LOOP APPROACHES TO GAUGE FIELD THEORIES

R. Loll

Basic mathematical and physical concepts in loop- and path-dependent formulations of Yang–Mills theory are reviewed and set into correspondence. We point out some problems peculiar to these non-local approaches, in particular those associated with defining structure on various kinds of loop spaces. The issues of classical loop equations, differential operators, lattice gauge theory, loop algebras, and canonical quantization are discussed in some detail, and the paper concludes with an extensive list of references.

In Memory of Prof. M. C. Polivanov

1. MOTIVATION

This paper contains an extended version of a set of lectures given at the Centro de Física da Matéria Condensada (CFMC) in Lisbon in June 1992. Its purpose is at least twofold. The first part (Secs. 2–9) introduces basic concepts about paths and loops and their associated (traced) holonomies in theories which have a connection one-form (gauge potential) as their basic configuration variable. In exhibiting their mathematical structure, I illustrate that there are many different loop spaces, which arise in a variety of ways in physical applications. It is also meant to be a collection of formulas frequently used in this field, which to my knowledge does not exist in such a condensed form anywhere in the literature. (The reader may also consult the review sections of [1] and [2] for general reference.)

In the second part (Secs.10–14) I describe past applications and the present status of path- and loop-dependent formulations of Yang–Mills theory, with the help of some selected topics: classical loop equations, differential operators, lattice gauge theory, loop algebras, and canonical quantization. The treatment is not meant to be exhaustive, but rather to summarize results and illustrate where problems arise. Particular emphasis is put on the way in which different mathematical structures do (or do not) occur in gauge theory and gravity. This will hopefully clarify how different approaches are related to different physical motivations and interpretations of the loops themselves, and thus aid further research in this direction. I point out a number of open problems of the loop formulation as I go along. There is an extensive list of references, providing a guide to the literature on the field.

2. PATHS AND LOOPS

Given a differentiable, simply connected manifold Σ of dimension d, a *path* in Σ (Fig. 1a) is a continuous map w from a closed interval of the real line $\mathbb{R}$ into Σ,

$$w\colon [s_1, s_2] \to \Sigma, \qquad s \mapsto w^\mu(s) \ . \tag{2.1}$$

As such it has the properties of a map between two differentiable manifolds, for example, (C^r) differentiability, piecewise differentiability or non-differentiability, and its tangent vector $\frac{dw}{ds}$ may vanish for some or all parameter values s.

Departamento de Fisica, Instituto Superior Tecnico, 1000-Lisboa, Portugal. Physics Department, Syracuse University, Syracuse, NY 13244, USA. Published in Teoretĭcheskaya i Matematicheskaya Fizika, Vol. 93, No. 3, pp. 481–505, December, 1992. Original article submitted September 7, 1992.

A *loop* in Σ (Fig. 1b) is a closed path, by which we shall mean a continuous map γ of the unit interval into Σ,

$$\gamma\colon [0,1] \to \Sigma, \qquad s \mapsto \gamma^\mu(s), \qquad \gamma(0) = \gamma(1) \ . \tag{2.2}$$

We will be using such closed paths in the construction of gauge-invariant quantities in pure gauge theory. Open paths play an important role in gauge theory with fermions, where natural gauge-invariant objects are open flux lines with quarks "glued to the endpoints" [3, 4].

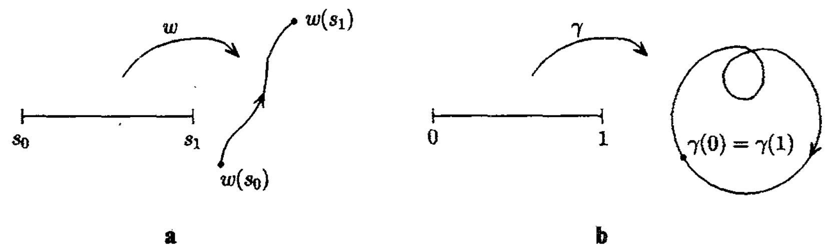

Fig. 1.

The manifold Σ may be the real vector space $\mathbb{R}^d$, possibly with a Euclidean or Minkowskian metric, but may also be non-linear and topologically non-trivial. In this case the possibility arises of having non-contractible loops, i.e., maps γ that cannot be continuously shrunk to a *point loop*

$$p_x(s) = x \in \Sigma, \qquad \forall s \in [0,1] \ . \tag{2.3}$$

Note that even if a path is closed it has a distinguished image point, namely, its initial and end point, $\gamma(0) = \gamma(1)$, and that each point in the image of γ is labelled by one or more (if there are self-intersections) parameter values s.

3. HOLONOMY IN GAUGE THEORY

Suppose in an open neighborhood V of Σ we are given a configuration $A \in \mathcal{A}$, the set of all *gauge potentials* $\Sigma \to \Lambda^1\mathbf{g}$, i.e., a smooth $\mathbf{g}$-valued connection one-form, with $\mathbf{g}$ denoting the Lie algebra of a finite-dimensional Lie group G. We have

$$A(x) = A_\mu(x)\, dx^\mu = A^a_\mu(x) X_a\, dx^\mu \ , \tag{3.1}$$

where X_a are the algebra generators in the fundamental representation of $\mathbf{g}$ ($a = 1, \ldots, \dim G$) and x^μ, $\mu = 1, \ldots, d$, a set of local coordinates on V.

The *holonomy* U_w of a path $w^\mu(s)$ with initial point s_0 and endpoint s_1 (whose image is completely contained in V) is the solution of the system of differential equations

$$\frac{dU_w(s,s_0)}{ds} = A_\mu(x)\, \frac{dw^\mu}{ds}\, U_w(s,s_0), \quad s_0 \le s \le s_1 \ , \tag{3.2}$$

with $x = w(s)$, subject to the initial condition

$$U_w(s_0,s_0) = e \ , \tag{3.3}$$

where e denotes the unit element in G. Note that this definition only makes sense for at least piecewise differentiable paths w. The solution of (3.2) is given by the *path-ordered exponential of A along w*,

$$U_w(s_1,s_0) = \mathrm{P}\exp ig \int_{s_1}^{s_2} A_\mu(w(t))\dot{w}^\mu(t)\, dt := \mathbf{1} + \sum_{n=1}^{\infty} (ig)^n$$
$$\times \int_{s_1}^{s_2} dt_1 \int_{s_1}^{t_1} dt_2 \cdots \int_{s_1}^{t_{n-1}} dt_n\, A_{\mu_1}(w(t_1)) \cdots A_{\mu_n}(w(t_n))\, \dot{w}^{\mu_1}(t_1) \ldots \dot{w}^{\mu_n}(t_n) \ . \tag{3.4}$$

The coupling constant g is necessary to render the argument of the exponential dimensionless. Note that from (3.4) follows the composition law $U_w(s_2, s_0) = U_w(s_2, s_1)U_w(s_1, s_0)$, for $s_0 \leq s_1 \leq s_2$. An alternative definition for U_w that does not need differentiability of the path and employs an approximation of w by n straight line segments $(x_i - x_{i-1})$ is as the limit

$$\begin{aligned} U_w(s_1, s_0) = \lim_{n\to\infty} &\big(1 + A(x_n)(x_n - x_{n-1})\big) \\ &\times \big(1 + A(x_{n-1})(x_{n-1} - x_{n-2})\big) \cdots \big(1 + A(x_1)(x_1 - x_0)\big) \ , \end{aligned} \tag{3.5}$$

with $\|x_i - x_{i-1}\| \to 0$ for the individual segments as n increases, and where $x_0 = w(s_0)$, $x_1, \ldots, x_n = w(s_1)$ is a set of $n+1$ points ordered along the path w (Fig. 2).

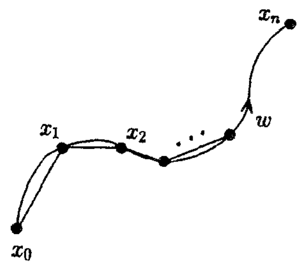

Fig. 2.

The holonomy U_w takes its values in G and transforms under gauge transformations (smooth functions $g\colon V \to G$) according to

$$U_w(s_1, s_0) \overset{g}{\longmapsto} g^{-1}\big(w(s_1)\big)U_w(s_1, s_0)g\big(w(s_0)\big) \ . \tag{3.6}$$

Note that this would not be true if we had allowed for discontinuities of the path w. The corresponding change of the gauge potential is straightforwardly computed from Eq. (3.2),

$$A_\mu(x) \overset{g}{\longmapsto} g^{-1}(x)A_\mu(x)g(x) + g^{-1}(x)\frac{dg}{dx^\mu} \ . \tag{3.7}$$

Another property of $U_w(s_1, s_0)$ following from (3.2) is its invariance under smooth orientation-preserving reparametrizations f of w, i.e.,

$$w(s) = w'\big(f(s)\big) \Longrightarrow U_w(s_1, s_0) = U_{w'}\big(f(s_1), f(s_0)\big) \ , \tag{3.8}$$

where $\frac{df}{ds} > 0$, $\forall s$. The term "non-integrable" (i.e., path-dependent) "phase factor" for U_w was introduced by Yang [5], in a generalization of its abelian version for $U(1)$-electromagnetism. It is a well-known result in mathematics that the connection A can (up to gauge transformations) be reconstructed from the knowledge of the holonomies of the *closed* curves based at a point $x_0 \in \Sigma$ (see [6, 7] for details on the concept of holonomy group and related issues).

