

3

The loop representation

3.1 Introduction

At the beginning of the 1970s gauge theories and in particular Yang–Mills theories appeared as the fundamental theories that described particle interactions. Two main perturbative results were established: the unification of electromagnetic and weak interactions and the proof of the renormalizability of Yang–Mills theory. However, the advent of proposals to describe strong interactions in terms of gauge theories — and in particular the establishment of QCD and the quark model for the hadrons — required the development of new non-perturbative techniques. Problems such as that of confinement, chiral symmetry breaking and the $U(1)$ problem spawned interest in various non-perturbative alternatives to the usual treatment of quantum phenomena in gauge theories. Both at the continuum and lattice levels various attempts were made [44, 48, 12, 49, 50] to describe gauge theories in terms of extended objects as Wilson loops and holonomies. Some of these treatments started at a classical level [44], with the intention of completely reformulating and solving classical gauge theories in terms of loops. Other proposals were at the quantum mechanical level; for instance, trying to find a Schwinger–Dyson formulation in order to obtain a generating functional for the Green functions of gauge theories using the Wilson loop. Among these latter proposals we find the loop representation [5, 34], based on constructing a quantum representation of Hamiltonian gauge theories in terms of loops. In this context, the main advantage of the loop representation was to do away with the first class constraint of the theories (the Gauss law), and therefore with the redundancy introduced by gauge symmetries. It allowed researchers to work directly in the space of physical states.

The idea that a non-perturbative quantization is possibly the only viable solution to the problems presented by the quantization of general

relativity is not new. However, the failure of various attempts based on perturbation theory over the last two decades has increased the belief that non-perturbative methods may be the only alternative to approach the quantization of gravity. In particular, the striking example of 2+1 gravity, which for many years was considered perturbatively as pathological as 3+1 gravity until it was proven by Witten [46] that it can be exactly quantized, has contributed to the belief that perturbative methods in general relativity can be quite misleading. Simultaneously, the introduction of a new set of variables by Ashtekar [51] that cast general relativity in the same language as gauge theories provided the natural framework for the introduction of loop techniques as a natural non-perturbative avenue for the quantization of Einstein's theory. As the Hamiltonian was the most promising scenario for the new variables, the loop representation appeared to be the most natural application of loop techniques to the problem [38, 39]. Moreover it was apparent from the beginning that the use of the loop representation allowed various new insights, in particular it revealed a new connection between general relativity and geometry, but now at a quantum level. Wavefunctions in the loop representation appeared in the pioneering work of Rovelli and Smolin as intimately related with various notions of mathematics, in particular those of the newly flourishing branch of knot theory. This connection was highlighted when the Jones polynomial was found to play the role of a possible state of quantum general relativity [52].

In this chapter we will briefly discuss various physical results that we will need, in combination with the loop techniques introduced in the first two chapters, to introduce the idea of a loop representation. These ideas will be used extensively to discuss the applications in subsequent chapters of the book. The level of rigor and depth that we will maintain in this chapter is only the one needed to discuss the applications. Many of the topics covered in this chapter would, in general, require a book by themselves if they were to be discussed in detail. The idea of this chapter is therefore to fix notation for the advanced reader and to introduce the beginner to these topics in order to allow a first reading of the rest of the book.

The organization of this chapter is as follows. We will start in section 3.2 with a discussion of the canonical formulation and quantization of field theories. The idea is to lay down the formalism that we will use to treat both Yang–Mills theories and gravity. In both cases we will be dealing with systems with constraints and we will briefly discuss their treatment. In section 3.3 we discuss Yang–Mills theories in the canonical formulation both at a classical and a quantum mechanical level, highlighting the role of the Gauss law. We will then discuss the role of Wilson loops as a basis of solutions of the Gauss law and their properties in section 3.4. In section

3.5 we will discuss, in general, the formulation of the loop representation and its implications. We analyze in some detail two possible definitions of the loop representation that we will use throughout the book. We will end with a summary and discussion in section 3.6

3.2 Hamiltonian formulation of systems with constraints

3.2.1 Classical theory

The subject of constrained Hamiltonian systems was pioneered by Dirac [27] in the 1950s and is well established by now. Abundant literature exists on the subject and treatments vary from elementary to very sophisticated, since the subject is endowed with a rich geometrical structure. The intention of this section is just to fix notation and to remind the reader briefly of the ideas involved. More extensive treatments can be found in [27, 28, 29, 30] and those who want to explore the geometrical framework are referred to [31, 32, 33].

Physical theories are not usually described in terms of the minimum possible number of variables. In general, descriptions are made in terms of quantities that present a certain degree of redundancy which results in the fact that the system is invariant under certain symmetries. For instance, one does not usually describe the free electromagnetic field in terms of the two helicity components of the electric field, but rather in terms of the vector potential. The resulting formulation is invariant under gauge transformations. What will happen in general is that given a set of initial data the end result of the evolution will not be unique but will lie on a set of equivalent physical configurations related by the symmetries of the theory. Systems as simple as the free relativistic particle are usually formulated with redundant variables due to the Lorentz symmetry which does not specify a unique choice of time.

We will assume that one has a Hamiltonian system (possibly with an infinite number of degrees of freedom), described by a set of canonical variables q_i and canonical momenta p_i with Poisson bracket relations,

$$\{q_i, p_j\} = \delta_{ij}. \quad (3.1)$$

When one formulates canonically a system with redundant variables and symmetries, the resulting canonical formulation has constraints. The constraints are a set of relations $\phi_m(p_i, q_i) = 0$, $i = 1, \dots, m$ among the canonical variables. Some constraints become manifest when one performs the Legendre transform from the Lagrangian formulation. These are called “primary” constraints. When one requires that these constraints be preserved by evolution, new constraints may appear, called “secondary”, which in turn have to be preserved by evolution and so on.

There is a further distinction between constraints. A constraint ϕ_k will be said to be of first class if its Poisson bracket with all the other constraints is a linear combination of the constraints,

$$\{\phi_k, \phi_i\} = C_{ki}^j \phi_j \quad \forall i. \tag{3.2}$$

Other constraints are called second class. In this book we will only discuss first class constraints. This is due to three reasons. First, there is a procedure (“Dirac brackets”) [27] to convert a set of second class constraints to first class ones by redefining the Poisson bracket structure of phase space. Second, most formulations of the theories of interest in this book such as Yang–Mills theories and general relativity only present first class constraints. Third, although certain gauge fixed formulations of gauge theories involve second class constraints, the loop formulation does not require any gauge fixing, since it is automatically gauge invariant.

The effect of having constraints in the theory is to restrict the dynamics to taking place on a surface $\hat{\Gamma}$ in the phase space Γ called the “constraint surface”. The dynamical trajectories on $\hat{\Gamma}$ are not well defined. Each dynamical evolution is represented by an infinite family of trajectories that are physically equivalent. This is the representation in this picture of what is usually called “gauge”. The family of trajectories are “gauge equivalent”. This is due to the fact that there is an ambiguity in extending quantities from $\hat{\Gamma}$ to Γ since two quantities that differ by a combination of constraints are equal on the constraint surface. In particular, the Hamiltonian is not well defined and two Hamiltonians differing by linear combinations of the constraints will generate two physically equivalent gauge related trajectories

$$H \sim H' = H + \lambda^m \phi_m, \tag{3.3}$$

where λ^m do not depend on the canonical variables.

After an infinitesimal amount of time, two equivalent dynamical evolutions which started from the same initial conditions differ by terms proportional to the commutators of the dynamical variables with the constraints. That is, one can view the commutator of any function of phase space with a constraint as a representation on phase space of the infinitesimal generator of the symmetry associated with the constraint,

$$\Delta f(p, q) = \{\lambda^m \phi_m(p, q), f(p, q)\} \Delta t. \tag{3.4}$$

Strictly speaking, these symmetries generated by the first class constraints of a theory should be called “gauge” symmetries of a theory. For the case of usual Yang–Mills theories on trivial fiber bundles the symmetries generated by the constraints coincide with the usual idea of gauge symmetries. In general, however, this equivalence is only local and global inequivalences may give rise to observable physical phenomena.

Non-vanishing quantities whose Poisson brackets with the constraints vanish on the constraint surface are called “observables” of the system, since they are the quantities that are invariant under the symmetries generated by the constraints.

3.2.2 Quantum theory

A procedure for quantizing Hamiltonian systems with first class constraints was first proposed by Dirac [27]. Although the original formulation was presented for systems with a finite number of degrees of freedom, it is readily generalizable to the case of field theories. The method consists basically of five steps.

- Select an algebra of quantities in the classical theory general enough to be able to express any quantity of physical interest in terms of the selected quantities. In the simplest examples, one usually takes the canonical coordinates with their Poisson relations as such an algebra.

- Represent this algebra as a set of operators acting on a functional space \mathcal{V} and promote the Poisson bracket relations to relations between commutators of operators. No particular restriction is imposed on the functional space at this level. Again, as an example one can choose representation on functionals of the configuration variables $\Psi[q]$, and represent the fundamental operators as $\hat{q}\Psi[q] = q\Psi[q]$, $\hat{p}\Psi[q] = -i\hbar(\delta\Psi[q]/\delta q)$ and their commutation relation as $[\hat{q}, \hat{p}] = i\hbar$. (From now on we will choose units such that $\hbar = 1$.)

- Promote the constraint equations to wave equations acting on the space of functions \mathcal{V} . This process is, in general, not unique, depending on regularizations and factor orderings. Moreover, it should be performed in such a way as to promote the classical Poisson brackets of the constraint to consistent commutation relations of the wave equations. The space of solutions to the wave equations will, in general, be a restriction of \mathcal{V} and will contain the wavefunctions of physical relevance: we call it $\hat{\mathcal{V}}$.

- Determine the evolution as a function of the parameter of evolution of the associated classical theory of the states (Schödinger picture) or observables (Heisenberg picture) with the use of either the Schrödinger equation for the states,

$$i\frac{\partial\Psi}{\partial t} = \hat{H}\Psi, \quad (3.5)$$

where \hat{H} is the Hamiltonian operator, or the Heisenberg equations for the observables. Notice that the evolution is unambiguous since in the previous point we imposed the constraints on the wavefunctionals. That is, adding a combination of constraints to the Hamiltonian does not change the evolution, since they annihilate the wavefunctionals.

- Introduce an inner product on $\hat{\mathcal{V}}$ such that it becomes a Hilbert space, the observables become self-adjoint operators and the wavefunctions of physical interest become normalizable.

With these steps completed one is in position to do physics by taking expectation values of physical observables using the inner product on the Hilbert space of wavefunctions. Notice that apart from some subtleties, this is what most physicists would recognize as the “usual” procedure of canonical quantization. However, several points need further comment.

First of all notice that in the first step we are allowing the use of a non-canonical algebra to perform the quantization. This is not, strictly speaking, what Dirac originally proposed, since he only considered the use of the algebra of canonical quantities. Allowing a non-canonical (and possibly overcomplete) algebra is more flexible in the sense that it can accommodate dynamical systems which do not naturally have a canonical algebra or situations where to use a canonical algebra is not convenient. It will be important in the formulation of the loop representation.

