CGPG-94/5-1
May 10, 1994

NEW LOOP REPRESENTATIONS FOR 2+1 GRAVITY

A. Ashtekar and R. Loll

Center for Gravitational Physics and Geometry
Physics Department, Pennsylvania State University
University Park, PA 16802, U.S.A.

Abstract

Since the gauge group underlying 2+1-dimensional general relativity is non-compact, certain difficulties arise in the passage from the connection to the loop representations. It is shown that these problems can be handled by appropriately choosing the measure that features in the definition of the loop transform. Thus, "old-fashioned" loop representations – based on ordinary loops – do exist. In the case when the spatial topology is that of a two-torus, these can be constructed explicitly; *all* quantum states can be represented as functions of (homotopy classes of) loops and the scalar product and the action of the basic observables can be given directly in terms of loops.

1 Introduction

Theories of connections are playing an increasingly important role in the current description of all fundamental interactions in Nature. The standard model of particle physics which encompasses the electro-weak and strong interactions is based on Yang-Mills theories. Classical general relativity (in three and four space-time dimensions) can also be formulated as a dynamical theory of connections. Finally, such theories are of interest from a mathematical viewpoint as well: many of the recent advances in the understanding of the topology of low-dimensional manifolds have come from theories of connections, in particular from the analysis of Yang-Mills instantons and expectation values of Wilson loop functionals in Chern-Simons theories.

In these theories, the configuration space generally arises as a space $\mathcal{A}$ of connection one-forms $A(x)$ on a Cauchy surface Σ, taking values in the Lie algebra of a gauge group G. The corresponding phase space $T^*\mathcal{A}$ is then naturally parametrized by canonically conjugate pairs of fields $(A(x), E(x))$, where E is a vector density of weight one on Σ, taking values in the dual of the Lie algebra of G, which can be thought of as a generalized "electric" field conjugate to the gauge potential A. Gauge invariance is enforced by a Gauss constraint $\mathrm{Div}_A E(x) = 0$. As a consequence, the physical configuration space is $\mathcal{A}/\mathcal{G}$, the quotient of $\mathcal{A}$ by the group $\mathcal{G}$ of local gauge transformations. The physical observables are the gauge-invariant functions on phase space.

Gambini and Trias [1] were the first to point out that a convenient set of such observables can be associated with loops, i.e. closed curves α in Σ as follows. Choose a representation of dimension N of the gauge group G, and set

$$\begin{aligned} (T^0(\alpha))\ [A] &= \frac{1}{N} Tr \left(\mathrm{P} \exp \oint_\alpha A \right) \\ (T^1(\alpha, s))^a\ [A, E] &= Tr \left(E^a(\alpha(s))\, \mathrm{P} \exp \oint_\alpha A \right). \end{aligned} \tag{1.1}$$

where P denotes path ordering along the loop α. $T^0(\alpha)$ is labelled by the loop α and represents a (gauge-invariant) configuration variable since it depends only on the connection A. $(T^1(\alpha, s))^a$ is labelled by a loop α and a point $\alpha(s)$ on α, and is a vector density at the point $\alpha(s)$. Being linear in the electric field, it represents a (gauge-invariant) momentum observable. These configuration and momentum variables are closed under the Poisson bracket and constitute a complete[1] set of functions on the phase space [2,3] (in the sense that their gradients suffice to span the tangent space of the phase space almost everywhere). Therefore, in

[1] Actually, these variables are *over*complete. This occurs because the variables $T(\alpha)$ for different α are not all independent; partly this is an expression of the fact that there are "many more" loops α than points

an algebraic approach to quantization [3], they can be chosen as the "elementary variables" which define the "basic" operators of the quantum theory. The task of quantization is then reduced to that of finding appropriate representations of the commutator algebra of these operators which mirror the Poisson bracket algebra of $T^0(\alpha)$ and $(T^1(\alpha, s))^a$.

The obvious way to represent states is by suitable functionals $\Psi[A]_{\mathcal{G}}$ of gauge equivalence classes of connections. This is the configuration or the connection representation. The operators $\hat{T}^0$ act by multiplication and the $\hat{T}^1$ by (Lie-)derivation. Over the past three years, the mathematical problems associated with these formal constructions have been analysed in detail. Specifically, integral and differential calculus has now been developed on the space $\mathcal{A}/\mathcal{G}$ of connections modulo gauge transformations, with the result that configuration representations can now be constructed rigorously in the case when the underlying gauge group is compact (for a summary, see, for example, [4]).

There is, however, another possibility: states can also be represented as suitable functions of closed loops. This is suggested by the possibility of making a "Fourier-type" transform from the connection representation to a loop representation via

$$\psi(\alpha) := \int_{\mathcal{A}/\mathcal{G}} d\mu([A]_{\mathcal{G}}) \;\; T^0_A(\alpha)[A]_{\mathcal{G}} \;\; \Psi[A]_{\mathcal{G}}, \tag{1.2}$$

where μ is a measure on $\mathcal{A}/\mathcal{G}$. This *loop transform* was first introduced in the context of Yang-Mills theories by Gambini and Trias [1] and later but independently in the context of general relativity by Rovelli and Smolin [5]. In both cases, however, it was a heuristic technique because the measure μ was not specified and the required integration theory did not exist. Nonetheless, it has played a powerful role as a heuristic device, especially in the context of general relativity. In particular, it has suggested how one may translate various operators acting on the connection states $\Psi[A]_{\mathcal{G}}$ to operators on the loop states $\psi(\alpha)$. This in turn suggested how to "solve" the quantum diffeomorphism constraint of general relativity. In the loop representation, one can write down the general solution to the diffeomorphism constraint as a loop functional in the image of the transform, which depends only on the (generalized) knot class to which the loop belongs.