The above description is valid only in a local neighborhood V of Σ. The appropriate global description is afforded by the mathematical theory of the principal fibre bundles $P\colon G \to P \to \Sigma$ over the base manifold Σ with typical fibre G, with A^P a connection one-form on P. Typically P has no global cross sections (this depends both on the group G and the manifold Σ), and then A^P can be identified with the one-form (3.1) only in a coordinate patch $V \subset \Sigma$, using a local cross section. The global implications of this fact are well known (see, for example, the discussion in [8]) and will not be addressed in this paper. However, there are important examples where global cross sections do exist (for instance, $G = SU(2)$ and any three-dimensional Σ) and hence all results described here are valid globally.

Note that for $F = 0$ (also called the case of a "flat connection"), where F is the field strength tensor,

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + ig\big[A_\mu(x), A_\nu(x)\big] \ , \tag{3.9}$$

U_w depends only on the homotopy of the path w.

4. WILSON LOOPS

In this section we will be working exclusively with loops, i.e., closed paths in Σ. For a loop γ based at x_0 there is a natural way of constructing a gauge-invariant quantity, namely,

$$T_A(\gamma) := \operatorname{Tr} U_\gamma = \operatorname{Tr} \operatorname{P} \exp \oint_\gamma A_\mu(x)\, dx^\mu \quad , \tag{4.1}$$

the ***traced holonomy*** or so-called *Wilson loop* (introduced by Wilson [9] as an indicator of confining behavior in lattice gauge theory). Under a gauge transformation, U_{γ,x_0} transforms homogeneously according to (3.5), but the matrices $g(x_0)$ and $g^{-1}(x_0)$ cancel each other upon taking the trace. Because of the cyclicity of the trace $T_A(\gamma)$ is independent of the choice of base point $x_0 = \gamma(0) = \gamma(1)$ of γ. Note that $T_A(\gamma)$ is invariant under orientation-preserving reparametrizations of γ and is therefore a function on *unparametrized loops*, i.e., reparametrization equivalence classes $\overline{\gamma}$, where any two members of $\overline{\gamma}$ are related by a smooth orientation-preserving bijection $b\colon S^1 \mapsto S^1$, $\frac{db}{ds} > 0$. Hence an unparametrized loop can be thought of as a continuous ordered set of points in Σ, with a (positive or negative) orientation assigned to each point in a continuous manner [as long as γ does not have any self-intersections or is a point loop (2.3)].

An unparametrized loop possesses certain (differential) topological properties (i.e., invariants under smooth diffeomorphisms of Σ), such as its number of self-intersection points, number of points of non-differentiability ("kinks"), its winding number, and its knot class [10, 11]. In concrete calculations involving an unparametrized loop $\overline{\gamma}$ one usually works with a chosen member γ from the equivalence class $\overline{\gamma}$ and then ensures the end result is independent of this choice. There have also been attempts to set up a loop calculus that is intrinsically reparametrization-invariant [12, 2].

5. THE SPACE OF ALL LOOPS

Since it will be of relevance to the field theoretic application later, let us try to give a mathematically meaningful definition of the "space of all loops." In mathematics, the *loop space* $\Omega_{x_0}X$ associated with some topological space X with distinguished base point x_0 is usually taken to be the function space

$$\Omega_{x_0}X = (X, x_0, x_0)^{(\mathrm{I},0,1)} \tag{5.1}$$

of continuous functions $\gamma\colon \mathrm{I} \mapsto X$ from the unit interval $\mathrm{I} = [0,1]$ to X such that $\gamma(0) = \gamma(1) = x_0$ [13]. In most of the following, we will omit the explicit reference to the base point x_0.

In order to have a notion of convergence for sequences of points in ΩX (i.e., sequences γ_i of functions, $i = 0, 1, 2, \ldots$) and a notion of continuity for functions $f\colon \Omega X \to Y$ into some space Y, one has to give ΩX a topological structure. A standard choice [13, 14] is the compact-open topology in which open sets are given by subsets of ΩX of the form

$$\begin{aligned} \Phi(J,O) := \big\{\, \gamma \mid \gamma(J) \subset O, J \text{ a closed interval in } [0,1], \\ O \text{ an open set in } X \,\big\} \quad , \end{aligned} \tag{5.2}$$

together with their unions and finite intersections [15]. It is straightforward to show that with this topology ΩX is Hausdorff (if we assume X to be Hausdorff), i.e., distinct points in ΩX possess disjoint neighborhoods. Note that ΩX is not, in general, a vector space since we cannot add elements of ΩX. However, if X is a linear space and if we choose $x_0 = 0$, we can obtain a linear structure on ΩX by defining addition and scalar multiplication pointwise:

$$\begin{aligned} (\gamma_1 + \gamma_2)(s) &:= \gamma_1(s) + \gamma_2(s) \quad , \\ (a\gamma)(s) &:= a \cdot \gamma(s), \quad a \in \mathbb{R}\,(\mathbb{C}) \end{aligned} \tag{5.3}$$

(the dot denoting scalar multiplication in X), which because of the continuity of these operations makes ΩX into a topological vector space. If X is a differential manifold Σ and we consider $\Omega\Sigma$ as consisting only of differentiable maps γ, $\Omega\Sigma$ can be made into an infinite-dimensional differential manifold (see [14] for a discussion of natural topologies on spaces of continuous and differentiable mappings). The finest topology one can impose on a loop space is the discrete topology, in which any element of $\Omega\Sigma$ is an open set. It has been used by Ashtekar and Isham in their treatment of representations of loop algebras [16, 17].

If for some reason one does not want to distinguish a base point in X, a natural loop space to work with is

$$\mathcal{L}X := \bigcup_{x\in X} \Omega_x X \ , \tag{5.4}$$

with $\Omega_x X$ as defined in (5.1). In physical applications one is usually interested in subspaces or quotient spaces of ΩX or $\mathcal{L}X$. Examples of the former are restrictions to the sets of loops without self-intersections, loops without kinks, or contractible loops. Typical quotient spaces are those of loops modulo orientation-preserving reparametrizations, or loops modulo constant translations as employed by Mensky [18–20] for the special case of $X = \mathbb{R}^n$. In these cases one has to check which properties of the original loop space (and the physical dynamics defined in terms of ΩX or $\mathcal{L}X$) are compatible with the restriction or the projection to the quotient space, respectively.

6. STRUCTURE ON LOOP SPACE

Let me emphasize that assigning well-defined mathematical properties to loop space is not a superfluous luxury but a necessity, for example, if one wants to set up a meaningful differential calculus on ΩX (cf. Sec. 11 below). From a physical point of view, it is of eminent importance to have additional (topological, algebraic, etc.) structure defined on ΩX, which is preserved (or approximately preserved) in the quantum theory.

A loop space ΩX is better than an arbitrary infinite-dimensional topological space because its elements can be composed, with the product map given by

$$(\gamma_1 \circ \gamma_2)(s) = \begin{cases} \gamma_1(2s), & 0 \leq s \leq \frac{1}{2} \\ \gamma_2(2s-1), & \frac{1}{2} \leq s \leq 1 \ . \end{cases} \tag{6.1}$$

This product is neither commutative nor associative, and we have neither a unit nor inverse loops in ΩX. (If we work with the loop space $\mathcal{L}X$, then composition is only defined if the end point of γ_1 coincides with the initial point of γ_2.) Taking the inverse γ^{-1} of a loop γ to be

$$\gamma^{-1}(s) := \gamma(1-s), \quad \forall s \ , \tag{6.2}$$

defines an involution on ΩX since $(\gamma^{-1})^{-1} = \gamma$, and $(\gamma_1 \circ \gamma_2)^{-1} = {\gamma_2}^{-1} \circ {\gamma_1}^{-1}$. Note that modifying a loop space to

$$\overline{\Omega}_{x_0} X := \bigcup_{\mathrm{I}} (X, x_0, x_0)^{([0,b],0,b)} \ , \tag{6.3}$$

with the union extending over all closed intervals $\mathrm{I} = [0, b]$, $b \geq 0$, and the product map defined by

$$(\gamma_1 \circ \gamma_2)(s) = \begin{cases} \gamma_1(s), & 0 \leq s \leq b_1, \\ \gamma_2(s), & b_1 \leq s \leq b_1 + b_2 \ , \end{cases} \tag{6.4}$$

which is associative (though not commutative), and with the unit given by the trivial loop $\gamma_0 \colon [0,0] \mapsto x_0$, $\overline{\Omega} X$ becomes a topological semigroup (see [21] for more comments on associativity, in the context of string theory).

One way ΩX may inherit structure is from X, for example, from a Riemannian metric on X. Bryzinski [22] and Schäper [23] describe the space of smooth (non-self-intersecting) loops as the space of embeddings of the circle S^1 into the manifold Σ. They use the fact that this space can be described as a (Fréchet) fibre bundle, with typical fibre $Diff^+(S^1)$, and the base space given by the unparametrized loops, in order to define further mathematical structures on it. (Their papers also contain references to the mathematical literature.) Another case we will be concerned with is when a quotient space $\Omega X/\sim$ acquires an algebraic structure via certain functions defined on $\Omega\Sigma$, as, for example, the Wilson loop function.

