Nuclear Physics B331 (1990) 80–152
North-Holland

LOOP SPACE REPRESENTATION OF QUANTUM GENERAL RELATIVITY

Carlo ROVELLI

INFN Sezione di Roma, Università di Roma, Roma, Italy and Dipartimento di Fisica, Università di Trento, Trento, Italy

Lee SMOLIN

Department of Physics, Yale University, New Haven, CT 06511, USA and Department of Physics, Syracuse University, Syracuse, NY 13210, USA and Dipartimento di Fisica, Università di Trento, Trento, Italy

Received 26 September 1988
(Revised 30 March 1989)

We define a new representation for quantum general relativity, in which exact solutions of the quantum constraints may be obtained.

The representation is constructed by means of a noncanonical graded Poisson algebra of classical observables, defined in terms of Ashtekar's new variables. The observables in this algebra are nonlocal and involve parallel transport around loops in a three-manifold Σ. The theory is quantized by constructing a linear representation of a deformation of this algebra. This representation is given in terms of an algebra of linear operators defined on a state space which consists of functionals of sets of loops in Σ. The construction is general and can be applied also to Yang–Mills theories.

The diffeomorphism constraint is defined in terms of a natural representation of the diffeomorphism group. The hamiltonian constraint, which contains the dynamics of quantum gravity, is constructed as a limit of a sequence of observables which incorporates a regularization prescription. We give the general solution of the diffeomorphism constraint in closed form. It is spanned by a countable basis which is in one-to-one correspondence with the diffeomorphism equivalence classes of multiple loops, which are a generalization of the link classes studied in knot theory. Then we explicitly construct, in closed form, a large space of solution of the entire set of constraints, including the hamiltonian constraint. These turn out to be classified by the ordinary knot and link classes of Σ.

The space of solutions that we find is a sector of the physical states space of nonperturbative quantum general relativity. The failure of perturbation theory is thus shown to be not relevant to the problem of the existence of a nontrivial physical state space in quantum gravity. The relationship between this new loop representation and the self-dual representation of Ashtekar is illuminated by means of a functional transform between states in the two representations. Questions of the completeness of the solution space, the meaning of the physical operators and the physical inner product, are discussed, but not, so far, resolved.

1. Introduction

In this paper we introduce a new approach to the quantization of general relativity* that allows us to obtain exact results about the structure of the physical state space of that theory**. This approach is an outgrowth of the reformulation of canonical general relativity of Ashtekar [3,4], and relies essentially on the simplification of the canonical theory achieved by the use of Ashtekar's new variables.

The main idea of this approach is to define a representation of quantum general relativity in terms of functionals over sets of loops. This idea was suggested by the discovery that when the hamiltonian constraint, or Wheeler–DeWitt equation [5], is reformulated in terms of Ashtekar's variables it admits a large class of exact solutions, which are related to loops in three dimensions [6]. In this paper we will show that, by the use of this new representation, called the *loop representation*, exact simultaneous solutions to all of the quantum constraint equations may be explicitly obtained.

The loop representation is obtained, following Isham's ideas [7,8], by quantizing a noncanonical Poisson algebra of nonlocal classical observables, which we call the $\mathscr{T}$ algebra. Thus, this work is a realization of the idea stressed by Isham that, at the nonperturbative level, where the governing symmetry is diffeomorphism invariance and not Poincaré covariance, the correct quantization of general relativity should be based on a noncanonical algebra and, consequently, should not involve the conventional Fock, or particle, structure.

Instead, the space of solutions of the constraints that we find using the loop representation turns out to be represented in terms of a countable basis of solutions which is in one-to-one correspondence with the knot and link classes of the spatial manifold. The knot and link classes are the topologically inequivalent ways in which a set of loops may be knotted and linked. Their classification forms the subject of a branch of mathematics called knot theory [9].

More explicitly, we describe below the *general solution to the (three-dimensional) diffeomorphism constraints* in terms of a countable basis associated with what we will call the *generalized* link classes of the manifold. These are a generalization of the usual knot and link classes which include the cases in which the loops intersect, overlap, and kink. A subset of these will be shown to be solutions to the hamiltonian constraint as well. We do not claim to have the general solution of the hamiltonian constraint.

There are many open problems associated with this approach, among which are the construction of a physical inner product and the algebra of physical observables.

* For a review of the present status of perturbative and nonperturbative methods in quantum gravity see Isham [1].

** A short account of the results of this paper appeared in Rovelli and Smolin [2]. Preliminary results appeared in ref. [16]. See also the contribution of the two authors to the proceedings of the Osgood Hill conference [2].

Thus, we do not now claim that the loop representation gives a complete quantization of general relativity.

However, we do claim that these results show that the failure of the perturbative quantization cannot be taken as an indication that quantum general relativity does not exist. Indeed, the existence, as well as the detailed structure, of the solutions that we find could not have been indicated by perturbative calculations. Thus, it seems that the perturbation expansion around flat Minkowski space does not reach the exact solutions and consequently does not give a reliable description of the short distance physics of the exact theory. Instead, as many people have argued, these results indicate that nonperturbative effects, coming from the strong coupling of the gravitational field at Planck scales, do drastically change the structure of the physical states in such a way to invalidate perturbative calculations based on a semiclassical approximation around a fixed classical metric.

Since perturbation theory does not capture the short distance physics correctly, it is not meaningful to study the limit in which the cutoff is removed perturbatively. Thus, the present results show that the failure of the perturbative approach does not bear on the issue of whether a quantum theory based on the dynamics of general relativity exists.

Here follows a brief overview of what is done in this paper:

In sect. 2 we work in the context of classical general relativity. We define the classical observables whose Poisson algebra defines the $\mathscr{T}$ algebra. These observables, that we denote generically as T observables, depend on loops in the three-dimensional space Σ. The simplest of them (which constitutes a maximal commuting subalgebra) is given by the trace of the holonomy of Ashtekar's connection $A_a(x)$ around a given loop γ

$$T[\gamma] = \mathrm{Tr}\,\mathrm{P}\exp\left(\oint_\gamma A\right), \tag{1}$$

(P means path ordered) and has an intuitive physical interpretation in terms of the parallel transport of a left-handed spinor in the gravitational field along γ. The other observables are obtained by inserting along the loops, in n different points, the conjugate Ashtekar variable $\tilde{\sigma}^a(x)$, (which is the densitized triad). We call T^n the observables obtained with n insertions. The Poisson brackets of the T's close and have the structure

$$\{T^n, T^m\} = T^{n+m-1}. \tag{2}$$

These Poisson brackets define the graded algebra $\mathscr{T}$. $\mathscr{T}$ may be graphically described in terms of breaking and rejoining of the loops at their intersection; for instance the Poisson bracket of a T with no insertion and a T with one insertion is described in fig. 1. This algebra codes the symplectic structure defined by the canonical Poisson brackets of classical general relativity.

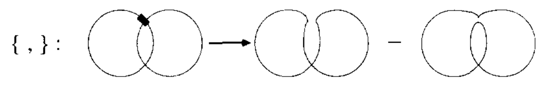

Fig. 1. An example of Poisson brackets between loop observables (the precise meaning of this drawing will be defined later in the paper).

The quantization is done by finding a linear representation of this algebra, or, more precisely, of a deformation of this algebra, in terms of an algebra of linear operators, which we call $\tilde{\mathscr{T}}$, on a space of functionals over the space of the sets of loops in Σ. The loop functionals represent unconstrained states of quantum gravity and the linear operators $\tilde{T} \in \tilde{\mathscr{T}}$ are the quantum operators corresponding to the classical observables T. This defines the "kinematics" of quantum gravity and is described in sect. 2.

Although the subject of this paper is general relativity, it is important to point out that the loop representation can be constructed for any theory in which a connection plays the role of a canonical coordinate. Thus it can be applied also to Yang–Mills theories. As we will describe in a separate paper, the techniques developed in this paper provide a systematic approach to the hamiltonian formulation of Yang–Mills theories in terms of loops functionals [12]*. A particularly illuminating example of the loop representation is also provided by the quantization of the free Maxwell theory. This is discussed in ref. [13].

In the context of Yang–Mills theories, the idea of expressing the quantum theory in terms of loop functionals is not new [10]; for example, it has been strongly advocated by Polyakov [11]. The loop representation differs from these earlier works in two principal ways. First, we work in a hamiltonian, rather than a path integral framework. Second, the loop representation makes crucial use of multiple loops.

We will assume in this paper that the three-manifold Σ is compact, and without boundary. In this case the dynamics of general relativity is given entirely by the constraints [14, 15]**. These are also constructed in sect. 3. The internal gauge SU(2) constraint of Ashtekar's theory is automatically taken into account in the quantization, because the T observables are SU(2) gauge invariant. The diffeomorphism constraint is defined by considering the linear representation of the Diff(Σ) group in the states space, defined by the natural action of the Diff(Σ) group on the loops. We demonstrate that the $\tilde{T}$ operators transform under such a group in the same way as the corresponding classical observables. We show that the generators of the spatial diffeomorphisms in the loop representation have the correct commuta-

* We may note that in Yang–Mills theories one does not have as strong a motivation for going to a loop basis as one has in gravity, since loop states are not eigenstates of the hamiltonian, as they are in general relativity. Nevertheless, as we discuss in ref. [12], the loop representation suggests a new approach to numerical computations in QCD.

** A review of the hamiltonian formulation of general relativity and of the issues raised by its quantization on a compact manifold is given by ref. [15].

tion relations among themselves and with the observables. As a result they can be, and are, identified with the diffeomorphism constraints of quantum gravity.

The hamiltonian constraint is defined in terms of the $\tilde{T}$'s. Classically it can be expressed as a limit of a suitable sequence of T observables. The quantum constraint will be defined by the limit of the corresponding quantum operators. We may note that this is the way that any operator product must be defined in a quantum field theory. What is nice about the loop representation is that it naturally provides a regularization for the hamiltonian constraint. This completes the definition of the loop representation.

In sect. 4 we discuss the relationship between the loop representation and the self-dual representation, which is the natural quantization diagonal in the Ashtekar connection. We express this relation in terms of a linear mapping, called $\mathscr{F}$, between the two representations. We also show that by introducing a measure on the space of the connections, this mapping may be expressed as a functional transform, which we call the loop transform. At least at a formal level, this transform provides an alternative route to the definition of the loop representation.

Then, in sect. 5, we study the solutions of the quantum constraint equations. We obtain the following results.

(i) The entire space of states annihilated by the constraints D_a, which generate spatial diffeomorphisms, is found in terms of an explicit countable basis. The elements of this basis are in one-to-one correspondence with the *generalized link classes* of Σ, which are the equivalence classes, under Diff(Σ) of sets of piecewise smooth loops in Σ.

(ii) Among these states are some which are also annihilated by the hamiltonian constraint, and are thus exact physical states of the gravitational field. Included in these is a sector whose basis is in one-to-one correspondence with the subset of the generalized link classes of Σ which are based on sets of smooth, nonintersecting loops. These are the well-known *ordinary link classes*, whose classification is the subject of knot theory.

Finally in the last section we discuss what remains to be understood before we can know whether or not the loop representation provides a completely satisfactory quantization of general relativity.

However, before beginning the technical work of constructing the loop representation, we conclude the introduction with a summary of the basic issues that motivate our approach to the quantization of general relativity.

1.1. NONPERTURBATIVE AND GENERALLY COVARIANT QUANTUM FIELD THEORY

The construction of a quantum field theory without a background geometry to describe gravitation raises several issues that are not faced in any flat space quantum field theory. In particular, what are the consequences of: (i). The absence of Poincaré invariance? (ii). The fact that the quantization and regularization must be done without a c-number background metric or connection structure? (iii). The requirement that the physical states are invariant under spatial diffeomorphisms? (iv). The fact that the local dynamics of the theory is driven by a hamiltonian constraint, rather than a hamiltonian? (v). The difficulties of the definition of diffeomorphism invariant physical observables?

From the point of view of the standard quantization schemes general relativity, treated nonperturbatively so that the metric is never split into a classical background and a quantum fluctuation, has very unusual features. For example, Poincaré invariance, which in one form or another is necessary for the definition of almost all forms of quantum field theory in four dimensions, is absent. The absence of the Poincaré group means the absence of the notion of particles, which are the irreducible representations of the Poincaré group in the Hilbert space. This disappearance of the particle is well known already in quantum field theory in curved space-time. The absence of particles and of Poincaré invariance means the absence, in the Hilbert space of the quantum theory, of a standard Fock structure, and therefore of the entire machinery of conventional quantum field theory. From this point of view, the failure of perturbative methods to yield any sensible generally covariant quantum field theories is not surprising.

As a second illustration of the way that a quantum theory of gravity must differ from the background dependent quantum field theories we are familiar with, let us consider an issue which is of crucial importance for understanding conventional quantum field theories: that of the short distance behavior. In an ordinary quantum field theory, the usual short distance divergences are measured in terms of the background metric. In nonperturbative quantum gravity, in which there is no background metric, there is no invariant c-number measure of the distance between two points, and questions about the short distance structure are much more difficult to formulate. Moreover, in quantum gravity the physical states have to be in the kernel of the diffeomorphism constraint, which means that they are diffeomorphism invariant. This means that any n-point functions, defined as the expectation value of products of local operators, must be constant as the n points are moved around by diffeomorphisms.

The contrast could thus not be stronger between the behavior of the two-point function in an ordinary quantum field theory, and in a diffeomorphism invariant theory without a background metric. In the first case we have

$$\langle 0|\phi(x)\phi(y)|0\rangle \sim 1/(x-y)^2, \tag{3}$$

where the norm is taken with respect to the fixed background metric. On the other hand, in quantum gravity we must have

$$\langle\psi|\phi(x)\phi(y)|\psi\rangle \sim \text{constant}, \quad \text{if} \quad x \neq y, \tag{4}$$

whenever $|\psi\rangle$ is any physical state, and ϕ any local operator. Thus, in nonperturbative quantum gravity the n-point functions cannot be used to extract any information corresponding to the usual notion of short distance behavior.

Of course, part of the point is that local observables such as $\phi(x)$ are not meaningful, as they do not commute with the constraints that generate diffeomorphisms. Thus, we cannot really conclude from this simple argument that the usual kinds of short distance structure will not be present in a quantum theory of gravity★. If we want to describe the short distance behavior of quantum general relativity nonperturbatively, we need first to have a nonperturbative characterization of the physically meaningful observables.

This raises what we believe is a crucial issue for constructing a satisfactory nonperturbative quantization of general relativity, which is the problem of the physical observables. We will discuss this problem in detail below, for now we may note that this problem has two aspects, both of which are highly nontrivial. These are: (i). What are the observables of the classical theory? (ii). Which (if any) classical observables can be translated into operators in the context of a regularization procedure that is nonperturbative *and* does not destroy diffeomorphism invariance?

The physical observables of the classical theory have to commute with the constraints, and, hence, be invariant under the full four-dimensional diffeomorphism group. An important corollary of this is that for the compact case any physical observable is also a constant of motion. In the case of a pure gravitational field, with no matter, and with a spatially compact topology, we know in fact of *not a single classical physical observable*. This is not really surprising, as to know a physical observable explicitly would be to know an explicit expression for a constant of the motion as a function of arbitrary initial data for the Einstein equations. But it highlights the fact that we actually have a very poor understanding of the general physical interpretation the full Einstein equations.

We will see that the fact that we do not know the classical physical observables causes a major difficulty in the construction of the quantum theory. We will find an algebra of physical observables, but we will not be able to give them an interpretation in terms of a correspondence with the observables of the classical theory.

The nontriviality of issue (ii) can be appreciated by noting that, in the familiar metric representation, an infinite number of three-dimensional diffeomorphism invariant observables can be written down explicitly as integrals over densities constructed from the phase space variables. However, every member of this set, save one, $\int_\Sigma \tilde{p}^{ab} q_{ab}$, is a nonpolynomial function of the basic observables. As we shall see, the loop representation provides a regularization procedure which is compatible

★ Note that if a perturbatively sensible quantum theory of gravity exists, one would expect its short distance structure to be conventional, the notion of distance being given in this case by the background metric around which the perturbation theory is defined. Thus, if one believes that the short distance structure of space-time is described by a perturbative theory, one might argue that the conventional short distance structure will be present in quantum gravity, and could be extracted from a diffeomorphism invariant state by using an appropriate physical observable. However, this kind of argument cannot be applied to general relativity, which we know must be quantized nonperturbatively, if it is to be quantized at all.

with three-dimensional diffeomorphism invariance, but it only can be applied to observables which are polynomial in the basis variables. Thus, even if the usual spatial diffeomorphism invariant observables constructed from the metric could be extended to physical observables of the theory, it is not at all clear that any of these could survive a passage to an algebra of physical observables in the quantum theory which is based on a nonperturbative regularization of the theory. Given these difficulties, the question of what the short distance structure of quantum general relativity is highly nontrivial.

Do these difficulties necessarily imply that quantum general relativity cannot exist? We do not think so. The standard Poincaré invariant quantum field theories are the only kind of quantum theory that has so far been fully understood in four dimensions, but it is very likely that these are not the only kinds that exist. A large part of the problem of quantum gravity is contained in the problem of finding a quantum mechanical description of a field theory which differs from conventional quantum field theory in that the governing symmetry principle is diffeomorphism *invariance* rather than Poincaré *covariance*.

Thus, we work in the following framework: We do not consider any modification to general relativity or to the basic axioms of quantum mechanics. Instead, we look for a new approach to the construction of a quantum field theory which is nonperturbative and allows the implementation of the full diffeomorphism invariance of the classical theory.

There are many ways in which one might try to achieve this. We choose to work in the framework of canonical quantization, as amended by Dirac for the case of systems with first class constraints. A program for the canonical quantization of a constrained theory with weakly vanishing hamiltonian, as in the case of general relativity, can be summarized in the following steps [4, 5, 7, 15].

(i) Choice of a preferred algebra of classical observables, that we call elementary observables, which is closed under Poisson brackets. The traditional choice is given by some set of canonically conjugate observables with canonical Poisson brackets.

(ii) Choice of a linear space $\mathcal{S}$, physically interpreted as "the space of the unconstrained quantum states" and a set of linear operators on $\mathcal{S}$ in one-to-one correspondence with the elementary observables. These are the elementary quantum operators.

(iii) Definition of the quantum operator constraints C_i, corresponding to the classical constraints.

(iv) Solution, on $\mathcal{S}$, of the quantum constraint equations*

$$C_i|\psi\rangle = 0. \tag{5}$$

* These must be appropriately modified to incorporate a suitable regularization procedure, as we discuss in detail below.

The linear subspace $\mathcal{S}_{\rm ph}$ of $\mathcal{S}$ of the solutions is interpreted as the space of the physical quantum states.

(v) Definition of the physical observables. These are operators on $\mathcal{S}$ which are well defined on $\mathcal{S}_{\rm ph}$; that is they commute with the constraint operators.

(vi) Choice of a scalar product on $\mathcal{S}_{\rm ph}$ such that the self-adjointness properties of the physical observables are assured. As discussed for instance in refs. [4, 17], it may be convenient mathematically to have a Hilbert structure on the space of unconstrained states, but this structure can have, in general, no physical significance. This is because there is generally no natural way to bring down the Hilbert structure of $\mathcal{S}$ to $\mathcal{S}_{\rm ph}$, where we need it for the definition of physical expection values [17]. In this paper we will not introduce any Hilbert structure on $\mathcal{S}$.

We refer to quoted references for a careful discussion of the many issues raised by such program. Here we want to discuss some specific questions raised by the attempt to implement this program in the case of the gravitational field.

1.1.1. Elementary observables. As is well known, the first attempt to realize the canonical quantization program for general relativity was the definition of the metric representation during the 60's [5, 18], in which the elementary observables are the three-metric and its conjugate momentum.

Ashtekar's reformulation of general relativity naturally suggested a different realization of the program in which the new variables $A_a(x)$ and $\tilde{\sigma}^a(x)$ are chosen to be the elementary operators [4]. This leads to the so-called self-dual representation, which we will briefly describe later.

A third representation is the loop representation, which we introduce in this paper, defined by using the T observables as the elementary observables.

The $\mathcal{T}$ algebra formed by the T's is, as we said, a noncanonical Poisson algebra. In quantum field theory the fixed time canonical commutation relations (CCR) are related to the Fock structure. Any representation of the CCR is equivalent to a representation defined by an infinite dimensional quasi-invariant measure [19]. However, very little is known about quasi-invariant measures besides the gaussian measures, which are canonically related to Fock space. If we are searching for a quantum theory without the standard Fock structure, we have either to look for non-gaussian representations of the CCR, or, as Isham has long argued [7], we have to change the departure point, and use a noncanonical algebra, as we do here.

1.1.2. Constraints. There are several ways to deal with the first class constraints and the gauge invariances that they generate. Since in what follows we will use a mixed approach, it is worthwhile to discuss this problem in detail. The first way is to define the constraints as operator functions of the elementary quantum observables, by choosing an ordering in the definition of the classical constraints as functions of the classical elementary observables [20]. The well-known issue of closure of the constraint algebra is then raised. This is the way in which we will define the hamiltonian constraint.

If we have control of the invariance group generated by the constraints, a powerful way to define the constraint operators is to define a representation of the group on $\mathscr{S}$ and then identify the quantum constraints as the generators of this representation [7, 8]. The constraint algebra is then naturally satisfied. What one has to worry about is, in this case, that the transformation properties of the other objects in the theory are the correct ones. This ensures the correctness of the algebra of the constraints and the observables together. We will follow this way when we give the definition of the constraint associated with the spatial diffeomorphisms.

A third possibility for dealing with the quantization of a closed algebra of constraints is simply to choose a set of elementary observables that commute with the constraints. Then the constraints can be simply forgotten in the rest of the theory, since such observables coordinatize the reduced classical system. For a detailed discussion of this see, for example, ref. [17]. We will deal in this way with the internal gauge constraint.

1.1.3. Divergences. Any formal definition of a quantum field theory contains infinite quantities. For the theory to be well defined these have to be eliminated from physical quantities in some way. In flat space quantum field theory a definite procedure exists that allows to subtract the divergencies order by order in a perturbation expansion.

There is a feature of the dynamics of general relativity that makes it, in this respect, very different from the flat space quantum field theories and which could make the problem of the infinities *less* problematic. This is because, as we shall argue on general grounds, it is possible that *general relativity needs a regularization, but it does not need a renormalization*; at least for the definition of the constraints and the physical state space.

Different kinds of infinities appear in the theory.

The first type comes from the singularities in observables defined at a point. These singularities can be treated, as usual, in the framework of the theory of distributions. We shall discuss how all singularities of this kind can be eliminated by suitably smearing the observables.

The second set of singularities comes from products of local operators, which are products of distributions. These products may be avoided by carefully using elementary observables which, suitably smeared, do not involve ill-defined quantities.

The crucial singular object in the theory is then the hamiltonian constraint. This situation is analogous to the one of flat space theories: The singularities in the definition of the fields operators are treated simply by interpreting the field as distributions and the real source of infinities is the hamiltonian, by virtue of its containing products of local operators.

The situation in gravity, however, is different from the one in flat space theories because the dynamical operator is a constraint. This means that we are not interested in knowing its action on all the states, we are only interested in knowing its kernel. Let us contrast the two equations

$$H|\psi\rangle = E|\psi\rangle, \qquad \mathscr{C}|\psi\rangle = 0. \tag{6,7}$$

The first, which defines the dynamics of an ordinary quantum field theory, requires both regularization and renormalization. If $H_\delta = \sum_i H_\delta^i$ is a sequence of regularized hamiltonians, containing a sum of regulated terms H_δ^i then the stationary state condition is

$$\lim_{\delta \to 0} \sum_i Z(\delta)_i H_\delta^i |\psi\rangle = E_{\text{renormalized}} |\psi\rangle, \tag{8}$$

where $Z(\delta)_i$ is a multiplicative renormalization, which often differs from term to term.