Recent mathematical developments have made such considerations rigorous in the case when the gauge group is compact. These results can be summarized as follows. In the connection representation, states are complex-valued functions on an appropriate completion $\overline{\mathcal{A}/\mathcal{G}}$ of $\mathcal{A}/\mathcal{G}$. This domain space of quantum wave functions, $\overline{\mathcal{A}/\mathcal{G}}$, carries a natural

---

x in Σ. In the algebraic quantization method discussed below, the relations among these variables have to be imposed in an appropriate fashion on the quantum algebra.

diffeomorphism-invariant measure μ_o which can be used to rigorously define the loop transform (1.2). Operators such as the $\hat{T}^0$ and the $\hat{T}^1$ can then be taken to the loop side and used in various constructions. Thus, in the case when the gauge group is compact, loop representations exist rigorously. [2]

Loop representations have several appealing features. For example, irrespective of the choice of the gauge group and the precise physics contained in the theory, quantum states arise as suitably regular functions on the loop space. That is, the domain space of quantum states in various physical theories is the same. The regularity conditions on loop states of course change from one theory to another. Nonetheless, since the domain space is "unified", mathematical techniques can be shared between the various theories. For diffeomorphism-invariant theories, one has the further advantage that the action of the diffeomorphism group is coded more easily in the loop space. In topological field theories where the connections under consideration are all flat, one encounters an interesting interplay between the quantum theory and the first homotopy group of the manifold Σ. Finally, if Σ is three-dimensional, one has an avenue to explore knot invariants via theories of connections.

It is therefore natural to ask if the loop representations can also be developed rigorously for physical theories – such as general relativity in three and four dimensions – in which the gauge group G is non-compact.

A number of difficulties arises immediately. First, if the gauge group is non-compact, the techniques [4] used to develop integration theory over $\overline{\mathcal{A}/\mathcal{G}}$ fail at a rather early stage. The problems here are not insurmountable. However, a number of new ideas are needed and in general the theory is likely to be considerably more complicated unless, as in four-dimensional general relativity, the gauge group is the complexification of a compact Lie group. It is therefore natural to first restrict oneself to a context where these difficulties do not arise. One such setting is provided by three-dimensional general relativity. Here, the connections of interest turn out to be flat and one can replace $\overline{\mathcal{A}/\mathcal{G}}$ by the moduli space of flat connections. Since these spaces are finite-dimensional, one does not have to develop the integration theory; the problems mentioned above do not arise. The moduli space has several "sectors". On the sector where the traces of holonomies T^0_α are all bounded, the transform can be defined and the loop representation can be constructed in a straightforward fashion [6]. However, it turns out that this sector does not correspond to "geometrodynamics". On the sector which does, a new difficulty arises: although the integration theory is straightforward, the traces of

[2] Note that each choice of measure defines a connection and an equivalent loop representation. (In the connection representation, the measure defines the inner product.) However, different choices of measures μ on $\overline{\mathcal{A}/\mathcal{G}}$ give rise to different representations. If the systems under consideration have an infinite number of degrees of freedom, the resulting representations are not generally expected to be unitarily equivalent. It is the underlying physics that must determine the appropriate measure and hence the appropriate representation.

holonomies T^0_α fail to be square-integrable with respect to "natural" measures, making the loop transform analogous to (1.2) ill-defined. Thus, on these physically interesting sectors, a new strategy is needed.

This problem was first pointed out by Marolf [7] who also suggested a way of tackling it in the special case where the manifold Σ is a torus T^2. The purpose of this paper is to suggest an alternative solution, which consists of suitably modifying the measure that appears in the transform. This solution is conceptually simpler in the sense that with the new measures various problems are avoided right from the beginning. There is also a technical simplification. While in the Marolf approach, one first restricts oneself to a suitable dense subspace of the connection Hilbert space, defines the transform and then extends it to the full Hilbert space, with the modified measures, the transform exists on the full Hilbert space from the outset. More importantly, in our approach *the final result is a genuine loop representation* in the sense that all states in the Hilbert space are represented as functions of loops. By contrast, in the Marolf approach the limiting states, which are not contained in the initial dense subspace, cannot in general be represented as functions of loops. This result had given rise to some concern about the utility and the role of loops in the representation of quantum states in the case when the gauge group is non-compact. Our analysis clarifies this issue and shows that "old-fashioned" loop representations, without the need of any smearing, do exist even in the geometrodynamical sector of three-dimensional gravity. Our solution does, however, have a drawback: our expressions for the $\hat{T}^1$-operators in the loop representation are more complicated. In spite of these differences, the final theory we obtain is unitarily equivalent to Marolf's for $\Sigma = T^2$. Therefore, the choice between the two strategies is primarily a matter of taste and convenience.

The outline of this paper is as follows. In Sec.2 we recall the basic structure of 2+1 gravity on space-times $M = \Sigma^g \times \mathrm{I\!R}$, with Σ^g a compact Riemann surface of genus g. Sec.3 discusses quantization in the connection representation and presents the general strategy for modifying the measure to make the loop transform well-defined. This strategy always leads to a "regular" loop representation. In the case when the sector of the moduli space of flat connections under consideration is compact and the traces of holonomies T^0_α are bounded functions, the modification of the measure is unnecessary. However, if one chose to follow this route, the resulting loop representation would be unitarily equivalent to that of [6]. In the non-compact case, on the other hand, the strategy appears to be essential to obtain a genuine loop representation. In Sec.4, the procedure is carried out in detail for the case when Σ is a two-torus. In particular, we present a family of new measures which make the loop transform well-defined and obtain the modified expressions of the $\hat{T}^0$- and $\hat{T}^1$-operators as well as the explicit form of the scalar product in the loop representation for this family. Sec.5 contains our conclusions. In the appendix we present some partial results for the genus-2 case.

2 Preliminaries

In this section, we will collect those results from the classical Hamiltonian formulation of the 2+1-dimensional (Lorentzian) general relativity that will be needed in the main part of the paper in Secs.3 and 4. The discussion will also serve to fix our notation. Note that this is not meant to be an exhaustive summary; further details may be found, for example, in [3,8,9].