In general, it is hard to find any kind of structure on an infinite-dimensional space like the loop space. Coming from a physics point of view may give us new ideas about what type of structures to look for and how to construct them.

7. IDENTITIES SATISFIED BY (TRACED) HOLONOMIES

We now give some more properties of the untraced and traced holonomies, which follow from their definitions (3.4) and (4.1). Take G to be the gauge group $GL(N,\mathbb{C})$ or one of its subgroups ($U(N)$, $SU(N)$, $SO(N)$, etc.) in its fundamental representation in terms of $N\times N$ (complex) matrices. For two loops γ_1 and γ_2 based at x, we have

$$U_{\gamma_1\circ\gamma_2}(A)=U_{\gamma_2}(A)\cdot U_{\gamma_1}(A),\quad \forall A\ , \tag{7.1}$$

with the loop product as defined in (6.1), the dot denoting matrix multiplication, and abbreviating $U_\gamma(0,1)=:U_\gamma$. In other words, the mapping U is compatible with the product structure on loop space. For the inverse loop γ^{-1} of γ, (6.2), we have

$$U_{\gamma^{-1}}=(U_\gamma)^{-1}\ . \tag{7.2}$$

Here and in the rest of this section, the dependence of U_γ and $T(\gamma)$ on A is understood. Since (7.2) holds also for open paths γ, we derive the important *retracing identity*

$$U_\gamma=U_{\gamma'},\quad \text{for}\quad \gamma'=\left((\gamma_1\circ w)\circ w^{-1}\right)\circ\gamma_2\ , \tag{7.3}$$

where $\gamma_1\circ\gamma_2=\gamma$, w is an open path "glued to γ" (see Fig. 3), and we have generalized the composition law (6.1) to open paths.

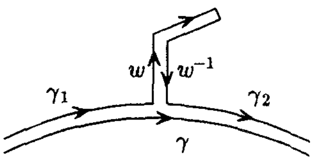

Fig. 3.

Although the composition of loops is non-associative, it is easily seen that $U_{\gamma'}$ in (7.3) is independent of the order of composition of the loop segments; for example, we could have chosen $\gamma'=\gamma_1\circ\left(w\circ(w^{-1}\circ\gamma_2)\right)$ instead. In fact, viewing U as a G-valued function on $\mathcal{A}\times\Omega\Sigma$ induces (as alluded to in Sec. 6) an equivalence relation between elements of $\Omega\Sigma$ via

$$\beta\sim\gamma\quad \text{if}\quad U_\beta=U_\gamma,\quad \forall A\in\mathcal{A}\ . \tag{7.4}$$

We denote the equivalence class of a loop γ by $\overline{\gamma}$. The composition law (6.1) is compatible with the equivalence relations ($\overline{\gamma}_1\circ\overline{\gamma}_2:=\overline{\gamma_1\circ\gamma_2}$ is independent of the representatives chosen), and induces a *group* structure on the set of equivalence classes $\Omega_{x_0}\Sigma/\sim$, with unit $\overline{p}_{x_0}$, the equivalence class of the constant loop (2.3). Thus we have

$$\begin{aligned}\overline{\beta}&=\overline{\gamma}\qquad \text{if}\ \beta=\gamma\text{mod retracing/reparametrization},\\ \overline{\gamma}\circ\overline{\gamma}^{-1}&=\overline{p}_{x_0},\\ (\overline{\alpha}\circ\overline{\beta})\circ\overline{\gamma}&=\overline{\alpha}\circ(\overline{\beta}\circ\overline{\gamma})\qquad (\text{associativity})\ .\end{aligned} \tag{7.5}$$

The group obtained this way is called the "group of loops" by Gambini and Trias [24]; it is non-abelian and not a Lie group (it also is mentioned in [25] in a somewhat different context). Furthermore, as was shown by Durhuus and Leinaas [26], it is not locally compact (in a natural topology), which means it is "very large" even locally.

There are analogous identities satisfied by the traced holonomies, characterizing them as a particular subset of complex-valued functions on $\Omega\Sigma \times \mathcal{A}$ (more precisely, $\Omega\Sigma \times \mathcal{A}/\mathcal{G}$ because of their gauge invariance, with $\mathcal{G}$ denoting the infinite-dimensional group of local gauge transformations). Independently of the gauge group G, we have the identities

$$T_A(\gamma_1 \circ \gamma_2) = T_A(\gamma_2 \circ \gamma_1) \tag{7.6}$$

because of the cyclicity of the trace, and again a retracing identity,

$$T_A(\gamma) = T_A(\gamma'), \quad \gamma, \gamma' \text{ related as in (7.3) .} \tag{7.7}$$

Another set of identities are the so-called Mandelstam constraints, whose form depends on the dimension N of the group matrices. They can be systematically derived from the identity of N-dimensional δ-functions,

$$\sum_{\pi \in \mathcal{S}_{N+1}} (-1)^{\sigma(\pi)}\, \delta_{i_1,\pi(j_1)} \cdots \delta_{i_{N+1},\pi(j_{N+1})} = 0, \quad i_k, j_k = 1, \ldots, N \ , \tag{7.8}$$

with the sum running over all permutations π of the symmetric group of order $N+1$, $\mathcal{S}_{N+1}$, and $\sigma(\pi)$ denoting the parity of the permutation. Contracting $N+1$ holonomy matrices U_γ with (7.8) results in a trace identity for (combinations of) $N+1$ loops. For $N=1$, we have

$$T_A(\alpha) T_A(\beta) - T_A(\alpha \circ \beta) = 0 \ , \tag{7.9}$$

and for $N=2$,

$$\begin{aligned} T_A(\alpha) T_A(\beta) T_A(\gamma) - T_A(\alpha\beta) T_A(\gamma) - T_A(\beta\gamma) T_A(\alpha) \\ - T_A(\alpha\gamma) T_A(\beta) + T_A(\alpha\beta\gamma) + T_A(\alpha\gamma\beta) = 0 \ , \end{aligned} \tag{7.10}$$

etc., where we have omitted the symbol $\circ$ denoting loop composition. Note that the Mandelstam identities are non-linear algebraic equations on the functions T. If we want to consider traced holonomies of specific subgroups of $GL(N, \mathbb{C})$, there will be more identities satisfied by T, for example, deriving from the condition $\det U_\gamma = 1$ (see [27, 28] and the next section for some selected cases). Again we may use the functions T to define an equivalence relation on the loop space $\Omega\Sigma$ by

$$\beta \sim \gamma \quad \text{if} \quad T_A(\beta) = T_A(\gamma), \quad \forall A \ . \tag{7.11}$$

The composition law for equivalence classes $\overline{\gamma}$,

$$\overline{\gamma}_1 \circ \overline{\gamma}_2 := \overline{\gamma_1 \circ \gamma_2} \ , \tag{7.12}$$

induces an abelian group structure on $\Omega\Sigma/\sim$ if, in addition, the relation $T_A(\alpha) = T_A(\beta) \Rightarrow T_A(\alpha \circ \gamma) = T_A(\beta \circ \gamma)$ is satisfied [28, 16].

8. EQUIVALENCE BETWEEN GAUGE POTENTIALS AND HOLONOMIES

The importance of the (traced) holonomies lies in the fact that one can reconstruct gauge-invariant information about the gauge potential A from them. More precisely, starting from a given configuration $A(x)$ defined on Σ, with $A(x)$ taking values in the defining representation of $GL(N, \mathbb{C})$ or one of its subgroups, and computing $U_\gamma(A)$ for all $\gamma \in \Omega\Sigma$ according to (3.4), one can reconstruct the original configuration A up to a gauge transformation $g(x)$ with $g(x_0) = e$ [29].

Our main interest here is the analogous reconstruction theorem for the traced holonomies. It turns out that, starting from a configuration $A(x)$ as above, and computing $T_A(\gamma)$ according to (4.1) for all $\gamma \in \Omega\Sigma$, one can in certain cases, depending on the group G, reconstruct the configuration A, again up to gauge transformations.

Alternatively, given any complex-valued function $F(\gamma)$ on loop space $\Omega\Sigma$ satisfying the Mandelstam identities of order N [and possibly some additional identities characterizing a specific subgroup of $GL(N, \mathbb{C})$], retracing and reparametrization invariance, Eq. (7.6), and appropriate smoothness conditions, one can construct $N \times N$ matrices $U_\gamma \in GL(N, \mathbb{C})$ (or of the subgroup in question) such that the traces of products $U_{\gamma_j} \cdot U_{\gamma_i} \cdot \ldots$ are exactly given by $F(\cdots \circ \gamma_i \circ \gamma_j)$, from which then the configuration A (modulo gauge transformations) can be computed as before.

The explicit construction of the holonomies from the traced holonomies has been given in the remarkable work of Giles [29]. What has not been proven is that by running through all admissible $F(\gamma)$'s one recovers indeed *all* possible holonomy configurations U_γ. In general, the U_γ's so obtained will form a subgroup of G. However, this can only happen if G is non-compact, a well-studied example being that of $G = SL(2, \mathbb{C})$ [30].

