guarantee the correct ordering and suffices to establish (4.4), since \( \Phi(C) \) is already known from above to be independent of the path \( \Sigma \) in \( G^1X \). Third, given (i) and (ii), there is a construction known to both physicists and mathematicians (though not yet up to the standards of rigour demanded by the latter) for a gauge potential \( A_\mu(x) \) which obeys (2.2) in relation to \( \Phi(C) \) as required, and is unique up to gauge invariance. What we need here is an extension of these arguments to the case when monopoles are present, and to ensure that the proper patching conditions are satisfied. Having done that, the proof of our assertion above is then complete.

Once proved, this generalisation of the Poincaré lemma establishes a one-one correspondence between the variables \( F_\mu(C|s) \) constrained by (A), (B) and (C) and the variables \( A_\mu(x) \) up to gauge equivalence. With this in hand, one can now proceed, at least in principle, to reformulate the whole theory in terms of \( \Phi_\mu(C|s) \) instead of \( A_\mu(x) \) as variables.

5 The Action Principle in Loop Space

In view of the result described in the last section, let us see how we may reformulate the action principle of pure gauge theories in terms of the variables \( F_\mu(C|s) \). We need first to express the usual Yang-Mills action:

\[
\mathcal{A} = \frac{-1}{\hbar^2} \int d^4x \text{Tr} \left[ F_{\mu\nu}(x) F^{\mu\nu}(x) \right]
\]

in terms of the new variables. The answer is as follows:

\[
\mathcal{A} = \frac{-1}{\hbar^2} \int d^4x \int_0^{s'} ds \text{Tr} \left[ F_{\mu}(C|s) \right] \left[ \frac{d\xi^2(s)}{ds} \frac{d\xi(s)}{ds} \right]^{-1}
\]

where the integral over loops is defined simply as an ordinary functional integral, thus:

\[
\int d^4\xi \rightarrow \int_0^{s'} ds \int d^4\xi(s)
\]

and \( \tilde{N} \) is just a normalisation factor:

\[
\tilde{N} = \frac{2\pi}{\hbar^2} \int_0^{s'} ds \int d^4\xi(s')
\]

The fact that we can write ordinary functional integrals for (5.3) and (5.4) instead of some horrible integrals over (unparametrised) loops with a complicated measure in a space with unfamiliar geometry is again a reflection of the virtue in working with parametrised loops in spite of the additional redundancy in variables so introduced.

That the action (5.2) is indeed equivalent to the original (5.1) can be seen as follows. Substituting the expression (2.6) for \( F_\mu(C|s) \) in terms of usual field variables into (5.2), one obtains:

\[
\mathcal{A} = \frac{-1}{\hbar^2} \int d^4x \int_0^{s'} ds \int d^4\xi \frac{d}{ds} \frac{d}{ds} \left[ s' \frac{d\xi(s)}{ds} \frac{d\xi(s)}{ds} \right]^{-1}
\]

Insert next the identity:

\[
\int d^4\xi \delta^4(\xi(s) - x) = 1
\]

into the integrand of (5.5) and recall (5.3), we have:

\[
\mathcal{A} = \frac{-1}{\hbar^2} \int d^4x \int_0^{s'} ds \int d^4\xi \left[ F_{\mu\nu}(\xi(s)) F^{\mu\nu}(\xi(s)) \right] \frac{d\xi(s)}{ds} \frac{d\xi(s)}{ds} \left[ s' \frac{d\xi(s)}{ds} \frac{d\xi(s)}{ds} \right]^{-1} \delta^4(\xi(s) - x)
\]
Integrating now over $d^4\zeta(s)$ by virtue of the δ-function gives:

$$\mathcal{A} = \langle \phi \bar{\phi} \rangle^{-1} \int d^4x \; \text{Tr} \left[ \frac{1}{\mathcal{L}_{\mu}^2} \right]$$

$$\int_0^2 \frac{d\lambda}{\lambda} \left[ \frac{d\zeta^0(s')}{ds'} \left( \frac{d\zeta^0(s)}{ds} \left( \frac{d\zeta^0(s)}{ds} \right)^{-1} \right) \right]_{\zeta(s) = x}$$

(3.8)

Notice now that $d\zeta^0(s)/ds$ in the last two factors represent the tangent vector to the loop C at the point $\zeta(s) = x$. When integrating over $d\zeta^0(s')$ for $s' \neq s$ while keeping $\zeta(s)$ fixed at $x$ one just averages over all directions of the tangent vector and gives for the last integral in (3.8) simply:

$$\left( \int_{s \neq s} d\zeta^0(s') \right) \cdot \frac{1}{\lambda} \zeta^0_x$$

(5.9)

or that:

$$\mathcal{A} = \langle \phi \bar{\phi} \rangle^{-1} \int d^4x \; \text{Tr} \left[ \frac{1}{\mathcal{L}_{\mu}^2} \right] \left( \int_0^2 \frac{d\lambda}{\lambda} \int_{s \neq s} d\zeta^0(s') \right)$$

(5.10)

which, by (5.4) is exactly (5.1) as desired. We have gone deliberately slowly over the various steps in the above demonstration to illustrate the fact that while operating in parameterised loop space $d^4\zeta$, we do not need to go beyond the usual rules of functional differentiation and integration, and can avoid much unnecessary ambiguities which may be encountered otherwise.