Since we are interested primarily in the canonical quantization of the spatially compact case, we will restrict ourselves to three-dimensional spacetimes M of the form $M = \Sigma^g \times \mathbb{R}$, where Σ^g is a two-dimensional compact Riemann surface of genus g. In the connection dynamics version, the Hamiltonian formulation of the theory leads to two sets of first-class constraints [8,6,3], one linear in momenta and the other independent of momenta. Consequently, the Dirac and the reduced phase space quantization methods lead to equivalent quantum theories. For convenience of presentation, we will use the reduced phase space method here. Because of the simplicity of the constraints, the reduced phase space is a cotangent bundle over a reduced configuration space, which in turn is just the moduli space $\mathcal{A}^F/\mathcal{G}$ of flat $SU(1,1)$-connections on Σ^g. We will first recall relevant facts about $\mathcal{A}^F/\mathcal{G}$ and then go on to discuss the structure of the reduced phase space $T^\star(\mathcal{A}^F/\mathcal{G})$.

A flat $SU(1,1)$-connection A on Σ^g is determined by the values of the $2g$ holonomies U_i, $i = 1 \ldots 2g$, around representatives α_i of the $2g$ homotopy generators $\{\alpha_i\}$, $i = 1 \ldots 2g$, on Σ^g. Explicitly, we have

$$U_i = \mathrm{P} \exp \oint_{\alpha_i} A. \tag{2.1}$$

Without loss of generality we will assume these representatives to be based at a fixed point $p \in \Sigma^g$ and evaluate the holonomies U_i at p. Because of the fundamental relation between the generators of the homotopy group, these holonomies are subject to the condition

$$U_1 U_2 U_1^{-1} U_2^{-1} U_3 U_4 U_3^{-1} U_4^{-1} \ldots U_{2g-1} U_{2g} U_{2g-1}^{-1} U_{2g}^{-1} = \mathbb{1}. \tag{2.2}$$

For computational purposes it is often useful to choose an explicit parametrization for the $SU(1,1)$-matrices. We will set

$$U_\alpha = \begin{pmatrix} \alpha_1 + i\alpha_2 & \alpha_3 + i\alpha_4 \\ \alpha_3 - i\alpha_4 & \alpha_1 - i\alpha_2 \end{pmatrix}, \tag{2.3}$$

with real parameters $\alpha_1, ..., \alpha_4$, subject to the condition $\alpha_1^2 + \alpha_2^2 - \alpha_3^2 - \alpha_4^2 = 1$. Our task is to construct the moduli space of flat connections, i.e. to determine the structure of the orbit space $\mathcal{A}^F/\mathcal{G}$. For this, we note that $\mathcal{G}$ now acts on the holonomies $(U_1, ..., U_{2g})$ at the base point p by the adjoint action of $U(p) \in SU(1,1)$ according to

$$(U_1, U_2, \ldots, U_{2g}) \to U(p) \cdot (U_1, U_2, \ldots, U_{2g}) \cdot U(p)^{-1}. \tag{2.4}$$

The moduli space is therefore obtained by incorporating (2.2) and taking the quotient with respect to (2.4).

A key point for us is that the moduli space contains a finite number of components, often referred to as "sectors". (For more precise statements, see, for example, [9].) This comes about because the isotropy group I of a holonomy U_i, i.e. the subgroup of $SU(1,1)$ leaving fixed a particular U_i under the adjoint action, is not universal but depends on (certain properties of) U_i. Let us associate a 3-vector $\vec{\alpha}_\perp := (\alpha_2, \alpha_3, \alpha_4)$, with each holonomy matrix U_α, and define its norm as

$$||\vec{\alpha}_\perp|| := \alpha_2^2 - \alpha_3^2 - \alpha_4^2. \tag{2.5}$$

We can then distinguish the following cases[3]:

i) $||\vec{\alpha}_\perp|| > 0$, i.e. $||\vec{\alpha}_\perp||$ is a timelike vector $\Longrightarrow I = SO(2)$, e.g. for $\alpha = (1, 0, 0)$,

$$I = \{ \begin{pmatrix} e^{i\omega} & 0 \\ 0 & e^{-i\omega} \end{pmatrix}, \ \omega \in [0, 2\pi] \}; \tag{2.6}$$

A general timelike vector can be obtained by conjugation of U_α with an arbitrary group element, in which case the isotropy group I likewise changes by conjugation. The same remark applies to the cases below.

ii) $||\vec{\alpha}_\perp|| < 0$, i.e. $||\vec{\alpha}_\perp||$ is a spacelike vector $\Longrightarrow I = \mathbb{R} \times \mathbb{Z}_2$, e.g. for $\alpha = (0, 1, 0)$,

$$I = \{ \begin{pmatrix} \epsilon \cosh \omega & \sinh \omega \\ \sinh \omega & \epsilon \cosh \omega \end{pmatrix}, \ \omega \in \mathbb{R}, \epsilon \in \mathbb{Z}_2 \}, \tag{2.7}$$

[3] This classification is equivalent to that of [9] which is based on the Lie algebra elements generating $U_i \in SU(1,1)$.

with group law $(\epsilon_1, \omega_1) \cdot (\epsilon_2, \omega_2) = (\epsilon_1 \epsilon_2, \epsilon_1 \omega_1 + \epsilon_2 \omega_2)$;

iii) $||\vec{\alpha}_\perp|| = 0$, but not $\vec{\alpha}_\perp = 0$, i.e. $||\vec{\alpha}_\perp||$ is a non-vanishing null vector $\Longrightarrow I = \mathbb{R} \times \mathbb{Z}_2$, e.g. for $\alpha = (1, 1, 0)$,

$$I = \{ \begin{pmatrix} \epsilon + i\omega & \omega \\ \omega & \epsilon - i\omega \end{pmatrix}, \ \omega \in \mathbb{R}, \epsilon \in \mathbb{Z}_2 \}, \tag{2.8}$$

with group law $(\epsilon_1, \omega_1) \cdot (\epsilon_2, \omega_2) = (\epsilon_1 \epsilon_2, \epsilon_1 \omega_2 + \epsilon_2 \omega_1)$. – The isotropy group of the null vector $\vec{\alpha}_\perp = 0$ is of course the entire group.