For the case where one does recover all of G this way, the importance of this equivalence proof lies in the fact that all gauge-invariant statements in terms of the connection A are, in principle, expressible in terms of the traced holonomies $T(\gamma)$, more precisely, the non-linear subspace of the loop functions which satisfy the Mandelstam constraints associated with the gauge group G.

Because of their importance in many applications, we will give the Mandelstam constraints for $G = SU(2)$ explicitly. One finds

$$T(\alpha \circ \beta) + T(\alpha \circ \beta^{-1}) = T(\alpha)T(\beta) \ , \tag{8.1}$$

$$T(\alpha) \in \mathbb{R}, \quad -2 \leq T(\alpha) \leq 2 \ . \tag{8.2}$$

The geometric loop configuration relevant for (8.1) is that of two loops α and β intersecting in a point, as illustrated in Fig. 4. Note that from the first equation it follows that $T(\alpha^{-1}) = T(\alpha)$, $T(p) = 2$ (where p is any loop from the reparametrization equivalence class $\bar{p}_{x_0}$ of the point loop) and $T(\alpha)^2 - T(\alpha^2) = 2$. The reality of $T(\alpha)$ and the inequality (8.2) are a result of unitarity.

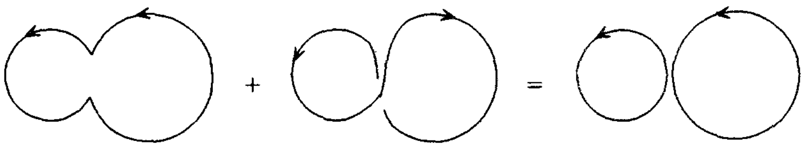

Fig. 4.

Again we see that the Mandelstam constraints associated with a given gauge group G via (7.11) may be used to define an equivalence relation on loop space (which is different for different G). In turn it is this quotient space of a loop space that must be used as a domain when one builds up a formulation of G-gauge theory entirely in terms of Wilson loops $T(\alpha)$.

Note that the Mandelstam relations above hold independently of A, and hence can be regarded as functions on $\Omega\Sigma$ alone, and not on $\Omega\Sigma \times \mathcal{A}/\mathcal{G}$ (this enabled us to interpret them as *constraints* on loop space functions, rather than as *identities* on the loop-dependent functions on the space of connections). However, for certain "degenerate" configurations $A(x)$ there may be more Mandelstam-type relations satisfied by the $T_A(\alpha)$ which depend explicitly on a configuration in $\mathcal{A}/\mathcal{G}$ and therefore cannot be interpreted in this way.

9. PHYSICAL INTERPRETATION OF HOLONOMY

It is well known that the existence of a connection on a manifold enables one to define the notion of *parallel transport.* For a field $\Psi(x)$ transforming according to the defining representation of the gauge group G,

$$\Psi(x) \xmapsto{g} g^{-1}(x)\Psi(x) \ , \tag{9.1}$$

we can compare fields at different points x and y by parallel-transporting $\Psi(y)$ from y back to x using the holonomy $U_w(s_1, s_0)$ along a path w with $w(s_1) = x$, $w(s_0) = y$, to obtain

$$\Psi_{\mathrm{par},w}(x) := U_w(s_1, s_0)\Psi(y) \ , \tag{9.2}$$

which has now the same behavior under gauge transformations as $\Psi(x)$. In the limit as the length of w goes to zero, this procedure leads to the definition of the covariant derivative of Ψ [31].

The holonomy of small closed loops measures the *curvature (or field strength) in internal space.* For a small square loop γ of side length ϵ in a coordinate chart around the point $x \in \Sigma$, the base point of γ, the holonomy U_γ can be expanded as

$$U_\gamma = \mathbf{1} + ig\, F^a_{\mu\nu}(x) X_a\, \epsilon^2 + O(\epsilon^3) \ , \tag{9.3}$$

where γ is defined by its four corners, $(x, x+\epsilon\mathbf{e}_\mu, x+\epsilon\mathbf{e}_\mu+\epsilon\mathbf{e}_\nu, x+\epsilon\mathbf{e}_\nu)$, with $\mathbf{e}_\mu$ denoting the unit vector in the μ-direction. Note that in non-abelian gauge theory, $F_{\mu\nu}(x) = F^a_{\mu\nu}(x)X_a$ is not an observable, since it transforms non-trivially under gauge transformations. Performing a similar expansion for the traced holonomy of the infinitesimal loop γ, we obtain

$$\begin{aligned} T_A(\gamma) = \operatorname{Tr} U_\gamma &= N + ig \sum_a F^a_{\mu\nu}(x) \operatorname{Tr} X_a\, \epsilon^2 \\ &+ (ig)^2 \sum_{a,b} F^a_{\mu\nu}(x) F^b_{\mu\nu}(x)\ \operatorname{Tr} X_a X_b\, \epsilon^4 + O(\epsilon^5) \ , \end{aligned} \tag{9.4}$$

(no sum over μ, ν). For a semi-simple Lie algebra $\mathbf{g}$ we can always find a basis of generators X_a such that $\operatorname{Tr} X_a X_b = \delta_{ab}$. For $G = SU(N)$, moreover, we have $\operatorname{Tr} X_a = 0$ because of unitarity, and the expansion therefore reduces to

$$T_A(\gamma) = N + (ig)^2 \sum_a F^a_{\mu\nu}(x) F^a_{\mu\nu}(x)\ \epsilon^4 + O(\epsilon^5) \ . \tag{9.5}$$

These expansions illustrate the way local gauge-invariant information about the curvature or field strength F is contained in $T_A(\gamma)$.

Having described the basic mathematical properties of the holonomy and the traced holonomy, there remains the question of the physical interpretation of the underlying loop space itself, and therefore of the way in which a physical theory is to be constructed on it. Formulating a theory in terms of non-local variables depending on loops is potentially very different from the usual local formulations. It makes a difference whether one wants to use loops as convenient auxiliary labels in an otherwise local formulation, or postulates them to be the basic entities of a new, intrinsically non-local description of gauge theory or gravity. There is a variety of ways in which past research has made use of path- and loop-dependent quantities, some examples of which will be described below. It is important to realize that different a priori physical interpretations of the role of loops will inevitably lead to different mathematical structures employed in their description, for example, the initial choice of a specific loop space or quotient of loop space. If one thinks of paths or loops as describing actual trajectories of charged particles [31, 9], say, one may work with smooth C^∞-loops in a (semi-)classical description and nowhere differentiable loops (which are supposed to give the main contribution to the Feynman path integral) in the quantum theory.

Many conceptual issues still remain unanswered. It is not clear whether a non-local description in terms of holonomies, as introduced above, is necessarily tied to the *quantum* aspects of the theory. Also, the non-linearities of Yang–Mills theory seem to play a role, since the abelian $U(1)$-theory does not require such a description. One may also ask how big loops are, and whether they have a preferred size in a given (gauge) theory. Is it physically sensible to treat the (traced) holonomies as the basic variables (Polyakov's "rings of glue" [32]), or are genuine physical observables again *composites* of the elementary Wilson loops? The following sections will deal with some issues peculiar to loop formulations of Yang–Mills theory. They are not intended as an exhaustive treatment, but as an introduction and a guide to the literature.

10. CLASSICAL LOOP EQUATIONS

In this section I give some examples of path-dependent formulations for gauge theory, starting from Mandelstam's early treatment of electrodynamics coupled to a scalar field, and its subsequent generalization to the non-abelian case. This illustrates how classical equations of motion for path-dependent variables may be obtained.

Mandelstam's description of "QED without potentials" [33] is interesting in the present context, because it is the first instance of the use of holonomy-dependent field variables. He demonstrates that a gauge-invariant and Lorentz-covariant formulation of electrodynamics (avoiding the unphysical negative-norm states of the Gupta–Bleuler approach) is possible, *provided* one allows for non-local matter field variables. Instead of using the gauge potential $A_\mu(x)$ and the charged scalar field $\phi(x)$, he works with the field strength $F_{\mu\nu}(x)$ and objects

$$\Phi(\gamma, x) := \phi(x) \exp\left\{ -ie \int_{-\infty}^{x} A_\mu \, d\gamma^\mu \right\} \tag{10.1}$$

as the basic variables, where γ denotes a space-like path in Minkowski space originating at spatial infinity and ending at the point x. These variables are obviously invariant under the gauge transformations

$$\Phi \to e^{ies}\Phi, \qquad A_\mu \to A_\mu + \partial_\mu s \ , \tag{10.2}$$

where s is an arbitrary scalar field satisfying the boundary condition $s(\pm\infty) = 0$. Mandelstam emphasizes that the need for a gauge-invariant formulation arises in the quantum theory, because only commutators between (gauge-invariant) observables are well defined. He argues that the inherent path dependence of the matter field variable is natural and corroborated by the Aharonov–Bohm effect.