In quantum gravity, on the other hand the physical state condition is simply

$$\lim_{\delta \to 0} \mathscr{C}^\delta |\psi\rangle = 0, \tag{9}$$

where $\mathscr{C}^\delta$ are the regulated operators. As the right-hand side is simply zero, renormalization may not be required to construct the solutions to this equation. This difference occurs for the following reason. Renormalization is necessary in an ordinary quantum field theory because the hamiltonian must be well defined on the whole Hilbert space. However, in quantum gravity the hamiltonian constraint needs only to be well defined on its kernel, but here it is represented just by the operator multiplication by zero. The states outside the kernel are unphysical, and while the regulated hamiltonian constraint must be defined on the whole state space, including the unphysical states, there is no reason for its limit, as the regularization is taken away, to be well defined anywhere except on the physical state space*.

In other words, because of the particular features of gravity, we do not need to renormalize $\mathscr{C}$. It is sufficient to regularize it in order to have a precise definition of its kernel.

This fact is hidden in the metric representation, in which the hamiltonian constraint is the sum of two terms. The second of these does not need to be regularized, therefore it is not clear how the sum behaves in the limit. In the self-dual representation, on the contrary, because the constraint consists of a single term, it is possible to define a regularization scheme in which $\mathscr{C}^\delta \Psi$ is proportional to $1/\delta$, and apply these ideas.

The situation is even better in the loop representation. As we shall see the hamiltonian constraint is *naturally* defined in the classical theory as a limit of observables which are linear combinations of the elementary observables of the loop representation. Having the dynamical operators defined as a limit of the elementary observables is unpleasant for the classical theory. But it turns out to be nice

* This idea was implicitly suggested in ref. [6].

quantum mechanically, since it provides a natural regularization scheme in which the dynamical operator is defined as a limit of elementary operators*.

1.1.4. Reality conditions. In the Ashtekar formalism one first treats complexified general relativity. In order to reduce the theory to the real sector the constraints must be supplemented by a reality condition.

The reality condition cannot be interpreted as a constraint on the state functions, in the manner of the other constraints. This is because in the quantum theory reality conditions are related to the Hilbert structure, rather than to the linear structure. The translation of the † operation, in other words, necessarily involves the inner product, as under quantization real quantities become hermitian operators.

For instance if we have a one-dimensional system, we can complexify the (x, p) phase space, then define the observables $a = x + ip$, $b = x - ip$, and quantize the system by looking for a linear representation of the a, b Poisson algebra. The real sector of the classical theory is given by the reality condition $b = \bar{a}$. In the quantum theory this amounts to choosing a Hilbert product such that the operator $\hat{b}$ is the conjugate of the operator $\hat{a}$.

Any quantization can be seen as the quantization of the complexified classical theory, up to the point in which the introduction of the Hilbert structure singles out the hermitian operators.

In the quantization of a constrained system the physical inner product is defined on the space of solutions to the constraints, and not on the larger state space. This means that any reality condition that the phase space variables of a theory satisfies may be ignored during the construction of the physical state space, and imposed afterwards as a condition on the inner product defined on the physical state space. The inner product on the physical state space must be chosen so that physical observables, whose classical counterparts are real when the classical reality conditions are imposed, are represented by hermitian operators. In the body of this paper we formulate the theory on the unconstrained function space, and solve the constraints. We thus will not introduce a Hilbert structure on the state space and the reality conditions will play no role.

1.1.5. Physical observables. The problem of the identification of the physical observables is a crucial one for any nonperturbative quantization of a gravitational theory. We do not solve this problem in this paper; rather the progress we are able to make concerning the nonperturbative structure of the physical state space serves to bring out the importance of this question.

Very briefly, the problem is that since we do not know any physical observables of the classical theory, we are not able to identify the physical operators of the

* A more subtle question is the choice of the topology in which to take the limit (9). If we had a Hilbert structure, the Hilbert norm would clearly be the correct choice, but, as we just discussed, we do not have one on the unconstrained states. In this paper we work in the pointwise topology, which is the only natural choice at this stage in the development of the theory. Whether or not this choice will turn out to lead to a completely satisfactory theory, with a physical inner product and a full set of physical states and operators, is presently an open question.

quantum theory with specific physical quantities. Furthermore, if we cannot do this we do not know how to implement the reality conditions of the theory, and we cannot identify the physical inner product of the theory. Without a physical inner product and an identification of at least some operators on the physical Hilbert space with physical observables of the classical theory, the theory cannot be given a physical interpretation. We will return to these issues more than once in the course of this paper.

However, it is worth mentioning that in spite of this difficulty there is one general and important conclusion we will be able to draw about the physical observables of quantum gravity at the nonperturbative level. We will show that the space of solutions to the spatial diffeomorphism constraints can be constructed explicitly in terms of a countable basis in which the basis elements are in one-to-one correspondence with the (generalized) link classes of the manifold. Any observable which commutes with the spatial diffeomorphism constraints must then be expressible as a matrix in this basis, which means that it is expressible in terms of linear operations on the link classes. This means that any diffeomorphism invariant observable of the theory measures topological information about the extended structures that are in correspondence with the basis states in the loop representation. Thus, at the nonperturbative level, the "short distance" structure of quantum general relativity must be described in terms of topological relations of nonlocal observables rather than metrical relations of local observables.

2. Loop variables in the classical theory

The aim of this section is to introduce a class of nonlocal observables on the phase space of general relativity which will be the starting point for the quantization in the next section.

The main properties of these observables, which will be denoted by the letter T, are the following. (i). They are invariant under SU(2) gauge transformations. (ii). They parametrize the gauge* constraint surface of the phase space, so that any gauge invariant observable may be expressed in terms of them. (iii). They form a closed Poisson algebra, which we call $\mathscr{T}$ algebra. (iv). They carry a realization of the diffeomorphism group of the three-manifold.

The T observables depend on a loop, or a set of loops, in the three-manifold, and on a number of points on the loops. We will denote by T^n the T observables that depend on n points on the loop. The set of T observables form a *graded algebra* under the Poisson bracket, where the grading is given by the nonnegative integer n.

We begin this section by briefly reviewing the Ashtekar formulation of general relativity. Then we introduce the observables T^0 and T^1. They form a preferred set

* In this paper we use the following nomenclature: *gauge* refers only to the local SU(2) invariance; *diffeomorphism* refers only to the spatial diffeomorphism group; and *physical*, as in "physical observable" means invariant under gauge transformations, diffeomorphisms and transformations generated by the hamiltonian constraint.

of the T variables in that their Poisson brackets close. We call the subalgebra that they define the small $\mathcal{T}$ algebra. At this stage we introduce a graphical notation which is useful for calculations. After this we introduce the entire class of T^n observables and compue the full $\mathcal{T}$ algebra. We then study how the usual local variables of general relativity, and in particular the three-metric, can be recovered from the T's, and, finally, we discuss the way in which the constraints are expressed in terms of these loop variables.

2.1. ASHTEKAR'S NEW VARIABLES

We briefly recall here the basic equations of Ashtekar's formulation of canonical general relativity [3, 4].

We assume, to begin with, a three-dimensional manifold Σ, whose topological and differentiable structure is fixed. In this paper it is assumed that Σ is compact, and without boundary*.

Ashtekar's variables are coordinates on a complexification of the phase space of general relativity, extended by the introduction of local frame fields in order to allow couplings to spinors. The fundamental variable in Ashtekar's formalism is a complexified SU(2) (or SL(2, $\mathbb{C}$)) connection, denoted $A_{aAB}(x)$. It may be geometrically interpreted as the projection into the three-manifold of the left-handed spin connection of the four-metric**. $A_{aAB}(x)$, being a complexified SU(2) connection, is symmetric in its spinor indices AB. The curvature of $A_{aAB}(x)$, called $F_{abAB}(x)$, also symmetric in the spinor indices, is related to the self-dual, or left-handed, piece of the riemannian curvature tensor.

Conjugate to $A_{aAB}(x)$ is the variable $\tilde{\sigma}^{aAB}(x)$. It is geometrically interpreted as the densitized inverse frame field. It is also symmetric in spinor indices. In what follows we often will not indicate the spinor indices. $\tilde{\sigma}^{a}(x)$ and $A_a(x)$ are treated as 2×2 matrices in multiplications and traces. Spinor indices are raised and lowered with the usual two-spinor antisymmetric object ϵ_{AB} and its inverse [4].

The relation with the standard canonical variables is the following. The three-metric q_{ab} is given by***

$$q(x)q^{ab}(x) = -\tfrac{1}{2}\mathrm{Tr}\left[\tilde{\sigma}^{a}(x)\tilde{\sigma}^{b}(x)\right], \tag{10}$$

where q is the determinant of q_{ab}. The relation between the connection and the

* The extension of the loop space formalism to manifolds with boundaries or asymptotic conditions is in progress.

** Lower case Latin letters from the beginning of the alphabet denote three-dimensional abstract indices, upper case Latin indices from the beginning of the alphabet are two-component abstract spinor indices.

*** The variable $\tilde{\sigma}^{a}(x)$ that we use in this paper is equal to $1/\sqrt{2}$ times the variable with the same name used in refs. [3, 4]. This is the reason for the $\frac{1}{2}$ in eq. (10), and for the absence of the $1/\sqrt{2}$ factor in eq. (12).

standard variables is that $A_a = \Gamma_a + (i/\sqrt{2})\Pi_a$, where Γ_a is the SU(2) spin connection on Σ compatible with $\tilde{\sigma}^a$, and Π_a is linearly related to the extrinsic curvature.

$A_{aAB}(x)$ is thus a function of both the three-metric $q_{ab}(x)$ and its conjugate momenta. In spite of this fact, the $A_{aAB}(x)$ are all simultaneously commuting in the Poisson brackets of general relativity*,

$$\{A_{aAB}(x), A_{bCD}(y)\} = 0. \tag{11}$$

The $\tilde{\sigma}^{aAB}$'s have vanishing Poisson brackets with each other, and satisfy

$$\{A_{aAB}(x), \tilde{\sigma}^{bCD}(y)\} = i\delta_a^b \delta_A^{(C}\delta_B^{D)}\delta^3(x,y). \tag{12}$$

Of particular interest in this paper will be quantities defined by parallel transport along loops in Σ using the connection $A_a(x)$. If γ is a parametrized loop in Σ, then we will define the SL(2, $\mathbb{C}$) matrix

$$U_\gamma(s,t) \equiv \mathrm{P}\exp\left(\int_{\gamma(s)}^{\gamma(t)} A\right), \tag{13}$$

where the notation

$$\int_\gamma A = \int A_a(\gamma(s))\dot{\gamma}^a(s)\,\mathrm{d}s \tag{14}$$

is used for the line integral of a one-form, and P means path ordered. By $U_\gamma(s) = U_\gamma(s,s)$ we denote the parallel transport all around the loop, also known as the holonomy, beginning and ending at s. By virtue of the fact that it lives in SL(2, $\mathbb{C}$), it satisfies the identity

$$U_\gamma(s)_{AB} = -U_{\gamma^{-1}}(s)_{BA}. \tag{15}$$

The main result of Ashtekar's formalism is that the form of the constraints is greatly simplified. There is a constraint which generates the SU(2) internal gauge transformations

$$\mathscr{G}(x) \equiv \mathscr{D}_a\tilde{\sigma}^a(x) \sim 0, \tag{16}$$

where $\mathscr{D}_a$ is the A_a covariant derivative. This is, of course, just the usual Gauss law constraint**.

* We denote Poisson brackets by { , }, and quantum commutators by [,].

** Up to this point there is a complete correspondence with the hamiltonian formalism for a (complexified) SU(2) Yang–Mills theory, the $\tilde{\sigma}^{aAB}$'s corresponding to the Yang–Mills electric field.

In the presence of the SU(2) constraint, the diffeomorphism and hamiltonian constraints of general relativity take the form

$$\mathscr{C}_a(x) = \mathrm{Tr}\left[F_{ab}(x)\tilde{\sigma}^b(x)\right] \sim 0, \tag{17}$$

$$\mathscr{C}(x) = \mathrm{Tr}\left[F_{ab}(x)\tilde{\sigma}^a(x)\tilde{\sigma}^b(x)\right] \sim 0. \tag{18}$$

In contrast to their form in the usual variables, these are both polynomial in $A_a(x)$ and $\tilde{\sigma}^a(x)$, as well as homogeneous in the latter.

These constraints define the complex version of general relativity. In order to have real minkowskian or euclidean general relativity we have to impose additional reality conditions. We demand that $\tilde{\sigma}^a$ is hermitian, namely $\tilde{\sigma}^{a\dagger} = \tilde{\sigma}^a$, where the $\dagger$ refers to a fixed hermitian conjugate operation in the spin space, and that its time derivative is also hermitian.

2.2. THE T^0 AND T^1 OBSERVABLES AND THE SMALL $\mathscr{T}$ ALGEBRA

We consider continuous, piecewise smooth, nondegenerate mappings α: $S_1 \to \Sigma$. Each such mapping gives us a parametrized closed curves in space. We call such curves loops, and we denote them by greek letters $\alpha, \beta, \gamma, \ldots$, or, in components, $\alpha^a(s), \beta^a(s), \ldots$, where the loop parameter s will always be considered modulo 2π, that is $s = s + 2\pi$. The inverse of a curve α is defined to be the curve $(\alpha^{-1})^a(s) = \alpha^a(2\pi - s)$.

The first observable that we introduce is T^0. T^0 depends on a loop γ and is defined to be simply the trace of the holonomy of Ashtekar's connection along γ*

$$T[\alpha] \equiv \mathrm{Tr}\, U_\alpha(s) = \mathrm{Tr}\,\mathrm{P}\exp\left(\oint_\alpha A\right). \tag{19}$$

The set of $T[\alpha]$'s coordinatize the phase space of general relativity in the following way. As shown in ref. [6], any holomorphic and gauge invariant functional of the A_a's can be determined by expressing it as a functional of the $T[\alpha]$. This is because any SU(2) connection is determined up to gauge transformations by the traces of its holonomy, and holomorphic functionals of an SL(2, $\mathbb{C}$) connection are determined by their values when the connection is restricted to lie in SU(2)**. Thus, the $T[\alpha]$ are holomorphic coordinates on the gauge constraint surface on the phase space of $(A_a, \tilde{\sigma}^a)$, which is a complexification of the phase space of general relativity.

* Note that the superscript 0 is suppressed. In general the n on a T^n is suppressed when it is indicated by the number of indices.

** Holomorpic is here defined in terms of a fixed hermitian structure on the spin bundle, as is explained in ref. [4].

In order to have a complete set of coordinates on the gauge constraint surface of the phase space we must add observables which involve $\tilde{\sigma}^a(x)$. We want them to be SU(2) gauge invariant, at the same time, however, we will require that they do not involve products of more than one $\tilde{\sigma}^a(x)$ at a single point. This is because we expect that in the quantum theory such coincident operator products will not be well defined. The simplest way of satisfying these requirements is to insert the $\tilde{\sigma}^a(x)$'s inside of traces of parallel transport around loops. We thus define a second set of observables called T^1 as follows. For any loop γ, and loop parameter s, $T^a[\gamma](s)$ is given by inserting $\tilde{\sigma}^a(x)$ along the holonomy of γ, at the point $x = \gamma(s)$. That is, we define

$$T^a[\gamma](s) \equiv \mathrm{Tr}\left[U_\gamma(s)\tilde{\sigma}^a(\gamma(s))\right]. \tag{20}$$

$T[\gamma]$ is invariant under reparametrization of γ; this is also true if the reparametrization changes the orientation, since the trace of an SL(2, $\mathbb{C}$) matrix is equal to the trace of its inverse. $T^a[\gamma](s)$ is not reparametrization invariant since it depends on a preferred value of the loop parameter s. However, it is reparametrization covariant. Let γ' be a reparametrization of γ with the same orientation, $\gamma'(s) = \gamma(f(s))$, $f'(s) > 0$. Then clearly

$$T^a[\gamma'](s) = T^a[\gamma](f(s)), \tag{21}$$

so that we can identify these two objects.

$T^a[\gamma](s)$, however, changes sign under reparametrizations of the loop that change its orientation. This follows from eq. (15) and from the symmetry of $\tilde{\sigma}^a(x)$ in the spinor indices. Therefore, the T^1 observable depends on *oriented unparametrized loops with a preferred point*. We will sometimes also associate an orientation also to the loop on which the T^0 operator depends; the orientation is arbitrary and no final result will depend on it, but it is useful to have an orientation in the course of calculations.

We now introduce a graphical notation that will be useful in the following. Let us denote a class of loops equivalent under reparametrization by a closed line. Since T^0 depends on single loops, but not on their parametrization, we can note a $T[\gamma]$ simply by means of the corresponding equivalence class of parametrized loops, as in fig. 2a. Our convention will be that the intersections and the points of nondifferentiability on the loops and their order (which are parametrization invariant concepts) are reproduced by the drawing. For instance a loop with one self-intersection is

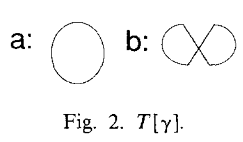

Fig. 2. $T[\gamma]$.

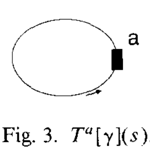

Fig. 3. $T^a[\gamma](s)$.

depicted in fig. 2b. Then we can denote the equivalence classes of oriented loops with one particular selected point by putting a dot in one point and an arrow that fixes the orientation. Using this we denote the T^1 as in fig. 3. We call the dots on the drawing of the loops which represent the insertion of a $\tilde{\sigma}^a(x)$ "hands".

Before introducing the T^n with $n > 2$ we discuss the properties of the T^0, T^1 set. Their main property is that they form a closed Poisson algebra.

We have, of course, for any two loops α and β,

$$\{T[\alpha], T[\beta]\} = 0. \tag{22}$$

To express the rest of the Poisson algebra we need to refer to loops that result from the breaking and joining of two loops that intersect at a point. Let α and β be two loops that intersect at the point $x = \alpha(\hat{s}) = \beta(\hat{t})$. We can construct a loop by starting from x, going first around α and then around β. We call this loop $\alpha \#_x \beta$. The subscript is needed because the two curves may intersect in more than one point. We will not write the subscript where the context is clear, and we also use $\alpha \#_s \beta$ for $\alpha \#_{\alpha(s)} \beta$. To complete the definition we need to specify how the loop $\alpha \#_x \beta$ is parametrized. We adopt the following convention. If $\hat{s}$ and $\hat{t}$ are the values of the parameters of α and β at the intersection, $\alpha(s + \hat{s})$ and $\beta(t + \hat{t})$ are parametrized loops that start and end at the intersection (recall that the loop parameters are defined modulo 2π). We define

$$\alpha \#_x \beta(u) = \begin{cases} \alpha(2u + \hat{s}) & \text{for} \quad 0 < u < \pi, \\ \beta(2u + \hat{t}) & \text{for} \quad \pi < u < 2\pi. \end{cases} \tag{23}$$

We will use the notation $u'(s)$ and $u(t)$ for the values of the parameter of $\alpha\#\beta$ that correspond to the points $\alpha(s)$ and $\beta(t)$. That is the functions u' and u are defined by $\alpha\#\beta(u'(s)) = \alpha(s)$, $\alpha\#\beta(u(t)) = \beta(t)$ and may be easily computed from eq. (23).

Now we can compute the Poisson algebra. The Poisson bracket of a T^0 with a T^1 gives a term proportional to a T^0 as follows. By using the identity

$$\begin{aligned} \left\{ \tilde{\sigma}^a_{AB}(x), U_\gamma^{CD}(0,s) \right\} &= -i \frac{\delta}{\delta A_a^{AB}(x)} \left[\mathrm{P}\exp\left(\int_0^s A_b(\gamma(u)) \dot{\gamma}(u)^b \, \mathrm{d}u \right) \right]^{CD} \\ &= -i \int_0^s \mathrm{d}u \, \delta^3(\gamma(u), x) \dot{\gamma}^a(u) U_\gamma(0,u)^C{}_{(A} U_\gamma(u,s)_{B)}{}^D, \end{aligned} \tag{24}$$

we have

$$\begin{aligned}\{T^a[\gamma](s), T[\eta]\} &= iU_\gamma(s)^{AB}\{\tilde{\sigma}^a_{AB}(\gamma(s)), T^0[\eta]\}\\ &= iU_\gamma(s)^{AB}\int \mathrm{d}t\,\delta^3(\gamma(s),\eta(t))\dot{\eta}^a(t)U_\eta(t)_{(AB)}\\ &= \frac{-i}{2}\int \mathrm{d}t\,\delta^3(\gamma(s),\eta(t))\dot{\eta}^a(t)\big[T^0[\gamma\#\eta] - T^0[\gamma\#\eta^{-1}]\big].\end{aligned} \quad (25)$$

The bracket is zero unless η intersects γ at the point s.

It is convenient to introduce a special notation for the singular terms that appear in the commutators of the $\mathscr{T}$ algebra. We define

$$\Delta^a[\gamma,\eta](s) \equiv \tfrac{1}{2}\int \mathrm{d}t\,\delta^3(\gamma(s),\eta(t))\dot{\eta}^a(t). \quad (26)$$

In subsect. 2.4 we show that these singularities are harmless and can be eliminated by interpreting the T observables as distributions, such that the singularities in the commutators are eliminated by averaging over suitable test functions. These singularities thus have the same character as the singularities in the canonical commutation relations $\{\phi(x),\pi(y)\} = \delta^3(x,y)$, which are present in every classical field theory.

It is also worth while, for later convenience, to introduce the notation

$$(\alpha\#\beta)^{><} = \alpha\#\beta, \qquad (\alpha\#\beta)^{\overset{\vee}{\wedge}} = \alpha\#\beta^{-1}, \quad (27)$$

and a symbol $\diamond$ that may take the value $><$ or $\overset{\vee}{\wedge}$. The reasons for this notation will become clear in a moment. Using this notation, the Poisson brackets (25) becomes

$$\{T^a[\gamma](s), T[\eta]\} = -i\sum_{\diamond}(-1)^{|\diamond|}\Delta^a[\gamma,\eta](s)T^0\big[(\gamma\#\eta)^{\diamond}\big], \quad (28)$$

where we have defined $|><| = 0$, $|\overset{\vee}{\wedge}| = 1$.

The Poisson bracket of two T^1's is again proportional to another T^1. We have

$$\begin{aligned} i\{T^a[\gamma](s), T^b[\eta](t)\} &= U_\gamma(s)^{AB}\big\{\tilde{\sigma}^a_{AB}(\gamma(s)), U_\eta(t)^{CD}\big\}\tilde{\sigma}^b_{CD}(\eta(t))\\ &\quad + U_\eta(t)^{CD}\big\{U_\gamma(s)^{AB}, \tilde{\sigma}^b_{CD}(\eta(t))\big\}\tilde{\sigma}^a_{AB}(\gamma(s))\\ &= \Delta^a[\gamma,\eta](s)\big(\mathrm{Tr}\big[U_\gamma(s)U_\eta(u,t)\tilde{\sigma}^b(\eta(t))U_\eta(t,u)\big]\\ &\qquad - \mathrm{Tr}\big[U_{\gamma^{-1}}(s)U_\eta(u,t)\tilde{\sigma}^b(\eta(t))U_\eta(t,u)\big]\big)\\ &\quad - \Delta^b[\eta,\gamma](t)\big(\mathrm{Tr}\big[U_\eta(t)U_\gamma(v,s)\tilde{\sigma}^a(\gamma(s))U_\gamma(s,u)\big]\\ &\qquad - \mathrm{Tr}\big[U_{\eta^{-1}}(t)U_\gamma(v,s)\tilde{\sigma}^a(\gamma(s))U_\gamma(s,u)\big]\big). \end{aligned} \quad (29)$$

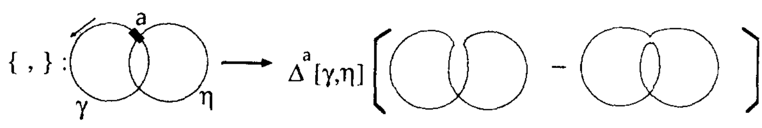

Fig. 4. Poisson brackets of a T^0 and a T^1.