In the sector that corresponds to geometrodynamics, all the U_i are boosts [10], corresponding to case ii) above. It turns out that this sector is identical with the well-known Teichmüller space $\mathcal{T}(\Sigma^g)$ associated with the Riemann surface Σ^g. This is a finite-dimensional space diffeomorphic to $\mathbb{R}^{6g-6}$ for $g > 1$ and R^2 for $g = 1$. Thus, the geometrodynamic sector of the reduced configuration space – on which we will focus from now on – is precisely the Teichmüller space $\mathcal{T}(\Sigma^g)$.

Let us now turn to the reduced phase space $T^*\mathcal{T}(\Sigma^g)$. We will give an explicit description of the moduli space of flat connections in terms of gauge-invariant loop variables.

We begin by recalling [6,11] that there is an (over)complete set of Dirac observables on the phase space $T^*(\mathcal{A}^F/\mathcal{G})$, given by the loop variables

$$\begin{aligned} T^0(\alpha)[A] &= \frac{1}{2} Tr\, U_\alpha \\ T^1(\alpha)[A, E] &= \oint_\alpha d\alpha^a\, \eta_{ab}\, Tr\,(E^b U_\alpha), \end{aligned} \tag{2.9}$$

where the canonical pairs $([A], [E])$ coordinatize the cotangent bundle $T^*\mathcal{A}^F$ of the space $\mathcal{A}^F$ of flat connections, $U_\alpha = \mathrm{P} \exp \oint_\alpha A$ is the holonomy around the loop α evaluated at the base point, and where η_{ab} is the totally anti-symmetric Levi-Civita density on Σ^g. (Note that we have exploited the fact that Σ^g is two-dimensional to integrate out $(T^1(\alpha, s))^a$ in (1.1) over the loop α to obtain a momentum observable $T^1(\alpha)[A, E]$; the vector density index a and the dependence on the marked point $\alpha(s)$ are lost in the integration.) Since all connections under consideration are flat, the variables (2.10) depend only on the homotopy of the loop α, and we may substitute α by its corresponding homotopy group element.

The loop variables T^I, $I = 0, 1$, form a closed Poisson algebra with respect to both the canonical symplectic structure on $T^*\mathcal{A}^F$, and to the induced symplectic structure on the reduced, physical phase space $T^*(\mathcal{A}^F/\mathcal{G})$. The algebra is given by

$$\begin{aligned}
&\{T^0(\alpha), T^0(\beta)\} = 0\\
&\{T^0(\alpha), T^1(\beta)\} = -\frac{1}{2}\sum_i \Delta_i(\alpha,\beta)\left(T^0(\alpha\circ_i\beta) - T^0(\alpha\circ_i\beta^{-1})\right)\\
&\{T^1(\alpha), T^1(\beta)\} = -\frac{1}{2}\sum_i \Delta_i(\alpha,\beta)\left(T^1(\alpha\circ_i\beta) - T^1(\alpha\circ_i\beta^{-1})\right)
\end{aligned} \tag{2.10}$$

.

The sums in (2.10) are over all points i of intersection of the loops α and β, with $\Delta_i(\alpha,\beta) = 1$ ($= -1$) if the two tangent vectors $(\dot\alpha, \dot\beta)$ form a right- (left-)handed dyad at i and zero if the tangent vectors are parallel. The algebra (2.10) is independent of the representatives chosen in the homotopy classes $\{\alpha\}$ and $\{\beta\}$, and the representatives can be chosen to originate and intersect at a fixed base point $p \in \Sigma$. For this reason we will from now on identify the loop composition $\circ_i$ with the group multiplication $\circ$ in $\pi_1(\Sigma^g)$.

Because of the identities that hold among the traces of 2×2-matrices, the T^I are not all independent. They are subject to the following algebraic relations:

$$\begin{aligned}
T^0(\alpha)T^0(\beta) &= \frac{1}{2}\left(T^0(\alpha\circ\beta) + T^0(\alpha\circ\beta^{-1})\right)\\
T^0(\alpha)T^1(\beta) + T^0(\beta)T^1(\alpha) &= \frac{1}{2}\left(T^1(\alpha\circ\beta) + T^1(\alpha\circ\beta^{-1})\right).
\end{aligned} \tag{2.11}$$

For a general gauge group, relations of this type are also known as Mandelstam constraints. Finally, as an aside, note that the norm (2.5) of $\vec\alpha_\perp$ is expressible in terms of the loop variable $T^0(\alpha)$ introduced in (2.9) as

$$||\vec\alpha_\perp|| = \ 1 - T^0(\alpha)^2, \tag{2.12}$$

which shows that the classification into timelike, spacelike and null rotations made earlier is gauge-independent.

Since the reduced configuration space, the Teichmüller space $\mathcal{T}$, is topologically trivial, one can attempt to find a global chart on it using the T^0-functions. To achieve this, it is necessary to eliminate the redundancy inherent in the Mandelstam constraints (2.11) and to re-express and solve the condition (2.2) on holonomies as conditions on the *traces* of the holonomies of the $2g$ homotopy generators.

The overcompleteness of the T^0-variables has already been discussed in a related case, namely that of $SU(2)$-holonomy variables of a lattice gauge theory [11]. In so far as the arguments there were based on the existence of constraints of the form (2.11) (which are the same for both $SU(2)$ and $SU(1,1)$ in their two-dimensional representations), they are equally valid in the present setting. Let us summarize: given a set of n basic loops α_i (here the $2g$ homotopy generators) and their associated holonomy matrices U_{α_i}, any gauge-invariant quantity $T^0(\gamma)$, where γ is a loop composed of the basic loops, can be expressed as an algebraic function of the variables

$$\begin{aligned} L_1(\alpha_i) &:= T^0(\alpha_i) \\ L_2(\alpha_i, \alpha_j) &:= \frac{1}{2}(T^0(\alpha_i \circ \alpha_j^{-1}) - T^0(\alpha_i \circ \alpha_j)). \end{aligned} \tag{2.13}$$