It could happen that when one performs the Legendre transform to determine the Hamiltonian the end result is a quantity that vanishes on the constraint surface. That is, the Hamiltonian of the theory is a combination of constraints. In this case the Schrödinger (or Heisenberg) equations simply say that the states (or observables) do not evolve with the classical parameter of evolution. In this case the notion of “time” in the system has to be retrieved in a different way. One possibility is to isolate one of the canonical variables as a “time” T and “deparametrize” the theory in such a way that the Hamiltonian constraint can be written as $H \equiv \pi_T - \bar{H}$, where π_T is the variable canonically conjugate to T (“energy”). Then one considers \bar{H} as a true non-vanishing Hamiltonian and T as an evolution variable. The evolution in the “time” T is generated with the Hamiltonian \bar{H} and its corresponding Schrödinger (or Heisenberg) equation. This procedure is generically by no means trivial and in many systems it is not known how to perform it in a consistent way. Many systems have vanishing Hamiltonians and almost any system can be written in such a way that the Hamiltonian vanishes (these are usually called “parametrized theories”). Other systems, however, come naturally “already parametrized”. An example of this behavior, as we shall see, is general relativity. Other examples are the relativistic free particle and string. A comprehensive discussion of these and other related issues is the review article by Kuchař [57].

Another thing that can happen is that the theory could have symmetries that are not reflected in the appearance of constraints. This is usually the case with global symmetries, like “large” gauge transformations or diffeomorphisms. We will largely ignore these in this book. In principle, one

should require that wavefunctions transform as unitary representations of the respective symmetry. This can put constraints on the inner product one selects, as has been emphasized by Peldán [58].

In the method presented above the first three steps contained specific proposals. Although in each of them one is faced with many inequivalent choices, one can always pick one of them and proceed. A different situation arises in the last step, where no prescription for the introduction of an inner product is made. The Dirac quantization procedure does not specify how to introduce an inner product and in this sense it is incomplete. This situation is particularly complex in systems where one does not have extra auxiliary structures that in some sense determine a preferred inner product. For instance, in usual field theories on a flat background the Poincaré invariance uniquely fixes the inner product. However, in the gravitational case, for instance, one does not have at hand such a guiding principle. There are proposals to extend Dirac's method of quantization in such a way as to have a program that chooses an inner product without resorting to any additional symmetries or structures. Among these proposals is that of Ashtekar [2] who suggests endowing the phase space with a star algebra structure which may be sufficient to fix the inner product. The issue of the inner product in non-linear field theories is by no means completely understood at present and in our book we will discuss it only tangentially.

3.3 Yang–Mills theories

Yang–Mills theories have proven to be very useful as descriptions of the physics of the elementary particles. An extensive literature has dealt with them from various viewpoints and at present there is a good understanding of many of their features. It is therefore reasonable to introduce Yang–Mills theories at this point to illustrate various concepts we will need in the rest of the book, especially in the applications to gravity. In particular Yang–Mills theories have proven an adequate ground to develop techniques related to loops. Many techniques and results that are only conjectured to hold for the gravitational case have actually been proved and exhaustively studied for the Yang–Mills case.

In this section we will introduce the canonical formulation of classical and quantum Yang–Mills theories in terms of the traditional variables. In particular we will study the meaning of the Gauss law as a constraint and generator of the gauge symmetries of the theory. In subsequent sections we will review these results in the language of loops.

3.3.1 Canonical formulation

Yang–Mills theories are based on an algebra valued connection \mathbf{A}_μ on a flat manifold $\eta_{\mu\nu}$ with an action

$$S \equiv \int d^4x \tilde{\mathcal{L}} = -\frac{1}{2} \int d^4x \sqrt{|\eta|} \eta^{\mu\lambda} \eta^{\nu\rho} \text{Tr}(\mathbf{F}_{\mu\nu} \mathbf{F}_{\lambda\rho}), \tag{3.6}$$

where $\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + i[\mathbf{A}_\mu, \mathbf{A}_\nu]$ and $[\ , \]$ is the commutator in the Lie algebra associated with the gauge group. We denote by $\sqrt{|\eta|}$ the square root of the absolute value of the determinant of the metric. Elements of the Lie algebra will be denoted with boldface characters. Sometimes it will also be convenient to introduce the notation in components in terms of the basis of generators of the Lie algebra, for instance,

$$\mathbf{A}_\mu = A_\mu^i \mathbf{X}^i, \tag{3.7}$$

where \mathbf{X}^i are the generators of the Lie algebra satisfying

$$[\mathbf{X}^i, \mathbf{X}^j] = C^{ij}_k \mathbf{X}^k, \tag{3.8}$$

where C^{ij}_k are the structure constants of the group in question.

We take as background metric $\eta = \text{diag}(-1, 1, 1, 1)$ and consider as configuration variables \mathbf{A}_0 and \mathbf{A}_a and compute their canonical momenta*, $\tilde{\pi}^0$ and $\tilde{\pi}^a$

$$\tilde{\pi}^0 \equiv \frac{\delta S}{\delta \dot{\mathbf{A}}_0} = 0, \tag{3.9}$$

$$\tilde{\pi}^a \equiv \frac{\delta S}{\delta \dot{\mathbf{A}}_a} = \sqrt{|\eta|} \eta^{ab} (\dot{\mathbf{A}}_b - \partial_b \mathbf{A}_0 + i[\mathbf{A}_0, \mathbf{A}_a]) = \tilde{\mathbf{E}}^a. \tag{3.10}$$

The momentum conjugate to \mathbf{A}_0 vanishes. This will be a primary constraint. We now perform the Legendre transform to define the Hamiltonian density $\tilde{\mathcal{H}} = 2\text{Tr}(\tilde{\mathbf{E}}^a \dot{\mathbf{A}}_a) - \tilde{\mathcal{L}}$,

$$\tilde{\mathcal{H}} = \text{Tr} \left((\tilde{\mathbf{E}}^a \tilde{\mathbf{E}}^b \eta_{ab} + \tilde{\mathbf{B}}^a \tilde{\mathbf{B}}^b \eta_{ab}) - \sqrt{|\eta|} \mathbf{A}_0 (\mathbf{D}_a \tilde{\mathbf{E}}^a) \right) \tag{3.11}$$

where $\mathbf{D}_a \equiv \partial_a + i[\mathbf{A}_a, \]$ is the gauge covariant derivative. We can now extend the Hamiltonian including the primary constraint $\tilde{\mathcal{H}}' = \tilde{\mathcal{H}} + \text{Tr}(\lambda_0 \tilde{\pi}^0)$, where λ_0 is a group-valued Lagrange multiplier.

* In this book we will use an overtilde to denote tensor densities of weight +1, and an undertilde for weight -1, a notation that is becoming standard. The only exceptions will be the Dirac delta function, which is a +1 density, but which we will denote as $\delta(x - y)$ to adhere to usual conventions, the square root of the determinant of metrics, since it is obvious, and — in an effort to try not to clutter the notation — the multitenor densities $X^{a_1 a_2 \dots a_n x_n}$, since their tensor density character has been abundantly emphasized.

To compute the equations of motion of this theory we take the Poisson brackets of the phase space variables with the Hamiltonian. In particular, one observes that the time evolution of $\tilde{\pi}_0$ is given by

$$\dot{\tilde{\pi}}^0 = -\frac{\delta\tilde{\mathcal{H}}}{\delta\mathbf{A}_0} = \mathbf{D}_a\tilde{\mathbf{E}}^a = \partial_a\tilde{\mathbf{E}}^a + i[\mathbf{A}_a, \tilde{\mathbf{E}}^a] = 0. \quad (3.12)$$

This equation guarantees the preservation in time of the primary constraint. It is in itself a new (secondary) constraint. It can be checked that this constraint is automatically conserved. Moreover, the primary and secondary constraints are first class, i.e.,

$$\{\mathcal{P}(\mu), \mathcal{P}(\lambda)\} = 0, \quad (3.13)$$

$$\{\mathcal{P}(\mu), \mathcal{G}(\lambda)\} = 0, \quad (3.14)$$

$$\{\mathcal{G}(\lambda), \mathcal{G}(\mu)\} = \mathcal{G}([\lambda, \mu]), \quad (3.15)$$

where we have introduced the notation of “smeared out constraints”,

$$\mathcal{G}(\lambda) \equiv \int d^3x \lambda D_a \tilde{\mathbf{E}}^a, \quad (3.16)$$

$$\mathcal{P}(\mu) \equiv \int d^3x \mu \tilde{\pi}^0. \quad (3.17)$$

From now on, every time we refer to a constraint as dependent on a parameter we will assume that the parameter has the needed index structure to be contracted with the constraint and an integration over the manifold has been performed. This enables us to avoid dealing with distributional expressions. Notice the geometric interpretation of the Gauss law as a generator of infinitesimal gauge transformations associated with the arbitrary group valued function λ ,

$$\{\mathcal{G}(\lambda), \mathbf{A}_a\} = \mathbf{D}_a \lambda, \quad (3.18)$$

$$\{\mathcal{G}(\lambda), \tilde{\mathbf{E}}^a\} = [\lambda, \tilde{\mathbf{E}}^a]. \quad (3.19)$$

With this point of view of the Gauss law as a generator of gauge transformations notice that one can interpret the commutator (3.15) in the following way: the commutator of the infinitesimal gauge transformation parametrized by λ and that parametrized by μ is an infinitesimal gauge transformation parametrized by $[\lambda, \mu]$. The primary constraint simply states that the zeroth component of the vector potential can be arbitrarily rescaled,

$$\{\mathcal{P}(\mu), \mathbf{A}_0(x)\} = -\mu(x). \quad (3.20)$$

This tells us that \mathbf{A}_0 and π^0 could be eliminated from the classical theory by appropriate rescalings. This fact will find a counterpart in quantum theory. This ends the discussion of the classical theory.

3.3.2 Quantization

We will now apply the program of quantization that we introduced in section 3.2.2 to the canonical formulation of Yang–Mills theories. We start by picking an algebra of classical quantities, in our case simply the canonical algebra in terms of the Poisson brackets,

$$\{A_a^i(x), \tilde{E}_j^b(y)\} = \delta_a^b \delta_j^i \delta(x - y), \tag{3.21}$$

$$\{A_0^i(x), \tilde{\pi}_j^0(y)\} = \delta_j^i \delta(x - y), \tag{3.22}$$

with all other brackets vanishing. We also pick a polarization for the wavefunctionals $\Psi[\mathbf{A}, \mathbf{A}_0]$ where by \mathbf{A} we mean the one form on the spatial surface with components \mathbf{A}_a

We now find a representation of the canonical algebra by defining,

$$\hat{\mathbf{A}}_a \Psi[\mathbf{A}, \mathbf{A}_0] = \mathbf{A}_a \Psi[\mathbf{A}, \mathbf{A}_0], \tag{3.23}$$

$$\hat{\mathbf{E}}^a \Psi[\mathbf{A}, \mathbf{A}_0] = -i \frac{\delta}{\delta \mathbf{A}_a} \Psi[\mathbf{A}, \mathbf{A}_0], \tag{3.24}$$

$$\hat{\mathbf{A}}_0 \Psi[\mathbf{A}, \mathbf{A}_0] = \mathbf{A}_0 \Psi[\mathbf{A}, \mathbf{A}_0], \tag{3.25}$$

$$\hat{\tilde{\pi}}^0 \Psi[\mathbf{A}, \mathbf{A}_0] = -i \frac{\delta}{\delta \mathbf{A}_0} \Psi[\mathbf{A}, \mathbf{A}_0]. \tag{3.26}$$

Notice that up to now we have performed several arbitrary choices, which in general would yield inequivalent quantum theories if performed in a different way. For example, we could have added the functional gradient of an arbitrary function of \mathbf{A} to the definition of the conjugate momenta $\hat{\mathbf{E}}^a$ and this would yield the same quantum commutator algebra.