The result is then

$$\{T^a[\gamma](s), T^b[\eta](t)\} = i\sum_{\diamondsuit}(-1)^{|\diamondsuit|}\Delta^b[\eta,\gamma](t)T^a\big[(\eta\#_t\gamma)^{\diamondsuit}\big](u(s))$$

$$-i\sum_{\diamondsuit}(-1)^{|\diamondsuit|}\Delta^a[\gamma,\eta](s)T^b\big[(\gamma\#_s\eta)^{\diamondsuit}\big](u(t)). \quad (30)$$

We call the Poisson algebra of the T^0, T^1 observables, defined by these equations, the small $\mathscr{T}$ algebra.

Now we show how these calculations may be performed in terms of the graphical notation that we introduced. The Poisson brackets are an operator that sends two observables, each parametrized by a loop, to a linear combinations of observables parametrized by loops that are related to the original ones by simple topological operations. The Poisson brackets of a T^0 and T^1 such that the hand indicating the presence of the $\tilde{\sigma}^a(x)$ is at one of the intersection points can be written* as in fig. 4. Similarly, the Poisson brackets of two T^1's such that each of the two $\tilde{\sigma}^a(x)$'s is at a point of intersection is drawn as in fig. 5. It is clear that there is a simple graphical rule at work here. The { , } operator acts at each point where there is a hand. There is an elementary action at any hand, which we now describe.

If the hand is not at an intersection point of the two loops, the result is zero. If the hand is on γ at a point of intersection with η and has index ^{a}, the action is the following. We obtain two new loops from γ and η by removing the hand, breaking each of them at the point of intersection, and rejoining each resulting leg of γ with one of η. In one of these two loops the orientations are consistent, in the other they clash so that we have to reverse the orientation of η before rejoining; we take the difference between the first loop and the second one and, finally, we multiply the result by $\Delta^a[\gamma,\eta]$.

This graphical rule expresses exactly the contents of eq. (24). It is summarized in fig. 6. Now the total action of the operator { , } is given by its action over the hand

* In the figures we use the following convention to distinguish the two ways in which the four legs arriving at an intersection can be rearranged. We consider the rearrangement in which the two legs coming from the right (and the two coming from the left) are connected (the > < figure) to be the one in which the orientation of the two loops matches naturally. Note that in fig. 4 the handed loop is oriented, while the orientation of the other loop is arbitrary; by reversing its orientation, the role of the two resulting loops is inverted, but also the Δ changes sign, and the overall result is invariant.

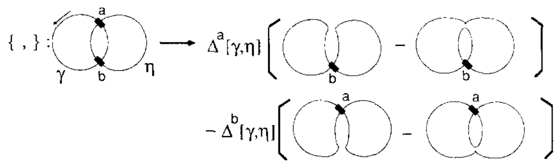

Fig. 5. Poisson brackets of two T^1's.

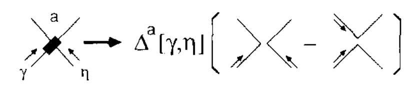

Fig. 6. The action of the grasp operator.

of the first loop (if it has a hand) minus its action over the hand of the second one (if it has a hand).

We use for compactness the following terminology. We say that a hand on a loop γ "sees" a loop η if γ and η intersect at the location of the hand. We say that the hand "grasps" the loop η to indicate the operation described in fig. 6, and we call the "grasp" of the hand over the loop η the result of the operation (the grasp is zero unless the hand sees η). Then we can re-express our result about the Poisson algebra in the following terms.

The Poisson brackets of two handed loops α and β is given by the grasp of the hand of α over β, minus the grasp of the hand of β over α.

Note that intersections which are not at the locations of the dots are not relevant.

At this point, the pictorial meaning of the notation $><$, $\overset{\vee}{\wedge}$ should be clear. These symbols refer exactly to the pictorical description, as in fig. 6, of the rearrangement of the legs at the intersection.

The last topic concerning the small $\mathcal{T}$ algebra that we need to discuss is the question of whether the T^0 and T^1 are enough to parametrize the gauge constraint surface of the phase space. If they are then, in principle, a complete quantization of the theory can be based on them. This issue is presently open; the problem is the following.

There exists a subspace of the phase space of complexified general relativity on which the mapping $(\tilde{\sigma}^a(x), A_a(x)) \mapsto (T^0, T^1)$ is degenerate: if $A_a(x) = 0$, then $T^0 = 2$ and $T^1 = 0$ whatever $\tilde{\sigma}^a(x)$ is. The intersection of this subspace with the constraint surface consists of the initial data for the self-dual sector of the theory. However, the intersection of this self-dual sector with the real slice of the phase space – that corresponding to initial data whose evolution is real – is exactly one point, which is the initial data for flat space-time.

Thus, there are two questions. First, is flat space-time the only point at which a coordinatization of the constraint surface by the functions T^0 and T^1 becomes degenerate? Second, even if this is the case, off the constraint surface the $A_a = 0$ subspace contains a nontrivial set of nonphysical initial data, so that the T^0 and T^1 clearly do not coordinatize the whole phase space off the constraint surface. Thus, even if the coordinatization they give of the constraint surface is nondegenerate, it is not clear that they contain a sufficient number of observables to construct the Dirac quantization of the theory*.

Even if the T^0 and T^1 do parametrize the constraint surface, the parametrization that they give will not be very convenient for the expression of some important observables such as the three-metric and the hamiltonian constraint. Thus, whatever the answer to the question of the completeness of T^0 and T^1 is, we will find it useful to have a larger algebra of observables.

2.3. THE GENERAL T OBSERVABLES AND THE FULL $\mathscr{T}$ ALGEBRA

In this subsection we introduce a more general set of loop observables, obtained by inserting more than one $\tilde{\sigma}^a(x)$ along the holonomy of a loop. We define T^n as a function of a loop and of n points on the loop fixed by loop parameter values $s_1, \dots, s_n$, which satisfy $0 < s_1 < \cdots < s_n \leqslant 2\pi$. We define

$$T^{a_1 \dots a_n}_{\text{ordered}}[\gamma](s_1 \dots s_n) \equiv \operatorname{Tr}\big[\tilde{\sigma}^{a_1}(\gamma(s_1)) U_\gamma(s_1, s_2) \tilde{\sigma}^{a_2}(\gamma(s_2)) \dots$$
$$\tilde{\sigma}^{a_n}(\gamma(s_n)) U_\gamma(s_n, s_1)\big]. \qquad (31)$$

For later convenience we adopt the convention that the order in which the s_i (and related a_i) are written as arguments of a T^n is irrelevant. If $P = (i, j, \dots, p, q)$ is a permutation of the first n natural numbers we define T^n (without the subscript *ordered*) to be

$$T^{a_1 \cdots a_n}[\gamma](s_1 \dots s_n) = \sum_P \theta(s_q - s_p) \dots \theta(s_j - s_i) T^{a_i \dots a_p}_{\text{ordered}}[\gamma](s_i \dots s_p), \qquad (32)$$

where $\theta(t) = 1$ for $t > 0$ and zero otherwise; T^n may be graphically represented as an oriented curve with n hands.

The main property of the T^n's is that they form a closed graded Poisson algebra, which has the structure

$$\{T^n, T^m\} \sim T^{n+m-1}. \qquad (33)$$

* In 2 + 1 gravity [37] one can demonstrate that the mapping $(\tilde{\sigma}^a(x), A_a(x)) \mapsto (T^0, T^1)$ is in fact degenerate *only* on the $A = 0$ subspace.

More precisely

$$
\begin{aligned}
&\left\{T^{a_1\dots a_n}[\gamma](s_1\dots s_n),\quad T^{b_1\dots b_m}[\eta](t_1\dots t_m)\right\}\\
&\quad= -i\sum_{k=1}^{n}\sum_{\diamondsuit}(-1)^{|\diamondsuit|}\Delta^{a_k}[\gamma,\eta](s_k)\\
&\qquad\times T^{a_1\dots \not{a}_k\dots a_n, b_1\dots b_n}\left[\left(\gamma\#_{s_k}\eta\right)^{\diamondsuit}\right]\left(u'(s_1)\dots \not{u}'(s_k)\dots u'(s_n), u(t_1)\dots u(t_m)\right)\\
&\qquad+i\sum_{k=1}^{m}\sum_{\diamondsuit}(-1)^{\diamondsuit}\Delta^{b_k}[\eta,\gamma](t_k)\\
&\qquad\times T^{b_1\dots \not{b}_k\dots b_m, a_1\dots a_n}\left[\left(\eta\#_{t_k}\gamma\right)^{\diamondsuit}\right]\left(u'(t_1)\dots \not{u}'(t_k)\dots u(t_m), u(s_1)\dots u(s_n)\right).
\end{aligned}
\tag{34}
$$

The slash over a term means that that term is not present: $(a_1\dots \not{a}_k\dots a_m) = (a_1\dots a_{k-1}a_{k+1}\dots a_n)$. Each of the two terms contains a sum over the hands of one of the loops, and, for every hand, the sum of the two ways to rearrange the legs. This result is obtained after some algebra, or much more easily by the graphical calculus described in the previous subsection. The full Poisson algebra can then be expressed in the following form.

The Poisson brackets of two handed loops α and β are given by the sum of all the grasps of the hands of α, over β, minus the sum of all the grasps of the hands of β over α.

For instance the Poisson brackets of a T^2 and a T^3 that intersect at two points, such that there is a hand of the T^2 at each intersection, are given in fig. 7.

We have shown that the T observables form a graded Poisson algebra. As we already said we call this algebra $\mathscr{T}$. This algebra is entirely expressible in term of breaking and rejoining of handed loops and, remarkably, the full symplectic structure of the phase space is coded in this algebra (more precisely, in a suitable closure of the algebra).

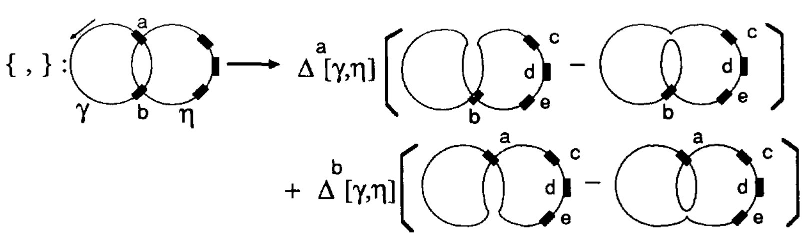

Fig. 7. Poisson brackets of a T^2 with a T^3 which intersect at the hands of the T^2.

2.4. SMEARING

In this subsection we show how the singularities present in the formal expression for the $\mathscr{T}$ algebra are avoided by smearing the T observables.

In any field theory the Poisson algebra of local field observables involve singular expressions. These are handled by interpreting the fields as three-dimensional distributions, to be defined by smearing them in three dimensions with smooth test functions. Now, the $T^a[\gamma](s)$ observable depends on the space point $\gamma(s)$, and we do find distributional expressions in its Poisson brackets. If we want to eliminate these singularities it is not enough to average in the loop parameter, as the singular expressions contain three-dimensional delta functions.

We will thus introduce a smearing or regularization procedure in which the T^n observables are averaged over both the loop parameter and the position of the loop in space. This will be sufficient to eliminate all singularities in the Poisson brackets of the $\mathscr{T}$ algebra.

To define the smeared observables we consider, instead of loops, two parameter congruences of loops $\gamma_\sigma(s)$, $\sigma = (\sigma_1, \sigma_2)$. We then define the smeared T^1 observables to be

$$T^1[\gamma_\sigma](f) = \int \mathrm{d}^2\sigma\, \mathrm{d}s f_a(\gamma_\sigma(s)) T^a[\gamma_\sigma](s), \tag{35}$$

where f is a one-form on Σ.

The Poisson brackets of the smeared T^1 with a T^0 are given by

$$\{T^1[\gamma_\sigma](f), T[\eta]\} = -i\int_\eta f\left(T[\gamma_\sigma \# \eta] - T[\gamma_\sigma \# \eta^{-1}]\right), \tag{36}$$

where the σ in the r.h.s. is a function of the integration variable. Note that the singular coefficient Δ has been replaced by the line integral of the one-form f along the grasped loop η.

For $n > 1$ we introduce an independent three-dimensional smearing for every hand. That is we consider $2n$-parameter families of loops $\gamma_{\sigma_1 \ldots \sigma_n}$ and define

$$T^n[\gamma_{\sigma_1 \ldots \sigma_n}](f) = \int \mathrm{d}^2\sigma_1 \ldots \mathrm{d}^2\sigma_n \int \mathrm{d}s_1 \ldots \mathrm{d}s_n$$
$$\times f^{(1)}_{a_1}\left(\gamma_{\sigma_1 \ldots \sigma_n}(s_1)\right) \ldots f^{(n)}_{a_n}\left(\gamma_{\sigma_1 \ldots \sigma_n}(s_n)\right) T^{a_1 \ldots a_n}[\gamma_{\sigma_1 \ldots \sigma_n}](s_1 \ldots s_n). \tag{37}$$

It is straightforward to verify that all the singularities of the $\mathscr{T}$ algebra are eliminated in this way.

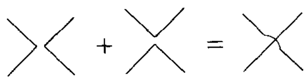

Fig. 8. $\delta_A{}^B\delta_C{}^D + \epsilon_{AC}\epsilon^{DB} = \delta_A{}^D\delta_C{}^B$.

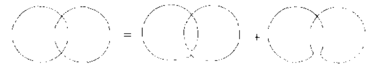

Fig. 9. The fundamental two-spinor identity as a condition on the T^0.

We may assume that the dependence of the loop on each pair of parameters σ_i is nondegenerate only in an interval around one of the hands (that is on the support of the corresponding f_a) and that these intervals do not overlap. In this way we keep the smearing of each hand separate from the others, and the regularized loop appears as one which is fattened in a neighborhood of each hand. Actually, in the classical theory it is not necessary to smear each hand independently. However, this will be essential in the quantum theory. We will thus describe the smearing procedure in detail only after the introduction of the quantum loop operators.

2.5. OTHER PROPERTIES OF THE T OBSERVABLES

The SL(2, $\mathbb{C}$) algebra in which the Ashtekar connection lives is characterized by the following algebraic identity which plays an important role in the loop formalism

$$\delta_A^B\,\delta_C^D + \epsilon_{AC}\epsilon^{DB} = \delta_A^D\,\delta_C^B\,. \tag{38}$$

By inserting this identity between the four legs of the loops arriving at one intersection we obtain the identity expressed in fig. 8. For instance, by using this identity we have the relation

$$T[\alpha\#\beta] + T[\alpha\#\beta^{-1}] = T[\alpha]T[\beta], \tag{39a}$$

which is graphically expressed in fig. 9. Therefore we have another interpretation of the two symbols $><$ and $\begin{smallmatrix}\vee\\ \wedge\end{smallmatrix}$. They refer respectively to $\delta_A^B\,\delta_C^D$ and $\epsilon_{AC}\epsilon^{BD}$. Their difference $\delta_A^D\,\delta_C^B$ represents the crossing of the two loops and may be represented with the symbol $\times$. In this spirit we will also use $(\alpha\#\beta)^{\times} = \alpha \cup \beta$.

Note that in the notation of fig. 8 the identity (38) is expressed very similar to the way in which the same identity is expressed in Penrose's spin network formalism [21].

The $\mathscr{T}$ observables satisfy also a second identity having to do with retracing over portions of a loop. We denote by $\rho \circ \sigma \circ \tau \ldots$ a loop obtained by joining the ends of the segments $\rho, \sigma, \tau \ldots$.

Let α be a loop that starts and ends in the point P, and ρ a segment with an end on P. We call the loop $\alpha \circ \rho \circ \rho^{-1}$ a *loop with a tail* (fig. 10a) and we have

$$T[\alpha \circ \rho \circ \rho^{-1}] = T[\alpha]. \tag{39b}$$

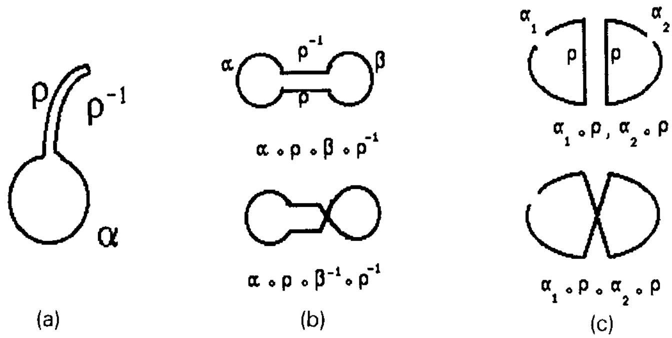

Fig. 10. (a) Loop with tail; (b), (c) eyeglass loops.

The identities (39a) and (39b) allow a complete characterization of T^0. In fact it is possible to show [22] that, given any function on the space of unparametrized loops which satisfy these identities, there is an SL(2, $\mathbb{C}$) connection such that the loop function is the holonomy of that connection. These two identities imply the two following relations between the T^0's. Let α and β be any two loops and let ρ be a segment that joins a point of α with a point of β. The first relation is

$$T[\alpha]T[\beta] = T\left[\alpha \circ \rho \circ \beta \circ \rho^{-1}\right] + T\left[\alpha \circ \rho \circ \beta^{-1} \circ \rho^{-1}\right]. \tag{40a}$$

We call the loops of the form $\alpha \circ \rho \circ \beta \circ \rho^{-1}$ *eyeglass loops* (see fig. 10b). The second relation is the following. Let α be any loop and P and Q two distinct points on α. Let α_1 and α_2 be the two segments of α separated by P and Q, and let ρ be another segment joining P and Q. Then

$$T[\alpha] = T[\alpha_1 \circ \rho]T\left[\alpha_2 \circ \rho^{-1}\right] - T\left[\alpha_1 \circ \rho \circ \alpha_2^{-1} \circ \rho\right]. \tag{40b}$$

We denote also the loops of the form $\alpha_1 \circ \rho \circ \alpha_2^{-1} \circ \rho$ and the set of the two loops $\alpha_1 \circ \rho, \alpha_2 \circ \rho^{-1}$, as eyeglass loops (see fig. 10c). Any eyeglass loop determines in a unique way a loop (or a set of two loops), obtained by cutting away the segments ρ and ρ^{-1}. We will say that the eyeglass loop is *related* to the loop (or the set of two loops) determined in this way. For instance $\alpha \circ \rho \circ \beta \circ \rho^{-1}$ is related to the double loop α, β, and $\alpha_1 \circ \rho \circ \alpha_2^{-1} \circ \rho$ is related to the loop α.

The two relations (40a, b) are important for the following reason. Since the value of T^0 on any loop with a tail is determined by the value on the corresponding loop without tail, we can eliminate the loops with a tail from the formalism and restrict our loop space to the space of the loops without tail, namely to the loops that are not of the form $\alpha \circ \rho \circ \rho^{-1}$. Then, we may also forget the retracing identity eq. (39b); however, because of the interplay between this identity and the identity (39a), there are some remaining relations between the T^0's on loops with no tail. It is possible to show that all these relations are given by eqs. (40a, b).

The full set of the T coordinatizes the gauge invariant phase space. Most of the interesting local observables of general relativity are at the border of the T set; that is they can be obtained as the limit of certain linear combinations of T's, as the loop shrinks down to a point. For instance the three-metric is related to the T^2 observables in the following way. Consider a loop $\gamma(x)$ which starts and ends at the point x and the sequence $\gamma^\delta(x)$ obtained by shrinking down $\gamma(x)$ to the point x: $\gamma^\delta(x) = \delta\gamma(x) + (1-\delta)x$, $\delta \in (0,1)$. Then we have that

$$\tilde{\tilde{q}}^{ab}(x) = -\tfrac{1}{2} \lim_{\delta \to 0} T^{ab}\big[\gamma^\delta(x)\big](s_1, s_2). \tag{41}$$

In the limit $\delta \to 0$ the U's in this expression go to the identity, so that only the $\tilde{\sigma}^a(x)$'s survive in the trace.

The hamiltonian constraint can be defined in terms of T^2's in a similar way. To define it we consider a fixed coordinate system in the neighborhood of a point x. Given these coordinates let us define a coordinate circle of radius δ beginning at x which lies in the a–b coordinate plane. We call this $\gamma^\delta_{ab}(x)$. Then we have

$$U_{\gamma^\delta_{ab}(x)}(s) = \mathbb{1} + \delta^2 F_{ab}(x) + \mathrm{o}(\delta^2). \tag{42}$$

In order to get the hamiltonian constraint we have to start from a combination of T^2's such that in the limit the first term that contains the identity cancels. If we take

$$\mathscr{C}^\delta(x) = T^{[ab]}\big[\gamma^\delta_{ab}(x)\big](\delta^2, 2\pi), \tag{43}$$

it follows that

$$\mathscr{C}(x) = \lim_{\delta \to 0} (1/\delta^2)\mathscr{C}^\delta(x). \tag{44}$$

In fact (we use γ for $\gamma^\delta_{ab}(x)$)

$$\mathscr{C}^\delta(x) = \mathrm{Tr}\big[\tilde{\sigma}^{[a}(\gamma(\delta^2))U_\gamma(\delta^2, 2\pi)\tilde{\sigma}^{b]}(\gamma(2\pi))U_\gamma(2\pi, \delta^2)\big]. \tag{45}$$

We can neglect the difference between $\tilde{\sigma}(x) = \tilde{\sigma}(\gamma(2\pi))$ and $\tilde{\sigma}(\gamma(\delta^2))$, which is of order δ^3, and for the same reason we can substitute $U_\gamma(2\pi, \delta^2)$ with the identity and $U_\gamma(\delta^2, 2\pi)$ with $U_\gamma(0)$. So we have

$$\mathscr{C}^\delta(x) = \mathrm{Tr}\big[\tilde{\sigma}^{[b}(x)\tilde{\sigma}^{a]}(x)U_{\gamma^\delta_{ab}(x)}(0)\big] + \mathrm{o}(\delta^2). \tag{46}$$

Now we can expand U as in eq. (42). The leading term is zero because of the antisymmetrization, thus

$$\mathscr{C}^\delta(x) = \delta^2\,\mathrm{Tr}\big[\tilde{\sigma}^b(x)\tilde{\sigma}^a(x)F_{ab}(x)\big] + \mathrm{o}(\delta^2), \tag{47}$$

from which eq. (44) follows.

We will not be directly interested in expressing the other constraints in terms of the T variables. Rather we will be interested in the transformation properties of the T's under the transformation generated by these constraints in the phase space. To give these transformation properties is equivalent to giving the constraints.

Under SU(2) gauge transformations the T observables are invariant. Under a diffeomorphism ϕ they transform as scalars as far as their dependence on the loop γ is concerned and as vector densities of weight 1 at the points $\gamma(s_i)$ and related index a_i. That is, in the new coordinate system $x'^a = \phi^a(x)$ they are given by

$$T'^{a_1 \dots a_n}[\gamma](s_1 \dots s_n) = J^{-1}(\gamma(s_1)) \dots J^{-1}(\gamma(s_n)) \times \frac{\partial \phi^{a_1}(\gamma(s_1))}{\partial x^{b_1}} \dots \frac{\partial \phi^{a_n}(\gamma(s_n))}{\partial x^{b_n}} T^{b_1 \dots b_n}[\phi^{-1} \cdot \gamma](s_1 \dots s_n), \tag{48}$$

where $(\phi \cdot \gamma)^a(s) = \phi^a(\gamma(s))$ and $J(x)$ is the jacobian of the coordinate transformation ϕ.