This reduces the number of loop variables to $n + \frac{n(n-1)}{2}$. A further reduction is provided by the following procedure. Fix two loops (which can be thought of as "projectors" in the Lie algebra [12]), say, α_1 and α_2. Then any point in the space $\mathcal{A}/\mathcal{G}$ can be described locally by the $3n-3$ variables

$$\begin{aligned} &L_1(\alpha_i), \quad i = 1, \ldots, n \\ &L_2(\alpha_1, \alpha_i), \quad i = 2, \ldots, n \\ &L_2(\alpha_2, \alpha_i), \quad i = 3, \ldots, n. \end{aligned} \tag{2.14}$$

For $2+1$-gravity this leaves us with a set of $6g-3$ variables to describe the space $\mathcal{A}^F/\mathcal{G}$. The fundamental relation (2.2) yields three additional constraints on the variables (2.14), unless $g = 1$, in which case one obtains only one additional condition. Thus we end up with $6g-6$ basic loop variables for $g \geq 2$ and 2 basic loop variables for $g = 1$, coinciding with the dimensionality of the Teichmüller spaces. Since moreover each space $\mathcal{T}(\Sigma^g)$ is contractible, there are no obstructions in principle to finding sets of loop variables that constitute a good global chart on it. Still those loop variables may not independently assume arbitrary values on the real line, due to the existence of inequalities among the variables (2.14) [13]. It is fairly straightforward to explicitly identify the true physical degrees of freedom in this manner. For general higher genus, one may follow the strategy for the genus-2 case discussed in the appendix.

We will conclude this discussion by giving the general form of a $SU(1,1)$-rotation U in the space-like sector in terms of the exponentiated connection components A_i, $i = 2,3,4$, which will be useful later (for simplicity, we consider only the part connected to the identity, where we have $T^0(\alpha) \geq 1$, $\forall \alpha$):

$$U_\alpha = \exp\Big(A_2 \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + A_3 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + A_4 \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \Big)$$
$$= \begin{pmatrix} \cosh A + i\frac{A_2}{A}\sinh A & \frac{A_3+iA_4}{A}\sinh A \\ \frac{A_3-iA_4}{A}\sinh A & \cosh A - i\frac{A_2}{A}\sinh A \end{pmatrix}, \tag{2.15}$$

where $A := \sqrt{-A_2^2 + A_3^2 + A_4^2}$. Because of the condition $||\vec{\alpha}_\perp|| < 0$, the square root is always well-defined. Given a second holonomy matrix U_β, obtained by exponentiating a connection with components B_i, one can compute the explicit expressions for the loop variables (2.14),

$$\begin{aligned} L_1(\alpha) &= \cosh A \\ L_1(\beta) &= \cosh B \\ L_2(\alpha,\beta) &= \frac{1}{AB}(A_2B_2 - A_3B_3 - A_4B_4)\sinh A \sinh B. \end{aligned} \tag{2.16}$$

3 Quantization

This section is divided into two parts. In the first, we recall quantization in the connection representation. This discussion will facilitate the introduction of the loop transform and also serve to bring out some subleties. In the second, we first point out the difficulties associated with the loop transform and then sketch our proposal for overcoming them. In the next section, this strategy is carried out in detail for the genus-one case .

3.1 Connection Representation

By a quantization of $2+1$-gravity we shall mean a representation of the algebra (2.10) as the commutator algebra of self-adjoint operators $\hat{T}^I$ on some Hilbert space, such that the conditions (2.11) hold (the products of operators on the left side being replaced by their anti-commutators).

In the connection representation, we can proceed as one normally does when dealing with quantum mechanics of systems whose configuration space is a manifold. We can take the states to be densities $\tilde{\Psi}[A]_{\mathcal{G}}$ of weight one-half on the reduced configuration space $\mathcal{T}$, and let the scalar product be the obvious one:

$$< \tilde{\Psi}|\tilde{\Phi} >:= \int_{\mathcal{T}} \overline{\tilde{\Psi}}[A]_{\mathcal{G}}\ \tilde{\Phi}[A]_{\mathcal{G}}, \tag{3.1}$$

where a choice of a volume element on $\mathcal{T}$ is not necessary because the integrand is a density of weight one. Since the T^0 are configuration variables, they are represented by multiplication operators. Similarly, being momentum variables, the T^1 are represented by Lie derivatives,

$$\begin{aligned}(\hat{T}^0_\alpha \circ \Psi)[A]_{\mathcal{G}} &= T^0_\alpha[A]_{\mathcal{G}} \cdot \Psi[A]_{\mathcal{G}} \\ (\hat{T}^1_\alpha \circ \Psi)[A]_{\mathcal{G}} &= \frac{\hbar}{i}\mathcal{L}_{X(\alpha)}\Psi[A]_{\mathcal{G}},\end{aligned} \tag{3.2}$$

where $X(\alpha)$ is the vector field on the reduced configuration space $\mathcal{T}$ which defines the classical momentum variable $T^1(\alpha)$. (Recall that T^1_α, being linear in momentum, is a contraction of $[E]$ with a vector field on $\mathcal{T}$. Thus, $X(\alpha)$ is the projection to $\mathcal{T}$ of the vector field δA on $\mathcal{A}^F$, defined by $\delta A = \oint_\alpha d\alpha^a\, \eta_{ab}\ \delta^2(x, \alpha(s))\ \mathrm{Tr}\, U_\alpha(s)(\delta/\delta A_b(x))$.) It is straightforward to verify that with this definition the $\hat{T}^I_\alpha$ satisfy the canonical commutation relations that arise from (2.10) and the algebraic conditions coming from (2.11).

In this description, the states naturally arose as scalar densities. Given a volume element, one can "de-densitize" them and represent them as the more familiar wave functions. This can be achieved using *any* volume element dV on $\mathcal{T}$. Furthermore, for any dV, the action of the operators $\hat{T}^I$ can be translated in a canonical fashion to $L^2(\mathcal{T}, dV)$.