We now promote the constraints to quantum equations and impose them on the wavefunctions. The primary constraint can be satisfied immediately, simply by noticing that it requires the wavefunctionals not to depend on \mathbf{A}_0 ,

$$\mathcal{P}(\mu) \Psi[\mathbf{A}, \mathbf{A}_0] = -i \int d^3x \mu \frac{\delta}{\delta \mathbf{A}_0} \Psi[\mathbf{A}, \mathbf{A}_0]. \tag{3.27}$$

First of all, notice that we have imposed the “smeared out” form of the constraint, and we will usually do this. This is equivalent to imposing the constraint point by point in the manifold since the equation should hold for an arbitrary smearing function μ . Moreover, it is instructive to view the action of the constraint in the following way. Consider the action of $(1 + i\epsilon \mathcal{P}(\mu))$ on a wavefunction in the limit $\epsilon \rightarrow 0$,

$$(1 + i\epsilon \mathcal{P}(\mu)) \Psi[\mathbf{A}, \mathbf{A}_0] = \Psi[\mathbf{A}, \mathbf{A}_0 + \epsilon \mu]. \tag{3.28}$$

We see that the quantum constraint acts as the infinitesimal generator on the wavefunctions of the symmetry that we mentioned in the classical theory: that the component \mathbf{A}_0 of the vector potential could be rescaled

arbitrarily. From now on we will therefore concentrate on functionals that only depend on the spatial part of the connection, $\Psi[\mathbf{A}]$.

Let us now focus on the Gauss law. We can promote it to a quantum operator in the following way,

$$\hat{\mathcal{G}}(\lambda)\Psi[\mathbf{A}] = -i \int d^3x \lambda^k \left(\partial_a \frac{\delta}{\delta A_a^k} + C_{lm}^k A_a^l \frac{\delta}{\delta A_a^m} \right) \Psi[\mathbf{A}]. \quad (3.29)$$

Notice that, in principle, there is a factor ordering ambiguity when representing the non-Abelian term of the covariant derivative. The reader may check that due to the symmetries of the structure constants of any compact group (in particular $SU(N)$) it is immaterial which ordering is picked for the non-Abelian term.

Let us again study the infinitesimal action of the Gauss law on wavefunctionals,

$$(1 + i\epsilon\mathcal{G}(\lambda))\Psi[\mathbf{A}] = \Psi[\mathbf{A} + \epsilon\mathbf{D}\lambda]. \quad (3.30)$$

We see that it acts as an infinitesimal generator of gauge transformations on the wavefunctionals. It is therefore immediate to solve the constraint. One just has to consider wavefunctionals which are gauge invariant functions of the connection and they will automatically be annihilated by the Gauss law.

Notice that the Gauss law, both at a classical and quantum mechanical level, only generates gauge transformations connected to the identity. “Large” gauge transformations are not included and their presence can give rise to observable physical effects. This is a generic feature of constrained systems. Constraints usually only generate local gauge symmetries. In the case of Yang-Mills theories the presence of large gauge transformations gives rise to the Θ -vacua, connected with the instanton structure of the theory [59]. Similar effects arise for gravity [61]. For both Yang-Mills and gravity “large” gauge transformations are responsible for the presence of fractional spin states [60].

One should now study the evolution of the wavefunctionals (let us adopt for the sake of argument the Schrödinger picture). For that we have to promote the Hamiltonian of the theory to an operator. This can be accomplished with a straightforward factor ordering (though a regularization is needed). One can then study the eigenstates and spectra of eigenvalues of the theory. In Yang-Mills theories the interpretation of the eigenvalues would be the masses of the particle spectra of the theory. This formulation would lead to a non-perturbative solution of Yang-Mills theories if one could implement the evolution equation and introduce an inner product. The treatment of this problem is involved and there is not a closed solution for it in the continuum, although lattice techniques have been applied to it. We will return to these issues in chapter 6.

3.4 Wilson loops

Observable quantities in gauge theories need to be gauge invariant. Wavefunctions in a quantum representation also need to be gauge invariant. It will therefore be very useful to introduce a set of quantities involving the connection \mathbf{A}_a in terms of which any gauge invariant quantity can be written. These objects are called Wilson loops, are gauge invariant under both small and large gauge transformations and are constructed taking traces of the holonomy,

$$W_{\mathbf{A}}(\gamma) = \text{Tr} \left[\text{P exp} \left(i \oint_{\gamma} dy^a \mathbf{A}_a \right) \right]. \quad (3.31)$$

The gauge invariance of these quantities follows immediately from the properties of the connection and the holonomy that were introduced in chapter 1 and the cyclicity of the traces. Because of this, they are observables in the canonical sense (they have vanishing Poisson brackets with all the constraints of the theory).

The objects are dependent on a loop and have a non-local dependence on the gauge connection. In general they are complex numbers. We can write the Wilson loop using the notation of chapter 2 as

$$W_{\mathbf{A}}(\gamma) = \text{Tr}(I) + \sum_{n=1}^{\infty} i^n \text{Tr}(\mathbf{A}_{a_1}(x_1) \dots \mathbf{A}_{a_n}(x_n)) X^{a_1 x_1 \dots a_n x_n}(\gamma). \quad (3.32)$$

Observe that the trace $\text{Tr}(A_{a_1}(x_1) \dots A_{a_n}(x_n))$ is cyclic in the indices $a_1 x_1 \dots a_n x_n$, and therefore the Wilson loop only depends on the cyclic portion of the multitangents. As we mentioned in chapter 2, this removes all information about the basepoint of the loop. That is, Wilson loops are functions of non-basepointed loops.

Wilson loops have two fundamental properties, the discussion of which will occupy the rest of this section:

- the Mandelstam identities;
- the reconstruction property.

The Mandelstam identities are a set of relations between Wilson loops which reflect the structure of the particular gauge group considered. The reconstruction property will tell us that given the Wilson loop functions evaluated for all possible loops we can reconstruct all the gauge invariant information present in the gauge connection. Both properties together will imply that Wilson loops constitute an overcomplete basis of solutions of the Gauss law constraint.

3.4.1 The Mandelstam identities

The Mandelstam identities are the reflection in the language of Wilson loops of the particular properties of the gauge group used to define the holonomies and of generic properties of traces. In terms of them we will see reflected group properties such as unitarity, the dimension of the representation and value of the determinant of matricial representations. They will allow us to express products of Wilson loops in terms of sums of products involving a smaller number of Wilson loops.

These identities were first introduced by Mandelstam [9] for the $O(3)$ group. Giles [35] extended them for groups $GL(N)$ and Gambini and Trias [34] extended them to the case of special and unitary groups. Loll [23] discussed the case of certain non-compact groups.

Let us consider gauge groups that admit fundamental representations in terms of $N \times N$ matrices, for instance, $GL(N)$, $SL(N)$, $U(N)$, $SU(N)$. The Mandelstam identities arise as a consequence of the properties of the traces of $N \times N$ matrices. There are two kinds of identities, called identities of the first and second kinds.

The Mandelstam identities of the first kind are a simple consequence of the cyclic property of the traces, which we mentioned in the previous section[†],

$$W(\gamma_1 \circ \gamma_2) = W(\gamma_2 \circ \gamma_1). \quad (3.33)$$

These identities hold for *any* gauge group of any dimension.

There are various identities of the second kind. The first family are a set of non-linear constraints that ensure that $W_{\mathbf{A}}(\gamma)$ is a trace of an $N \times N$ matrix. They can be obtained in the following way.

Observe first that in N dimensions any object with $N + 1$ totally anti-symmetric indices vanishes,

$$\delta_{[B_1}^{A_1} \delta_{B_2}^{A_2} \cdots \delta_{B_{N+1}}^{A_{N+1}}] = 0. \quad (3.34)$$

Then contract this with $N + 1$ holonomies,

$$H(\gamma_1)_{A_1}^{B_1} \cdots H(\gamma_{N+1})_{A_{N+1}}^{B_{N+1}} \quad (3.35)$$

where $A_1, B_1, \dots, A_{N+1}, B_{N+1}$ are matrix indices in the matricial representation of the group. The result is an identically vanishing sum of products of traces of products of holonomies. From here one can work out explicitly the identities for any order. For example, if $N = 1$, as in a

[†] In this section and the following we will omit writing the dependence of the Wilson loop on the connection \mathbf{A} since the results proven will not depend on the choice of a particular connection

$U(1)$ group, the identity reads

$$W(\gamma_1)W(\gamma_2) - W(\gamma_1 \circ \gamma_2) = 0. \tag{3.36}$$

There is a compact way of writing this identity for an arbitrary order in terms of the quantities M_K , depending on K loops and defined by the following recurrence relations

$$(K + 1)M_{K+1}(\gamma_1, \dots, \gamma_{K+1}) \equiv W(\gamma_{K+1})M_K(\gamma_1, \dots, \gamma_K) - M_K(\gamma_1 \circ \gamma_{K+1}, \gamma_2, \dots, \gamma_K) - \dots - M_K(\gamma_1, \gamma_2, \dots, \gamma_K \circ \gamma_{K+1}), \tag{3.37}$$

$$M_1(\gamma) \equiv W(\gamma). \tag{3.38}$$

In terms of the M s, the identity for an $N \times N$ matrix group can be written as

$$M_{N+1}(\gamma_1, \dots, \gamma_{N+1}) = 0. \tag{3.39}$$

Notice that for the case of $N \times N$ matrices it is also true that

$$M_L(\gamma_1, \dots, \gamma_L) = 0 \quad \forall L > N + 1. \tag{3.40}$$

An immediate consequence of the recurrence relation (3.37), obtained identifying the loop $N + 1$ with ι (the identity loop), is

$$(N + 1)M_{N+1}(\gamma_1, \dots, \gamma_N, \iota) = (W(\iota) - N)M_N((\gamma_1, \dots, \gamma_N) = 0, \tag{3.41}$$

from which we see that

$$W(\iota) = N. \tag{3.42}$$

Let us examine another example, for 2×2 matrices. One can expand the product of three traces in terms of two,

$$W(\gamma_1)W(\gamma_2)W(\gamma_3) = W(\gamma_1 \circ \gamma_2)W(\gamma_3) + W(\gamma_2 \circ \gamma_3)W(\gamma_1) + W(\gamma_3 \circ \gamma_1)W(\gamma_2) - W(\gamma_1 \circ \gamma_2 \circ \gamma_3) - W(\gamma_1 \circ \gamma_3 \circ \gamma_2). \tag{3.43}$$

For instance $SU(2)$, $SU(1, 1)$ and other groups that admit fundamental representations in terms of 2×2 matrices give rise to Wilson loops that satisfy the identity (3.43). These groups also admit other identities that reflect other properties apart from the 2×2 matricial nature of their representation.