Finally we need to introduce a generalization of the definition of the T loop observables which will be needed in what follows.

First of all we define a *multiple loop* to be a set containing a finite number of (continuous, piecewise smooth) loops. We use the notation $\{\eta\}$ to indicate a multiple loop formed by single loops $\eta_1 \dots \eta_n$ and we denote the space of multiple loops by $\mathcal{M}$. $\mathcal{M}$ is a graded space which is the direct sum of the space of the single loops, the subspace of the multiple loops consisting of two loops, the subspace of the triple loops, and so on. For later convenience we add to $\mathcal{M}$ a point representing the multiple loop formed by zero loops.

In what follows we will often need to use multiple loops instead of loops. A multiple loop can be seen as a direct generalization of a single loop, obtained by relaxing the continuity condition to the condition of the absence of end points. In particular we assume that the multiple loop has a unique parameter $s \in (0, 2\pi n)$ which "jumps" from one point to the other. In this spirit we will use the notation

$$\int_{\{\eta\}} = \int_{\eta_1} + \cdots + \int_{\eta_n}. \tag{49}$$

Note that this notation allows some, but not all, of the expressions of this section to be immediately generalized to multiple loops. An example of this is

$$\Delta^a[\gamma, \{\eta\}](s) = \sum_i \Delta^a[\gamma, \eta_i](s), \tag{50}$$

where the only nonvanishing term of the sum is the one corresponding to the loop in which s lies.

In conclusion we have shown that general relativity may be formulated in terms of the T variables, defined in eqs. (19) and (31). The symplectic structure is given by the graded algebraic structure $\mathscr{T}$ given in eq. (34), the hamiltonian constraint is given in eq. (44), the transformation properties under transformation generated by the other constraints are well defined. The relation with the metric formalism is expressed in eq. (41).

3. Definition of the loop representation

We now are ready to come to the main point of this paper, which is the construction of a representation for canonical quantum general relativity as a representation of the $\mathscr{T}$ algebra described in the last section. We will start by constructing the linear space of functionals on which the representation is defined. This will be a space of functionals on the space of multiple loops. An element of this space, which represents an unconstrained quantum state of the theory, may then be thought of as an assignment of amplitudes for each collection of loops on the three-manifold Σ.

It is important to note that the loop representation is not a space of functionals over a classical configuration space. This is related to the fact that it is associated with the quantization of a noncanonical observable algebra.

Given the definition of the representation space we will go on to construct the algebra of linear operators on this space. These will form a linear representation of a deformation of the Poisson algebra $\mathscr{T}$. The operators are denoted $\tilde{T}$ and are interpreted as the quantum operators corresponding to the classical observables T. The correspondence between the classical and the quantum theory is thus established by means of this correspondence between these algebras.

After introducing the operator algebra we come back to the definition of the state space and specify more precisely the state space by imposing some restrictions on the space of the loop functionals. These conditions are necessary if the quantum algebra is to be an irreducible representation of a deformation of the classical algebra.

Finally we complete the definition of the theory by defining the quantum operators that express the constraints.

We construct the quantum observable algebra in two steps, as we did for the classical Poisson algebra. We begin with the observables in the small $\mathscr{T}$ algebra. The quantum commutator algebra of these observables is isomorphic to the classical Poisson algebra. After this we define the general $\tilde{T}$ observables. In the general case we will find that the commutator algebra of the $\tilde{T}$ defines a linear representation of a deformation of the classical $\mathscr{T}$ algebra.

3.1. THE STATE SPACE AND THE QUANTIZATION OF THE SMALL $\mathcal{T}$ ALGEBRA

Let us consider a linear space of complex functionals on $\mathcal{M}$, the space of sets of piecewise smooth continuous loops. We shall denote this space by $\mathcal{S}$; its elements will be denoted by $\mathcal{A}[\{\eta\}]$. Any $\mathcal{A}[\{\eta\}]$ may be expressed in terms of its "components" $\mathcal{A}_n[\eta_1, \ldots, \eta_n]$ on the subspaces of $\mathcal{M}$ of the sets composed of n loops. Thus $\mathcal{A}[\{\eta\}]$ has the form

$$\mathcal{A} = \{\mathcal{A}_0, \mathcal{A}_1(\alpha), \mathcal{A}_2(\beta_1, \beta_2), \mathcal{A}_3(\gamma_1, \gamma_2, \gamma_3), \ldots\}. \tag{51}$$

Note that these amplitudes depend on *unordered* sets of loops. Thus, in eq. (51) the $\mathcal{A}_n$'s are symmetric in their entries.

We now introduce an algebra of operators on $\mathcal{S}$, which provides a faithful representation of the $\mathcal{T}$ algebra. We begin with the observable T^0. The quantum operator $\tilde{T}[\gamma]$ corresponding to the observable $T[\gamma]$ is defined by

$$\tilde{T}[\gamma]\mathcal{A}[\{\eta\}] \equiv \mathcal{A}[\gamma \cup \{\eta\}], \tag{52}$$

where $\gamma \cup \{\eta\}$ is the collection of loops which is the union of γ and $\{\eta\}$. Thus, $\tilde{T}[\gamma]$ is a kind of lowering* operator.

Next we consider the T^1 observable. We define

$$\tilde{T}^a[\gamma](s)\mathcal{A}[\{\eta\}] = \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\gamma, \{\eta\}](s) \mathcal{A}\left[(\gamma \# \{\eta\})^{\diamond}\right]. \tag{53}$$

Note that this action is defined in terms of the elementary grasp operations defined in fig. 5 of the previous section. Here, however, the grasp is on the argument of the loop functional, and the coefficient $(-1)^{|\diamond|}\Delta^a[\gamma, \{\eta\}](s)$ multiplies the value of the loop functional. In the quantum context we will use the expression "grasp" in this way. That is we express the content of eq. (53) by saying that the action of the quantum observables corresponding to the handed loop γ is given by the grasp of the hand on the loop functional.

Let us compute the commutator algebra of these operators. First of all it is straightforward to show that the $\tilde{T}^0$ operator commutes with itself. We have

$$\tilde{T}[\alpha]\tilde{T}[\beta]\mathcal{A}[\{\eta\}] = \mathcal{A}[\alpha \cup \beta \cup \{\eta\}] = \tilde{T}[\beta]\tilde{T}[\alpha]\mathcal{A}[\{\eta\}], \tag{54}$$

* If $\mathcal{A}$ has support only on sets of n loops then $\tilde{T}[\gamma]\mathcal{A}$ defined by eq. (52) has support only on sets of $n-1$ loops.

because the action of $\cup$ is commutative. The commutator of $\tilde{T}^0$ and $\tilde{T}^1$ is given by

$$\begin{aligned}
&\left[\tilde{T}[\alpha], \tilde{T}^a[\beta](s)\right] \mathscr{A}[\{\eta\}] \\
&\quad = \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\beta, \{\eta\}](s) \mathscr{A}\left[\alpha \cup (\beta \# \{\eta\})^{\diamond}\right] \\
&\qquad - \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\beta, \alpha \cup \{\eta\}](s) \mathscr{A}\left[(\beta \# (\alpha \cup \{\eta\}))^{\diamond}\right].
\end{aligned} \tag{55}$$

Note that

$$\Delta^a[\beta, \alpha \cup \eta](s) = \Delta^a[\beta, \alpha](s) + \Delta^a[\beta, \eta](s), \tag{56}$$

where the first or the second term are different from zero only if the hand of β sees α or η respectively. Thus the second term in the commutator may be decomposed as

$$\begin{aligned}
&\hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\beta, \alpha \cup \{\eta\}](s) \mathscr{A}\left[(\beta \# (\alpha \cup \{\eta\}))^{\diamond}\right] \\
&\quad = \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\beta, \alpha](s) \mathscr{A}\left[(\beta \# \alpha)^{\diamond} \cup \{\eta\}\right] \\
&\qquad + \Delta^a[\beta, \{\eta\}](s) \mathscr{A}\left[\alpha \cup (\beta \# \{\eta\})^{\diamond}\right].
\end{aligned} \tag{57}$$

The last term cancels with the first term of the commutator and we obtain the final result

$$\left[\tilde{T}[\alpha], \tilde{T}^a[\beta](s)\right] = \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\beta, \alpha](s) \tilde{T}\left[(\beta \# \alpha)^{\diamond}\right]. \tag{58}$$

In words, the difference between first adding a loop α to the argument of the functional and then grasping it by an operator, and first grasping it and then adding the loop α, is given by the possibility that the grasped loop is α itself. Therefore, the difference of the two operations is the same thing as adding the grasp of α.

A similar calculation shows that

$$\begin{aligned}
\left[\tilde{T}^a[\gamma](s), \tilde{T}^b[\eta](t)\right] &= \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^a[\gamma, \eta](s) \tilde{T}^b\left[(\gamma \#_s \eta)^{\diamond}\right](u(t)) \\
&\quad - \hbar \sum_{\diamond} (-1)^{|\diamond|} \Delta^b[\eta, \gamma](t) \tilde{T}^a\left[(\eta \#_t \gamma)^{\diamond}\right](u(s)).
\end{aligned} \tag{59}$$

All these commutators reproduce $i\hbar$ times the corresponding classical Poisson

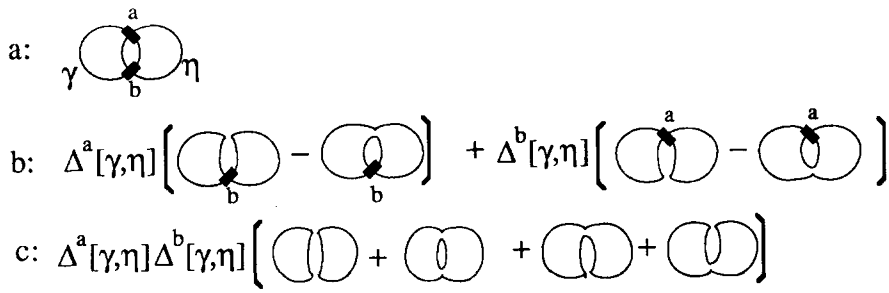

Fig. 11. The results of two hands grasping (b) one at a time, and (c) simultaneously.

brackets given in eqs. (28) and (30) of the previous section. Thus the set of the operators $\tilde{T}^0$ and $\tilde{T}^1$ form a linear representation of the small $\mathscr{T}$ algebra.

3.2. THE QUANTIZATION OF THE FULL $\mathscr{T}$ ALGEBRA

In order to represent the full $\mathscr{T}$ algebra by linear operators on $\mathscr{S}$ we need to generalize our notation on the breaking and rejoining of loops at intersections to the case in which the breaking and rejoining happens *simultaneously* at two or more intersections. This is illustrated by fig. 11. In fig. 11a we have two loops, one with two hands and the other without hands. The loops intersect twice, and there is one hand at each intersection. In fig. 11b we draw the result of grasping with the hands *one at a time* (such as in the case of classical Poisson brackets), while fig. 11c shows the result of grasping with the two hands *simultaneously*. In general we call grasp of n hands one at the time the sum of the n grasps of the n hands, this is the sum of the $2n$ terms obtained by breaking and rejoining at one intersection at a time. The simultaneous grasp of n hands is the sum of 2^n terms obtained by breaking and rejoining *at all the hands* at the same time.

Given a loop α with n hands at the points $s_1 \ldots s_n$ and a loop β which intersects α at each of these points we denote by

$$\left(\alpha\#_{s_1 \ldots s_n}\beta\right)^{(\Diamond_1 \cdots \Diamond_n)} \tag{60}$$

the loop obtained by substituting the n intersections with the n alternatives $(\Diamond_1 \ldots \Diamond_n)$. We do not need to specify how the resulting loop is parametrized, since the parametrization will not play any role in the future. We do maintain the notation, introduced in the previous section, that $u'(s)$ and $u(t)$ indicate the parameters of the composed loop that correspond to the points $\alpha(s)$ and $\beta(t)$.

Our notation for loops generalizes in a straightforward way to multiple loops. The reader may note that the result of a simultaneous grasp will often be a multiple loop even when the two original loops are single loops.

We may now introduce the quantum loop operators with more than one hand. We define

$$\tilde{T}^{a_1 \dots a_n}[\gamma](s_1 \dots s_n)\mathscr{A}[\{\eta\}]$$
$$\equiv \hbar^n \sum_{\diamond_1} \dots \sum_{\diamond_n} (-1)^r \Delta^{a_1}[\gamma,\{\eta\}](s_1) \dots \Delta^{a_n}[\gamma,\{\eta\}](s_n) \mathscr{A}\left[(\gamma \#_{s_1 \dots s_n}\{\eta\})^{(\diamond_1 \dots \diamond_n)}\right], \tag{61}$$

where r will be defined in a moment. In words, the action of a many handed loop quantum operator is proportional to the simultaneous grasp of all its hands. That is, it is given by the sum of all the 2^n possible ways in which one can simultaneously substitute all of the intersections at which there is a hand with one of the two ways to join the legs of the two intersecting loops to each other. Note that the result of the grasp is zero unless *all* the hands of γ see $\{\eta\}$.

The sign of each term in eq. (61) is determined by r, which has to do with the orientations of the two loops. It is defined as follows. Each of the two original loops has (or can be given) an orientation. Each term of the grasp is a (multiple) loop made out of segments of the original loops, thus each of these segments has an orientation. In order to assign an orientation consistently to each of the loops some of the orientations of the segments may have to be reversed. r (or more properly $r((\gamma \#_{s_1 \dots s_n}\{\eta\})^{(\diamond_1 \dots \diamond_n)})$) is defined as the number of these segments whose orientations must be reversed*.

The computation of the commutator algebra of the $\tilde{T}$ operators is a tedious but straightforward exercise in combinatorics. We give here the main idea of the calculation.

We must compute the difference between acting on a loop functional first with a handed loop α, with n_α hands and then with a handed loop β with n_β hands, and acting in the opposite order. In the first case the hands of β may grasp both the loop α or the loop functional. In the second case they are the hands of β which may grasp both α or the loop functionals. All the terms in which both loops grasp only the loop functional cancel among themselves.

Let us evaluate the commutator at a loop η. Consider the first half of the commutator, which is given by first grasping η with the hands of α and then by grasping the resulting functional with the hands of β. In the second step we must evaluate the action of the hands of β on terms in which the loop functional is evaluated at $\alpha \# \eta$. For each hand there are two possibilities: it may grasp either a segment of $\alpha \# \eta$ which comes from α or one which comes from η. Thus we can express the result of the grasp as the sum of terms which is organized according to

* In the case of grasps of a single hand we have $r((\gamma \# \{\eta\})^{\diamond}) = |\diamond|$; therefore the previous way to keep track of the sign is a particular case of the present one.

how many of the hands of β grasp α and how many grasp η. As we already said the term in which *all* the hands of β grasp η cancels with analogous term of the second half of the commutator. Let us then consider the case in which *all but one* of the hands of β grasp η and one hand grasps α. The corresponding term is a term in which η is grasped by all the hands of α and β, save one of β. Thus it is equivalent to the action, on η, of the quantum loop operator with $n_\alpha + n_\beta - 1$ hands obtained by grasping α with one of the hands of β. Thus we have found a first term in the commutator: the commutator of a $\tilde{T}^n$ and a $\tilde{T}^m$ contains a term of the kind $\tilde{T}^{n+m-1}$ obtained by grasping one of the loops with one of the hands of the other one. Note that this is exactly the same term which is on the r.h.s. of the Poisson brackets of the corresponding classical observables, but multiplied by $i\hbar$, since any grasp in the quantum context contains also a multiplication by $\hbar$. A calculation shows also that the $(-1)^r$ rule reproduces the signs which result from the classical calculation.

However, the term proportional to $\hbar$ is only the first term in the quantum commutator. There are also terms of higher order in $\hbar$. For example, consider the case in which *two* hands of β grasp α and the others grasp η. This case produces the term of the kind $\tilde{T}^{n_\alpha + n_\beta - 2}$ obtained by the *simultaneous* grasp of two hands of β over α. As two hands have acted to yield the operator this term is multiplied by $\hbar^2$. This continues up to the term of the kind $\hbar^{n_\beta} \tilde{T}^{n_\alpha + n_\beta - n_\beta}$ in which all the hands of β grasp α.

The total commutator is given by the formula

$$
\begin{aligned}
&\left[\tilde{T}^{a_1 \dots a_n}[\alpha](s_1 \dots s_n), \quad \tilde{T}^{b_1 \dots b_m}[\beta](t_1 \dots t_m)\right] \\
&\quad = \sum_{p=1}^{n} \hbar^p \left\{ \sum_{k_1=1}^{n} \dots \sum_{k_p = k_{p-1}+1}^{n} \sum_{\diamondsuit_{k_1}} \dots \sum_{\diamondsuit_{k_p}} (-1)^r \Delta^{a_{k_1}}[\beta, \alpha](t_{k_1}) \dots \Delta^{a_{k_p}}[\beta, \alpha](t_{k_p}) \right. \\
&\qquad\qquad \times \tilde{T}^{a_1 \dots \not{a}_{k_1} \dots \not{a}_{k_p} \dots a_n, b_1 \dots b_m}\left[\left(\beta \#_{s_{k_1} \dots s_{k_p}} \alpha\right)^{(\diamondsuit_{k_1} \dots \diamondsuit_{k_p})}\right] \\
&\qquad\qquad \left. \times \left(u'(s_1) \dots \not{u}'(s_{k_1}) \dots \not{u}'(s_{k_p}) \dots u'(s_n), u(t_1) \dots u(t_m)\right) \right\} \\
&\quad - (\alpha \leftrightarrow \beta, a \leftrightarrow b, s \leftrightarrow t, n \leftrightarrow m) .
\end{aligned}
\tag{62}
$$

An example is given in fig. 12. We call this algebra the $\tilde{\mathscr{T}}$ algebra. Note that $\tilde{\mathscr{T}}$ has the structure (assuming $n < m$)

$$
[\tilde{T}^n, \tilde{T}^m] = \hbar \tilde{T}^{n+m-1} + \hbar^2 \tilde{T}^{n+m-2} + \dots + \hbar^m \tilde{T}^n, \tag{63}
$$

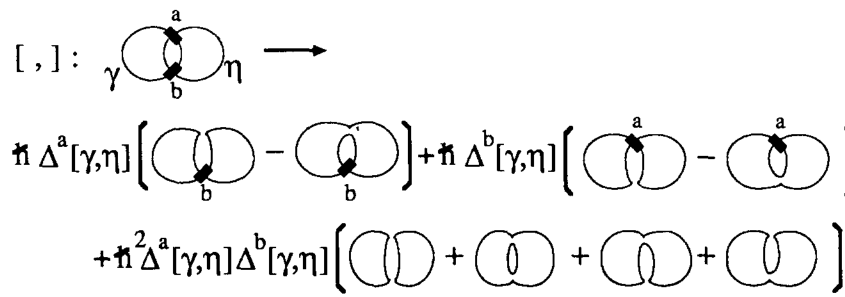

Fig. 12. Commutator of a $\tilde{T}^2$ with a $\tilde{T}^3$. Compare with fig. 5.

and that it is related to the classical Poisson algebra $\mathscr{T}$ by

$$\lim_{\hbar \to 0} (1/i\hbar)[\tilde{T}^n, \tilde{T}^m] = \{\widetilde{T^n, T^m}\}. \tag{64}$$

In other words, $\tilde{\mathscr{T}}$ is a deformation of $\mathscr{T}$, $i\hbar$ being the deformation parameter. Thus, the linear operators $\tilde{T}$ form a faithful representation of a deformation of the Poisson algebra of the classical T observables. They therefore provide a quantization of the phase space of general relativity in which they may be considered to be the quantum operators corresponding to the classical T observables [7].

3.3. RESTRICTIONS IMPOSED ON THE SPACE OF THE STATES

The classical loop observables are not independent. There are relations among them which follow from their definition in terms of parallel transport. In the quantum theory these relations may be realized by imposing suitable restrictions on the space of states.

A first set of these relations follows from the invariance of the holonomy under reparametrizations. We may implement these relations in the quantum theory simply by demanding that the loop functionals $\mathscr{A}[\{\eta\}]$ are *invariant under reparametrizations*, i.e. they depend only on the unparametrized multiple loops. It is straightforward to verify that all the formulas that we have written are consistent with this restriction.

A second set of relations among the T's are those that are consequences of the two-spinor identity (39a). We can realize these relations in the quantum theory by imposing directly a corresponding condition on the states. To do this it is helpful to realize that although the identity (39a) is a *nonlinear* condition on the basic loop observables it can be realized as *linear* restriction on states which are functionals of *multiple* loops. We require that, for any pair of loops γ and η and any intersection point involving them,

$$\mathscr{A}[\gamma\#\eta,\dots] + \mathscr{A}[\gamma^{-1}\#\eta,\dots] - \mathscr{A}[\gamma \cup \eta,\dots] = 0, \tag{65}$$

where ... refers to any other loops that may be present in the argument of the loop functional. In general, using the $\times$ symbol defined in sect. 2,

$$\mathscr{A}\left[(\gamma\#\eta)^{\times}\right] - \mathscr{A}\left[(\gamma\#\eta)^{><}\right] + \mathscr{A}\left[(\gamma\#\eta)^{\substack{\vee \\ \wedge}}\right] = 0. \tag{66a}$$

It is straightforward to demonstrate that on the space of the states that satisfy this restriction the loop operators satisfy the spinor identity (39a).

A final set of restrictions follows from the identities (39) illustrated in fig. 10. Referring again to that figure we impose the conditions,

$$\mathscr{A}\left[\alpha \circ \rho \circ \rho^{-1},\dots\right] = \mathscr{A}[\alpha,\dots]. \tag{66b}$$

By using eqs. (66a, b) we obtain the counterpart of eqs. (40a, b) for the loop states. These are

$$\mathscr{A}[\alpha, \beta] = \mathscr{A}\left[\alpha \circ \rho \circ \beta \circ \rho^{-1}\right] + \mathscr{A}\left[\alpha \circ \rho \circ \beta^{-1} \circ \rho^{-1}\right], \tag{67a}$$

$$\mathscr{A}[\alpha] = \mathscr{A}\left[\alpha_1 \circ \rho, \alpha_2 \circ \rho^{-1}\right] - \mathscr{A}\left[\alpha_1 \circ \rho \circ \alpha_2^{-1} \circ \rho\right]. \tag{67b}$$

We may note that the first of these equations implies that the value of any loop state is determined by its value on the single loops.

3.4. THE DIFFEOMORPHISM CONSTRAINT

There is a natural action of the diffeomorphism group of Σ on the loop space $\mathscr{M}$. For any $\phi \in \mathrm{Diff}(\Sigma)$ it is given by

$$(\phi \cdot \{\gamma\})(s) = \{\phi(\gamma(s))\}. \tag{68}$$

This action induces a natural linear representation, which we call U, of the diffeomorphism group on the space of the loop functionals on $\mathscr{M}$, by

$$U(\phi)\mathscr{A}[\{\eta\}] = \mathscr{A}\left[\phi^{-1} \cdot \{\eta\}\right]. \tag{69}$$

If ϕ_t is a one-parameter group of diffeomorphisms generated by the vector field v on Σ, the generators $D(v)$ of the representation U are given by

$$D(v)\mathscr{A}[\{\eta\}] \equiv \frac{\mathrm{d}}{\mathrm{d}t} U(\phi_t)\mathscr{A}[\{\eta\}]|_{t=0}. \tag{70}$$

The operators $D(v)$ are unbounded and we do not expect them to be defined on the entire space $\mathscr{S}$, but only in some "dense" domain. This domain consists essentially of the loop functionals which are differentiable in the directions tangent to the orbits of the diffeomorphism group in $\mathscr{M}$. (Recall that we did not impose any general differentiability requirement on the loop functionals.)