It turns out that the Teichmüller space $\mathcal{T}$ admits a natural volume element. To see this, we first note that there is a natural symplectic structure Ω on the space of connections $\mathcal{A}$ given by:

$$\Omega(\delta A, \delta A') = \int_\Sigma d^2x\ \eta^{ab}(\delta A)^i_a(\delta A')_{bi}, \tag{3.3}$$

for any two tangent vectors δA, $\delta A'$ to $\mathcal{A}$. This form Ω can be pulled back to the space $\mathcal{A}^F$ of flat connections. Being gauge-invariant, the pull-back in turn projects down to Teichmüller space, where it coincides with the well-known Weil-Petersson symplectic form[4] ω. This gives rise to a natural volume element dV on $\mathcal{T}$, namely, the $(3g-3)$-fold exterior product $dV_o = \omega \wedge \ldots \wedge \omega$. It is then natural to represent states as complex-valued functions $\Psi[A]_{\mathcal{G}}$ and define the inner product as

[4] A global coordinate chart on Teichmüller space which is convenient for our purposes is given by the Fenchel-Nielsen coordinates (see, for example, [14]). For $g > 1$, they are a set of length and angle coordinates $[l_i, \tau_i]$, $i = 1 \ldots 3g-3$, associated with a pants decomposition of the Riemann surface along a set of $3g-3$ minimal geodesics. Moreover, they are canonical coordinates for the symplectic form ω, i.e. $\omega = \sum_{i=1}^{3g-3} dl_i \wedge d\tau_i$.

$$< \Psi|\Phi >:= \int_{\mathcal{T}} dV_o \; \overline{\Psi}[A]_{\mathcal{G}} \; \Phi[A]_{\mathcal{G}}. \tag{3.4}$$

On this Hilbert space, the $\hat{T}^0_\alpha$ can be represented as before by multiplication operators. They are densely defined and symmetric. Normally, the definition of the momentum operators $\hat{T}^1(\alpha)$ would require a modification: if the Lie derivative of the volume element with respect to the vector fields $X(\alpha)$ does not vanish, we would have to add a multiple of the divergence of the vector field to ensure that the resulting operator is symmetric. However, it turns out [15] that the vector fields $X(\alpha)$ are in fact the Hamiltonian vector fields on $\mathcal{A}^F$ for the symplectic structure ω, where the Hamiltonians are simply the functions T^0_α on $\mathcal{T}$. Hence, in particular, the Lie derivative of the Liouville volume element dV_o with respect to any vector field $X(\alpha)$ vanishes. Therefore, we can continue to represent $\hat{T}^1(\alpha)$ simply by the Lie derivative. Thus, the representation of the basic operators is the same as in (3.2), although the states are now wave functions $\Psi[A]_{\mathcal{G}}$ on $\mathcal{T}$ rather than half-densities $\tilde{\Psi}[A]_{\mathcal{G}}$. This formulation of the connection representation will constitute the starting point for the discussion of the loop transform in the next subsection.

3.2 The loop transform

We are now ready to construct the loop representation. The key idea is to define this representation through a loop transform of the type (1.2). In the present case, using the fact that the reduced configuration space can be identified with the Teichmüller space $\mathcal{T}$, we can simplify the transform to

$$\psi(\alpha) = \int_{\mathcal{T}} dV \; T^0(\alpha)[l,\tau]\Psi[l,\tau], \tag{3.5}$$

where dV is a volume element on $\mathcal{T}$ and we have used the Fenchel-Nielsen coordinates [14] to parametrize $\mathcal{T}$ globally. Thus, just as in the general context of Sec.1 the transform needed a measure on $\overline{\mathcal{A}/\mathcal{G}}$, the transform now requires the introduction of a volume element on $\mathcal{T}$. Could we not have avoided the introduction of this ad-hoc structure? After all, the transform has the form of an inner product of $T^0(\alpha)$ with a wave function of connections Ψ and the connection representation *could* be constructed intrinsically (i.e. without any additional structure such as the volume element) if the states were represented by densities $\tilde{\Psi}$ of weight one-half on $\mathcal{T}$. Unfortunately, even if we replaced the $\Psi[A]$ in (3.5) by $\tilde{\Psi}$, *because the integral kernel of the transform,* $T^0(\alpha)$, *is a function* rather than a density of weight one-half, we would still need an additional structure (say, a fiducial density of weight one-half) to make the integral well-defined. Thus, while the connection representation itself does not require the choice of a volume element, the passage to the loop representation does[5].

We saw at the end of Sec.3.1 that there *is* a natural volume element dV_o on $\mathcal{T}$ which arises from the Weil-Petersson symplectic form. Therefore, a simple solution to the problem would be to just choose this dV_o for the required volume element in (3.5). This strategy would work if we were interested in the "time-like" sector of 2+1 gravity [6]. However, as noted in Sec.2, in this paper we are interested in the "space-like" sector which corresponds to geometrodynamics. It is this choice that led us to take the Teichmüller space $\mathcal{T}$ as the reduced configuration space. Now, $\mathcal{T}$ is non-compact and in general the $T^0(\alpha)$, being unbounded, fail to be square-integrable on $(\mathcal{T}, dV_o)$. Hence, for a general quantum state Ψ in the connection representation – which belongs to $L^2(\mathcal{T}, dV_o)$ – the integral in the transform would not be well-defined. In the general setting considered here, this is the problem that was first noted by Marolf [7] in the explicit context of the torus topology for Σ.

A way out would be to first restrict the transform to a dense subspace D of $L^2(\mathcal{T}, dV_o)$ – such as the one spanned by the smooth wave functions $\Psi[l,\tau]$ of compact support – on which the transform *is* well-defined, obtain the loop states and then take the Cauchy completion of this space. While this procedure appears to be simple at first sight, a detailed examination [7] shows that there are two key problems. First, for the loop representation to exist, the dense space D has to satisfy *three* conditions: i) the integral on the right of (3.5) must be well-defined for all Ψ in D; ii) on D, the transform should be faithful; and iii) D should remain invariant under the action of the $\hat{T}^I$-operators. Although one does expect such dense subspaces D to exist, already in torus case it is a quite non-trivial problem to find them. The second and conceptually more important problem is that when one takes the Cauchy completion of the image of D, one finds that it admits states which cannot, in a natural way, be represented as functions of loops. Consequently, the sense in which such a representation can be considered a "loop representation" becomes rather obscure.