Bialnycki-Birula [34] extended these ideas to formulate what one might call "Yang–Mills theory without field strengths." Since the non-abelian gauge field interacts with itself, there is in a first step no need to couple it to an external matter field in order to make the theory non-trivial. Since the field variables $F_{\mu\nu}(x)$ are themselves not observables (i.e., not gauge-invariant), Bialnycki-Birula introduces in analogy with (10.1) the gauge-invariant fields

$$\mathcal{F}_{\mu\nu}(\gamma, x) := \mathrm{P}\exp\left\{ -ig \int_{-\infty}^{x} A_\mu d\gamma^\mu \right\} F_{\mu\nu}(x) \ \mathrm{P}\exp\left\{ ig \int_{x}^{-\infty} A_\mu d\gamma^\mu \right\} \ , \tag{10.3}$$

which makes use of the holonomies of the space-like paths γ and γ^{-1}, as illustrated by Fig. 5.

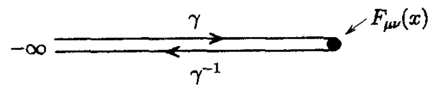

Fig. 5.

The usual classical Yang–Mills equations of motion

$$\mathcal{D}_\mu F_a^{\mu\nu}(x) = \big(\partial_\mu \delta_{ac} - g\epsilon_{abc}A_{b\mu}(x)\big) F_c^{\mu\nu}(x) = 0 \ , \tag{10.4}$$

with ϵ_{abc} denoting the structure constants of $\mathbf{g}$, translate to

$$\partial_\mu(x)\mathcal{F}^{\mu\nu}(\gamma,x)=0 \ , \tag{10.5}$$

where the differential $\partial_\mu(x)$ acts on the holonomy $U_{\gamma,x}$ as an "endpoint derivative" (see Eq. (11.1) for a definition). The Bianchi identities (which are satisfied automatically in the connection formulation) have to be imposed as separate equations,

$$\partial_\lambda(x)\mathcal{F}_{\mu\nu}(\gamma,x)+\partial_\mu(x)\mathcal{F}_{\nu\lambda}(\gamma,x)+\partial_\nu(x)\mathcal{F}_{\lambda\mu}(\gamma,x)=0 \ . \tag{10.6}$$

It may be somewhat surprising that Eqs. (10.5) and (10.6), unlike Eqs. (10.4), are *linear* in the basic field variables. Mandelstam [35] writes: "The field equations are simpler *in appearance* than the Maxwell equations of electrodynamics, since there is no additional current term." He is obviously aware of the fact that difficulties associated with the non-linear functional form of the Yang–Mills equations must be hidden in (10.5), however, neither spelling out in what sense non-triviality arises, nor making any further use of the equations (10.5) classically (they are used to obtain equations for the corresponding path-dependent Green's functions in the quantum theory though).

Different (gauge-*co*variant) field variables are used by Polyakov [36, 31], namely,

$$\mathcal{F}_\mu(\gamma,s):=U_\gamma(0,s)\,F_{\nu\mu}(\gamma(s))\,\dot\gamma^\nu(s)\,U_\gamma(s,0) \ , \tag{10.7}$$

where γ is a closed loop, s some intermediate parameter value, $s\in[0,1]$, and $U_\gamma(0,s)$ the holonomy along the portion $[0,s]$ of the loop γ (Fig. 6).

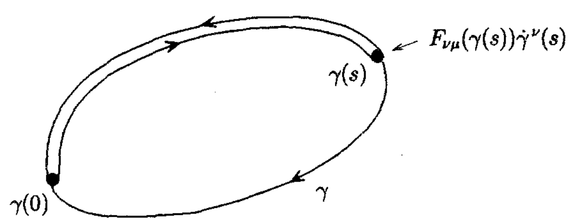

Fig. 6.

The field equations

$$\frac{\delta}{\delta x^\mu(s)}\,\mathcal{F}^\mu(\gamma,s):=\dot\gamma^\lambda(s)\frac{\delta}{\delta\sigma^{\lambda\mu}(\gamma(s))}\mathcal{F}^\mu(\gamma,s)=0 \tag{10.8}$$

correspond to the Yang–Mills equations projected onto the tangent vector $\dot\gamma^\nu$ of γ at s. The differential operator $\frac{\delta}{\delta\sigma^{\lambda\mu}(\gamma(s))}$ is the projection of the area derivative of definition (11.2), obtained by adding an infinitesimal loop in the μ-direction at the point s. Note that $\mathcal{F}_\mu$ still depends on the curve parameter s; parametrization invariance and the (projected) Bianchi identities have to be imposed as separate equations in addition to (10.8). The reason why this reformulation of Yang–Mills theory was thought to be appealing is its close formal resemblance with the two-dimensional non-linear σ-model (see also the work of Aref'eva [37–39] for a related treatment). It turned out that this simple analogy does not hold, essentially because the theory is not defined on ordinary space, but on loop space. The loop equations (10.8), unlike their analogues for the classical non-linear σ-model, do not contain enough information for constructing a set of conserved currents [40]. In Polyakov's words, the (up-to-date) failure of this approach "has to do with the difficulties we experience in treating equations in loop space, or, what is more or less the same, with string dynamics" [31].

A further example of loop-dependent equations of motion for Yang–Mills theory, this time in terms of the traced holonomies (4.1), is afforded by the classical Makeenko–Migdal equations

$$\partial^\mu(x)\,\frac{\delta}{\delta\sigma^{\mu\nu}(x)}\,T(\gamma)=0 \ , \tag{10.9}$$

with $\partial^\mu(x)$ denoting the ordinary differential on space-time and $\frac{\delta}{\delta\sigma^{\mu\nu}(x)}$ Mandelstam's area derivative (see the next section for a definition). What is puzzling about this equation is its linearity in the field variable T; again the non-linearities seem to have disappeared. In fact, there are spurious solutions to (10.9) [2], which have to be eliminated by other means. This feature is attributed "partly to the presence of the Mandelstam constraints," which have not been taken into account in the derivation of the equation. Migdal concludes that "the classical loop dynamics is quite complicated and implicit, but presumably it is irrelevant as well as the classical colour dynamics," the latter statement being, of course, a debatable issue.

11. DIFFERENTIAL OPERATORS

In order to define equations of motion in a path-dependent approach, one needs to have a notion of differentiation. The properties of differential operators are intimately tied to the space they act upon; hence path and loop derivatives assume different meanings in different contexts. Unfortunately those distinctions are often not clearly stated in the literature, nor are the relevant spaces of loop functions and the underlying loop spaces.

Recall that for differentiation on some space X to be well defined, one needs at least some topological vector space structure on X [8]. If one is lucky, X can be made into a Banach space (i.e., a complete normed vector space), in which case most of the differential calculus on $\mathbb{R}^n$ can be generalized in a straightforward way to X. Also, the norm induces a translationally invariant metric and a natural topology on X. For more general topological vector spaces one may still be able to define differentiation, but there are, in general, *no inverse and implicit function theorems and no theorems on the existence and uniqueness of solutions of differential equations.*

In physical applications, X is usually some space of loop functions or functionals, i.e., essentially an infinite-dimensional function space with a vector space structure, but is rarely given any further structure, for example, a topology. A similar statement concerns the loop and path spaces themselves. In some of the previous sections we described the problem of "giving structure" to these spaces, which in turn is an obstruction to defining a meaningful differential calculus on them.

In the following I will describe some typical path-dependent differential operators, and outline some of the problems associated with them. For a function $F(w, x)$ depending on an (unparametrized) path w, with x denoting one of its endpoints, we define an "endpoint derivative"

$$\partial_\mu(x)\, F(w,x) := \lim_{dx_\mu \to 0} \frac{F(w', x + dx_\mu) - F(w,x)}{dx_\mu} \ , \tag{11.1}$$

where w' is obtained by adding to w an infinitesimal straight line element dx_μ in the μ-direction. It is used, for example, by Bialnycki-Birula [34] and Mandelstam [35, 41] for the special case where F is the holonomy $U_{\gamma,x}$ of a path starting at spatial infinity and going to the point x. Gambini and Trias call it "Mandelstam's covariant derivative" and use it in a generalized context where F is a G-valued function on the set of open paths modulo reparametrizations and retracing [24].

Another frequently used differential operator is the so-called area derivative. For a path-dependent function $F(w)$ it is usually defined as

$$\frac{\delta}{\delta\sigma^{\mu\nu}(x)} F(w) := \lim_{d\sigma^{\mu\nu} \to 0} \frac{F(w \circ_x \gamma_{\mu\nu}) - F(w)}{d\sigma^{\mu\nu}} \ , \tag{11.2}$$

where $\gamma_{\mu\nu}$ is an infinitesimal planar loop in the μ–ν-plane attached (by path composition) to the path w at the point x on w. In a local coordinate chart $\{x_\mu\}$, the area of the small loop is given by

$$d\sigma^{\mu\nu} = \frac{1}{2} \int_{\gamma_{\mu\nu}} x^\mu \, dx^\nu \ , \tag{11.3}$$

which is antisymmetric in μ and ν. For the special case where $F(w) = U_w$, the holonomy of a path w with initial point x_0 and endpoint x_1, we have

$$\frac{\delta}{\delta\sigma^{\mu\nu}(x)} U_{w;x_0,x_1} = U_{w;x_0,x}\, F_{\mu\nu}(x)\, U_{w;x,x_1} \ , \tag{11.4}$$

whence the area derivative is automatically antisymmetric in μ and ν. For more general cases of loop functions F, one may have to introduce an explicit antisymmetrization in (11.2) in order for the definition to make sense. Further discussions about path and area derivatives, in the context of both gauge theory and canonical gravity (where, in addition, one must take into consideration diffeomorphism invariance) can be found in [42, 43, 44, 2, 32].