The $D(v)$'s defined in this way satisfy the algebra of the diffeomorphism group

$$[D(v), D(w)] = D([v, w]). \tag{71}$$

We are interested in the commutation relations of the quantum loop observables $\tilde{T}$ with the $D(v)$ or, equivalently, in the transformation properties of the loop observables under the representation U. We use the notation

$$\phi \cdot \tilde{T}^{a_1 \dots a_n}[\gamma](s_1 \dots s_n) \equiv U(\phi)\tilde{T}^{a_1 \dots a_n}[\gamma](s_1 \dots s_n)U(\phi^{-1}). \tag{72}$$

It is straightforward to show that

$$\phi \cdot \tilde{T}[\gamma] = \tilde{T}[\phi^{-1} \cdot \gamma]. \tag{73}$$

The calculation is

$$\begin{aligned}\phi \cdot \tilde{T}[\gamma]\mathscr{A}[\eta] &\equiv U(\phi)\tilde{T}[\gamma]U(\phi^{-1})\mathscr{A}[\eta] = U(\phi)\mathscr{A}[\gamma \cup (\phi \cdot \eta)] \\ &= \mathscr{A}[(\phi^{-1} \cdot \gamma) \cup \eta] = \tilde{T}[\phi^{-1} \cdot \gamma]\mathscr{A}[\eta].\end{aligned} \tag{74}$$

It is slightly more complicated to show that

$$\begin{aligned}\phi \cdot \tilde{T}^{a_1 \dots a_n}[\gamma](s_1 \dots s_n) = J^{-1}(\gamma(s_1)) \dots J^{-1}(\gamma(s_n)) \frac{\partial \phi^{a_1}(\gamma(s_1))}{\partial x^{b_1}} \dots \\ \frac{\partial \phi^{a_n}(\gamma(s_n))}{\partial x^{b_n}} \tilde{T}^{b_1 \dots b_n}[\phi^{-1} \cdot \gamma](s_1 \dots s_n);\end{aligned} \tag{75}$$

(see eq. (48)). Eqs. (73) and (75) show that the quantum operators $\tilde{T}$ transform under the representation U of the diffeomorphism group exactly as the corresponding classical observables transform under diffeomorphisms. Since the Poisson brackets of the diffeomorphism constraints generate infinitesimal diffeomorphisms on the classical phase space, it follows that the commutator of the $D(v)$ with all the observables of the theory reproduces exactly the Poisson brackets of the diffeomorphism constraints with the corresponding classical observables. Thus we can identify the $D(v)$ operators with the quantum diffeomorphism constraints, or more precisely, with the quantum operator corresponding to the smeared form

$$\int_\Sigma \mathrm{d}^3x\, v^a(x) C_a(x) \tag{76}$$

of the diffeomorphism constraints.

3.5. THE HAMILTONIAN CONSTRAINT

Finally, we come to the definition of the hamiltonian constraint. The hamiltonian constraint is, in Ashtekar's formalism, a quadratic expression of the conjugate momenta, $\tilde{\sigma}^a(x)$. As such, it cannot be expressed directly as an operator in the quantum theory, it must be expressed as the limit of a sequence of regulated operators. Therefore, in order to define the hamiltonian constraint in the quantum theory additional structure must be introduced to allow us to define a suitable regularization procedure. From this point of view the situation is the same as in flat space quantum field theories in which a regularization procedure must be introduced in order to define the hamiltonian. What is different is that we must introduce a suitable regularization in the absence of a background metric, in a way that does not destroy the diffeomorphism invariance of the theory.

As we already discussed in the introduction, the classical loop formulation provides a natural solution to this problem. In fact, the classical hamiltonian constraint is already defined in eqs. (43) and (44), as the limit of the observables $\mathscr{C}^\delta(x)$. In the quantization introduced earlier in this section, each of these objects is represented by a well-defined quantum operator,

$$\tilde{\mathscr{C}}^\delta(x) = \tilde{T}^{[ab]}\left[\gamma^\delta_{ab}(x)\right](\delta^2, 2\pi). \tag{77}$$

Thus the loop representation directly gives us a natural definition of a *regulated* quantum hamiltonian constraint.

According to the discussion contained in the introduction we implement the quantum hamiltonian constraint in the form

$$\lim_{\delta \to 0} \tilde{\mathscr{C}}^\delta \mathscr{A}[\{\eta\}] = 0. \tag{78}$$

In sect. 5 we will study a class of solutions to this equation.

Before accepting this definition of the quantum hamiltonian constraint, we must discuss its compatibility with the diffeomorphism constraints. Since additional structure has been added in order to define the regulated operators $\tilde{\mathscr{C}}^\delta$, what conditions should be satisfied regarding the action of the diffeomorphisms on them?

A non-diffeomorphism invariant structure is necessary to define the regulated constraint operator since it is necessary to replace the local observable $\mathscr{C}(x)$ by an extended object, defined in terms of some extended structures on space. This is unavoidable since regularization necessarily involves some form of point splitting or short distance cutoff. In our construction of the regulated hamiltonian constraint, given by eq. (43), we make use of a specific set of loops, which are defined in terms of a particular coordinate system. This is the way in which our regularization of the hamiltonian constraint breaks the diffeomorphism invariance.

Therefore we must assure ourselves that, nevertheless, the space of solutions is diffeomorphism invariant. This amounts to showing that to impose the constraint in one particular set of coordinates is equivalent to imposing it any other coordinate system. This will be true if the regulated hamiltonian constraint is transformed linearly into itself under the representation U of the diffeomorphisms group.

This can be demonstrated in the following way: Let us fix a coordinate system O and define $\tilde{\mathscr{C}}^{\delta}(x)$ with respect to it. If we call $\phi \cdot \tilde{\mathscr{C}}^{\delta}(x)$ the regulated hamiltonian constraint defined in the coordinate system ϕO, we need to show is that

$$\phi \cdot \tilde{\mathscr{C}}^{\delta}(x) = U(\phi^{-1})\tilde{\mathscr{C}}^{\delta}(x)U(\phi). \tag{79}$$

But this relation follows from all the definitions.

In subsect. 5.3 we shall see that the class of solutions to the regulated hamiltonian constraint that we will find is diffeomorphism invariant: the result of applying a diffeomorphism to any functional $\mathscr{A}$ which satisfies eq. (78) is another functional which satisfies eq. (78).

3.6. SMEARING IN THE QUANTUM THEORY

In subsect. 2.4 we showed that the distributional singularities contained in the Poisson algebra of the loop observables may be eliminated by smearing over a space of test functions. This may also be done for the quantum algebra. The situation is more complicated because the simultaneous grasps produce terms containing products of delta functions. These are prevented from coinciding by the requirement that the loop parameters of the hands must be unequal. To construct a completely nonsingular algebra we must be careful to define suitable spaces of congruences and test functions in such a way that these simultaneous delta function singularities are always completely eliminated by the integrations over the test functions.

By smearing the $\tilde{T}^n$ operators in the same way as the classical observables, the delta distributions in the Δ-functions only appear inside the corresponding integrals; however, an inspection of the integrals shows that if two of the integration regions could overlap some degeneracy may appear such that some of the delta distributions which result are not compensated by a corresponding integration.

In order to avoid this problem we have to assume that the integration regions corresponding to different hands on the same loop do not overlap. This requirement can be satisfied as long as there is an open interval in the loop parameter separating each insertion of a $\tilde{\sigma}^a$ in T^n. We defined the quantum loop algebra so that this would be the case.

Intuitively we want to substitute for a loop with n hands a loop which has "fattenings" in n regions, put each hand in a region, and require that the regions do not overlap. Each fattening is described by a (two-dimensional) congruence of loops and each congruence depends on a different two-dimensional parameter σ_i.

We can obtain this by considering only n-dimensional congruences in which the dependence on the ith parameter σ_i is trivial everywhere except on an interval γ^i of the loop parameter. We require that the intervals do not overlap and that in the interval the loop actually spans a congruence as the parameter varies.

More precisely we consider mappings $\tilde{\gamma}^n$ from $\mathrm{R}^{2n} \times \mathrm{S}^1$ into Σ, which we denote $\gamma_{\sigma_1 \dots \sigma_n}(s)$, with the following characteristics. We call γ^i the interval of S^1 in which the dependence from σ^i is nontrivial. We assume that each γ^i is connected and disconnected from the others γ^j, and that the restriction of $\tilde{\gamma}^n$ to any γ^i is injective.

In terms of these objects it is now easy to define the smeared quantum operators. What we have to do is simply to make sure that any hand $\tilde{\sigma}^a(\gamma(s))$ stays in its interval γ^i. We obtain this by inserting in the definition of the smeared operator an universal smooth function $R^n_\gamma(s_1 \dots s_n)$ which is defined to be zero unless each s_i is in the corresponding γ^i. Finally the definition of the smeared quantum loop observable is

$$\tilde{T}[\tilde{\gamma}_n](f) = \int \mathrm{d}^2\sigma_1 \dots \mathrm{d}^2\sigma_n \int \mathrm{d}s_1 \dots \mathrm{d}s_n \, R^n_{\tilde{\gamma}}(s_1 \dots s_n)$$

$$\times f_{a_1 \dots a_n}\big(\gamma_{\sigma_1 \dots \sigma_n}(s_1) \dots \gamma_{\sigma_1 \dots \sigma_n}(s_n)\big) \tilde{T}^{a_1 \dots a_n}[\gamma_{\sigma_1 \dots \sigma_n}](s_1 \dots s_n). \quad (80)$$

Note that the overlap of two integration regions is still allowed if the two fattenings of the loop belong to the two different arms of a self-intersection of the loop. An inspection of the commutators shows that this case is safe and introduces no additional singularities.

We leave to the reader as an exercise to compute the commutator of two of these operators and verify that the result is given by a finite dimensional integral over smeared operators of the same form.

4. Transformation theory

4.1. RELATION BETWEEN THE LOOP REPRESENTATION AND THE SELF-DUAL REPRESENTATION; THE BRA SPACE

In this section we describe the relationship between the loop and the self-dual representations. We expect such a relationship to exist because they both form a representation of the same linear algebraic structure. We will proceed by looking for a linear mapping $\mathscr{F}$, defined between the two, that preserves this structure. We will find that we can represent this mapping as an integral transform from the state space of the self-dual representation to the state space of the loop representation.

The linear equivalence that we are considering should not be confused with unitary equivalence. In general, we may expect that there exist infinitely many unitary inequivalent representations of any infinite dimensional algebra of observables. Linear mappings can sometimes be defined between certain kinds of extensions of the irreducible representation spaces; such mappings typically send a state defined in one space to an "infinite norm state" of the other space. In field theory the Bogoliubov transformations are examples of such mappings. However, we are not concerned here with problems related to unitary equivalence, since we work with the general philosophy that the physical Hilbert structure will only be imposed on the space of solutions of the constraint equations. Also, we are not concerned in this section with giving an exact or rigorous characterization of the function spaces involved in the loop representation. These questions are both interesting and important, but they are beyond the scope of this paper. Rather, we are more interested here in gaining some intuition about the physical meaning of the structures that we have built*.

Several aspects of the loop representation are clarified by having an explicit expression for its relation with the self-dual representation. This is because in the self-dual representation the connection with the basic, but noninvariant observables, such as the metric, extrinsic curvature and connection is more direct. At the same time, the structure of the diffeomorphism invariant states and observables is transparent in the loop representation, but very difficult to see in the self-dual representation.

Our philosophy in this section is then to try to understand the two representations as two realizations of the same structure, in the spirit of Dirac [23]. This will be done by establishing a mapping between the two representations by choosing a complete set of commuting operators and identifying its eigenstates, in each representation, with each other.

A maximum set of commuting observables that are well defined in both representations is given by the T^0 observables. We will therefore study the eigenvector equation

$$T[\alpha]\Psi_i = c_i[\alpha]\Psi_i, \tag{81}$$

where i is an index that labels the eigenstates, in each of the representations, and we will identify the corresponding Ψ_i's with each other.

Before beginning we need to discuss an important point. Because we have no inner product in the unconstrained theory, there is no canonical mapping between states and linear functionals on the states. We therefore have the possibility of identifying the state space of the loop representation with either the state space of the self-dual representation, or its dual. It will turn out that the loop space is naturally identified with the dual, or bra space, of the self-dual representation.

We begin by reviewing the self-dual representation and by describing the T operators in it, and in its conjugate representation. Then we discuss the eigenvalue

* We assume that a topological structure has been given to the state space of the self-dual representation, and make use of this structure in our definitions.

equation (81) first in the self-dual representation and then in the loop representation. Having the solutions we can define the mapping that identifies the two sets of eigenstates. We then discuss how to express this mapping in the form of an integral transform, and, finally, we discuss the possibility of using the mapping itself as the instrument for *defining* the loop representation.

4.2. THE SELF-DUAL REPRESENTATION

The variable $A_a(x)$ has, when expressed in terms of the traditional variables of canonical relativity, the form of a function of the three-metric plus i times a function of the conjugate momenta. It is thus a nonlinear analogue of the Bargmann coordinate $z = q + ip$ on the complexified phase space of a finite dimensional system. This analogy guides the construction of the self-dual representation, which is the representation in which the elementary observables chosen for the quantization are the Ashtekar variables $A_a(x)$ and $\tilde{\sigma}^a(x)$. The space of the states is defined to be the linear space, which we call S, of holomorphic functionals of the connection $\Psi[A]$. Holomorphic means in this context that

$$\frac{\partial \Psi[A]}{\partial A_a^\dagger(x)} = 0, \tag{82}$$

where † is defined with respect to a fixed hermitian conjugate operation in the spin space. To the elementary observables $A_a(x)$ and $\tilde{\sigma}^a(x)$ correspond the quantum operators defined, respectively, as the multiplicative operator $A_a(x)$ and the functional derivative operator

$$\hat{\tilde{\sigma}}^a(x)\Psi[A] \equiv \frac{\hbar}{i}\frac{\partial}{\partial A_a(x)}\Psi[A]. \tag{83}$$

We use a hat to indicate the operators in the self-dual representation and to distinguish them from operators in the loop representation, which are indicated by a tilde.

The definition of any other observable in terms of the basic observables requires a choice of the factor ordering. We define the quantum operators $\hat{T}$ which quantize the classical observables T in the self-dual representation by requiring that all the $\hat{\tilde{\sigma}}^a(x)$ stand to the right of all the $A_a(x)$.

We now come to the definition of the conjugate representation. As we said in the previous paragraph, we will not be concerned with mathematical rigour, and assume that a topological structure has been assigned to S. The conjugate representation is then defined in terms of the dual S* of the states space S, which is the set of continuous linear functionals Φ: S → C.

On S* a representation of the observables' algebra conjugate to the one on S is defined: to any operator $\hat{O}$ on S is associated the operator $\hat{O}^*$ on S* defined by

$$\hat{O}^*\Phi(\Psi) \equiv \Phi(\hat{O}\Psi). \tag{84}$$

This equation defines a representation of the observables on S*. More precisely it is an antirepresentation, since

$$\begin{aligned}[\hat{O}^*, \hat{O}^{*\prime}]\Phi(\Psi) &= (\hat{O}^*\hat{O}^{*\prime} - \hat{O}^{*\prime}\hat{O}^*)\Phi(\Psi)\\ &= \hat{O}^{\prime *}\Phi(\hat{O}\Psi) - \hat{O}^*\Phi(\hat{O}'\Psi) = \Phi(-[\hat{O}, \hat{O}']\Psi).\end{aligned} \tag{85}$$

4.3. SELF-DUAL REPRESENTATION; COHERENT STATES

We now study eq. (81) in the self-dual representation. In this section we will use the following notation* for the trace of the holonomy of the connection A along a loop α

$$H[\alpha, A] = \mathrm{Tr}\left(\mathrm{P}\exp\oint_\alpha A\right). \tag{86}$$

Since the operator $\hat{T}^0[\alpha]$ is diagonal in the self-dual representation

$$\hat{T}[\alpha]\Psi[A] = H[\alpha, A]\Psi[A], \tag{87}$$

its eigenstates are "delta-functionals". This is the first point in which the conjugate representation turns out to be useful. We consider eq. (81) in the conjugate self-dual representation

$$\hat{T}^*[\alpha]\Phi_i = c_i[\alpha]\Phi_i. \tag{88}$$

This equation has a well-defined meaning and it is straightforward to see that it is solved by the delta distributions $\delta_{\bar{A}}$ which** are defined, for any given connection $\bar{A}_a(x)$ by

$$\delta_{\bar{A}}(\Psi) = \Psi[\bar{A}]. \tag{89}$$

* The notation is a bit redundant, since we have $T[\alpha] = H[\alpha, A]$, but we want to have a notation for the holonomy of a connection that distinguishes between cases in which we consider it as an observable on the phase space ($T[\alpha]$) and cases in which it has to be regarded just as a function of two variables ($H[\alpha, A]$).

** In this section the notation $\bar{A}$ does not mean the complex conjugate of A. It simply indicates a particular connection. We use the notation $\bar{A}$ instead of A when the connection is the index that labels the different eigenstates.

In fact, for any Ψ in S,

$$\hat{T}^*[\alpha]\delta_{\bar{A}}(\Psi) = \delta_{\bar{A}}(\hat{T}[\alpha]\Psi)$$
$$= \delta_{\bar{A}}(H[\alpha, \cdot]\Psi) = H[\alpha, \bar{A}]\Psi[\bar{A}] = H[\alpha, \bar{A}]\delta_{\bar{A}}(\Psi). \quad (90)$$

Therefore the index i, labeling the independent eigenstates, runs over the space of the connections $\bar{A}$. The simultaneous eigenvectors of all the $T^*[\alpha]$ operators are

$$\Phi_{\bar{A}} = \delta_{\bar{A}} \quad (91)$$

and the eigenvalue of $T^*[\alpha]$ corresponding to $\delta_{\bar{A}}$ is $H[\alpha, \bar{A}]$.

Since $H[\alpha, A]$ is gauge invariant, the simultaneous eigenspaces of all the $T[\alpha]$ are not one dimensional. The eigenspace corresponding to the eigenvalue $H[\alpha, \bar{A}]$ contains all the $\delta_{\bar{A}}$'s with $\bar{A}$ in the same gauge orbit as $\bar{A}$. This is, in fact, the only degeneracy, as the self-dual representation is defined to be the space of holomorphic functionals of A_a. Thus, any functional in the space must be completely determined by its value when the connection is restricted to be in SU(2). However, the $H[\alpha, A]$ uniquely determines an SU(2) connection up to gauge equivalence, so that on the space of holomorphic and gauge invariant functionals the simultaneous eigenspaces of $H[\alpha, A]$ are one-dimensional. Therefore the simultaneous eigenspaces of all the $T[\alpha]$'s uniquely define a basis for the space of the gauge invariant states. They can be labeled, for instance, by the connections $\bar{A}$ that satisfy a gauge condition that uniquely fixes a gauge. We assume in what follows that such condition has been chosen and the $\bar{A}$'s satisfy it.

These delta functions states have some interesting properties, which are relevant for discussing their "physical interpretation"*.

Recall that the self-dual representation is a holomorphic representation. The multiplicative operator A is therefore a creation operator, as are all the functions of A, including the $T[\alpha]$. In the same sense that $A_a(x)$ creates elementary excitations of the Ashtekar connection which are localized at the point x, the $T[\alpha]$ create excitations of the connection that are localized on the loop α. When we go to the conjugate representation these operators (which are not physical and must not be thought as self-adjoint, because the Ashtekar connection is complex), become destruction operators. The corresponding eigenstates, being eigenstates of a destruction operator, are "coherent states"**. Therefore they have a corresponding classical configuration, which is given by the corresponding eigenvalues. The bra-states $\delta_{\bar{A}}$

* Note that, as these are unconstrained states for quantum gravity they cannot be interpreted as physical states of the gravitational field. However, they are physical states for Yang–Mills theory, thus the following discussion can be understood as giving us some intuition as to how loop states represent gauge invariant quantum states of a connection.

** Since we are not in a Hilbert space, this notion of coherent states should not be taken too literally.

are therefore the "coherent states" corresponding to the classical configurations $\bar{A}_a(x)$ (A has a real part that contains information about the metric and an imaginary part that contains information about the momentum). The relation between the delta distributions and the bra coherent states is not surprising if we recall that one of the properties of the coherent statesin the holomorphic representation is exactly to act as delta functions; for instance for a one-dimensional system we have, for a coherent state $|c\rangle$, with $c = x + ip$,

$$\langle c|\psi\rangle = \psi(c). \tag{92}$$

for all Bargmann states $|\psi\rangle$.

4.4. LOOP REPRESENTATION; EXPONENTIAL STATES

Now let us study the eigenvalue equation (81) in the loop representation. It is given by

$$\tilde{T}[\alpha]\mathscr{A}_i[\{\eta\}] = \mathscr{A}_i[\alpha \cup \{\eta\}] = c_i[\alpha]\mathscr{A}_i[\{\eta\}]. \tag{93}$$

The eigenvector equation (93) is solved by any loop functional $\mathscr{A}[\{\eta\}]$ which has the form

$$\mathscr{A}_n[\{\eta_1 \dots \eta_n\}] = \prod_{I=1,n} c[\eta_I] \tag{94}$$

for any generic functional of single loops $c[\eta]$*. Recall that the loop states are functional of multiloops and we are using the notation

$$\mathscr{A}[\{\eta\}] = (\mathscr{A}_0, \mathscr{A}_1[\eta], \mathscr{A}_2[\{\eta_1, \eta_2\}], \dots). \tag{95}$$

In fact, if $\mathscr{A}[\{\eta\}]$ is given by eq. (94)

$$\tilde{T}^0[\alpha]\mathscr{A}[\{\eta\}] = \mathscr{A}[\alpha \cup \{\eta\}] = \prod_{\eta_1 \in \alpha \cup \{\eta\}} c[\eta_i] = c[\alpha]\mathscr{A}[\{\eta\}]. \tag{96}$$

Let us pause a moment to consider these loop functionals. They are given, in components by

$$\mathscr{A}[\{\eta\}] = (1, c[\alpha], c[\alpha_1]c[\alpha_2], \dots). \tag{97}$$

They are thus analogous to the exponential states of a symmetric Fock space. This is not surprising, since the exponential states in the Fock theory are the eigenvectors of the lowering operator, and $\tilde{T}^0$ is in fact a lowering operator in the Fock like structure of the loop representation. By appropriating the notation for the Fock

* Here I labels the different loops in the multiloop $\{\eta\}$.

exponential states we may introduce the notation exp(c) for the functional defined in eq. (94), and call the states in the image of exp, exponential states. exp is a mapping from the space of the functionals of a single loop to the space of the functionals of multiple loops.

In the standard Fock spaces, as ψ varies in the Hilbert space, the exp(ψ)'s span (or, actually, overspan) the symmetric Fock space. We expect something analogous to happen here, if it does, we could actually *define* the loop space as the linear span of the exponential states.

The situation is however somewhat complicated by the two-spinor condition

$$\mathscr{A}[\gamma\#\delta]+\mathscr{A}[\gamma\#\delta^{-1}]=\mathscr{A}[\gamma\cup\delta]. \tag{98}$$

This is a *linear* condition on the space of functionals of multiple loops, but it becomes a *nonlinear* condition on the functionals $c[\alpha]$ of a single loop if we impose it on the eigenstates of eq. (93), that is on the exponential states. In other words, from eq. (98) and

$$\mathscr{A}=\exp(c), \tag{99}$$

the nonlinear condition

$$c[\gamma\#\delta]+c[\gamma\#\delta^{-1}]=c[\gamma]c[\delta] \tag{100}$$

follows.