Our proposal therefore is to try a new strategy. The key idea is to exploit the freedom in the choice of the volume element dV. Since we are regarding $\mathcal{T}$ as a manifold, volume elements correspond to $(6g-6)$-forms on $\mathcal{T}$. Hence, any two are related by a (suitably regular) function. Thus, we can set $dV = m[l,\tau]dV_o$ for some non-negative, smooth function $m[l,\tau]$. Following the terminology common in physics, we will refer to $m[l,\tau]$ as a "measure". The idea therefore is to choose an appropriately damped measure to make the loop transform well-defined.

---

[5] One might imagine defining the transform intrinsically by using densities of weight one (rather than one-half) as connection states. However, one would still need a volume element to decide which of these densities of weight one are normalizable, i.e. qualify to feature in the transform in the first place.

What conditions does $m[l,\tau]$ have to satisfy? First, as already noted, it should ensure a sufficient damping so that the loop transform is well-defined. More precisely, we will require that $m[l,\tau]$ be such that the traces of holonomies $T^0(\alpha)$ are in $L^2(\mathcal{T}, dV = m\, dV_o)$. Then, if we *define* the connection representation using dV – which we are free to do – we will be led to a well-defined loop transform. However, we also need the action of the operators $\hat{T}^I(\alpha)$ on loop states to be well-defined and manageable. For a general measure $m[l,\tau]$, the action of the $\hat{T}^I(\alpha)$ reduces to

$$
\begin{aligned}
(\hat{T}^0(\alpha)\Psi)[l,\tau] &= T^0(\alpha)[l,\tau]\Psi[l,\tau] \\
(\hat{T}^1(\alpha)\Psi)[l,\tau] &= -i\hbar(\mathcal{L}_{X(\alpha)} + \frac{1}{2}\mathcal{L}_{X(\alpha)} \ln m)\Psi[l,\tau],
\end{aligned}
\tag{3.6}
$$

where, as before, $\mathcal{L}_{X(\alpha)}$ is the Lie derivative along the vector field $X(\alpha)$ corresponding to the loop momentum observable $T^1(\alpha)$, and where the second term on the right-hand side of the relation for $\hat{T}^1$ compensates for the fact that the Lie derivative of the volume element $dV = m\, dV_o$ with respect to $X(\alpha)$ may not vanish.

These operator actions can be translated into the loop representation via the transform (3.5). In order that the action of the resulting operators be manageable – so that in the final picture the loop representation can exist in its own right – it is necessary that the term $\hat{T}^I(\beta)T^0(\alpha)$ in

$$
\begin{aligned}
\left(\hat{T}^I(\beta)\psi\right)(\alpha) &:= \int_{\mathcal{T}} dV\, m[l,\tau]\, T^0(\alpha)\left(\hat{T}^I(\beta)\Psi[l,\tau]\right) \\
&= \int_{\mathcal{T}} dV\, m[l,\tau]\, \left(\hat{T}^I(\beta)T^0(\alpha)\right)\Psi[l,\tau]
\end{aligned}
\tag{3.7}
$$

be expressible as some linear combination $\sum a_j T^0(\alpha_j)$ (where α_j denotes any homotopy group element). For $\hat{T}^I = \hat{T}^0$ this condition is automatically fulfilled, thanks to the algebraic relation (2.11). For $I = 1$, the term under consideration takes the form

$$
(\mathcal{L}_{X_\beta} + \frac{1}{2}\mathcal{L}_{X_\beta} \ln m)T^0(\alpha) = \{T^1(\beta), T^0(\alpha)\} + \frac{1}{2}\{T^1(\beta), \ln m\}T^0(\alpha),
\tag{3.8}
$$

where the curly brackets denote the Poisson brackets on phase space. The first term on the right-hand side is already in the required form, again due to the algebraic relations (2.11). Next, it follows from (2.10) and (2.11) that the second term, $\{T^1(\beta), \ln m\}$, would have the desired form if it were expressible as a linear combination of the T^0. Thus, the action of $\hat{T}^1$ would be manageable on loop states if the measure $m[l,\tau]$ were of the form

$$m = \exp\Big(\sum_i b_i\, T^0(\alpha_i)\Big), \tag{3.9}$$

for some fixed real constants b_i and some fixed homotopy generators α_i. (Note that we could also have chosen to use for $m[l,\tau]$ the exponential of a product of $T^0(\alpha_i)$ since the product can always be re-expressed as a sum, using (2.11).)

Our strategy is therefore to use a measure $m[l,\tau]$ of the form (3.9) (both in the definition of the connection representation and) in the definition of the loop transform. The key question is whether one can choose a finite number of b_i and α_i such that the measure damps sufficiently fast for the transform (3.5) to be well-defined for *any* element α of the homotopy group. At a heuristic level, it would seem that the freedom in the choice of α_i is so large that it should be easy to meet this damping condition. However, because we do not have sufficient control over the behaviour of traces of holonomies $T^0(\alpha)$ on the Teichmüller spaces of higher genus, we have been able to explicitly demonstrate the existence of the measures of the required type only in the $g=1$ case. However, if b_i and α_i can be chosen to ensure the existence of the integral in (3.5) for all α, the existence of a loop representation with the required properties *is* ensured. In particular, all normalizable states in such a representation would arise as functions of loops; in contrast to [7], generalized loops would not be necessary.