Notice that because we are working with non-basepointed loops, the composition of two loops $\gamma_1 \circ \gamma_2$ in general is not well defined. For the remainder of this section whenever a composition of two loops appears, we will assume an arbitrary basepoint has been chosen to perform the composition. One simply links both loops to the basepoint through arbitrary retraced paths. The Mandelstam identities are independent of the basepoint chosen to define the composition of the loops.

Another identity appears for special groups, i.e., groups that admit fundamental representations in terms of matrices of unit determinant. As

was proved in reference [34], for a group with fundamental representation in terms of $N \times N$ matrices of unit determinant the following identity in terms of the M s holds:

$$M_N(\gamma_1 \circ \gamma, \gamma_2 \circ \gamma, \dots, \gamma_N \circ \gamma) = M_N(\gamma_1, \gamma_2, \dots, \gamma_N) \tag{3.44}$$

from which it trivially follows that $M_N(\gamma, \gamma, \dots, \gamma) = 1$. These identities allow us, for a special group, to express the product of N Wilson loops in terms of that of $N - 1$ by taking $\gamma = \gamma_i$ for some i in equation (3.44). For example, for any special 2×2 matrix group (such as $SU(2)$, $SL(2, C)$, etc),

$$M_2(\gamma_1, \gamma_2) = M_2(\gamma_1 \circ \gamma_2^{-1}, \iota) \tag{3.45}$$

and

$$M_2(\gamma_1, \gamma_2) = \frac{1}{2}(W(\gamma_1)W(\gamma_2) - W(\gamma_1 \circ \gamma_2)), \tag{3.46}$$

$$M_2(\gamma_1 \circ \gamma_2^{-1}, \iota) = \frac{1}{2}(W(\gamma_1 \circ \gamma_2^{-1})W(\iota) - W(\gamma_1 \circ \gamma_2^{-1})), \tag{3.47}$$

therefore,

$$W(\gamma_1)W(\gamma_2) = W(\gamma_1 \circ \gamma_2^{-1}) + W(\gamma_1 \circ \gamma_2). \tag{3.48}$$

Finally, we will discuss the Mandelstam identities of the second kind that reflect the fact that a group is unitary. That is, if the group admits a fundamental representation in terms of unitary $N \times N$ matrices, the Wilson loops satisfy

$$W(\gamma) = W^*(\gamma^{-1}), \tag{3.49}$$

where $*$ indicates the complex conjugate.

In general, apart from the Mandelstam identities, Wilson loops satisfy a series of inequalities. For instance, for unitary groups, the following inequality holds trivially

$$|W(\gamma)| \leq |W(\iota)| = N. \tag{3.50}$$

These inequalities contain additional information that is not present in the identities we discussed previously. For instance, all the identities we have discussed so far are the same for the groups $SU(2)$ and $SU(1, 1)$. It is by considering inequalities in terms of the Wilson loops that one can determine which of these two groups is being considered. A discussion of inequalities and their consequences can be found in reference [23].

Let us end by summarizing the Mandelstam identities for the group $SU(2)$, which we will use extensively in this book:

Identity of the first kind,

$$W(\gamma_1 \circ \gamma_2) = W(\gamma_2 \circ \gamma_1). \tag{3.51}$$

Identity of the second kind,

$$W(\gamma_1)W(\gamma_2) = W(\gamma_1 \circ \gamma_2^{-1}) + W(\gamma_1 \circ \gamma_2). \quad (3.52)$$

From here it is immediate to prove, choosing $\gamma_1 = \iota$,

$$W(\gamma) = W(\gamma^{-1}), \quad (3.53)$$

and from this and the unitarity property it follows that $W(\gamma)$ is real and less than or equal to 2 in absolute value.

In spite of their simple appearance, successive combinations of the Mandelstam identities can lead to very non-trivial relations among Wilson loops. In any formulation in which one wants to use the Wilson loops as basic variables, these relations imply an additional complication in the formulation of the theory, since there is no definite way to determine which are the freely specifiable functions [24]. In particular there is no systematic way of generating the set of all possible relations between products of Wilson loops that are derived from the Mandelstam identities [24]. An important development in this area is the recognition by Rovelli and Smolin that spin networks might be used to characterize a complete set of independent products of Wilson loops [146]. We will return to these issues when we discuss the loop representation.

3.4.2 Reconstruction property

In the previous section we introduced a set of identities satisfied by Wilson loops. In this section we will study the opposite question: to what extent does a prescribed function of loops, satisfying the Mandelstam identities, qualify to become a Wilson loop? In particular, can we reconstruct the holonomy given such a function?

This question is of great importance. From the results of chapter 1 we have seen that one could use holonomies to describe gauge theories since they embody all the gauge invariant information of the connection. What we are about to do is to show that all the information present in a holonomy can be reconstructed from the Wilson loops. That is, the Wilson loops will acquire a status of fundamental variables in themselves since we will be able to reconstruct all the gauge invariant information of a theory from them. This step will be of fundamental importance in following sections where we will formulate a quantum representation purely in terms of loops.

The proof that this can actually be accomplished, i.e., that given a function of loops satisfying the Mandelstam constraints one can reconstruct the gauge invariant information encoded in it is the subject of the so called “reconstruction theorems”. The idea is the following. Given a function $W(\gamma)$, satisfying the Mandelstam constraints (3.33), (3.39) it is

possible to construct explicitly a set of $N \times N$ matrices \mathbf{H}_γ defined modulo a similarity transformation, such that their traces are $W(\gamma)$. The first such theorem was proved by Giles [35] for the case of $U(N)$. Loll considered the cases of $SU(2)$ [25]. Ashtekar and Lewandowski [40] refined many of the proofs presented early in the literature and introduced several generalizations. Here we will discuss a simplified derivation assuming the Mandelstam identities for a 2×2 matrix group and we will follow the presentation of Giles. An elegant, short, alternative derivation of the reconstruction property has recently been presented in reference [40].

One starts by defining an algebra associated with the group of loops. It is constructed in the following way. Take the group of loops \mathcal{L}_o . Define a formal sum and product by a complex number law for elements of \mathcal{L}_o . Construct then an algebra $F\mathcal{L}_o$ by appending to the elements of \mathcal{L}_o all their possible finite complex linear combinations. The product law of the algebra will be induced by the composition law of \mathcal{L}_o .

We now consider the extension of the notion of Wilson loop to this algebra. For those elements of $F\mathcal{L}_o$ belonging to \mathcal{L}_o it is defined in the usual way. For linear combinations of them it is given by

$$W(a_1\gamma_1 + a_2\gamma_2) = a_1W(\gamma_1) + a_2W(\gamma_2). \quad (3.54)$$

Notice that $\gamma_1, \gamma_2 \in \mathcal{L}_o$ and therefore the $W(\gamma_{1,2})$ are well defined. From now on we will use the same notation for elements of $F\mathcal{L}_o$ and elements of \mathcal{L}_o , it will be clear from the context to which we are referring.

This algebra is isomorphic to a complexification of the algebra \mathcal{E}_D that we introduced in chapter 2, obtained by allowing the multitensor densities $E_D^{a_1x_1 \dots a_nx_n}$ that satisfied the differential constraint to become complex-valued.

We want to see if these extended Wilson loops can be obtained as traces of “extended” holonomies $H(\gamma)$ in the sense introduced in chapter 2 (traces of linear combinations of holonomies are allowed). We would like to think of $H(\gamma)$ as representations of $F\mathcal{L}_o$. Notice that $F\mathcal{L}_o$ is associated with an infinite-dimensional group (in particular because \mathcal{L}_o is) whereas the vector space of extended holonomies is finite-dimensional (they are 2×2 matrices in our simplified derivation). Therefore many elements of $F\mathcal{L}_o$ are represented by the same matrix. We now introduce an equivalence relation such that two elements of $F\mathcal{L}_o$ are equivalent if they lead to the same matrix. We are then able to establish a correspondence between equivalence classes of elements of $F\mathcal{L}_o$ and the matrices.

We say that $\gamma_1 \sim \gamma_2$ if

$$W(\gamma_1 \circ \zeta) = W(\gamma_2 \circ \zeta) \quad \forall \quad \zeta. \quad (3.55)$$

By the definition (3.54) it is obvious that the equivalence relation defined is compatible with the sum and product times a complex number.

We now prove that it is compatible with the product law of the algebra. Suppose $\gamma_1 \sim \gamma_2$ and $\eta_1 \sim \eta_2$. Then

$$\begin{aligned} W(\gamma_1 \circ \eta_1 \circ \zeta) &= W(\gamma_2 \circ \eta_1 \circ \zeta) = W(\eta_1 \circ \zeta \circ \gamma_2) \\ &= W(\eta_2 \circ \zeta \circ \gamma_2) = W(\gamma_2 \circ \eta_2 \circ \zeta). \end{aligned} \tag{3.56}$$

We denote by $F\mathcal{L}_o/K$ the algebra of equivalence classes of extended loops, K being the kernel of the equivalence relation.

We now use the Mandelstam identity of the second kind (3.39), to derive an explicit form for the matrix associated with an equivalence class belonging to $F\mathcal{L}_o/K$. Let us explicitly consider the identity for the case of 2×2 matrices already introduced in equation (3.43) (notice that we do not, at this stage, know the dimension of the representation and we will *prove* that the representation is (2×2) dimensional based on this identity),

$$\begin{aligned} W(\gamma_1)W(\gamma_2)W(\zeta) &= W(\gamma_1 \circ \gamma_2)W(\zeta) + W(\gamma_1)W(\gamma_2 \circ \zeta) \\ &+ W(\gamma_2)W(\gamma_1 \circ \zeta) - W(\gamma_1 \circ \gamma_2 \circ \zeta) - W(\gamma_2 \circ \gamma_1 \circ \zeta). \end{aligned} \tag{3.57}$$

We will interpret this identity in the following way. Consider two elements of the algebra γ_1 and γ_2 . The identity should hold for arbitrary ζ . This means that the identity between elements of $F\mathcal{L}_o$ (3.57) induces an identity between equivalence classes given by

$$\begin{aligned} (W(\gamma_1)W(\gamma_2) - W(\gamma_1 \circ \gamma_2))\iota - W(\gamma_1)\gamma_2 - W(\gamma_2)\gamma_1 \\ + \gamma_1 \circ \gamma_2 + \gamma_2 \circ \gamma_1 = 0, \end{aligned} \tag{3.58}$$

where ι should be understood as the identity element of $F\mathcal{L}_o$.