Having expressed the action in terms of $\mathcal{F}_\mu(C|x)$, we may now formulate the action principle as the requirement that $\mathcal{A}$ be stationary under variations in the field variables $\mathcal{F}_\mu(C|x)$. However, $\mathcal{F}_\mu(C|x)$ being not free variables, but subject to the conditions (A) - (C), it becomes a constrained variational problem. The constraints are complicated, especially (A), which gives one constraint for every parameterised surface $\Sigma$, which is a horrendous number. Nevertheless, they are all of the so-called 'isoperimetric' type for which the Lagrange method of undetermined multipliers should apply. The problem then becomes in principle straightforward. We simply introduce one such multiplier $\lambda_C$ for each constraint labelled by $\Sigma$ and some other appropriate multipliers to account for the simpler constraints (B) and (C).

We next extremise the action (5.2) under these constraints to obtain the Euler-Lagrange equations in terms of these multipliers, and if we are then able to eliminate all those multipliers from the resulting equations, we shall obtain the equations of motion for the system in terms of our loop variables.

In practice, however, because of the large number of Lagrange multipliers involved, the process is rather complicated. Nevertheless, after some considerable effort, which we will not here reproduce, we in fact succeeded to carry it out, obtaining as the result the equation:

$$\frac{\delta}{\delta \zeta^0(s)} \mathcal{F}_\mu(C|x) = 0$$

(5.11)

which is to hold for all (C|x).

That (5.11) is just the loop space version of the Yang-Mills equation\(^\dagger\) can be seen by referring to (3.12), which for $\kappa = 1$ gives:

$$\frac{\delta}{\delta \zeta^0(s_2)} \mathcal{F}_\mu(C|x)$$

(5.12)

The usual Yang-Mills equation clearly implies via (5.12) that (5.11) is satisfied. Conversely, the validity of (5.11) for all (C|x), and hence for all directions of the tangent $d\zeta^0(s)/ds$ at $s$ in (5.2), implies also that the usual Yang-Mills equation holds.

That the action (5.1) should lead to the Yang-Mills equation is of course nothing new. However, the fact that we are now able to derive exactly the same result from an action principle formulated entirely in terms of $\mathcal{F}_\mu(C|x)$ serves as a valuable check that this formulation is indeed equivalent to the conventional one in terms of $A_\mu(s)$ as variables.

We conclude this lecture with two remarks:

First, the fact that one has needed such effort just to derive the Yang-Mills equation in loop space is a reminder of our observation in the beginning that the loop space formulation is not in general a very practical approach to gauge theories because of the intrinsic constraints on the loop...
variables. But, if for some reason, a loop space formulation becomes necessary or desirable as we believe it is for the theory of nonabelian monopoles, then the constraints are essential, for otherwise we shall not obtain the correct result. The simplicity of the answer, however, belies the extreme complications of the constraints to which the system is subjected.

Second, we were led to employ the connection in loop space $F_\mu$ (Cl's) as field variables by considerations in our problem on nonabelian monopoles as explained in our other contribution to this school. This choice may seem surprising in view of the fact that $F_\mu$ (Cl's), being labelled by $C$ and $\epsilon$, are even more numerous than the phase factors $\Phi(C)$, which are labelled only by $C$ and already highly redundant. At first sight, it would appear that $\Phi(C)$ may be a simpler variable to use. We have tried this, but without success. Indeed, as far as we are aware, this is the first time that anyone has succeeded in deriving the equations of motion from an action principle entirely in terms of loop variables. We do not understand clearly the reason behind this, but believe that it is due to the relatively simple form the constraints take in terms of $F_\mu$ (Cl's), and their seemingly intimate relationship to the monopole problem via the generalised Poincaré lemma. Had we used $\Phi(C)$ as variables, we would have subjected them to, instead of (A), (B) and (C) in section 4., the requirement of (i) parametrisation independence and (ii) the composition law (4.4). Now the constraint (4.4), being non-linear, is not easily implemented. In contrast, the constraints (A), though numerous, are, apart from the exponentiation, essentially linear in the variables $F_\mu$ (Cl's). So it is, once again, as already happened before in introducing the parametrised loop space, profitable to enlarge our space of variables so as to simplify its structure, even at the cost of multiplying the number of constraints by a large factor.

Acknowledgement

We are deeply grateful to our Polish hosts for their kind hospitality. Most of the original work described in this lecture was done in collaboration with Peter Scharbach.
References


Figure Captions

Fig. 1. Schematic representation of the Bohm-Aharonov experiment.
Fig. 2. Illustrations for two versions of the loop derivative.
Fig. 3 Illustration for the formula (2.8)
Fig. 4. A loop in loop space $Q^4$ viewed in ordinary space-time $X$. 