Let me again emphasize that for general loop functions there is no reason for the limits in (11.1) and (11.2) to be well defined and exist. Furthermore, if we talk about "infinitesimal loops," this implies that we have chosen some topology on loop space which tells us about small variations, i.e., what it means for two loops to be infinitesimally close to each other. Our "intuitive" notion of closeness of two paths is that coming from viewing them as embedded in the manifold Σ and using the Euclidean metric of $\mathbb{R}^n$ in local charts of Σ. However, this may not be the appropriate thing to do. For example, in the context of the "group of loops" introduced in Sec. 7 above, two paths are considered equivalent if they are in the same class with respect to retracing. This leads to a generalized and non-local (with respect to $\mathbb{R}^n$) notion of closeness. The natural loop space metric and topology in this context are discussed by Durhuus and Leinaas [26]. Since the issue of giving topological structures to infinite-dimensional spaces is mathematically involved, there is a genuine need for physical arguments to restrict the possible choices. This seems to be the only way to obtain meaningful equations of motion in a path-dependent approach to gauge theory.

12. LATTICE GAUGE THEORY

The only way of obtaining quantitative results about the behavior of non-abelian gauge theory, such as the values of hadron masses, and testing the hypothesis of quark confinement, is in a regularized version of the theory, where continuous space-time is approximated by a finite hypercubic lattice. The basic gauge field variables in this case are the link holonomies U_l, i.e., gauge potentials integrated over elementary lattice links. Wilson, who first proposed this approach to gauge theory [9], identified closed flux lines on the lattice as natural gauge-invariant objects, and introduced a discretized, Euclidean form of the Yang–Mills action, which in the continuum limit (as the length a of lattice links goes to zero) reduces to the ordinary one. For gauge group $SU(N)$, it is given essentially as the sum over all lattice plaquettes (elementary square loops made up of four contiguous oriented links) P of the traces of the corresponding holonomies $U(P)$,

$$S_W(U_l) = -\frac{1}{Ng^2}\sum_P (N - \mathrm{Tr}\ U_P). \tag{12.1}$$

The plaquette holonomy U_P is to be thought of as the product of the four link holonomies U_l associated with P (see Fig. 7, where we have $U_P = U_{l_1}U_{l_2}U_{l_3}U_{l_4}$).

The underlying physical interpretation for holonomies of closed paths is that of weight factors in the Feynman path integral, associated with classical quark trajectories. More precisely, in order to compute the current–current propagator for quark fields between two points 0 and x in space-time, one has to average over all possible classical quark trajectories and classical gauge field configurations. The relevant quark configurations are pairs of quarks created at the origin 0 and annihilated at x (Fig. 8a),
the weight for such a pair being exactly the holonomy U_γ along the loop formed by the pair of quark trajectories. In the strong coupling limit of the lattice gauge theory ($g \to \infty$), one can produce arguments that the gauge field average of $U\gamma$ for a fixed lattice loop γ behaves as $\exp icA(\gamma)$, with $A(\gamma)$ denoting the area enclosed by the loop γ, and constant c. This suggests confining behavior for quarks, because large areas $A(\gamma)$ (corresponding to the quarks being far apart) are strongly suppressed in the sum over all paths, whereas narrow "flux tubes" (Fig. 8b) are favored [9]. Note that this argument was made within the pure gauge theory.

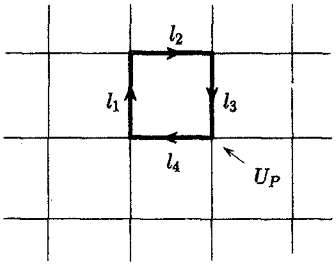

Fig. 7.

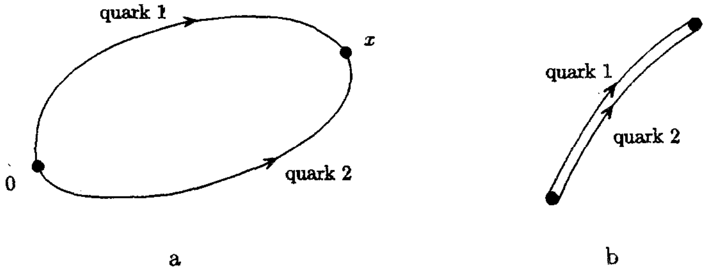

Fig. 8.

In the Hamiltonian formulation of the lattice theory due to Kogut and Susskind, the (spatial) link holonomies again play an important role [3]. Together with an appropriate set of conjugate momentum variables they form a closed Poisson-bracket algebra, which is of the form of a semi-direct product. This algebra is then quantized, leading to a representation where the quantized link holonomies $\widehat{U}_l$ are diagonal. The wave functions in the Hilbert space are not gauge-invariant, and the physical subspace has to be projected out by imposing the Gauss law constraint as a Dirac condition on quantum states. It turns out that gauge-invariant states can be labelled by closed paths of links on the lattice.

However, with growing lattice size, selecting the physical subspace and calculating the action of the Hamiltonian on it become quickly involved. Furthermore this set of variables does not seem particularly suited for treating the weak coupling limit. For that reason, there have been various alternative proposals to formulate the Hamiltonian lattice theory in an explicitly gauge-invariant manner [45–48]. They involve Hilbert spaces of gauge-invariant states labelled by loops, and the action of the Hamiltonian operator typically results in geometric deformations or rearrangements (fusion, fission, etc.) of these loop arguments.

Unfortunately there is now another drawback, since such bases of loop states are vastly overcomplete (due to the Mandelstam constraints), which renders the physical interpretation of the geometric picture of "loops interacting through the action of the Hamiltonian" rather obscure. The problem is how to isolate in an efficient way a set of independent loop states, since the number of loops on the lattice grows very fast with growing lattice size. A frequently used approximation employs a truncated loop basis, which only considers states labelled by loops that are shorter than a given number n of links [45, 48], and calculates the spectrum of the Hamiltonian in this restricted context. This seems to work reasonably well, at least for some sectors of the theory. However, what one ideally would like to have is a formulation directly in terms of the physical degrees of freedom, which is gauge-invariant and has a minimal redundancy with respect to the Mandelstam constraints. Such a set of variables has been given in [49], and recent results, obtained in a Lagrangian context, show that it is indeed possible to perform calculations directly in terms of these independent loop variables [50].

13. LOOP ALGEBRAS

This section should more precisely be called "algebras of loop- and path-dependent functions." The algebras we shall be concerned with involve holonomies and are defined on the configuration or phase space of Yang–Mills theory. The first such algebra was written down by Mandelstam in a Lagrangian framework [35]. Splitting $\mathcal{F}_{\mu\nu}(\gamma,x) = \mathcal{F}^a_{\mu\nu}(\gamma,x)X_a$ [cf. (10.3)] into its components $\mathcal{F}_{ij}$ and $\mathcal{F}_{0i}$, $i = 1, 2, 3$, he finds for the equal-time commutators

$$\begin{aligned}
\left[\mathcal{F}^a_{ij}(\gamma,x),\mathcal{F}^b_{kl}(\gamma',x')\right] &= 0 \ , \\
\left[\mathcal{F}^a_{0i}(\gamma,x),\mathcal{F}^b_{jk}(\gamma',x')\right] &= -i\delta_{ab}(\delta_{ik}\partial_j - \delta_{ij}\partial_k)\delta^3(x-x') \\
&\qquad + i\epsilon_{abc}\int_{\gamma'} d\xi_i\, \delta^3(x-\xi)\mathcal{F}^c_{jk}(\gamma',x') \ , \\
\left[\mathcal{F}^a_{0i}(\gamma,x),\mathcal{F}^b_{0j}(\gamma',x')\right] &= i\epsilon_{abc}\int_{\gamma'} d\xi_i \delta^3(x-\xi)\mathcal{F}^c_{0j}(\gamma',x') \\
&\qquad + i\epsilon_{abc}\int_{\gamma} d\xi_j\, \delta^3(x'-\xi)\mathcal{F}^c_{0i}(\gamma,x) \ .
\end{aligned} \tag{13.1}$$

The relations (13.1) define a closing algebra (if we include the unit generator appearing on the right-hand side of the second equation), i.e., the commutator of two $\mathcal{F}$'s is again proportional to an algebra generator. A characteristic feature is the fact that the structure constants of the algebra are distributional and depend on the path configurations appearing as arguments of the $\mathcal{F}$-variables.