We will now use this identity to determine the eigenvalues of a matrix associated with the loop γ . To this end, we put $\gamma_1 = \gamma_2 = \gamma$ in (3.58) and get

$$\frac{1}{2}(W(\gamma)^2 - W(\gamma^2))\iota - W(\gamma)\gamma + \gamma^2 = 0. \tag{3.59}$$

This second order relation can be factorized as

$$(\gamma - \lambda_1\iota)(\gamma - \lambda_2\iota) = 0, \tag{3.60}$$

where $\lambda_1 + \lambda_2 = W(\gamma)$ and $\lambda_1\lambda_2 = \frac{1}{2}(W(\gamma)^2 - W(\gamma^2))$. If we now want to represent γ by a matrix, we see that it has at most two different eigenvalues. Therefore, this proves that a 2×2 representation suffices.

Let us now assume[†] that for at least one γ , which we will call γ_0 , $\lambda_1 \neq \lambda_2$. We have therefore established the form of the matrix $H(\gamma_0)$ associated with a particular loop γ_0 , and it is in diagonal form. Notice that because

[†] This assumption is not really needed, see Giles [35] for the exceptional case in which no such element exists.

holonomies are defined modulo a single similarity transformation at the basepoint, it is impossible to diagonalize the holonomies simultaneously for all possible γ s.

We will now determine the matrix element associated with an arbitrary element γ of the algebra. With this aim we define the elements of the algebra,

$$\phi_1 = \frac{\gamma_0 - \lambda_1 \iota}{\lambda_1 - \lambda_2}, \tag{3.61}$$

$$\phi_2 = \frac{\gamma_0 - \lambda_2 \iota}{\lambda_1 - \lambda_2}, \tag{3.62}$$

which behave as projectors, $\phi_1 \phi_2 = \phi_2 \phi_1 = 0$, $\phi_i^2 = \phi_i$, $\phi_1 + \phi_2 = \iota$. The reader can check by applying the definition (3.54) that $W(\phi_i) = 1$. The idea of introducing these elements is that in a matricial representation they will behave as projectors on the one-dimensional eigenspaces associated with each eigenvalue.

We now apply these projectors. Given an arbitrary element η of the algebra we define its “components” η_{ij} by

$$\eta_{ij} = \phi_i \eta \phi_j. \tag{3.63}$$

As can be readily seen from their definition and the definition of the projectors, these “components” satisfy

$$\eta = \sum_{i,j=1}^2 \eta_{ij}, \tag{3.64}$$

$$(\eta_1 \eta_2)_{ij} = \sum_{k=1}^2 (\eta_1)_{ik} (\eta_2)_{kj}, \tag{3.65}$$

$$W(\eta_{ij}) = \delta_{ij} W(\eta_{jj}). \tag{3.66}$$

We will now use these definitions to compute the “diagonal” elements of the algebra η_{ii} . They are given by

$$\eta_{ii} = W(\eta_{ii}) \phi_i \tag{3.67}$$

(no sum over i is assumed). Let us prove this for the “11 component”, the proof being totally analogous for the other component. We apply the Mandelstam identity to the following elements of the algebra, ϕ_2 , η_{11} and an arbitrary element ζ ,

$$\begin{aligned} W(\phi_2)W(\eta_{11})W(\zeta) &= W(\phi_2 \circ \eta_{11})W(\zeta) + W(\phi_2)W(\eta_{11} \circ \zeta) \\ &+ W(\eta_{11})W(\phi_2 \circ \zeta) - W(\phi_2 \circ \eta_{11} \circ \zeta) - W(\eta_{11} \circ \phi_2 \circ \zeta), \end{aligned} \tag{3.68}$$

and observing that $\phi_2 \circ \eta_{11} = 0$ and $W(\phi_2) = 1$,

$$W(\eta_{11})W(\zeta) - W(\eta_{11} \circ \zeta) - W(\eta_{11})W(\phi_2 \circ \zeta) = 0, \tag{3.69}$$

which implies the following relation among equivalence classes (since ζ is arbitrary),

$$\eta_{11} = W(\eta_{11})(\iota - \phi_2) = W(\eta_{11})\phi_1, \quad (3.70)$$

and therefore this is the expression of the “11 component” of the element η . All this language in terms of the elements of the algebra has a natural counterpart in terms of the actual components of the representation in terms of 2×2 matrices $H(\eta)$. In particular, the diagonal components of the matrix are therefore given by

$$H(\eta)_{ii} = W(\eta_{ii}). \quad (3.71)$$

The non-diagonal elements are not uniquely determined. Remember that to perform the construction we chose a particular γ_0 represented by a diagonal matrix. There exist similarity transformations that maintain the diagonality of $H(\gamma_0)$ but change the non-diagonal components of the representations of a generic element η .

To determine the non-diagonal components, Giles [35] introduces a procedure based on picking a second specific loop η_0 and fixing the value of some off-diagonal components of its matrix representation. In this way, the freedom to perform similarity transformations is frozen. In the 2×2 case, one needs to fix one component, say, $H(\eta_0)_{12} = 1$. The other component of this matrix is determined by

$$H(\eta_0)_{21} \equiv W(\phi_1 \circ \eta_0 \circ \phi_2 \circ \eta_0). \quad (3.72)$$

This completes the determination of all the matrix elements of the fixed element η_0 . The matrix elements of an arbitrary element η are given by

$$H(\eta)_{21} = W(\eta_{21} \circ \eta_0), \quad (3.73)$$

$$H(\eta)_{12} = W(\eta_{12} \circ \eta_0) / H(\eta_0)_{21}. \quad (3.74)$$

With this construction one actually has a representation of the algebra,

$$H(\gamma \circ \eta)_{ij} = \sum_k H(\gamma)_{ik} H(\eta)_{kj}, \quad (3.75)$$

which can be verified by combining the following expression (which is a consequence of equation (3.67)),

$$W(\eta_{jj} \circ \gamma_{jj}) = W(\eta_{jj})W(\gamma_{jj}), \quad (3.76)$$

and equations (3.72), (3.74) and (3.73).

Let us review what has been accomplished so far. We have established a procedure to reconstruct a holonomy given a set of quantities that satisfies the Mandelstam identities. In particular, this proves that one can reconstruct a holonomy from Wilson loops. The holonomy so constructed constitutes a representation of the group of loops the traces of which satisfy the Mandelstam identities.

The representation found only reproduces the Mandelstam identities of a general (2×2) matrix, the ones we used explicitly in the reconstruction. One could extend the method to take into account more specific Mandelstam identities (or inequalities). For instance, if one applies the above construction to a set of $SU(2)$ Wilson loops satisfying identities (3.48), (3.49) one does not necessarily end up with an $SU(2)$ holonomy but with a holonomy that satisfies the said identities. This could be accomplished, for instance, by an $SU(1, 1)$ holonomy as well.

An important point to notice is that the reconstructed holonomy from an arbitrary set of functions satisfying the Mandelstam identities will in general not correspond to a usual smooth connection, but rather to a generalized (“distributional”) connection. Because of this, when we formulate gauge theories purely in terms of loops, as we will do in the following sections, the formulation will usually correspond to this kind of generalized connections. If one wished to work with genuine connections one could do so by requiring extra conditions on the Wilson loops or in the case of loop representations on the corresponding wavefunctions.

Another point is the relation between the formalism introduced for the reconstruction theorem and that of the extended loop group introduced in chapter 2. As we pointed out at the beginning, the starting algebra defined on loops is isomorphic to a complexification of \mathcal{E}_D , the algebra of multitensor densities that satisfy the differential constraint. This helps to elucidate the nature of the algebra introduced by Giles, in the sense that it includes objects that are more general than loops, as is obvious due to the isomorphism with \mathcal{E}_D . In fact, the reconstruction theorems naturally work on \mathcal{E}_D , allowing us in general to reconstruct the gauge covariant matrix associated with any multitensor density $E_D^{a_1 x_1 \dots a_n x_n}$. In particular, one gets the generalized holonomies associated with the elements of the SeL group.

3.5 Loop representation

The results we introduced in the previous section show that Wilson loops are an overcomplete basis of solutions of the Gauss law. In other words, any gauge invariant function (and therefore any physically interesting quantity) can be expressed as a combination of products of Wilson loops. It is therefore natural to try to build a quantum representation purely in terms of loops. Two different constructions have been introduced that allow us to define a quantum representation for gauge theories purely in terms of loops. In the first one a transform is defined between the connection and loop representations. This procedure allows us to convert any gauge invariant operator or wavefunction into a corresponding object

in the loop representation. The second procedure is to introduce at a classical level an algebra of quantities parametrized by loops and take this algebra as the one to be represented in the first step of the Dirac quantization procedure. The resulting quantum representation is the loop representation.

Let us consider an analogy with a finite-dimensional system which clarifies the ideas underlying the loop representation. Suppose one is quantizing the non-relativistic free particle in one spatial dimension. Classically, the system is defined in terms of the canonical coordinates x and p , with Poisson brackets $\{x, p\} = 1$. Quantum mechanically, we take as the space of wavefunctions the functions of x , $\Psi(x)$. We will now construct a new representation for the system applying the ideas we will use to construct the loop representation.

Let us start by considering a transform approach. We consider a basis of states $W_k(x) = \exp(ikx)$, parametrized by a continuous variable k . Any wavefunction can be expanded in terms of this basis. We introduce a k -representation with wavefunctions $\Psi(k)$ given by the integral

$$\Psi(k) \equiv \int dx W_k^*(x) \Psi(x). \quad (3.77)$$

This equation is just the Fourier transform, and the reader may immediately recognize the k -representation as the ordinary momentum representation. The basis of states is an improper basis in the sense that it is not normalizable. Any operator in the position representation \hat{O}_x with a specific order in the canonical variables acting on functions $\Psi(x)$ can be translated into the k -representation by

$$\hat{O}_k \Psi(k) \equiv \int dx W_k^*(x) \hat{O}_x \Psi(x) = \int dx (\hat{O}_x^\dagger W_k(x))^* \Psi(x). \quad (3.78)$$

As an example of the use of the transform, let us consider the transform of a set of quantities that we will use in what follows. They are defined as $T^0(k) = \exp(ikx)$ and $T^1(k) = p \exp(ikx)$. It is immediate to see that one can express any classical quantity in terms of the T s. They satisfy a non-canonical algebra,

$$\{T^0(k_1), T^0(k_2)\} = 0, \quad (3.79)$$

$$\{T^1(k_1), T^0(k_2)\} = -ik_2 T^0(k_1 + k_2), \quad (3.80)$$

$$\{T^1(k_1), T^1(k_2)\} = i(k_1 - k_2) T^1(k_1 + k_2). \quad (3.81)$$

We now introduce a quantum representation of the algebra via the Fourier transform

$$\hat{T}^0(k_1) \Psi(k) = \int dx \exp(-ikx) \exp(ik_1 \hat{x}) \Psi(x) = \Psi(k - k_1) \quad (3.82)$$

$$\hat{T}^1(k_2)\Psi(k) = -i \int dx \exp(-ikx) \frac{d}{dx} (\exp(ik_2\hat{x})\Psi(x)) = k\Psi(k - k_2) \quad (3.83)$$

and it can be seen that the non-canonical classical Poisson algebra is reproduced by the quantum commutator algebra. Notice that k_1 and k_2 are arbitrary parameters.