Therefore the eigenstates of the $\tilde{T}^0$ are the functionals exp(c), where c satisfies the nonlinear condition (100).

Now in ref. [22] it is shown that every loop functional $c[\alpha]$ that satisfies this nonlinear condition, the retracing condition (66b) and a suitable continuity condition can be written as the trace of the holonomy of an SL(2,$\mathbb{C}$) connection*. Thus, we have

$$c[\alpha]=H[\alpha,\bar{A}]. \tag{101}$$

Therefore we have obtained the result that in the loop representation the simultaneous eigenstates of the $T^0[\alpha]$ observables are given by the states

$$\mathscr{A}_{\bar{A}}=\exp H[\,\cdot\,,\bar{A}] \tag{102}$$

and the corresponding eigenvalues are $H[\alpha,\bar{A}]$, which are the same set of eigenvalues found in the self-dual representation.

Note that, because of the definition of the holonomy of a multiple loop, eq. (102) can be written as

$$\mathscr{A}_{\bar{A}}[\{\eta\}]=H[\{\eta\},\bar{A}], \tag{103}$$

* We are not interested here in the continuity condition, since we can always restrict ourselves, for the present purposes and at our level of rigour, to the "dense" subset in which such condition holds.

where

$$H[\{\eta\},\bar{A}] \equiv \prod_{\eta\in\{\eta\}} H[\eta,\bar{A}]. \tag{104}$$

The holonomies $H[\{\eta\},\bar{A}]$ are seen here, for any fixed connection $\bar{A}$, as loop functionals. These loop functionals are the eigenvectors of the $\tilde{T}^0$ operators. They represent, in the loop representation, the same states that are represented by the $\delta_{\bar{A}}$'s in the conjugate self-dual representation. Thus they represent coherent states. Each one of them is characterized by a connection $\bar{A}$ and corresponds to the classical gravitational configuration with the three-metric and its conjugate momentum given by the real and the imaginary part of $\bar{A}$. Since by definition these loop states span the loop states space, we have obtained an intuitive picture of the loop state space.

4.5. THE MAPPING $\mathscr{F}$

At this point we can put together the results of the two previous subsections. Following the program outlined in the beginning of this section, we can identify the eigenvectors of the two representations characterized by the same eigenvalues

$$\delta_{\bar{A}} \leftrightarrow H[\{\eta\},\bar{A}]. \tag{105}$$

We may introduce a linear mapping $\mathscr{F}$ from S* into the loop functionals space that realizes this identification. We define a map

$$\mathscr{F}:\ \Phi \mapsto \mathscr{A} \tag{106}$$

by

$$\mathscr{A}[\{\eta\}] = \mathscr{F}(\Phi)[\{\eta\}] = \Phi(H[\{\eta\},\cdot]). \tag{107}$$

Note that in this equation the holonomy $H[\{\eta\},A]$ is seen, for any multiple loop $\{\eta\}$, as a functional of the connection. Φ acts on this functional and gives, for any $\{\eta\}$, a complex number. That is, it gives a loop functional. The mapping $\mathscr{F}$ defined in eq. (107) is linear and continuous. It is straightforward to verify that it realizes the identification (105), that is it sends the T^0 eigenvectors in the conjugate self-dual representation to the corresponding eigenvectors in the loop representation.

Let us now consider how operators transform under the action of $\mathscr{F}$. We will define a self-dual operator $\hat{O}$ and a loop operator $\tilde{O}$ to be *equivalent* if

$$\mathscr{F}\hat{O}^* = \tilde{O}\mathscr{F}. \tag{108}$$

We now show that the loop observables and constraints are equivalent to the corresponding operators in the self-dual representation. That is, $\mathscr{F}$ is in fact the mapping that realizes the identification of the two representations.

$\hat{O}$ and $\tilde{O}$ are equivalent if, for any Φ,

$$\mathscr{F}\hat{O}^*\Phi = \tilde{O}\mathscr{F}\Phi. \tag{109}$$

Using the definitions (84) and (107), this is the condition,

$$\Phi(\hat{O}H[\alpha, \cdot]) = \tilde{O}\Phi(H[\alpha, \cdot]). \tag{110}$$

$\tilde{O}$ can be brought inside the parentheses, since it acts only on the loop variables, and since this equation has to hold for any Φ, we have the result that $\hat{O}$ and $\tilde{O}$ are equivalent if

$$\hat{O}H[\{\alpha\}, A] = \tilde{O}H[\{\alpha\}, A], \tag{111}$$

where the first operator acts on the A argument, while the second acts on the loop argument. Eq. (111) is the main tool used in the study of the relationship between the two representations.

Note that not all the self-dual operators admit an equivalent. An operator that admits an equivalent will be called transferable. In particular the elementary operators of the self-dual representation $\tilde{\sigma}$ and A do not admit equivalents.

A class of operators that are transferable are the T^n operators. In the self-dual representation, we may choose for the $\hat{T}^n$ operators the ordering given by putting all the $\tilde{\sigma}^*$ operators *on the left* of the A. Note that the $\hat{T}^n$ operators are well defined and do not need to be regularized, because we have built into their definition the condition that the hands, which correspond, in the self-dual representation, to functional derivative operators, cannot coincide.

It is straightforward to verify that the T^n operators satisfy eq. (111). For instance consider T^0. We have

$$\hat{T}[\alpha]H[\{\gamma\}, A] = H[\alpha, A]H[\{\gamma\}, A] = H[\alpha \cup \{\gamma\}, A] = \tilde{T}[\alpha]H[\{\gamma\}]. \tag{112}$$

The reader may verify that the same is true for the other T^n.

The fact that the $\mathscr{T}$ algebra is transferable can be used to elucidate some of the properties of $\mathscr{F}$. First we note that any transferable operator sends the kernel of $\mathscr{F}$ into itself*. In fact, if $\hat{O}$ is transferable then $\tilde{O}$ exists such that $\mathscr{F}\hat{O}^* = \tilde{O}\mathscr{F}$; if Φ in the kernel of $\mathscr{F}$, we have $0 = \tilde{O}\mathscr{F}\Phi = \mathscr{F}(\hat{O}^*\Phi)$; that is $\hat{O}^*\Phi$ is also in the kernel of $\mathscr{F}$. Assume now that the space of the self-dual states S is restricted in such a way that we have an irreducible representation of the gauge invariant algebra of observables. One can always do this without losing the physical content of the

* One can also demonstrate the vice versa: if an operator sends the kernel of $\mathscr{F}$ into itself, then it is transferrable.

classical limit (indeed, one *has* to restrict the state space to an irreducible representation in order to avoid adding extra degrees of freedom). Then by assumption there is no nontrivial invariant subspace under the algebra. Since the kernel of $\mathscr{F}$ is an invariant subspace, it has to be trivial (or the whole space, but this is not the case, because the image of $\mathscr{F}$ is not empty). Therefore we do not "lose" any state going from the self-dual representation to the loop representation. Thus, a left inverse mapping $\mathscr{F}^{-1}$ exists on the image of the irreducible sector such that

$$\mathscr{F}^{-1}\mathscr{F}=1\,. \tag{113}$$

The constraints in the self-dual representation are related to the constraints in the loop representation by eq. (111), in the same way as the observables T^n. As the constraints are the central operators in quantum gravity, it is interesting to study how they behave under $\mathscr{F}$ in more detail.

The self-dual internal gauge constraint is, because it annihilates the holonomy, equivalent to the null operator in loop space. This is because, as we already said, the transform $\mathscr{F}$ "picks up" only the SU(2)-gauge invariant content of the self-dual wave function. Therefore the internal gauge constraint is automatically solved in going to the loop representation.

For the diffeomorphism and the hamiltonian constraints there is an issue related to the ordering problem. We choose for the conjugate constraints $\hat{C}^*$ the ordering with all the $\tilde{\sigma}^*$ operators *on the left* of the A operator (this is the ordering in which the constraint algebra formally closes [4]). Then $D_a(x)$ is given, in the self-dual representation, by

$$D_a(x)=\mathrm{Tr}\,F_{ab}(x)\frac{\delta}{\delta A_b(x)}\,, \tag{114}$$

which is the generator of the diffeomorphisms, and the hamiltonian constraint $\mathscr{C}(x)$ (in the unregularized form) is given by

$$\mathscr{C}(x)=\mathrm{Tr}\,F_{ab}(x)\frac{\delta}{\delta A_a(x)}\frac{\delta}{\delta A_b(x)}\,. \tag{115}$$

The loop generator of the diffeomorphisms is the equivalent of the operator $D_a(x)$. In fact, as the holonomy of a shifted A is like the holonomy of the same A for a loop shifted in the opposite direction, it is straightforward to verify that

$$\int_\Sigma \mathrm{d}^3x\, v^a D_a(x)H[\alpha, A]=D(v)H[\alpha, A]\,, \tag{116}$$

where in the left-hand side we have the self-dual operator and in the right-hand side the loop operator.

Finally, the equivalence of the regulated loop and the regularized self-dual hamiltonian operators follows directly from the fact that they are defined in terms of the T^n operators, which are equivalent in the two representations.

4.6. INTEGRAL REPRESENTATION OF THE TRANSFORM

In this subsection we give a formal integral representation of the mapping $\mathscr{F}$. The definition is formal because we ignore issues related to the exact definition of the spaces of functionals on which the integral is defined.

The idea is to represent a dense subspace in S* by functionals of the connection by introducing a generic measure $\mu[A]$. μ is a technical device with no direct physical meaning. We stress the fact that it should not be confused with the measure that defines the physical Hilbert structure. In particular we do not require that the observables are self-adjoint with respect to the Hilbert structure defined by this measure, nor that it is invariant under gauge or coordinate transformations. For instance, we can fix a coordinate system on Σ and choose for μ a gaussian measure formally given by

$$\mathrm{d}\mu[A] = \mathrm{e}^{-S[A]}\,\mathrm{d}[A], \tag{117}$$

where $S[A]$ is a quadratic functional of the connection and its derivatives. This is a well-defined measure for any choice of such a quadratic functional. We do not need to specify it here more precisely because, as we shall see, the essential results are, in a sense that will be specified, independent of the choice of this measure. Given μ, a subspace of S* can be represented by functionals $\Phi[A]$. These are defined by

$$\Phi(\Psi) = \int \mathrm{d}\mu[A]\,\Psi[A]\Phi[A]. \tag{118}$$

By using this representation of the distributions Φ, the transform $\mathscr{F}$ can be written as

$$\mathscr{A}[\{\gamma\}] = \int \mathrm{d}\mu[A]\,H[\{\gamma\},A]\,\Phi[A], \tag{119}$$

which is the original formula for the transform $\mathscr{F}$ given in ref. [16].

It should be emphasized that $\mathscr{F}$ is well defined from the conjugate self-dual states to loop states without any measure; the measure is needed only to express the conjugate self-dual states in terms of functionals $\Phi[A]$, and therefore express $\mathscr{F}$ as an integral transform.

The operators are represented on the $\Phi[A]$'s in the following way. O^* is represented by the operator $O^\dagger$ adjoint to O with respect to the chosen measure μ, that is the operator that satisfies

$$\int \mathrm{d}\mu[A]\,\Phi[A](O\Psi[A]) = \int \mathrm{d}\mu[A]\,(O^{\dagger}\Phi[A])\Psi[A]. \tag{120}$$

The $\Phi[A]$ that represents a given Φ depends, obviously, on the choice of μ. Similarly the notion of adjoint depends on the choice of μ. In particular the adjoint of A_a is A_a itself and the adjoint of $\hat{\tilde{\sigma}}^a$ is

$$\hat{\tilde{\sigma}}^{\dagger a}(x) = -\frac{\delta}{\delta A_a(x)} + \frac{\delta\, \mathrm{d}\mu[A]/\delta A_a(x)}{\mathrm{d}\mu[A]}. \tag{121}$$

The second term comes from the derivative of the measure in the integration by parts. We assume that the definition of S is such that the boundary term in the integration by parts vanishes. If the measure has the form (117) we have

$$\hat{\tilde{\sigma}}^{\dagger a}(x) = -\frac{\delta}{\delta A_a(x)} - \frac{\delta S[A]}{\delta A_a(x)}. \tag{122}$$

Note that the multiplicative term added to the (functional) derivative does not modify the basic commutation relations. Recall that to add to $\hat{\tilde{\sigma}}^a(x) = \delta/\delta A_a(x)$ the functional gradient of any functional of A is exactly the freedom allowed in the definition of the fundamental observables. The freedom to redefine the measure is therefore related to the freedom to redefine the operator in the conjugate self-dual representation by adding a gradient.

Representations defined by operators in which a different gradient term has been added to the functional derivatives define, in general, unitarily inequivalent representations. That is, the corresponding operators are self-adjoint with respect to different Hilbert structures, represented on the functional space by inequivalent measures, and the functional spaces that represent the states, which are the corresponding L_2 spaces, are different. Thus, if the transform, and its functional representation (119) are to be understood rigorously, it will be necessary to give a more precise definition of the spaces of functionals we are working with, and to understand the issues of the irreducibility of these spaces as representations of the algebra of observables. These questions are important but beyond the scope of this paper.

The integral (119) may be explicitly computed in the linearized theory, where there is a natural Poincaré invariant choice for $\mu(A)$. By doing so a complete representation of the Fock space of linearized general relativity in terms of loop functionals may be obtained [13]. This result gives us confidence in the transform and the loop representation.

4.7. ALTERNATIVE DEFINITION OF THE LOOP REPRESENTATION

The formulas of the previous two sections offer an alternative way to define the loop representation which does not directly involve any reference to the classical $\mathcal{T}$ algebra.

Assume that we first quantize the theory in the self-dual representation. We may then define the corresponding conjugate representation. Then, by means of the transform $\mathcal{F}$ defined in eq. (107), we obtain the states space of the loop representation, and, by means of eq. (108), or its simpler equivalent eq. (111), we "bring" observables and constraints to the loop space. In this way we construct the loop representation as a "transformed form" of the self-dual representation.

This is analogous to defining the momentum representation of finite dimensional quantum mechanics by first defining the coordinate representation and then the Fourier transform. The analogy is even stronger if we compare our theory with the quantum theory of a free relativistic particle. This is defined by the Klein–Gordon equation, which can be seen as the quantum hamiltonian constraint corresponding to the classical constraint $p^2 - m^2 = 0$, which defines the classical dynamics. The solutions of the Klein–Gordon equation are certain exponentials $\psi_k(x) = e^{ikx}$. It is very convenient to "go to the k basis", where k is the label of the exponentials. This is obtained by means of the Fourier transform which is an integral transform with the exponentials as the integral kernel. The observables are "brought" to the k-representation by constructing their equivalents under the Fourier mapping. Equivalent observables are defined as those that have the same action on the exponentials. For instance, the momentum operator is given in the two representations by, respectively, $-i\,\mathrm{d}/\mathrm{d}x$ and k, and the relation between the two is given by

$$-i\frac{\mathrm{d}}{\mathrm{d}x}e^{ikx} = k\,e^{ikx}. \tag{123}$$

Now, since, in the self-dual representation, a class of solutions of the hamiltonian constraint is given by products of traces of the holonomies around certain classes of loops, it is natural, by analogy, to try to "go to the loop basis". Following the analogy, one can introduce an integral transform, with the holonomy as the integral kernel, and use it to define the loop representation and to "bring" the operators to the loop space. Note the strict analogy between the form of eq. (119) and the Fourier transform and between eqs. (111) and (123).

This was the way in which the loop representation was originally discovered. The $\tilde{T}^n$ operators, for instance, which may seem to be pulled from the air in sect. 3, were obtained originally, through eq. (111), by looking for transferable operators. This construction of the loop representation is more intuitive, and the introduction of the loops is directly motivated by the discovery of the loop solutions of the self-dual hamiltonian constraint. On the other hand, the quantization by means of the $\mathcal{T}$ algebra is much more direct, and avoids the long path through the self-dual representation and most of the difficulties related to the infinite dimensional character of the mapping $\mathscr{F}$.

From this point of view the loop representation is nothing but a change of basis in the state space of the quantum theory: we use a set of states on which the hamiltonian constraint acts in a simple way as the new basis elements. This is analogous to the change of basis from the functional formulation of a free field theory to a Fock basis, or from the Schrödinger formulation of the hydrogen atom to the energy–angular momentum basis.

A change of basis may be nontrivial for two reasons. From the mathematical side we know that a formal change of basis leads in general to inequivalent structures. From the physical side we expect that a new basis which simplifes the dynamics can give us new concepts to describe the physics, in the same sense in which the Fock basis in the quantized Maxwell field provides us with the concept of photons.

5. Solutions

In this section we study the solutions of the quantum constraints in the loop representation.

5.1. DIFFEOMORPHISM INVARIANT STATES AND GENERALIZED KNOT CLASSES

We begin with the diffeomorphism constraint. We want to find the general solution to

$$D(v)\mathscr{A}[\{\eta\}]=0. \tag{124}$$

Since $D(v)$ is the generator of the action of the diffeomorphism group on the loop space $\mathscr{M}$, eq. (124) is equivalent to the requirement that the state $\mathscr{A}[\{\eta\}]$ is constant along the orbits of this action. We denote by $K(\{\eta\})$ the orbit in which the multiple loop $\{\eta\}$ lies. Thus, the general solution to eq. (124) is

$$\mathscr{A}[\{\eta\}]=\mathscr{A}[K(\{\eta\})]. \tag{125}$$

Let us now study these orbits.

First of all, as ϕ sends a multiple loop composed of n loops into a multiple loop composed of the same number of loops, a first characterization of the orbits is given by the number of loops. A second diffeomorphism invariant of the loop is the number of intersections. A third is the numbers of points of discontinuity of the first derivatives of each loop*. The orbits are thus distinguished by a string of integers that codes these characteristics.

In addition, sets of smooth and nonintersecting loops fall into equivalence classes under the diffeomorphism group which are the knot and link classes of the loops.

The knot classes of Σ are the equivalence classes of smooth loops in Σ under an operation known as *ambient isotopy* [9]. It is then a standard theorem in knot theory

* We have already required that the loops be nondegenerate, so there are no points where the tangent vectors vanish.

that the equivalence classes of smooth loops with respect to diffeomorphisms are the same as the equivalence classes under ambient isotopy [9].

For the case of *multiple* smooth nonself-intersecting loops there is, beside the knotting of each single loop, the additional phenomenon of the linking of different loops, which is also invariant. The orbits of the diffeomorphisms in the space of these multiple loops are called the *link classes* of the manifold. We use the notation L for the link classes of the manifold.

The generalization of knot theory to include also loops with corners is rather trivial. The generalization to include intersecting loops (or graphs) is less trivial, and has already been considered by some authors [24, 29].

In this paper we will call the orbits of the diffeomorphism group on Σ the *generalized link classes of* Σ.

Functionals in the loop representation that depend only on the generalized link class of the loops will then be solutions to the diffeomorphism constraints if it is true that the condition that the states be constant under the orbits of the diffeomorphisms is compatible with the conditions we imposed on the states space in subsect. 3.3.

The first condition on the states is reparametrization invariance. Since the image of the loop is transformed by a diffeomorphism in a way that is independent of its parametrization it is clear that reparametrizations commute with the action of the diffeomorphisms, and therefore that the orbits of the diffeomorphisms are well defined on the space of unparametrized loops.

The second condition is the two-spinor identity (66). The diffeomorphisms preserve this condition, because it is a topological relation, expressed by the breaking and rejoining at the intersections, and these operations are diffeomorphisms invariant. Therefore the condition is well defined on the orbits K. The third condition, the retracing identity (67) is preserved under diffeomorphisms, for the same reason.

The general solution to the diffeomorphism constraint is thus given by functionals of the form (125), with $K(\{\eta\})$ denoting the generalized link class of η in Σ, which, in addition, satisfy the two-spinor identity (66a) and the retracing identity (66b).

It is important to stress the fact that the generalized link classes form a *denumerable* set. The loop space $\mathscr{M}$ is then, on the one hand, rich enough that a complete algebra of quantum observables for general relativity can be defined on it, but on the other hand, simple enough that, using it, the complete solution to the diffeomorphism constraints can be given in terms of a denumerable basis.

5.2. SOLUTIONS TO THE HAMILTONIAN CONSTRAINT: FORMAL CALCULATION

We now show that the regulated hamiltonian constraint has a nontrivial space of solutions. In particular, we want to show that if $\mathscr{A}$ has support only on smooth loops, without intersections, then it is in the kernel of the hamiltonian constraint, in the sense that

$$\lim_{\delta \to 0} \frac{1}{\delta^2} \tilde{\mathscr{C}}^{\delta}(x) \mathscr{A}[\eta] = 0. \tag{126}$$

Thus, we will recover the results of ref. [6], using the loop space representation and regularization.

We denote by $\mathscr{M}_0$ the subspace of $\mathscr{M}$ formed by the multiple loops which are nonintersecting and smooth everywhere. We denote by $\mathscr{M}_0^{\text{ext}}$ the subspace of $\mathscr{M}$ formed by multiloops which are composed by smooth nonintersecting loops and/or by eyeglass loops that are related to smooth nonintersecting loops (see subsect. 2.5). We will first show that a loop functional that has support on $\mathscr{M}_0$ is in the kernel of the hamiltonian constraint. These loop functionals do not represent states because they cannot satisfy eqs. (66a); however, by suitably extending these functionals to $\mathscr{M}_0^{\text{ext}}$, we shall obtain states that are in the kernel of the hamiltonian constraint.

We do the calculation in two steps. First we consider the regularized hamiltonian constraint without the smearing described at the end of sect. 3. Using it we obtain our result by a formal calculation in which we deal with the Δ-distributions. Then in the next subsection we repeat the same calculation in the context of the completely well-defined regularization given by both the δ regularization of the hamiltonian and by the congruence-smearing of the loops.

We begin with the regularized unsmeared hamiltonian constraint. Its action on any loop functional $\mathscr{A}[\{\eta\}]$ gives

$$\tilde{\mathscr{C}}^{\delta}(x) \mathscr{A}[\{\eta\}] = \sum_{i=1}^{4} c_i \Delta^{[a}[\gamma_{ab}^{\delta}(x)](\delta^2) \Delta^{b]}[\gamma_{ab}^{\delta}(x)](2\pi) \mathscr{A}[\eta_i], \tag{127}$$

where η_i are four multiloops and c_i four finite coefficients, which are given by the action of the T^2's on $\gamma_{ab}^{\delta}(x)$ on η.

We now consider separately two cases. In the first case the loop η is in $\mathscr{M}_0$, in the second η has corners or intersections so that it is not in $\mathscr{M}_0$. On $\mathscr{M}_0$ it is not difficult to show that $\tilde{\mathscr{C}}^{\delta}(x)\mathscr{A}[\{\eta\}]$ is $\mathrm{O}(\delta^3)$. Since a loop in $\mathscr{M}_0$ is smooth, the first of the two factors of Δ may be expanded in the loop parameter around the origin. If η has no self-intersections, then the leading term is equal to the second Δ factor and the two cancel by the antisymmetrization in the vector indices. The next to leading term is then $\mathrm{O}(\delta^3)$ so that in the limit we have zero. Note that if the loop has a corner, this expansion is not possible, since the η in Δ may be discontinuous in the region, where we want to expand it.

Now consider the value of eq. (127) on the loops η with intersections or corners. One may show that if α has a corner or an intersection, then $\alpha \# \beta$ also has a corner or an intersection for any β in $\mathscr{M}$, since the operation of breaking and joining cannot subtract a corner or intersection.