In the Hamiltonian formulation in terms of traced holonomies, one finds a similar algebra structure after introducing conjugate holonomy variables on the Yang–Mills phase space (in the $A_0 = 0$-gauge) coordinatized by the pairs $\left(A^a_i(x), E^i_a(x)\right)$, with canonical Poisson brackets $\left\{A^a_i(x), E^j_b(y)\right\} = \delta^a_b\delta^j_i\delta^3(x-y)$. They depend on a loop γ, a marked point $\gamma(s)$ on γ, and both the gauge potential and the generalized electric field, and are defined as [28, 51]

$$T^i_{A,E}(\gamma,s) := \operatorname{Tr} U_\gamma(s,s)E^i\left(\gamma(s)\right) \ . \tag{13.2}$$

For the special case of $SU(2)$, computing the Poisson brackets on phase space of these loop variables leads to

$$\begin{aligned}
&\left\{T(\gamma),T(\gamma')\right\} = 0 \ , \\
&\left\{T^i(\gamma,s),T(\gamma')\right\} = -\Delta^i(\gamma',\gamma(s)) \\
&\qquad \times \left(T(\gamma\circ_s\gamma') - T(\gamma\circ_s\gamma'^{-1})\right) \ , \\
&\left\{T^i(\gamma,s),T^j(\gamma',t)\right\} = -\Delta^i(\gamma',\gamma(s)) \\
&\qquad \times \Big(T^j\left(\gamma\circ_s\gamma',u(t)\right) + T^j\left(\gamma\circ_s\gamma'^{-1},u(t)\right)\Big) \\
&\qquad + \Delta^j\left(\gamma,\gamma'(t)\right)\Big(T^i\left(\gamma'\circ_t\gamma,v(s)\right) - T^i\left(\gamma'\circ_t\gamma^{-1},v(s)\right)\Big) \ .
\end{aligned} \tag{13.3}$$

The structure constants Δ are again distributional:

$$\Delta^i(\gamma,x) = \oint_\gamma dt\, \delta^3\left(\gamma(t),x\right)\dot\gamma^i(t) \ . \tag{13.4}$$

In its general form for gauge group $SU(N)$, the algebra (13.3) was first introduced by Gambini and Trias [28]. For the case of $G = SL(2,\mathbb{C})$ [for which the algebra coincides with (13.3)], it was later rediscovered by Rovelli and Smolin in a loop approach to canonical quantum gravity [51]. Note that the algebra has the form of a semi-direct product, with the abelian subalgebra of the traced holonomies $\{T\}$ acted upon by the non-abelian algebra of the T^i-variables, and is similar in structure to the Lagrangian algebra (13.1). This is not true for $N \neq 2$, for which the right-hand sides of the Poisson brackets in (13.3) are *not* linear in T. In order to make the algebra (13.3) non-distributional, one has to "smear out" the loop variables appropriately. One way of doing this is to integrate (13.2) over a ribbon or "strip" R [52, 16], i.e., a non-degenerate one-parameter family $\gamma_t(s) =: R(s,t)$ of loops, $t \in [0,1]$, according to

$$T(R) := \int_0^1 dt \int_0^1 ds\, R'^i(s,t)\dot R^j(s,t)\epsilon_{ijk}T^k\left(\gamma_t(s)\right) \ . \tag{13.5}$$

The dot and the prime denote differentiation with respect to s and t respectively, and ϵ_{ijk} is the totally antisymmetric ϵ-tensor in three dimensions. The resulting algebra of the loop and ribbon variables, $T(\gamma)$ and $T(R)$, has real structure constants that can be expressed in terms of the intersection numbers of the loops and ribbons appearing in their arguments. Although the algebraic structure of the relations (13.4) can be neatly visualized by "cutting and gluing" of diagrams of loops and ribbons, a more physical interpretation of it has not been found so far. Part of the problem is again the overcompleteness of these non-local variables, which affects also the ribbon variables $T(R)$. There have been attempts of integrating the algebra (13.4) with the help of a formal group law expansion [53], and thus possibly to obtain a new infinite-dimensional group structure on the space of three-dimensional loops. Unfortunately one does not get very far this way, again because of the lack of a topological and differentiable structure on loop space.

The main utilization of the loop algebras described above is their importance for canonical non-standard quantization schemes, which will be the subject of the next section. (There are also straightforward lattice analogues of the algebra (13.4), which have been employed in [47, 48].) Another example of a non-trivial loop algebra, involving non-intersecting loops, is due to 't Hooft [54, 55]. He supplements the Wilson loops $T(\gamma)$ by a set of dual loop operators $\overline{T}(\gamma)$, corresponding to magnetic flux lines. Both the T- and the $\overline{T}$-operators commute among themselves, but the commutation relation between a $T(\gamma)$ and a $T(\gamma')$ depends on the linking number of γ and γ'. This algebra is used to extract the qualitative behavior of different phases of Yang–Mills theory (see also the discussions by Mandelstam [41] and Gambini and Trias [56], and the work by Hosoya and Shigemoto [57, 58] for the idea of duality between electric and magnetic flux lines).

14. CANONICAL QUANTIZATION

The only non-perturbative quantization schemes put forward in the loop formulation are Hamiltonian and operator-based. "Non-perturbative" in this context means "not resorting to a perturbation expansion in terms of the gauge potentials A_μ," which would defeat the purpose of a pure loop approach. Indeed, the hope in such non-local formulations is often for an *in*equivalence of the quantum theory and the usual local field-theoretic quantization. Unfortunately, a corresponding, alternative "loop perturbation theory" has not yet been developed. One also has to decide how to treat the Mandelstam constraints in the quantization, for example, whether to solve them before or after the quantization, and different choices may well lead to inequivalent quantum theories.

All existing canonical quantizations for Yang–Mills theory in the loop formulation postulate the existence of (self-adjoint) operator analogues of a set of basic loop variables (such as the traced holonomies and appropriate momentum variables), defined on some "Hilbert space" of loop functionals, such that their commutation relations are preserved in the quantum theory. Most of them are defined at a formal level, in the sense that there is no proper Hilbert space structure, and the wave functions are just elements of some linear function space, depending on loops. Some of them have in common that the action of the operator analogue $\widehat{T}(\gamma)$ of the traced holonomy in this function space is given by multiplication by $T(\gamma)$. Note that we cannot employ a Schrödinger-type quantization, because the algebra relations of the basic variables of the theory [for example, (13.3)] are not of the form of canonical commutation relations.

Gambini and Trias were the first to write down an algebra of quantum operators, realizing the algebra (13.3). Wave functions in their approach are labelled by individual loops and sets of loops, and there is a "vacuum state," the no-loop state $|0>$, which is annihilated by the momentum operators $\widehat{T}^i$. In this aspect their representation is similar to the highest-weight representation of the loop algebra proposed by Aldaya and Navarro-Salas [59].

For reasons inherent in the loop formulation of canonical gravity, Rovelli and Smolin [51] quantize an extended set of $SL(2,\mathbb{C})$-loop variables $T^{i_1\cdots i_n}$, which depend on n "electric field" variables inserted into the traced holonomy, and are a straightforward generalization of the expression (13.2). It turns out that in order to obtain a closed Poisson algebra one has to include the infinite tower of these generalized holonomy variables, for any n. The resulting algebra has a graded structure, schematically given by

$$\begin{aligned} &\{T^0, T^0\} = 0 \ , \\ &\{T^m, T^n\} \sim T^{m+n-1}, \quad m+n>0 \end{aligned} \tag{14.1}$$

where m and n denote the numbers of electric field insertions. This algebra contains the algebra (13.3) as a closed subalgebra. The corresponding quantum algebra obtained in [51] is isomorphic to (14.1), with the exception of higher-order correction terms proportional to $\hbar^k$, $k \geq 2$, appearing on the right-hand sides of the commutators. (See also the work by Rayner [60] on this particular quantum representation.)

These higher-order terms do not appear in the quantization proposed in [53], where the semi-direct product structure of the algebra (13.3) is exploited, using methods from the theory of unitary irreducible representations of semi-direct product algebras. This theory is very powerful in finite dimensions, giving a complete classification and construction of such representations. Some of the formalism can be applied to the infinite-dimensional case too, although there is, of course, no reason to expect that similarly strong results will hold.

Unfortunately, all of the above-mentioned quantum representations remain very formal, since there is no well-defined Hilbert space structure, and defining differential operators on spaces of loop wave functions meets exactly the same problems described in Sec. 11 above.

A somewhat different approach is followed by Ashtekar and Isham [16]. They start from the abelian algebra of the traced holonomies $T(\gamma)$, endow it with the structure of a C*-algebra, and then look for its cyclic representations, using Gel'fand spectral theory. The Hilbert spaces involved are given by spaces of square-integrable functions on the space of ideals of the C*-algebra, whose mathematical structure has not yet been fully explored. From a mathematical point of view, this to date seems to be the most rigorous representation theory of a loop algebra, although appropriate conjugate momentum variables have not been included so far.

Beyond these mainly kinematical considerations of how to quantize algebras of basic phase space variables, hardly anything is known about a proper formulation of the Hamiltonian dynamics, i.e., about how to express the (quantum) Hamiltonian of Yang–Mills theory in terms of the non-local (quantized) loop variables. On the other hand, we know from finite-dimensional examples that the quantization of non-canonical commutation relations at a kinematic level usually is non-unique, and we expect the situation to be much worse in the present, field-theoretic case (see also [61] on the ambiguity of field-theoretic quantizations). Whichever representation theory we come up with for the loop algebra, we will need further physical criteria to decide which of the multitude of possible representations is physically relevant, for example, by selecting those in which the Hamiltonian assumes a simple form.