It is important to notice that the action of these operators may be translated to the action on a space of kets $|k\rangle$,

$$\hat{T}^0(k_1)\Psi(k) = \langle k | T^0(k_1) | \Psi \rangle = \langle k - k_1 | \Psi \rangle, \quad (3.84)$$

thus,

$$\hat{T}^{0\dagger}(k_1)|k\rangle = |k - k_1\rangle = \exp(-ik_1\hat{x})|k\rangle. \quad (3.85)$$

Therefore,

$$\hat{T}^0(k_1)|k\rangle = \exp(ik_1\hat{x})|k\rangle = |k + k_1\rangle, \quad (3.86)$$

and analogously we find

$$\hat{T}^1(k_2)|k\rangle = \exp(ik_2\hat{p})|k\rangle = |k + k_2\rangle. \quad (3.87)$$

Notice that there is a factor ordering involved in the quantum algebra. The resulting ordering in the ket space representation is the *opposite* than the one in the space of wavefunctions.

Now consider a gauge theory (for instance $SU(2)$) in three dimensions described by canonical coordinates \mathbf{A}_a and \mathbf{E}^a with the usual Poisson brackets. Quantum mechanically, we consider wavefunctions of the connection, $\Psi[A]$. An (overcomplete) basis of states is given by the Wilson loops $W_\gamma[A]$. Again, the basis is parametrized by a continuous parameter, in this case the loop γ . The loop representation is defined in terms of the transform,

$$\Psi(\gamma) \equiv \int dA W_{\mathbf{A}}^*(\gamma)\Psi[\mathbf{A}]. \quad (3.88)$$

Again we can transform any operator by using the transform. Notice an important difference. In the case of the free particle we chose a basis of functions $\exp(ikx)$ whereas in the gauge theory case we chose a basis of *solutions* of the Gauss law $W_{\mathbf{A}}(\gamma)$. That is, by going to the loop representation one has automatically solved the Gauss law. Similar situations could arise in the case of the free particle (i.e., by choosing a basis of solutions of the Schrödinger equation) but we will not pursue these here, their meaning being quite transparent. Notice another crucial difference: while the transform used in the free particle case is a well known Fourier transform, the one used for the gauge theory case is only formal. Very

little is known about integration in spaces of connections and the theory of measure in this case is not well developed in general (for further developments see references [36, 40, 66, 67]). We will return to these issues in the following sections. Notice that the introduction of the loop transform can be thought of as performing an inner product in the connection representation between a wavefunction $|\Psi\rangle$ and elements of a basis $\langle\gamma|$. Therefore we can write

$$\Psi(\gamma) = \langle\gamma|\Psi\rangle = \int DA \langle\gamma|A\rangle \langle A|\Psi\rangle \quad (3.89)$$

through the introduction of the identity $1 = \int DA |A\rangle \langle A|$ which means that having a correct definition of the transform is equivalent to having an inner product in the connection representation.

Finally, it is not strictly true that for an arbitrary gauge group *single* Wilson loops are a basis of gauge invariant functions, but rather one needs to consider products of Wilson loops. This can be readily done, and the resulting wavefunction is a function of multiloops. We will discuss this in section 3.5.3.

Let us now explore the second approach, i.e., quantizing a non-canonical algebra of quantities. Again we consider the free particle and on the classical phase space we define the quantities $T^0(k) = \exp(ikx)$ and $T^1(k) = p \exp(ikx)$ which satisfy the non-canonical algebra discussed above. It is evident that one can express any classical quantity of interest in terms of this algebra. If one has a well defined transform, as is the case for the Fourier transform, one could proceed as before and find a quantum realization of this non-canonical algebra using the transform. It is therefore evident that the quantization that one would achieve coincides with the one that was introduced before via the transform. If one does not have a well defined transform at hand one can propose a quantum realization of the algebra and check that one reproduces the classical algebra at the quantum commutator level.

In this particular case we would propose

$$\hat{T}^0(k_1)\Psi(k) = \Psi(k - k_1), \quad (3.90)$$

$$\hat{T}^1(k_2)\Psi(k) = k\Psi(k - k_2) \quad (3.91)$$

and check that this representation reproduces the non-canonical algebra through quantum commutators. Notice that a choice of factor ordering must be made in the process. One can find the quantum expression for any classical quantity simply by writing the classical expression in terms of the T s and translating with due care for factor orderings.

Again a very similar construction (at least formally) can be performed for a gauge theory. Consider the set of classical quantities

$$T^0(\gamma) = W(\gamma), \quad (3.92)$$

$$T^a(\gamma_x^x) = \text{Tr}(H_{\mathbf{A}}(\gamma_o^x) \tilde{E}^a(x) H_{\mathbf{A}}(\gamma_x^o)), \quad (3.93)$$

where $W(\gamma)$ is the Wilson loop and $H_{\mathbf{A}}(\gamma_o^x)$ is the holonomy along the loop γ from the basepoint o to x . These quantities satisfy a closed non-canonical Poisson algebra and can be promoted to a quantum operator algebra. The result would be the loop representation. However, various detailed issues have to be discussed and we will postpone their treatment until section 3.5.2.

In spite of the appeal of these simplified analogies, the fact that gauge theories are infinite-dimensional systems of a non-Abelian nature implies that all the steps described above are considerably more involved. We will discuss these points in detail in the following chapters. Here we will discuss the definitions. In later chapters of this book the applications of the loop representation for gauge theories and general relativity will be explored in detail.

3.5.1 The loop transform

As we mentioned before, the loop transform involves a functional integral in the space of connections modulo gauge transformations. This makes it considerably more involved from a technical point of view than the transforms among representations of ordinary quantum mechanics which we discussed as an analogy. Little is known about integration theory in non-linear spaces both from a mathematical and a physical point of view.

The loop transform was introduced for the treatment of gauge theories in the early 1980s by Gambini and Trias [62]. At that time the approach was to assume that the transform existed and study *a posteriori* the physical implications of its existence. In a sense, a high degree of assurance of its existence was obtained through this approach, since it was proven in very concrete situations that results obtained via the transform coincided with those obtained via more traditional techniques. An important arena for this kind of test was the application of loop techniques in the lattice [95, 109]. In this case the loop transform is rigorously defined for any Yang–Mills theory in terms of the Haar measure of the group. For the case of general relativity, the loop transform in terms of Ashtekar variables was first introduced by Rovelli and Smolin [39] in again the same spirit. Only recently have studies of some mathematical rigor been performed on its existence. The main effort in this area is the result of the collaboration of Ashtekar, Isham, Lewandowski, Marolf, Mourão and Thiemann [36, 40, 203] and the work by Baez [66]. A particularly readable account from the point of view of physicists is given in reference [204], the pattern of which we follow in this section.

In this book we will use the transform as a heuristic tool to derive re-

sults. The correctness or otherwise of such results will be judged through their consistency both among themselves and with facts known through other means, not through the rigor of their derivation via the transform. We will discuss in each case in detail which are the arguments and consistency checks that support that result. In this section, however, we would like to give a brief glimpse of some of the mathematical developments that are taking place to put the loop transform and the results derived through it on a solid mathematical ground. It is yet to be seen if the particular results presented in this book will survive in the form presented when a rigorous operational definition of the transform is found.

The key idea that allows the definition of a measure of integration on the non-linear space of connections modulo gauge transformations is the use of the Wilson loop as a projection operator. This allows the definition of the so called “cylindrical measures”, which reduce the infinite-dimensional integral to a finite set of integrals over the gauge group. By demanding consistency of the various projections one ends with a theory of integration in infinite-dimensional spaces. Let us discuss in detail how this is accomplished. To investigate the ideas in a simpler context we discuss the definition of a measure in an infinite-dimensional but linear space, that of a Klein–Gordon field.

Consider a scalar field ϕ in flat spacetime satisfying the Klein–Gordon equation. The classical configuration space of such a theory is the set of all smooth field configurations on a spatial manifold that fall off appropriately at infinity. Quantum states for the theory are functions on the space of classical configurations $\Psi(\phi)$. One would like to introduce an inner product through an expression of the type $\int D_\mu \phi \bar{\Psi}(\phi) \Phi(\phi)$ where the integral ranges over the configuration space and our task is to introduce a suitable measure μ to perform the integral.

In order to do so we need to consider some particular functions on the configuration space. Possibly the simplest kind of function we can introduce are the functionals F defined by test functions $f(x)$ (the set of which is called Schwarz space) of the spatial manifold which we convolute with the classical configurations,

$$F_f(\phi) \equiv \int d^3x f(x)\phi(x) \quad (3.94)$$

and we require that $f(x)$ have appropriate regularity and falloff conditions such that the integral is well defined. With the above definition of the functionals F we are now in a position to introduce the idea of cylindrical functions. Consider a finite-dimensional subspace V_n of the Schwarz space and a basis of functions in it (e_1, \dots, e_n). Given a classical configuration $\phi(x)$ we can define its “projection” on the finite-dimensional subspace

which yields a set of n numbers,

$$(F_{e_1}(\phi), \dots, F_{e_n}(\phi)). \quad (3.95)$$

A function of the classical configuration space is called cylindrical with respect to V_n if its dependence on the classical configurations is through the set of n numbers that we introduced above, for some set of e_i s. That is, $g(\phi)$ is cylindrical iff

$$g(\phi) = G(F_{e_1}(\phi), \dots, F_{e_n}(\phi)) \quad (3.96)$$

for some function of n real variables G .

A cylindrical measure μ is a measure that allows us to integrate cylindrical functions. Each of these measures is defined by an infinite consistent family of measures $\{\mu_{e_1, \dots, e_n}\}$ each defined on all finite-dimensional spaces \mathbf{R}^n associated with each basis of vectors (e_1, \dots, e_n) . With these measures the integral of a cylindrical function is simply defined as an integral on \mathbf{R}^n ,

$$\int D_\mu(\phi)g(\phi) = \int G(x_1, \dots, x_n)D_{\mu_{e_1, \dots, e_n}}(x_1, \dots, x_n). \quad (3.97)$$

The key issue is that the above expression has to be well defined and consistent for any set V_n that one chooses. This restricts considerably the choice of the family of measures, imposing a set of consistency conditions. First consider the case of a function that is cylindrical with respect to two subspaces V_n and V'_m that are disjoint. Such functions are necessarily constants, so the integrals of such constants with μ_{e_1, \dots, e_n} and $\mu_{e'_1, \dots, e'_m}$ should be the same, which fixes a normalization condition for the measures. Next consider a function $g(\phi)$ that is cylindrical with respect to two subspaces $V_n \subset V'_m$. Such a function has associated with it two functions of n and m real variables $G(x_1, \dots, x_n)$ and $G'(x_1, \dots, x_m)$ that define it as a cylindrical function with respect to both spaces. Since the basis of V_n will be a linear combination of the basis of V'_m one can figure out the precise relationship between G and G' . Since the integral of G with the measures μ_{e_1, \dots, e_n} has to be the same as the integral of G' with the measure $\mu_{e'_1, \dots, e'_m}$ this imposes a consistency condition on the elements of the family $\{\mu_{e_1, \dots, e_n}\}$.