Thus we have the following result. If we apply the hamiltonian operator to a functional $\mathscr{A}[\{\eta\}]$ with support only on $\mathscr{M}_0$, then $\lim_{\delta\to 0}\tilde{\mathscr{C}}^{\delta}(x)\mathscr{A}[\{\eta\}]$ is zero both on $\mathscr{M}_0$ and outside it. Thus, it vanishes everywhere in $\mathscr{M}$. Thus, we have shown that using the loop representation, the hamiltonian constraint may be regularized in such a way that it has a nontrivial kernel, and that any loop functional with support on $\mathscr{M}_0$ is in this kernel. As we said, a loop functional $\mathscr{A}$ that has support on $\mathscr{M}_0$ cannot satisfy eq. (66a) and therefore does not represent a state. We will now construct an extension $\mathscr{A}_{\text{ext}}$ of the functional $\mathscr{A}$, which has support on $\mathscr{M}_0^{\text{ext}}$, satisfies eq. (66) and is still in the kernel of the hamiltonian constraint. Let us call any state that has support on $\mathscr{M}_0^{\text{ext}}$ and satisfies the conditions

$$\mathscr{A}_{\text{ext}}\left[\alpha\circ\rho\circ\beta\circ\rho^{-1}\right]=\mathscr{A}_{\text{ext}}\left[\alpha\circ\rho\circ\beta^{-1}\circ\rho^{-1}\right],$$

$$\mathscr{A}_{\text{ext}}\left[\alpha_1\circ\rho,\alpha_2\circ\rho^{-1}\right]=-\mathscr{A}_{\text{ext}}\left[\alpha_1\circ\rho\circ\alpha_2^{-1}\circ\rho\right],$$

an *extended* state. Note that the value of any extended state $\mathscr{A}_{\text{ext}}$ on the eyeglass loops is uniquely determined by its restriction on $\mathscr{M}_0$, namely by its value on the smooth nonintersecting loops. In fact, it follows from the definition and from eqs. (66a, b) that

$$\mathscr{A}_{\text{ext}}\left[\alpha\circ\rho\circ\beta\circ\rho^{-1}\right]=\tfrac{1}{2}\mathscr{A}_{\text{ext}}[\alpha,\beta],$$

$$\mathscr{A}_{\text{ext}}\left[\alpha_1\circ\rho\circ\alpha_2^{-1}\circ\rho\right]=-\tfrac{1}{2}\mathscr{A}_{\text{ext}}[\alpha].$$

Thus we have one extended state $\mathscr{A}_{\text{ext}}$ with support on $\mathscr{M}_0^{\text{ext}}$ for every state $\mathscr{A}$ with support on $\mathscr{M}_0$. We call $\mathscr{A}_{\text{ext}}$ the extension (to the eyeglass loops) of $\mathscr{A}$. It is straightforward to show that these extended functionals are still in the kernel of the hamiltonian constraint. This follows from the fact that the hamiltonian constraint operator commutes with the conditions that define the extended states, so that, for instance

$$\tilde{\mathscr{C}}^{\delta}\mathscr{A}_{\text{ext}}\left[\alpha\circ\rho\circ\beta\circ\rho^{-1}\right]=\tfrac{1}{2}\tilde{\mathscr{C}}^{\delta}\mathscr{A}_{\text{ext}}[\alpha,\beta]\xrightarrow[\delta\to 0]{}0.$$

Thus, the extended functionals are states that solve the hamiltonian constraint. Therefore, we have the result that for every functional on the space of smooth nonintersecting loops we have one solution of the hamiltonian constraint.

This calculation is, however, formal because the Δ's in the r.h.s. of the eq. (127) are distributions. Thus the action of the unsmeared hamiltonian constraint is not well defined unless all of the distributional singularities can be eliminated by appropriate integrations. In subsect. 3.6 we showed that this is the case for all elements of the $\tilde{\mathscr{T}}$ algebra. However, for concreteness we will, in the next subsection, indicate the details of the calculation when the action of the T^2's in the definition of $\tilde{\mathscr{C}}^{\delta}$ have been regularized by integrating over the space of test functions, as described in subsect. 3.6. Before this, however, three comments are in order.

(i). There is an analogy with the Fourier transform representation of the Klein–Gordon equation, in which the solutions are given by the wave functions with support on the Lorentz hyperboloid. From this point of view $\mathscr{M}_0$ as a subspace of $\mathscr{M}$ is analogous to the Lorentz hyperboloid as a subspace of momentum space.

(ii). We may expect, from the results of refs. [6] and [25], that the space of solutions is larger and includes also some $\mathscr{A}[\{\eta\}]$ which have support on intersections, providing that certain restrictions are satisfied.

(iii). Finally, as is noted in the introduction, the limit $\delta \to 0$ we have built into the construction of the hamiltonian constraint operator is defined by the point wise topology in $\mathscr{M}$. If we had a Hilbert structure on the space of the unconstrained states we could have imposed a stronger requirement using the Hilbert norm. However, as we have emphasized, the physical inner product is only available on the space of solutions to the constraint, it is thus not available for use here. However, it is possible that another, nonphysical inner product could be used to give a stronger definition of this limit. We may note that this situation is, to our knowledge, new in context of nonperturbative analysis of four-dimensional quantum field theories.

5.3. SOLUTIONS TO THE HAMILTONIAN CONSTRAINT: REGULARIZED CALCULATION

In order to demonstrate that the right-hand side of eq. (127) does in fact vanish as a distribution, we have to smear the hamiltonian constraint, along the lines described in sect. 3, and verify that the smeared hamiltonian constraint does annihilate any states whose support is restricted to smooth, nonself-intersecting loops.

The first step in doing this is to substitute for the loop $\gamma^{\delta}_{ab}(x)$ a suitable two-dimensional congruence $\tilde{\gamma}^{\delta}_{ab^2}$. (In what follows we will, for ease of notation, drop the $_{ab}$ and the (x) wherever they are not important.)

In order to mimic the definition of the unsmeared loop we choose, for every δ, a congruence with the following characteristic. $\gamma^{\delta}_{0,0}(s)$ is given by the loop γ^{δ} which is in the definition of the unsmeared hamiltonian constraint. γ^1, that is the first fattening of the congruence, is centered at $s=0$ and is contained in the interval $[-\frac{1}{4}\delta^2, \frac{1}{4}\delta^2]$. γ^2, that is the second fattening, is centered at $s=\delta^2$ and it is contained in the interval $[\delta^2-\frac{1}{4}\delta^2, \delta^2+\frac{1}{4}\delta^2]$. Moreover, we assume that the width of the fattening is also of the order δ^2.

In terms of this congruence we define the smeared regulated hamiltonian constraint

$$\tilde{\mathscr{C}}^{\delta} = \sum_{ab} T\left[\gamma^{\delta}_{2[ab]}(x)\right](f^{ab}), \tag{128}$$

where the test function is given in term of a generic smooth scalar function $f(x, y)$ by

$$f_{cd}^{ab}(x, y) = f(x, y)\, \delta_c^a \delta_d^b . \tag{129}$$

Now we show that for every test function $f(x, y)$ the result of the action of this operator on any loop functional with support on $\mathscr{M}_0$ goes pointwise to zero as δ goes to zero.

By applying $\tilde{\mathscr{C}}^\delta$ to a state $\mathscr{A}[\{\eta\}]$ we get the following integral

$$\int \mathrm{d}s\, \mathrm{d}t f(\eta(s), \eta(t))\, \dot{\eta}^{[a}(s) \dot{\eta}^{b]}(t) R(s, t) \sum_i c_i \mathscr{A}[\eta_i], \tag{130}$$

where $R(s, t)$ is zero unless $\eta(s)$ and $\eta(t)$ are in the regions of the first and second fattening of $\tilde{\gamma}^\delta(x)_{ab}$ respectively.

As in the unsmeared calculation we can expand $\dot{\eta}^b(t)$ around $t = s$,

$$\dot{\eta}^b(t) = \dot{\eta}^b(s) + t \frac{\mathrm{d}}{\mathrm{d}t} \dot{\eta}^b(t)|_{t=s} + \mathrm{O}(t^2). \tag{131}$$

Since nothing depends on the parametrization of η we can work for simplicity in a parametrization in which the η-parameter measures the proper lengths in a cartesian metric in the coordinate chart in which the hamiltonian constraint is defined. With other parametrizations one would have to add some suitable jacobian to the considerations that follow. The function R fixes the domain of integration in s and t in two regions of Σ which are separate but which become close, when δ goes to zero, as δ^3. Thus in the expansion of $\dot{\eta}^b(t)$ the t in front of the derivative is of order δ^3. Thus we can drop this term, and the ones of higher order in t, since we are taking the limit of $1/\delta^2$ times the action of the regulated operator. But to first order in t the integral we are considering is zero because of the antisymmetrization. This completes our calculation. We have shown that the formal result of the previous subsection is true also in the context of a completely well-defined regularization of the infinities.

5.4. PHYSICAL STATES

We now may combine the results of the previous subsections to exhibit some physical states of the gravitational field.

Consider the class of solutions of the hamiltonian constraint given by the extended functionals with support on $\mathscr{M}_0^{\mathrm{ext}}$. Note that this class is *transformed into itself by the diffeomorphism constraint*. This result is general: if a loop functional $\mathscr{A}[\{\eta\}]$ is in the kernel of the hamiltonian constraint, then so is $U_\phi \mathscr{A}[\{\eta\}]$. This means that the transformation properties of the hamiltonian constraint under diffeomorphisms are correct; that is, the commutator of the hamiltonian constraint with the diffeomorphism constraint is proportional to the hamiltonian constraint itself. Because of that the hamiltonian constraint is well defined on the orbits K.

The generalized link classes K which are in $\mathscr{M}_0$ are the ones formed by smooth nonself-intersecting loops. These are the ordinary link classes L.

Thus, we have the following result. Let $\mathscr{A}$ be an element of $\mathscr{S}$ defined such that,

$$\mathscr{A}[\{\eta\}]=0, \tag{132}$$

unless $\{\eta\}$ is smooth and nonintersecting, and, when it does not vanish,

$$\mathscr{A}[\{\eta\}]=\mathscr{A}[\mathrm{L}[\{\eta\}]] \tag{133}$$

where $\mathrm{L}[\{\eta\}]$ is any ordinary link class of Σ. Then $\mathscr{A}[\{\eta\}]$, extended to the eyeglass loops, is a physical state of the gravitational field.

We have one independent state of this kind for every ordinary link class L of Σ. For instance we can consider a basis in this space formed by states $\mathscr{A}_{\mathrm{L}}[\{\eta\}]$ which are defined to be 1 if $\{\eta\}$ is in L and zero otherwise.

To summarize, we have found a large space of physical states. A basis of this space is in one-to-one correspondence with the link classes of Σ. These states may be written in closed and explicit form in the loop representation. Furthermore, the action on any of these states of any (regularized) operator in the unconstrained operator algebra, $\tilde{\mathscr{T}}$, may be computed explicitly*. We call this space $\mathscr{S}^0_{\mathrm{phys}}$.

$\mathscr{S}^0_{\mathrm{phys}}$ forms a sector of the nonperturbative state space of quantum gravity, which we call the no intersections sector.

As we already said, the results of ref. [6] suggest that besides $\mathscr{S}^0_{\mathrm{phys}}$ there may be other sectors related to loops with intersections that satisfy suitable algebraic conditions at the intersections.

5.5. AN ALTERNATIVE FORM FOR THE HAMILTONIAN CONSTRAINT; THE SHIFT OPERATOR

The reader may be struck by how complicated, and cumbersome, the expression of the hamiltonian constraint is in the loop representation, compared to the other elements of the formalism. While the formalism for the hamiltonian constraint we have been discussing works, in the sense that it can be fully regularized in a way that is consistent with diffeomorphism invariance, one would like to know whether the hamiltonian constraint might have an expression in the loop representation which is more natural. In this section we would like to make a tentative proposal concerning this question. We will introduce a simple loop operator which has the same action as the regulated hamiltonian constraint, when it is evaluated on loops

* Of course, if the operator is unphysical its action will take the state out of the physical state space.

with corners but with no intersections. Whether or not the definition of this operator can be extended in such a way that its action agrees with that of the hamiltonian constraint on all loop functionals is presently an open problem.

However, even if such an extension is not possible, the form of the operator we will introduce is rather suggestive, and helps to clarify why the loops with no corners and no intersections are singled out by the hamiltonian constraint. Finally, the form of this operator, if it can be extended to the general case, suggests the existence of an extremely simple structure behind the dynamics of general relativity.

We start from the self-dual hamiltonian constraint $\mathscr{C}(x)$. We consider the smeared form

$$\mathscr{C}(f)=\int \mathrm{d}^3x f(x)\mathscr{C}(x), \tag{134}$$

and we regularize it by point splitting the two functional derivatives. We do this by defining a universal regularizing function $z_\epsilon(x, y)$ such that

$$\lim_{\epsilon\to 0} z_\epsilon(x, y)=\delta^3(x, y). \tag{135}$$

For instance, we could use

$$z_\epsilon(x, y)=\epsilon^{-3/2}\,\mathrm{e}^{-|x-y|^2/\epsilon}. \tag{136}$$

Then we can write

$$\mathscr{C}(f)=\lim_{\epsilon\to 0}\mathscr{C}_\epsilon(f), \tag{137}$$

where

$$\mathscr{C}_\epsilon(f)=\int \mathrm{d}^3x\,\mathrm{d}^3y f(x) z_\epsilon(x, y)\mathrm{Tr}\left[F_{ab}(x)\frac{\delta}{\delta A_a(x)}\frac{\delta}{\delta A_b(y)}\right]. \tag{138}$$

We then use eq. (111) to find the loop equivalent of $\mathscr{C}_\epsilon(f)$. By computing its action on the holonomy we get

$$\mathscr{C}_\epsilon(f)H[\gamma, A]=\int \mathrm{d}s\int \mathrm{d}t f(\gamma(s)) z_\epsilon(\gamma(s),\gamma(t))\dot\gamma^a(s)\dot\gamma^b(t)$$
$$\times F^k_{ab}(\gamma(s))\epsilon^{kij}\,\mathrm{Tr}\left[U(t,s)\tau^i U(s,t)\tau^j\right]. \tag{139}$$

Here τ^i are the Pauli matrices multiplied by $-\frac{1}{2}i$. Using

$$\epsilon^{ijk}\tau^j{}_A{}^B\tau^k{}_C{}^D=i\tau^i{}_A{}^D\delta^B_C-i\tau^i{}_C{}^B\delta^D_A \tag{140}$$

we obtain

$$\mathscr{C}(f)H[\gamma,A]$$
$$= i\int \mathrm{d}t \int \mathrm{d}s f(\gamma(s)) z_\epsilon(\gamma(s),\gamma(t))\dot{\gamma}^a(s)\dot{\gamma}^b(t)$$
$$\times\left[\mathrm{Tr}\, U(t,s)\mathrm{Tr}(F_{ab}(\gamma(s))U(s,t)) - \mathrm{Tr}\, U(s,t)\,\mathrm{Tr}(F_{ab}(\gamma(s))U(t,s))\right]. \tag{141}$$

We are interested in the limit of this formula for small ϵ. There are two possibilities for a nonvanishing integral. Either x is on γ and is not an intersection point for γ. Or the point x is an intersection point for the loop γ. In the first case, for small enough ϵ, s and t have to be close. In the second case $\gamma(s)$ and $\gamma(t)$ may coincide even if s is different than t.

Let us forget for the moment the possibility that there is an intersection, and assume that s and t are close. Without loss of generality, we may assume that $s > t$. Then for small ϵ it follows that $U(t,s) = \mathit{identity}$ and $U(s,t) = U(s)$. Then the first term in the brackets is zero, because the trace of the τ contained in F vanishes. In the second term the first trace is the trace of the identity, that is 2, and we obtain

$$\mathscr{C}_\epsilon(f)H[\gamma,A] = 2i\int \mathrm{d}t \int \mathrm{d}s f(\gamma(s)) z_\epsilon(\gamma(s),\gamma(t))\dot{\gamma}^a(t)\dot{\gamma}^b(s)$$
$$\times \mathrm{Tr}(F_{ab}(\gamma(s))U(s,s+1)) + \mathrm{O}(\epsilon^2). \tag{142}$$

Now we note that

$$\int \mathrm{d}s\, \delta^3(x,\gamma(s)) f^a(s) \frac{\delta}{\delta\gamma^a(s)} H[\alpha,A]$$
$$= \int \mathrm{d}s\, \delta^3(x,\gamma(s)) f^a(s)\dot{\gamma}^a(s)\mathrm{Tr}(F_{ab}(\gamma(s))U(s,s)). \tag{143}$$

where $\delta/\delta\gamma^a(s)$ is the loop derivative, which is discussed, for example, in refs. [12] and [17]. Therefore we have the result

$$\mathscr{C}_\epsilon(f)H[\gamma,A] = \tilde{\mathscr{C}}_\epsilon(f)H[\gamma,A], \tag{144}$$

where

$$\tilde{\mathscr{C}}_\epsilon(f) = 2i \int \mathrm{d}s \int \mathrm{d}t f(\gamma(s)) z_\epsilon(\gamma(s), \gamma(t)) \dot{\gamma}^a(t) \frac{\delta}{\delta\gamma^a(s)} \tag{145}$$

is an operator that acts only on the loop variable. Thus, acting on a restricted domain of loop functionals which vanish on intersecting curves, eq. (145) is, up to terms of $O(\epsilon)$ the transform of the regularized hamiltonian constraint. Thus, on this restricted space of functionals, it may be considered to be a form of the regularized hamiltonian constraint.

The case in which the transform is evaluated on sets of loops which intersect is more complicated, and we do not have, at the present time, a complete analysis of this case. It does seem possible that an extension of (145) in which the action of $\tilde{\mathscr{C}}_\epsilon(f)$ is supplemented by combinatorial operators which rearrange the legs at the intersections will work in the general case.

In spite of the fact that (145) is the transform of the hamiltonian constraint only on a restricted space of loop functionals, it has some extremely interesting properties, which we would like to describe here.

The operator $\tilde{\mathscr{C}}_\epsilon(f)$ is divergent as $1/\epsilon$ in the limit $\epsilon \to 0$. The simplest way to renormalize it is to introduce a *multiplicative* renormalization and multiply it by ϵ. The operator

$$\epsilon\tilde{\mathscr{C}}_\epsilon(f) \tag{146}$$

is well defined in the limit. Note that by multiplying the function $z_\epsilon(x, y)$ by ϵ we essentially obtain a one-dimensional delta function instead of a three-dimensional one. Thus, by renormalizing the operator in this way we are changing the operator from a density in the three-dimensional space to a density in the one-dimensional space of the loop parameter. The latter is well defined in the loop representation, however, we may have trouble with the commutator of the diffeomorphism constraint with the *renormalized* operator. (As (145) is the transform of the regulated hamiltonian constraint, it still should behave suitably under spatial diffeomorphisms. It is only the action of taking the limit above, that may be problematical*.)

We proceed to take the limit $\epsilon \to 0$. In the limit one of two integrations can be performed and we get the *renormalized* hamiltonian constraint

$$\mathscr{S}(f) = \lim_{\epsilon \to 0} \epsilon\tilde{\mathscr{C}}_\epsilon(f) = \int \mathrm{d}s f(\gamma(s)) \dot{\gamma}^a(s) \frac{\delta}{\delta\gamma^a(s)}. \tag{147}$$

* We may note that this form of *renormalization* of the hamiltonian constraint is only possible in the loop representation, where operators may depend explicitly on loops. A similar renormalization would not be possible, for example, in the self-dual representation, as it is improper to allow the regularization procedure which renders an operator meaningful to depend on the state that it is acting on.

Let us now compute the commutator of the regularized hamiltonian constraints. We have

$$
\begin{aligned}
&[\tilde{\mathscr{C}}_\epsilon(f), \tilde{\mathscr{C}}_\epsilon(g)] \\
&\quad = \int \mathrm{d}s\, \mathrm{d}t\, \mathrm{d}u\, \mathrm{d}w f(\gamma(s)) z_\epsilon(\gamma(s), \gamma(t)) \dot{\gamma}^a(s) \frac{\delta}{\delta\gamma^a(t)} \\
&\qquad \times g(\gamma(u)) z_\epsilon(\gamma(u), \gamma(w)) \dot{\gamma}^b(u) \frac{\delta}{\delta\gamma^b(w)} - (f \leftrightarrow g) \\
&\quad = \int \mathrm{d}s\, \mathrm{d}t\, \mathrm{d}u\, \mathrm{d}w \Big[f(\gamma(s))\, \delta_a g(\gamma(u))\, \delta(t-u) \dot{\gamma}^a(s) z_\epsilon(\gamma(s), \gamma(t)) z_\epsilon(\gamma(u), \gamma(w)) \\
&\qquad\qquad + f(\gamma(s)) g(\gamma(u)) z_\epsilon(\gamma(s), \gamma(t)) z_\epsilon(\gamma(u), \gamma(w)) \dot{\gamma}^a(s) \frac{\mathrm{d}}{\mathrm{d}w} \delta(t-w)\, \delta_a^b \Big] \\
&\qquad \times \frac{\delta}{\delta\gamma^b(w)} - (f \leftrightarrow g) \\
&\quad = \int \mathrm{d}s\, \mathrm{d}u\, \mathrm{d}w \big(f(\gamma(s))\, \partial_a g(\gamma(u)) - (g(\gamma(s))\, \partial_a f(\gamma(u)) \big) \\
&\qquad \times \dot{\gamma}^a(s) \dot{\gamma}^b(u) z_\epsilon(\gamma(s), \gamma(u)) z_\epsilon(\gamma(u), \gamma(w)) \frac{\delta}{\delta\gamma^b(w)}\,. \qquad (148)
\end{aligned}
$$

If we now take the $\epsilon \to 0$ limit we obtain

$$\int \mathrm{d}s \left[f(\gamma(s)) \partial_a g(\gamma(s)) - (g(\gamma(s)) \partial_a f(\gamma(s)) \right] \dot{\gamma}^a(s) \dot{\gamma}^b(s) \frac{\delta}{\delta\gamma^b(s)}, \qquad (149)$$

which is the regularized loop form of the observable

$$\int \mathrm{d}^3x \left[f(x)\, \partial_a g(x) - g(x)\, \partial_a f(x) \right] \tilde{\sigma}^a(x) \tilde{\sigma}^b(x) D_b(x), \qquad (150)$$

which is on the right-hand side of the classical commutator.

We may note that this result depends on our taking the commutator of the regularized constraints before we take the limit $\epsilon \to 0$. If we had taken the limit first, and thus taken the commutator of the *renormalized* operators (147), we would not have gotten the right answer.

A similar situation holds concerning the commutator of the diffeomorphism constraints with our new form of the hamiltonian constraint. An explicit calculation shows that diffeomorphism constraints generate, acting on the *regulated* operator $\tilde{\mathscr{C}}_\epsilon(f)$, the following action of the diffeomorphism group. The function $f(x) z_\epsilon(x, y)$ transforms as a scalar in x and y. This is correct, as $f(x)$ should be a density of weight -1 because $z_\epsilon(x, y)$, which goes in the limit $\epsilon \to 0$, to a three-dimensional delta-function, must, for consistency, be a density of weight 1.

Thus, the commutator of the regularized hamiltonian constraints, eq. (145), with themselves and with the diffeomorphism constraints are correct. However, once one takes the limit $\epsilon \to 0$ the renormalized form of the constraint, (147) has incorrect commutators with itself and with the diffeomorphism constraints.