ACKNOWLEDGEMENTS

I wish to thank the members of the Theoretical Gravitation Group at CFMC, especially J. Mourão, for their hospitality, and B. Brügmann for his comments on a preliminary manuscript of these notes.

REFERENCES

[1] Gu C.-H., *On classical Yang–Mills fields*, Phys. Rep. **80** (1981), 251–337.
[2] Migdal A.A., *Loop equations and 1/N expansion*, Phys. Rep. **102** (1983), 199–290.
[3] Kogut J., Susskind L., *Hamiltonian formulation of Wilson's lattice gauge theories*, Phys. Rev. **D11** (1975), 395–408.
[4] Fort H., Gambini R., *Lattice QED with light fermions in P representation*, Phys. Rev. **D44** (1991), 1257–1262.
[5] Yang C.N., *Integral formalism for gauge fields*, Phys. Rev. Lett. **33** (1974), 445–447.
[6] Kobayashi S., Nomizu K., *Foundations of differential geometry*, Vol.1, Interscience, New York, 1969.
[7] Lichnerowicz A., *Global theory of connections and holonomy groups (French edition published in 1955)*, Noordhoff International Publishing, 1976.
[8] Choquet-Bruhat Y., DeWitt-Morette C. (with M. Dillard-Bleick), *Analysis, manifolds and physics*, revised edition, North–Holland, Amsterdam, 1982.
[9] Wilson K., *Confinement of quarks*, Phys. Rev. **D10** (1974), 2445–2459.
[10] Kauffman L.H., *On knots*, Princeton University Press, 1987.
[11] Kauffman L.H., *Knots and Physics*, World Scientific, Singapore, 1991.
[12] Gambini R., Trias A., *Chiral formulation of Yang–Mills equations: A geometric approach*, Phys. Rev. **D27** (1983), 2935–2939.
[13] Adams J.F., *Infinite loop spaces*, Princeton University Press, 1978.
[14] Michor P.W., *Manifolds of differentiable mappings*, Shiva Publishing Limited, Orpington, 1980.
[15] Lipschutz S., *General topology*, McGraw–Hill, New York, 1965.

[16] Ashtekar A., Isham C.J., *Representations of the holonomy algebras of gravity and non-Abelian gauge theories*, Preprint Imperial/TP/91–92/14 and SU–GP–91–12–2, 1991, 43pp.
[17] Isham C.J., *Loop algebras and canonical quantum gravity*, Proceedings of the 1991 AMS Summer Research Conference on Mathematical Methods in Classical Field Theory (to appear).
[18] Mensky M.B., *Group of parallel transports and description of particles in curved space-time*, Lett. Math. Phys. **2** (1978), 175–180.
[19] Mensky M.B., *Application of the group of paths to the gauge theory and quarks*, Lett. Math. Phys. **3** (1979), 513–520.
[20] Mensky M.B., *The group of paths*, Nauka, Moscow, 1983 (Russian).
[21] Stasheff J., *Differential graded Lie algebras, Quasi–Hopf algebras and higher homotopy algebras*, Preprint UNC–MATH–91–3.
[22] Bryzinski J.M., *The Kaehler geometry of the space of knots in a smooth threefold*, Penn State pure mathematics report No. PM93, 1990.
[23] Schäper U., *Geometry of loop spaces. I. A Kaluza–Klein type point of view*, Preprint Freiburg THEP 91/3, March 1991, 41pp.
[24] Gambini R., Trias A., *Geometrical origin of gauge theories*, Phys. Rev. **D23** (1981), 553–555.
[25] Jackiw R., *Gauge-covariant conformal transformations*, Phys. Rev. Lett. **41** (1978), 1635–1638.
[26] Durhuus B., Leinaas J.M., *On the loop space formulation of gauge theories*, Physica Scripta. **25** (1982), 504–510.
[27] Gliozzi F., Virasoro M.A., *The interaction among dual strings as a manifestation of the gauge group*, Nucl. Phys. **B164** (1980), 141–151.
[28] Gambini R., Trias A., *Gauge dynamics in the C-representation*, Nucl. Phys. **B278** (1986), 436–448.
[29] Giles R., *Reconstruction of gauge potentials from Wilson loops*, Phys. Rev. **D24** (1981), 2160–2168.
[30] Goldberg J.N., Lewandowski J., Stornaiolo C., *Degeneracy in loop variables*, Comm. Math. Phys. (to appear).
[31] Polyakov A.M., *Gauge fields and strings*, Harwood Academic Publishers, Chur, 1987.
[32] Polyakov A.M., *Gauge fields as rings of glue*, Nucl. Phys. **B164** (1979), 171–188.
[33] Mandelstam S., *Quantum electrodynamics without potentials*, Ann. Phys. (N.Y.) **19** (1962), 1–24.
[34] Bialnycki-Birula I., *Gauge-invariant variables in the Yang–Mills theory*, Bull. Acad. Polon. Sci. **11** (1963), 135–138.
[35] Mandelstam S., *Feynman rules for electromagnetic and Yang–Mills fields from the gauge-independent field-theoretic formalism*, Phys. Rev. **175** (1968), 1580–1603.
[36] Polyakov A.M., *String representations and hidden symmetries for gauge fields*, Phys. Lett. **82B** (1979), 247–250.
[37] Aref'eva I.Ya., *The gauge field as chiral field on the path and its integrability*, Lett. Math. Phys. **3** (1979), 241–247.
[38] Aref'eva I.Ya., *Quantum contour field equations*, Phys. Lett. **93B** (1980), 347–353.
[39] Aref'eva I.Ya., *The integral formulation of gauge theories – strings, bags or something else*, Lectures given at the 17th Karpacz Winter School 1980.
[40] Gu C.-H., Wang L.-L.Ch., *Loop-space formulation of gauge theories*, Phys. Rev. Lett. **25** (1980), 2004–2007.
[41] Mandelstam S., *Charge-monopole duality and the phases of non-Abelian gauge theories*, Phys. Rev. **D19** (1979), 2391–2409.
[42] Blencowe M.P., *The Hamiltonian constraint in quantum gravity*, Nucl. Phys. **B341** (1990), 213–251.
[43] Brügmann B., Pullin J., *On the constraint of quantum gravity in the loop representation*, Preprint Syracuse SU–GP–91/8–5.
[44] Durhuus B., Olesen P., *Eigenvalues of the Wilson operator in multicolor QCD*, Nucl. Phys. **B184** (1981), 406–428.
[45] Furmanski W., Kolawa A., *Yang–Mills vacuum: An attempt at lattice loop calculus*, Nucl. Phys. **B291** (1987), 594–628.
[46] Gambini R., Leal L., Trias A., *Loop calculus for lattice gauge theories*, Phys. Rev. **D39** (1989), 3127–3135.
[47] Rovelli C., Smolin L., *Loop representation for lattice gauge theory*, Preprint Pittsburgh and Syracuse 1990.
[48] Brügmann B., *The method of loops applied to lattice gauge theory*, Phys. Rev. **D43** (1991), 566–579.
[49] Loll R., *Independent SU(2)-loop variables and the reduced configuration space of SU(2)-lattice gauge theory*, Nucl. Phys. **B368** (1992), 121–142.
[50] Loll R., *Yang–Mills theory without Mandelstam constraints*, Syracuse preprint SU–GP–92, 21pp.
[51] Rovelli C., Smolin L., *Loop space representation of quantum general relativity*, Nucl. Phys. **B331** (1990), 80–152.
[52] Rovelli C., *Ashtekar formulation of general relativity and loop-space non-perturbative quantum gravity: A report*, Class. Quantum Grav. (1991), 1613–1675.
[53] Loll R., *A new quantum representation for canonical gravity and SU(2) Yang–Mills theory*, Nucl. Phys. **B350** (1991), 831–860.
[54] 't Hooft G., *On the phase transition towards permanent quark confinement*, Nucl. Phys. **B138** (1978), 1–25.
[55] 't Hooft G., *A property of electric and magnetic flux in non-Abelian gauge theories*, Nucl. Phys. **B153** (1979), 141–160.
[56] Gambini R., Trias A., *On confinement in pure Yang–Mills theory*, Phys. Lett. **141B** (1984), 403–406.
[57] Hosoya A., *Duality for the Lorentz force in loop space*, Phys. Lett. **92B** (1980), 331–332; Erratum **96B** (1980), 444.
[58] Hosoya A., Shigemoto K., *Dual potential and magnetic loop operator*, Prog. Theor. Phys. **65** (1981), 2008–2022.
[59] Aldaya V., Navarro-Salas J., *New solutions of the Hamiltonian and diffeomorphism constraints of quantum gravity from a highest weight loop representation*, Phys. Lett. **259B** (1991), 249–255.
[60] Rayner D., *Hermitian operators on quantum general relativity loop space*, Class. Quant. Grav. **7** (1990), 651–661.
[61] Ashtekar A., Isham C.J., *Inequivalent observable algebras: another ambiguity in field quantisation*, Phys. Lett. **274B** (1992), 393–398.