An example of a family of measures that is compatible with the consistency conditions introduced above is given by appropriately chosen Gaussian measures on \mathbf{R}^n . The well known quantum field theory of free fields is based on such measures. One can obtain the Fock representation by taking the Cauchy completion with respect to the inner product defined by the measure of the space of cylindrical functions on the classical configuration space. The hope is that the quantum field theory of interacting fields will arise from non-Gaussian cylindrical measures, as

has been shown in several particular cases [37]. It is important to notice that the above mentioned Cauchy completion leads to a quantum theory defined — in the case of field theories — by functions on an enlargement of the classical configuration space. This amounts to considering not only regular functions of the classical configurations but also distributions. We will see in chapter 11 that the consideration of distributional fields in the loop transform poses new challenges for the regularization of the theory in the loop (and extended) representation.

How do these constructions apply to gauge theories? For the case of Maxwell theory the construction is basically the same as above. The reason is that for an Abelian theory the space of connections modulo gauge transformations is again a linear space and one simply repeats the above construction considering functions of the classical configuration space given by the magnetic fields.

For the non-Abelian case the configuration space is a non-linear space. The way around this problem is to exploit the properties of holonomies to provide an analogue of the functionals introduced above. Given a fixed finite set of independent[§] loops β_1, \dots, β_n we now say a function $g(\mathbf{A})$ of the space of connections modulo gauge transformations is cylindrical with respect to this set of loops if and only if it depends on the connection through the value of the holonomies associated with the β_i s,

$$g(\mathbf{A}) = G(H_{\mathbf{A}}(\beta_1), \dots, H_{\mathbf{A}}(\beta_n)), \quad (3.98)$$

where G is a function defined on n copies of the gauge group.

A cylindrical measure is defined in a way analogous to that used before as a consistent family of measures $\mu_{\beta_1, \dots, \beta_n}$ on the n th tensor power of the gauge group. Again, there are consistency conditions to be met, which are more involved than in the simple example described previously. The remarkable fact is that there exist consistent families which define measures. An example of this is given by n copies of the Haar measure defined on the gauge group. Since this measure is defined without the introduction of any background structure it is diffeomorphism invariant.

We therefore have not only succeeded in introducing in a rigorous way a measure on the space of connections modulo gauge transformations but the measure is diffeomorphism invariant. It is therefore the kind of measures one would expect to be useful for analyzing problems in diffeomorphism invariant theories such as quantum gravity.

As we will see in chapter 7, for quantum gravity there is an additional complication in the sense that the gauge group is a complexified version of

[§] By independent loops we mean loops that have at least a segment that is not shared by the other loops with at most a finite number of intersections with the other loops.

$SU(2)$. It is remarkable that in an unrelated development, Hall [211] introduced a generalization of the Gaussian measure for complexified gauge groups. His motivation (the Bargmann representation of the harmonic oscillator, see chapter 4) is basically the root of the complex nature of the gravitational variables. Therefore by replacing the Haar measure by the Hall measure in the discussion above one can have a measure that is appropriate for the gravitational case. Development in this area is very rapid at present and may allow us to put on a solid ground many results that in this book we can only present formally.

It is yet to be seen if these kinds of measures produce physical theories of interest or if they are just mathematical curiosities. However, one cannot overstate the fact that until recently there were almost no measures known in non-linear infinite dimensional spaces and with these developments one may be able to gain enough experience to define measures that yield physical theories of relevance.

There has been a rapid development of these ideas. In particular rigorous definitions of the constraints and states of quantum gravity for the Euclidean case (where the theory is real) are currently under study. Many of the rigorous results provide a formal setting for the ideas we will discuss in chapters 7, 8, 9 and 10 [203].

3.5.2 The non-canonical algebra

There is an alternative procedure for introducing a loop representation that avoids having to go through an intermediate representation. The way to proceed is to go back to step one of the canonical quantization procedure we introduced in section 3.2.2 and pick a different classical algebra to quantize. We introduce the following quantities on the classical phase space of any gauge theory (or general relativity written in terms of Ashtekar’s variables),

$$T(\gamma) \equiv \text{Tr}(H_{\mathbf{A}}(\gamma)) = W_{\mathbf{A}}(\gamma), \tag{3.99}$$

$$T^a(\gamma_x^x) \equiv \text{Tr}(H_{\mathbf{A}}(\gamma_o^x) \tilde{\mathbf{E}}^a(x) H_{\mathbf{A}}(\gamma_x^o)), \tag{3.100}$$

$$T^{ab}(\gamma_y^y, \gamma_x^x) \equiv \text{Tr}(H_{\mathbf{A}}(\gamma_o^x) \tilde{\mathbf{E}}^a(x) H_{\mathbf{A}}(\gamma_x^y) \tilde{\mathbf{E}}^b(y) H_{\mathbf{A}}(\gamma_y^o)), \tag{3.101}$$

where $\gamma_x^x \equiv \gamma_x^o \circ \gamma_o^x$ and generically,

$$T^{a_1 \dots a_n}(\gamma_{x_1}^{x_2}, \dots, \gamma_{x_n}^{x_1}) \equiv \text{Tr}(H_{\mathbf{A}}(\gamma_o^{x_1}) \tilde{\mathbf{E}}^{a_1}(x_1) H_{\mathbf{A}}(\gamma_{x_1}^{x_2}) \dots H_{\mathbf{A}}(\gamma_{x_{n-1}}^{x_n}) \tilde{\mathbf{E}}^{a_n}(x_n) H_{\mathbf{A}}(\gamma_{x_n}^o)), \tag{3.102}$$

where $\gamma \equiv \gamma_o^{x_1} \circ \dots \circ \gamma_{x_n}^o$.

Notice that if the loop has multiple points the quantities depend on what sort of partition of the loop one performs and care should be taken to keep track of these dependences.

We see that the quantities $T(\gamma)$ are our well known Wilson loops and the other T s consist of “breaking up” the holonomy at points x_i , inserting an electric field and continuing the holonomy until back at the basepoint. It can be checked that these quantities are gauge invariant, i.e., they commute with the Gauss law. Generically we will speak of the T s with k electric fields inserted as “ T^k ”. The T^k s behave as multivector densities on the indices a_1, \dots, a_k at the points of the manifold x_1, \dots, x_k . They were first introduced by Gambini and Trias [34] (in their notation they were called W s, naturally extending the notation of the Wilson loops, although we have favored here the notation that has become standard among relativists, introduced by Rovelli and Smolin [38]).

As we argued before, the Wilson loops contain enough information to construct any gauge invariant function of the connection. By introducing T s of higher order the expectation is that one would be able to construct any quantity depending on the electric fields, and therefore have an algebra of classical quantities which is sufficiently general to express any quantity of physical interest in terms of them. We have already shown examples of how to construct quantities of physical interest in terms of the Wilson loops, for instance,

$$\Delta_{ab}(\gamma_o^x)W_{\mathbf{A}}(\gamma) = \text{Tr}(\mathbf{F}_{ab}(x)H_{\mathbf{A}}(\gamma_x^x)). \quad (3.103)$$

So we see we can retrieve information about the F_{ab}^i . One can also retrieve information about momentum dependent quantities from the T s of higher order, for instance, a trace of two electric fields,

$$\text{Tr}(\tilde{\mathbf{E}}^a(x)\tilde{\mathbf{E}}^b(x)) = \lim_{\gamma \rightarrow x} T^{ab}(\gamma)(x, y), \quad (3.104)$$

where by $\lim_{\gamma \rightarrow x}$ we mean the limit in which we shrink the loop to a point at x (and consequently the point y tends to x). In terms of the Ashtekar new variables for general relativity this trace plays the role of the spatial metric.

We will not by any means prove here that one could reconstruct any quantity of physical relevance in terms of these quantities. It suffices to realize that most quantities that one is usually interested in can be written as limits of the T s and that therefore they seem to span the classical phase space of the theory of interest.

An interesting point is that the T^0 s with the T^1 s close an algebra, the “small T algebra”. Let us compute it in an explicit fashion for an $SU(N)$ Yang–Mills theory. Because the Poisson bracket of \mathbf{A}_a with itself is zero it is immediate that

$$\{T(\gamma), T(\eta)\} = 0. \quad (3.105)$$

In order to obtain the Poisson bracket of T with T^1 we compute, start-

ing from the canonical brackets (3.21),

$$\{H_A(\gamma)^A{}_B, \tilde{E}_j^a(x)\} = i(X^j)^C{}_D H_A(\gamma_o^x)^A{}_C H_A(\gamma_x^o)^D{}_B \oint_\gamma dy^a \delta(x - y), \tag{3.106}$$

where the indices A, B, \dots refer to the fundamental representation of $SU(N)$ and run from 1 to N , and $(X^j)^A{}_B, j = 1, \dots, N^2 - 1$ are the generators of the algebra satisfying (3.8). Combining this equation with the following identity for the generators of $SU(N)$,

$$\sum_{i=1}^{N^2-1} (X^i)^A{}_B (X^i)^C{}_D = \delta^C{}_B \delta^A{}_D - \frac{1}{N} \delta^A{}_B \delta^C{}_D, \tag{3.107}$$

we get

$$\{T^a(\gamma_x^x), T(\eta)\} = -i \left(T(\gamma_x^x \circ \eta_x^x) - \frac{1}{N} T(\gamma) T(\eta) \right) X^{ax}(\eta), \tag{3.108}$$

where $X^{ax}(\eta)$ is the multitangent of order one. Notice that the Poisson bracket vanishes if η and γ do not have a common point.

Similarly, for two T^1 's,

$$\begin{aligned} \{T^a(\gamma_x^x), T^b(\eta_y^y)\} = & -i \left(T^b(\eta_y^x \circ \gamma_x^x \circ \eta_x^y) - \frac{1}{N} T(\gamma) T^b(\eta_y^y) \right) X^{ax}(\eta) + \\ & + i \left(T^a(\gamma_x^y \circ \eta_y^y \circ \gamma_y^x) - \frac{1}{N} T(\eta) T^a(\gamma_x^x) \right) X^{by}(\gamma). \end{aligned} \tag{3.109}$$

In the general $SU(N)$ case, in the right-hand side of the Poisson brackets we have products of the elements of the non-canonical algebra. It is only for the case of $SU(2)$ that we can rearrange these terms as linear superpositions of elements of the non-canonical algebra. This means that if one wants to find a quantum representation, one needs to consider a non-canonical algebra incorporating products of the T s. As a consequence, wavefunctions in the loop representation so constructed will have to depend on more than one loop. We will return to these issues in the next section.