Finally, we may study the action of the renormalized operator $\mathscr{S}$. We recall that in this discussion we are disregarding the intersections, thus we will evaluate its action only on nonintersecting loops. Note that we can write

$$\mathscr{S}\mathscr{A}[\{\eta\}] = \lim_{h\to 0} \frac{1}{h}\left(\mathscr{A}[\{\eta_h\}] - \mathscr{A}[\{\eta\}]\right), \tag{151}$$

where

$$\eta_h^a(s) = \eta^a(s) + h\dot{\eta}^a(s). \tag{152}$$

If η is everywhere differentiable,

$$\eta^a(s) + h\dot{\eta}^a(s) = \eta^a(s+h). \tag{153}$$

If, on the contrary, $\eta^a(s)$ is nondifferentiable at s, that is there is a "corner", then eq. (153) is not true. In order to get a feeling of the action of $\mathscr{S}$ it is helpful to consider its action on the functionals of generic loops η that can be closed or open. Then we can say in general that η_h is the loop obtained by "shifting" η along its tangent. If η is closed and differentiable, this action is just a reparametrization, if it is open it is just shifted ahead towards one of the ends, and if there is a corner, η is broken in the point of the corner. $\mathscr{S}$ is therefore the operator that generates this shift. We call it the shift operator (see fig. 13). Since the state functions are reparametrization invariant loop functionals we have that for any differentiable loop η, with no intersections,

$$\mathscr{S}\mathscr{A}[\{\eta\}] = 0. \tag{154}$$

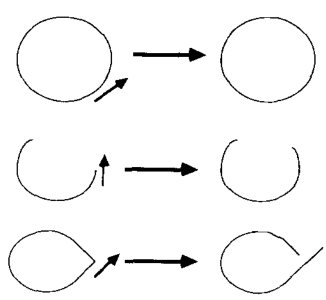

Fig. 13. The action of the shift operator.

Therefore we immediately have the result that the states that have support only on differentiable nonself-interesting loops are in the kernel of the hamiltonian operator.

The shift operator is a very natural operator on the loop space. Both the diffeomorphisms and the shift operator are first order operators, and, more precisely, are Lie derivatives on the loop manifold [31]. Recall that to any vector field $V[\gamma]^a(s)$ is associated the Lie derivative operator

$$\mathscr{L}_V = \int \mathrm{d}s\, V[\gamma]^a(s) \frac{\delta}{\delta\gamma^a(s)}\,. \tag{155}$$

The diffeomorphism constraint is the Lie derivative corresponding to the vector fields

$$V_{\vec{f}}[\gamma]^a(s) = f^a(\gamma(s)) \tag{156}$$

and the shift operator is the Lie derivative of the vector field

$$V_f[\gamma]^a(s) = f(\gamma(s))\dot{\gamma}^a(s)\,. \tag{157}$$

The regularized hamiltonian operator is also a Lie derivative. It corresponds to the vector field

$$V_{f,\epsilon}[\gamma]^a(s) = \int \mathrm{d}t f(\gamma(t))\dot{\gamma}^a(t) z_\epsilon(\gamma(t),\gamma(s))\,. \tag{158}$$

Thus, we may summarize the results of this section by saying that, at least on the restricted set of loop functionals which have support only on nonintersecting loops, the transform takes an operator which is a second order functional differential operator in the self-dual representation to an operator which is a first order functional differential operator in the loop representation. To compute this transform correctly, the operator in the self-dual representation must first be regularized by splitting the points at which the functional derivatives, $\delta/\delta A_a(x)$ act. After the transform we have a first order functional differential operator; however, it is still regulated, by means of a smearing of the vector field on loop space on which it depends to an expression which is nonlocal in terms of the loop parameter.

In the self-dual representation we do not know of a simple way to understand the action of the regularized hamiltonian constraint in the limit in which the regularization is removed. However, in the loop representation, at least in the restricted class of functionals we have considered, the action of the hamiltonian constraint does have a simple meaning when the regularization is removed. It is, up to an infinite multiplicative renormalization, proportional to a simple geometrical operator.

These remarks are very incomplete. It certainly seems that the simple structure revealed by the appearance of the shift operator deserves to be better understood. In particular we find it very striking that the action of the hamiltonian constraint can be coded (up to the action on the intersections) in an extremely natural operator like the shift operator. If an extension of the shift operator can be invented that, acting on a general state in the loop representation, reproduces the action of the regularized hamiltonian constraint, this extension would have to contain, through the transform, the dynamical content of the Einstein equations. It is thus possible that the full complexity of the Einstein equations may be coded in the simple action of shifting the loops along themselves.

5.6. STRUCTURE OF THE PHYSICAL STATES SPACE $\mathscr{S}^0_{\text{phys}}$ AND EXAMPLES

The sector of the physical states space $\mathscr{S}^0_{\text{phys}}$ that we have discovered inherits from knot theory a rich structure. We close this section with a few general remarks about the structure of this sector of the solution space.

First of all $\mathscr{S}^0_{\text{phys}}$ has the same Fock-like structure as the full $\mathscr{S}$. That is, there is a quantum number that counts the number of loops that can be used to give the physical state space a graded, or Fock-like structure. The zero level of this structure is formed by a single state. It is possible to show, under some symmetry assumption on the measure in the transform $\mathscr{F}$, that this level zero state is related to the constant state in the self-dual representation. The fact that the constant functional is a solution to all the constraints in the self-dual representation was already noted in ref. [6], where it was called, with some trepidation, the "vacuum state".

The expression vacuum should not be taken literally. It is not the state of minimum energy, since the theory has no preferred time, and therefore no natural notion of energy. It is also not an invariant state under rigid motions, since there is not any group of rigid motions in the theory, and there is no reason to believe that this state is associated with any flat metric with respect to which a rigid invariance group could be defined. It is also not clearly characterized by a notion of spatial homogeneity, since it is difficult to express this notion in the presence of the three-dimensional diffeomorphisms invariance, which is a property of all physical states. It may however be characterized as the physical state of minimal complexity in the loop representation.

The nth level of the Fock-like structure is then given by linear combinations of knots formed by n loops. Within each level the knots may still be ordered in terms of growing complexity. For instance the simplest one loop knot is the unknot. The complexity of a knot can be defined in several ways, one way involves the expression of the knot in terms of operations of the braid group, another involves the reduced, two-dimensional knot diagrams [9].

It is interesting to note that there is a natural implementation of chirality in knot theory. The simplest single loop knot after the unknot is the so-called trefoil knot, represented in fig. 14. It exists in two inequivalent versions; the right-handed and the left-handed one [9].

Fig. 14. Right and left trefoil knots.

As an example of a state in $\mathscr{S}^0_{\text{phys}}$ which is not diagonal in the number of loops quantum number, consider any ordinary link invariant $\mathscr{P}$ [9,26], and extend it to $\mathscr{M}$ by assigning the value 0 at any loop not in $\mathscr{M}_0$. Then

$$\mathscr{A}_{\mathscr{P}}[\{\eta\}] \equiv \mathscr{P}[\{\eta\}], \tag{159}$$

extended to the eyeglass loops, is a physical state of quantum gravity.

These states are particularly interesting for two reasons. First because the link polynomials are the principal tools used in attempts to construct a complete classifications of knots and links. If a, presumably denumerably infinite, set of link polynomials is discovered which classifies the links completely they could be used to define a useful complete basis for the no intersection sector of the physical state space. A second reason for interest in such states is the recent mathematical results that associate classes of link polynomials with conformal quantum field theory and integrable statistical mechanical models in two dimensions [26]. These results, together with the results of this section yield an unexpected connection between two dimensional physics and quantum gravity, as any system which yields a link invariant may be associated with an exact, nonperturbative, physical state of the gravitational field. It is tempting to speculate that this result is more than accidental. For example, one might conjecture that it is an indication of a connection between string theory and quantum general relativity at the nonperturbative level.

6. Conclusions and reflections

The results described in this paper were obtained without any physical input or hypothesis besides those contained in the basic principles of general relativity and quantum mechanics. The developments that made these results possibles are primarily formal: Ashtekar's reformulation of general relativity, Isham's studies on the possibility and opportunity of quantizing noncanonical algebras of observables, and, on the mathematical side, the development of knot theory.

The present formulation of the loop representation is not yet a complete theory of quantum gravity. Two basic elements are still missing for the definition of a complete theory. These are the definition of the physical scalar product and the definition of a class of physical observables. We begin these remarks with a discussion of these two issues.

Let us recall that the identification of physical observables is an unsolved problem already in *classical* general relativity. This is not surprising, since in the canonical framework the entire dynamics is contained in the constraints, so that the problem of finding classical observables that commute with the constraints is in fact closely related to the problem of finding the general solution to the equations of motion. The inadequacy of our understanding of pure general relativity and of its physical observables cannot be over stressed.

At the same time, we do know operators which are defined over the space of physical states that we have found. The "number of loops", "self-linking number", and so on, are all quantum numbers which define operators diagonal in the loop representation. But we are not able to express these operators in terms of the elementary operators. We therefore lose the connection with the physical interpretation of the theory, which is given by the fact that the elementary operators are related to classical observables to which operational procedures are attached.

At the present time we do not know the solution to this difficulty. Possible directions in which one might be sought are the following:

(i). Study the algebra of operators which are naturally defined on the physical states, and try to deduce from it directly a physical interpretation. This would be a departure from the textbook quantization procedure we have followed up to now.

(ii). Study the algebra of physical observables in terms of some suitable expansion. The strong coupling expansion is probably the most natural one in this context [27, 28]. Since we have exact solutions for the states, but do not know the observables, we have to expand the observables. One could start by defining observables that commute with the constraints only up to a certain order in the expansion parameter*.

(iii). Extend the formalism to the asymptotically flat case. This would provide us with an explicit algebra of known physical observables, such as energy and momentum, defined at spatial infinity.

(iv). Add matter. The presence of matter partially simplifies the issues concerned with the physical observables. To understand why matter could help, note that it is, in principle, possible to characterize a physical observable and to fix its operational definition even without having its explicit value as a function in the phase space. Now in *pure* general relativity, in the absence of matter, to do even this is extremely difficult. In the presence of matter, on the contrary, it is not difficult to operationally define physical observables. For instance, if we have a stone with a clock on it coupled to general relativity, then the value of the Ricci scalar where the stone is and at a fixed clock time is a well-defined diffeomorphism invariant observable, even if we do not know its explicit expression as a function on the space of the initial data. The addition of matter will certainly be a necessary step for the development of the loop representation.

* Global observables of this kind may be found in general relativity, for instance $P = \int_\Sigma q_{ab}\tilde{p}^{ab}$ commutes with *all* the constraints up to terms of the order $O(1/G)$.

(v). Finally it is possible that if the general solution to the Einstein equations for a spatially compact universe were known, one could express the space of parameters of these solutions in terms of topological quantities that have to do with the linking and knotting of some natural set of curves in the three-manifolds. This would directly provide the physical basis for the topological observables suggested by the present work.

The problem of the scalar product is closely related to the problem of the physical observables, since the main condition that we must impose on the physical Hilbert structure is that real physical observables should be represented by self-adjoint operators.

Note that the link classes form a *countable* set. Therefore the problem of fixing the Hilbert structure may be simplified by the fact that we have a discrete basis. This problem is then, perhaps, less burdened than one might at first think by subtleties concerning functional measures, allowing the problem to be formulated more directly in physical terms. For example, one could assume that the knot classes form a natural basis in the physical Hilbert space, and postulate that this basis is orthonormal. This would directly fix the Hilbert structure. The consequences of this hypothesis have not yet been studied.

Besides these basic open problems a number of points concerning the structures described remain unclear. Among these are:

(a). The question of the existence of other sectors of physical states. We have reasons to suspect the existence of additional sectors involving intersecting loops. These can very likely be studied by extending directly the methods developed in this paper.

(b). The exact definition of the space state, including the complete characterization of the functionals $\mathscr{A}[\{\eta\}]$ is still an open problem. This is related to the question of the reducibility or irreducibility of the representation of the algebra of observables that we have defined. In this context it would be useful to understand whether or not the T^0, T^1 observables by themselves are enough to describe the phase space of general relativity, as we discussed in sect. 2.

(c). The commutator algebra of the constraints has to be studied more thoroughly. The existence of a well-defined set of solutions to the full set of constraints assures us that there are no anomalies proportional to the identity operator, but there remains, at the present time, the possibility that terms appear in the commutator of two hamiltonian constraints which are of higher order in $\hbar$, and are proportional to operators that have the property of annihilating the space of solutions of the hamiltonian constraint.

(d). The shift form of the hamiltonian operator is very suggestive, but still not really understood. It could, perhaps, simplify the study of the physical sectors involving intersecting loops.

(e). We would like to understand more deeply the appearance of the link classes upon application of the diffeomorphism constraints. It is tempting to conjecture that this has something to do with the representation theory of the three-dimensional diffeomorphism group. Indeed, one may conjecture that there is an irreducible representation of the diffeomorphism group of a three-manifold Σ associated with each of its knot and link classes [29]. These would be constructed by defining a suitable linear function space over the configuration space of the curves in each class, the diffeomorphisms having a natural linear action on such spaces. Unfortunately, at present very little is known about the representation theory of the diffeomorphism groups of three-dimensional manifolds. But there are some results which are very suggestive in this regard [30].

This brings us to our final remarks. The reader who has followed us up until this point is undoubtedly ruminating over the following issue: how is it possible to find a large class of exact solutions to the dynamical equations of a quantum theory when comparably few exact solutions to the classical theory are known? While we have nothing definitive to say about this, we think the following considerations are relevant.

First of all, it must be stressed that we have so far only one sector of the physical state space. While this sector is infinite dimensional, and intricately structured, it may still correspond to only a small, or degenerate, set of the full set of solutions.

But even if we are able to construct the full space of physical states, the physical observables are still missing. In this sense the dynamics of the quantum theory is contained in the constraints twice. Only the first of these two problems is significantly addressed by the results we have described. It could be that the difficult problems of quantum gravity, from the present point of view, are concerned with the construction of the physical observables and the physical Hilbert structure. In particular, while it is possible that the divergences found in perturbative approaches are completely tamed in a nonperturbative approach such as our's, it is also possible that they are lurking ahead in the problems of the physical observables and Hilbert structure*.

The problem lies in finding physical operators to which we can give a physical interpretation. In turn, this problem reflects, as we stressed above, our very poor understanding of the classical observables of general relativity. Thus, we may be faced with an ironic situation in which the problem of giving a physical interpretation to the operator algebra which results from a completely nonperturbative quantization of general relativity rests, ultimately, on our gaining a deeper understanding of the classical theory.

From an optimistic point of view, it is possible that the dynamics of general relativity is based on a simple and elegant mathematical structure that is still largely uncomprehended, and that this simplicity becomes more apparent in the quantum

* However, note that all of the dangerous operator products in the dynamics are contained in the hamiltonian constraint. The hamiltonian, when it exists, is, besides the constraints, only a surface term linear in the basic observables [4]. Thus, no divergences are lurking there!

theory (certainly, the existence of the Ashtekar formalism, as well as the results of Penrose, Newman and others concerning self-dual solutions that, partially, inspired it, are very suggestive in this regard). This would suggest that new concepts are needed to describe what quantum general relativity has to say concerning the structure of space-time at the Planck length. The classification of physical states in terms of knot and link invariants may represent a first step in this direction.

At the very least the results we have described show that nonperturbative quantum general relativity has a rich and nontrivial structure and that it is possible to develop techniques and concepts to explore this structure.

Note added

After the submission of this work two interesting papers by E. Witten appeared which bear on the work described here. In the first [32], Witten shows that, in the language used here, the loop transform of $\exp(ib\int_\Sigma Y(A))$, where $Y(A)$ is the Chern–Simons form of an SU(2) connection, is equal to the Jones polynomial [26], which is an important invariant of links. (In this connection, see also ref. [33].) In Witten's paper a relationship between Chern–Simons theories and rational conformal field theories is introduced, so that it has been conjectured [34] that to every rational conformal field theory is associated a Chern–Simons theory. This strengthens the conjecture we made in sect. 5 and in ref. [2] that a large class of conformal field theories will be associated with quantum states of $3+1$ quantum gravity, by means of their association with link invariants. In Witten's second paper [35], $2+1$ dimensional general relativity is quantized by expressing it as a Chern–Simons theory. Witten's formulation of $2+1$ dimensional general relativity is, as is described in ref. [36], the analogue of Ashtekar's formulation of $3+1$ dimensional general relativity. $2+1$ dimension general relativity can also be quantized in the loop representation, thus providing an example that illuminates many of the results described here [37]. Finally, another application of the methods described here is to the quantization of the Gowdy models [38], which are reductions of general relativity to a $1+1$ dimensional field theory by the imposition of two spacelike Killing fields.

We are grateful, first of all, to Abhay Ashtekar, for his help, criticism, suggestions, and encouragement, which have been so essential to us during the last year. We are grateful also to Louis Crane who contributed with many discussion to the ideas contained in this work and to Chris Isham for his precise criticism of our mathematical naiveté and for several crucial discussions on the general problems of nonperturbative quantization of general relativity.

We would like to thank also Michael Atiyah, Viqar Husain, Ted Jacobson, Karel Kuchar, Miguel Virasoro and Richard Woodard for discussions, comments and criticism, and Luca Bombelli for a careful reading of the manuscript. In addition, CR would like to thank Professor Thomas Appelquist for the hospitality of the physics department at Yale University, where this work was begun, and both of us would like to thank Professor Marco Toller for the hospitality of the physics department at the Università di Trento, where this work was completed.

LS was supported by National Science Foundation grants to Yale University to Syracuse University and by a grant from the Research Corporation.

References

[1] C.J. Isham, GRG11 review talk, preprint Imperial/TP/85-86/39, 1986

[2] C. Rovelli and L. Smolin, Phys. Rev. Lett. 61 (1988) 1155;
C. Rovelli and L. Smolin, *in* Proc. of the Osgood Hill Conf. Conceptual Problems in Quantum Gravity, Boston, 1988, ed. A. Ashtekar and J. Stachel (Birkhauser, Basel).

[3] A. Ashtekar, Phys. Rev. Lett. 57 (1986) 2244; Phys. Rev. D36 (1987) 1587

[4] A. Ashtekar, New perspectives in canonical gravity (with invited contributions) (Bibliopolis, Naples, 1988)

[5] B.S. DeWitt, Phys. Rev. 160 (1967) 1113

[6] T. Jacobson and L. Smolin, Nucl. Phys. B299 (1988) 295

[7] C.J. Isham, *in* Relativity groups and topology II, Les Houches, 1983, ed. B.S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984)

[8] C.J. Isham and A.C. Kakas, Class. Quantum Grav. 1 (1984) 621

[9] L.H. Kauffman, Formal knot theory, and On knots (Princeton Univ. Press, Princeton, 1987) and refs. therein

[10] A.M. Polyakov, Phys. Lett. B82 (1979) 247;
A.M. Polyakov, Nucl. Phys. B164 (1979) 171;
Y. Nambu, Phys. Lett. B80 (1979) 372;
Yu. M. Makeenko and A.A. Migdal, Phys. Lett. B88 (1979) 135;
F. Gliozzi, T. Regge and M.A. Virasoro, Phys. Lett. B81 (1979) 178;
M. Virasoro, Phys. Lett. B82 (1979) 436;
K. Wilson, Phys. Rev. D10 (1974) 247;
J. Kogut and L. Suskind, Phys. Rev. D11 (1975) 395;
A. Jevicki and B. Sakita, Phys. Rev. D22 (1980) 467;
B. Sakita, *in* Field theory in elementary particles, ed. B. Kursunoglu and A. Perlmutter (Plenum, New York, 1983);
W. Furmanski, A. Kolawa, Nucl. Phys. B291 (1987) 594

[11] A.M. Polyakov, Gauge fields and strings (Harwood Academic publishers, Switzerland, 1987)

[12] C. Rovelli and L. Smolin, Loop representation of Yang Mills theories: lattice Syracuse and Roma preprint, in preparation;
C. Rovelli, Loop representation of Yang Mills theories: continuum theory (Univ. of Rome) preprint in preparation

[13] A. Ashtekar and C. Rovelli, Quantum Faraday lines, loop representation of Maxwell theory, Syracuse preprint in preparation (1989)

[14] P.M. Dirac, Can. J. Math. 2 (1950) 129; Proc. R. Soc. A246 (1958) 326 and 333; Phys. Rev. 114 (1959) 924;
P.G. Bergmann et al., Phys. Rev. 80 (1950) 81;
R. Arnowit, S. Deser and C.W. Misner, Phys. Rev. 113 (1959) 745; 116 (1959) 1322; 117 (1960) 1595

[15] K. Kuchar, *in* Quantum Gravity 2, ed. C.J. Isham, R. Penrose and D.W. Sciama (Oxford Univ. Press, Oxford, 1982)

[16] C. Rovelli, The loop space representation, ch. V.8 of New perspectives in canonical gravity, op. cit.

[17] C. Rovelli, Il Nuovo Cimento B100 (1987) 343

[18] J.A. Wheeler, *in* Battelle Rencontres 1967, ed. C. DeWitt and J.A. Wheeler (Benjamin, New York, 1968)
[19] I.M. Gel'fand, N.Ya. Vilenkin, Generalized functions, Vol. 4 (Academic Press, New York, 1964)
[20] P.M. Dirac, Lectures on quantum mechanics, Belfer Graduate School of Science (Yeshiva University, New York, 1964)
[21] R. Penrose, Application of negative dimensional tensors, *in* Combinatorial mathematics and its applications, ed. D.J.A. Welch (Academic Press, New York, 1971)
[22] F. Gliozzi and M.A. Virasoro, Nucl. Phys. B164 (1980) 141
[23] P.M. Dirac, The principles of quantum mechanics (Claredon Press, Oxford, 1930) first edition
[24] L. Kauffman, private communication
[25] V. Husain, Nucl. Phys. B313 (1989) 711
[26] V.F.R. Jones, Invent. Math. 72 (1983); Bull. Am. Math. Soc. 12 (1985) 103;
L.H. Kauffman, Statistical mechanics and the Jones Polynomial, preprint, Univ. of Illinois at Chicago (1987); Topology 26 (1987) 395;
Y. Akutsu and M. Wadati, J. Phys. Soc. Jpn. 56 (1987) 839;
A. Kuniba, Y. Akutsu and M. Wadati, J. Phys. Soc. Jpn. 55 (1986) 3285;
Y. Akutsu and M. Wadati, Commun. Math. Phys. (1988);
J. Fröhlich, Statistics of fields, the Yang–Baxter equation, and the theory of knots and links, ETH preprint, 1987
[27] M. Henneaux, M. Pilati and C. Teitelboim, Phys. Lett. B110 (1982) 123
[28] C. Rovelli, Phys. Rev. D35 (1987) 2987
[29] L. Smolin, Knot theory, loop space and the representation theory of the diffeomorphism group, ch. V.6 of New perspectives in canonical gravity, op. cit.
[30] Ismaliov, Funct. Anal. Appl. 9 (1975) 144;
A.M. Vershik, I.M. Gelfand and M.I. Graev, Russ. Math. Surv. 30 (1975) 1
[31] Y. Choquet-Bruhat and C. DeWitt-Morette, Analysis, Manifolds and Physics (North-Holland, Amsterdam, 1982)
[32] E. Witten, Quantum field theory and the Jones polynomials, IAS preprint HEP-88/33
[33] L. Smolin, Invariants of links and critical points of the Chern–Simon path integral, Syracuse University preprint, 1989
[34] G. Moore and N. Seiberg, preprint IASSNS-HEP-89/6
[35] E. Witten, Nucl. Phys. B311 (1988/89) 46
[36] I. Bengtsson, Yang–Mills theory and general relativity in three and four dimensions, preprint Inst. Theor. Phys. S-41296 Goteborg, Sweden (1989)
[37] A. Ashtekar, V. Husain, C. Rovelli, J. Samuel and L. Smolin, Class. and Quant. Grav. 6 (1989) L185
[38] V. Husain and L. Smolin, Exactly solvable quantum cosmologies from two Killing field reductions of general relativity, Syracuse preprint (1989)