For the $SU(2)$ case the algebra can be written in a very compact fashion. The Poisson brackets of the T 's are a linear combination of T 's evaluated on loops obtained from the original ones through very simple rules of fusion and rerouting through the intersection of the loops. The result is zero if the loops do not intersect. The action can be understood in a simple fashion through a graphical representation as shown in figure 3.1. The explicit form of the algebra is,

$$\{T^a(\gamma_x^x), T(\eta)\} = \frac{i}{2} \sum_{\epsilon=-1}^1 \epsilon X^{ax}(\eta) T(\gamma \circ \eta^\epsilon) \tag{3.110}$$

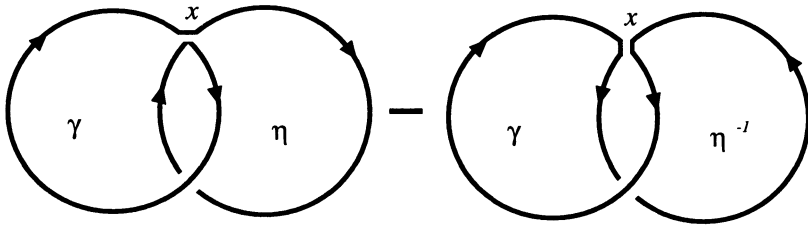


Fig. 3.1. The graphical representation of the commutator between T^1 and T^0 . The commutator is zero if the “hand” (the point at which one inserted the triad) of the loop γ does not “grab” the loop η . The figure shows the two reroutings that arise in the two terms that result from the commutator.

$$\begin{aligned} \{T^a(\gamma_x^x), T^b(\eta_y^y)\} &= -\frac{i}{2} \sum_{\epsilon=-1}^1 \epsilon X^{ax}(\eta) T^b(\eta_y^x \circ (\gamma_x^x)^\epsilon \circ \eta_y^y) \\ &\quad + \frac{i}{2} \sum_{\epsilon=-1}^1 \epsilon X^{by}(\eta) T^a(\gamma_x^y \circ (\eta_y^y)^\epsilon \circ \gamma_x^x), \end{aligned} \tag{3.111}$$

where η^ϵ represents either η or η^{-1} .

If one wants to consider higher order T s, one needs T s of arbitrarily large order in order to close the algebra, so strictly speaking it is not closed or only closes in a completion. For instance, for the $SU(2)$ case, the Poisson brackets are schematically

$$\{T^n, T^m\} \sim T^{n+m-1}. \tag{3.112}$$

The detailed commutation relations can be seen in reference [38].

The need to consider the infinite family of T s to attain closure is just another manifestation of the overcompleteness of the loop basis. Although we know that we have “too many” loops, we are forced to include them all to span the classical phase space of the theory. It is tempting to try to construct a quantum theory by only representing the “small” algebra of T and T^1 . Unfortunately it is not clear if these quantities are enough to span the classical phase space of gauge theories. There is a certain sense in which they do, though technicalities arise for the case of non-compact groups[68]. Even if they did in some particular cases, they are not very convenient for expressing some quantities of physical relevance, such as the Hamiltonian of Yang–Mills theories (and general relativity). Therefore from a practical point of view one resorts to the higher T s to express quantities of interest.

Let us now sketch the quantization of this non-canonical algebra for

the $SU(2)$ case. The idea is that we have now completed step one of the Dirac quantization procedure introduced in section 3.2.2: we have picked a (non-canonical) algebra of classical quantities that (modulo subtleties) spans the classical phase space of the theory. We now move on to the second step of the quantization program: to find a representation of this algebra in terms of operators acting on a space of wavefunctions. We pick wavefunctions of loops $\Psi(\gamma)$ and we represent the T operators by

$$\hat{T}(\eta)\Psi(\gamma) \equiv \Psi(\eta \circ \gamma) + \Psi(\eta \circ \gamma^{-1}), \quad (3.113)$$

$$\hat{T}^a(\eta)(x)\Psi(\gamma) \equiv - \sum_{\epsilon=-1}^1 \epsilon X^{ax}(\gamma)\Psi(\gamma \circ \eta^\epsilon). \quad (3.114)$$

These kinds of expressions face regularization difficulties. They could be regularized by considering, for instance, “thickened out” loops and defining a regularized T^1 operator via a two parameter congruence of loops. A discussion of this can be found in references [2, 69].

Similar expressions for the quantum representation of the higher order T s can be seen in reference [39]. One can check that these quantum operators satisfy quantum commutation relations that in the limit $\hbar \rightarrow 0$ (the T^n s have a prefactor of \hbar^n if one does not set \hbar to one as we have been doing) reproduce the classical commutation relations mentioned above. All this is discussed in reference [39].

The resulting quantum theory is the loop representation that we introduced before. One can check all this — at least heuristically given the various ill-defined constructions that are involved — by formally using the transform. One can represent the T operators in the connection representation (using an appropriate factor ordering) and then transform them into operators in the loop representation. One immediately finds that the representation introduced above corresponds to ordering the electric fields to the left in the connection representation.

Is it preferable to introduce the loop representation via a quantization of a non-canonical algebra or via a transform? At this moment this is largely a matter of choice. Both definitions, as we have seen, face various points where ill-defined mathematical operations are rampant. In fact, it is not difficult to see that many of these difficulties are somewhat connected. The important point that we have shown in this section is that there is nothing “strange” about the loop representation. It is a quantum representation that can be obtained directly, applying the traditional Dirac quantization procedure. It is by no means “subordinated” to the connection representation and has an existence on the same footing as any other quantum representation. The main difference between the loop representation and other more traditional ones is the use of an overcomplete non-canonical set of operators.

3.5.3 Wavefunctions in the loop representation

Now that we have introduced the loop representation, what about the wavefunctions in such a representation. Will any function of a loop do as a wavefunction or are there other requirements to be met?

As we discussed in section 3.5.1 wavefunctions in the loop representation can be thought of as transforms of functionals of connections weighed by products of Wilson loops,

$$\Psi(\gamma_1, \dots, \gamma_n) = \int d\mathbf{A} W_{\mathbf{A}}^*(\gamma_1) \cdots W_{\mathbf{A}}^*(\gamma_n) \Psi[\mathbf{A}]. \quad (3.115)$$

An immediate property that follows from the fact that the Wilson loops appear in the transform as a product is that wavefunctions are symmetric under interchange of arguments,

$$\Psi(\gamma_1, \dots, \gamma_i, \dots, \gamma_j, \dots, \gamma_n) = \Psi(\gamma_1, \dots, \gamma_j, \dots, \gamma_i, \dots, \gamma_n). \quad (3.116)$$

Wavefunctions in the loop representation will inherit a series of properties of Wilson loops. To begin with, they are functions with domain in the group of loops \mathcal{L}_o . Since the Wilson loops are traces of holonomies, they are actually functions of conjugacy classes of the group of loops; for example, for a function of a single loop,

$$\Psi(\gamma) = \Psi(\eta \circ \gamma \circ \eta^{-1}) \quad \forall \eta. \quad (3.117)$$

For functions of multiloops a similar expression holds at each entry. It is immediate from the previous expression that

$$\Psi(\gamma \circ \eta) = \Psi(\eta \circ \gamma). \quad (3.118)$$

It is here that the machinery introduced in chapter 1 and 2 will become useful, since we will all the time be operating on functions of the group of loops. In previous approaches wavefunctions in the loop representation were considered as functionals of parametrized curves with additional restrictions and functional derivatives played the role of differential operators. The consistency of this approach is delicate since one must ensure that the application of differential operators preserves the conditions imposed on the functional space. These issues are automatically taken care of by considering functions on the group of loops and the corresponding differential operators discussed in chapter 1.

Another important property is that wavefunctions inherit the Mandelstam identities among Wilson loops that we discussed in section 3.4.1.

To begin with, the Mandelstam identities relate products of Wilson loops of different orders. In particular for any group of $N \times N$ matrices, this allows us to express a product of Wilson loops in terms of expressions involving at most N factors and consequently to reduce any wavefunction to one depending on at most N loops.

Let us now discuss in detail the implications of the Mandelstam identities for the case of two-dimensional special groups (such as $SU(2)$, $SL(2, C)$, etc). In this case, the fundamental identity reads

$$\begin{aligned} \Psi(\gamma_1, \dots, \gamma_i, \gamma_j, \dots, \gamma_n) &= \Psi(\gamma_1, \dots, \gamma_i \circ \gamma_j, \dots, \gamma_n) \\ &+ \Psi(\gamma_1, \dots, \gamma_i \circ \gamma_j^{-1}, \dots, \gamma_n). \end{aligned} \quad (3.119)$$

An important consequence of this identity is that it will make it possible to express *any functional of an arbitrary number of loops in terms of a functional of a single loop*. That is, in these cases one can construct a loop representation considering functions of a single loop.

That wavefunctions depend on a single loop does not imply that they are unconstrained, since many identities for wavefunctions of single loops can be derived from (3.119). Consider expression (3.119) for the case of two entries and put $\gamma_j = \iota$ and $\gamma_i = \eta$. Then

$$\Psi(\eta, \iota) = 2\Psi(\eta), \quad (3.120)$$

which implies, considering (3.119) with $\gamma_i = \iota$, $\gamma_j = \eta$, that

$$\Psi(\eta) = \Psi(\eta^{-1}). \quad (3.121)$$

Finally applying (3.119) to

$$\Psi(\gamma \circ \eta, \beta) = \Psi(\eta \circ \gamma, \beta), \quad (3.122)$$

we get

$$\Psi(\gamma \circ \eta \circ \beta) + \Psi(\gamma \circ \eta \circ \beta^{-1}) = \Psi(\eta \circ \gamma \circ \beta) + \Psi(\eta \circ \gamma \circ \beta^{-1}). \quad (3.123)$$

Given this set of identities one can reconstruct the identities for multi-loops.

On a practical note, although these identities are fundamental in the sense that any other can be derived from them, they can imply very non-trivial relations between wavefunctions even at the single loop level.

Apart from these identities, as we mentioned in section 3.4.1 there are inequalities in terms of holonomies that reflect properties of the group (for instance that tell us if the group is $SU(2)$ rather than $SU(1, 1)$). At the moment the treatment of these inequalities is unclear. For instance, it is not established if they imply any restrictions on the wavefunctions. They imply restrictions on the quantities that one quantizes. This would not be the first time that a quantization was attempted in terms of variables that satisfy inequalities. For instance, usual quantizations of gravity based on metric variables have to deal with the fact that the metric of space must have a Euclidean signature. Or in a more simplified situation, consider the quantization of the hydrogen atom in the position representation in spherical coordinates, where the radial variable has to be positive definite. Dealing with the detailed problems posed by the fact that one is

quantizing in a representation where variables satisfy inequalities requires a degree of sophistication of the theory that has not yet been attained. For a deeper discussion of the problems of inequalities and quantization see reference [41].

3.6 Conclusions

In this chapter we introduced several physical techniques for the analysis of gauge theories. In terms of these, many of the notions of loops that we introduced in the first two chapters find a natural application. We introduced the loop representation and have shown that wavefunctions in the loop representation are simply functions of the group of loops. To develop in some detail the relationships known at present between the loop techniques of chapters one and two and the physical theories of chapter three is the subject of the rest of this book. In chapter 4, 5 and 6 of the book we will apply these techniques to gauge theories. In chapters 7-11 we will apply them to general relativity in terms of Ashtekar's variables.