4 The torus case

Let us briefly review the explicit structure of the reduced phase space $T^*\mathcal{T}$ in the case when the two-manifold Σ is a torus. (For further details, see [3,7].) We are interested in the sector where the holonomies U_i, $i=1,2$ of both homotopy generators (a_1,a_2) are rotations about spacelike axes in the three-dimensional Minkowski space. The fundamental relation (2.2) implies that U_1 and U_2 commute and are therefore rotations about the same axis. Without loss of generality we may choose this axis to lie along the vector $(0,1,0)$, which corresponds to setting $A_2 = A_4 = 0$ in the holonomy matrix (2.15). The reduced configuration space is therefore the two-dimensional space $\mathcal{T}$ of flat connections on a torus T^2 and can be parametrized by $\vec{a} \in \mathbb{R}^2$, with opposite signs identified ($\vec{a} \sim -\vec{a}$), i.e. $\mathcal{T} = \mathbb{R}^2/\mathbb{Z}_2$. The corresponding reduced phase space is its cotangent bundle, parametrized globally by the canonical variable pair $(\vec{a},\vec{p})$.

The loop variables $T^i(\alpha)$ depend only on the homotopy class $\{\alpha\}$ of the loop α, which for $\Sigma = T^2$ can be labelled by two integers $\vec{n}$, characterizing the decomposition $\{\alpha\} = n_1\{\alpha_1\} + n_2\{\alpha_2\}$. The variables T^0 and their associated momentum variables T^1 form an overcomplete set of observables on phase space. Their explicit form is

$$
\begin{aligned}
T^0(\vec{k})[\vec{a}] &= \cosh(\vec{k}\cdot\vec{a}) \\
T^1(\vec{k})[\vec{a},\vec{p}] &= \sinh(\vec{k}\cdot\vec{a})\,\vec{k}\times\vec{p},
\end{aligned}
\tag{4.1}
$$

where $\vec{k}\times\vec{p} = k_1 p_2 - k_2 p_1$. Their Poisson algebra can be written down explicitly:

$$
\begin{aligned}
\{T^0(\vec{m}), T^0(\vec{n})\} &= 0 \\
\{T^0(\vec{m}), T^1(\vec{n})\} &= -\frac{1}{2}\,(\vec{m}\times\vec{n})\,\Big(T^0(\vec{m}+\vec{n}) - T^0(\vec{m}-\vec{n})\Big) \\
\{T^1(\vec{m}), T^1(\vec{n})\} &= -\frac{1}{2}\,(\vec{m}\times\vec{n})\,\Big(T^1(\vec{m}+\vec{n}) - T^1(\vec{m}-\vec{n})\Big).
\end{aligned}
\tag{4.2}
$$

The overcompleteness of these variables is due to the Mandelstam constraints (2.11), which now simplify to

$$
\begin{aligned}
T^0(\vec{m})T^0(\vec{n}) &= \frac{1}{2}\Big(T^0(\vec{m}+\vec{n}) + T^0(\vec{m}-\vec{n})\Big) \\
T^0(\vec{m})T^1(\vec{n}) + T^0(\vec{n})T^1(\vec{m}) &= \frac{1}{2}\Big(T^1(\vec{m}+\vec{n}) + T^1(\vec{m}-\vec{n})\Big).
\end{aligned}
\tag{4.3}
$$

Note, as an aside, that in the "time-like" sector where the moduli space of connections is compact, the relevant loop observables are obtained from (4.1) by substituting the hyperbolic functions with the corresponding trigonometric functions, and the relation (4.2) and (4.3) remain the same.

For the two generators of the homotopy group $\pi_1(T^2)$, $\alpha_1 = (1,0)$ and $\alpha_2 = (0,1)$, one finds

$$
\begin{aligned}
T^0(\alpha_1)[\vec{a}] &= \cosh a_1 \\
T^0(\alpha_2)[\vec{a}] &= \cosh a_2 \\
L_2(\alpha_1,\alpha_2)[\vec{a}] &= -\sinh a_1 \sinh a_2.
\end{aligned}
\tag{4.4}
$$

It follows that $L_2(\alpha_1,\alpha_2)$ together with one of $T^0(\alpha_1)$, $T^0(\alpha_2)$ parametrize $\mathcal{T}$ globally and would therefore constitute a good choice of independent loop variables in terms of which all other $T^0(\alpha)$ can be expressed. At first sight, the easiest choice for an independent set may seem to take $T^0(\alpha_1)$ and $T^0(\alpha_2)$. However, they do not form a good global chart (a similar statement holds for the genus-2 case, see the appendix).

In the coordinates a_i on $\mathcal{T}$, the natural Liouville volume element is simply $dV_o = da_1 da_2$. Our objective is to choose an appropriate measure $m(a_i) = \exp -M(a_i)$ such that the transform and the resulting loop representation are well-defined. To follow the strategy outlined in Sec.3.2, let us begin by writing the analogs of (3.5) -(3.9) explicitly. First, we have

$$\psi(\vec{n}) = < T^0(\vec{n}), \Psi > = \int_0^\infty da_1 \int_0^\infty da_2 \; e^{-M(\vec{a})} \, T^0(\vec{n}) \; \Psi(\vec{a}), \tag{4.5}$$

using the scalar product notation. (Note that since the $T^0(\vec{n})$ are all real, complex conjugation is unnecessary in the integral.) According to our reasoning in Sec.3, the function M on $\mathcal{T}$ should have the form $M = \sum b_i T^0(\alpha_i)$ (c.f. (3.9)), where the b_i and the α_i are such that holonomies around arbitrary loops are square-integrable with respect to $dV = (\exp -M) dV_o$. In the case of the torus, an obvious choice is

$$M = c\Big(T^0(\vec{q}_1) + T^0(\vec{q}_2)\Big), \tag{4.6}$$

where $c \in \mathbb{R}$, $c > 0$, and $\vec{q}_1$ and $\vec{q}_2$ are linearly independent homotopy classes. It turns out that for *any* positive c and any $\vec{q_1}$, $\vec{q}_2$, (4.6) leads to a loop representation with all the desired properties discussed in Sec.3. More precisely, we have the following:

1. It is straightforward to verify that every $T^0(\vec{k})$ of (4.1) belongs to $L^2(\mathcal{T}, dV)$. Choosing this Hilbert space as the space of states in the connection representation, it follows that the transform, being simply the inner product, is a well-defined, continuous map from the connection states to loop states. In particular, while in the connection representation the states Ψ are equivalence classes of functions on $\mathcal{T}$ (where two are equivalent if they differ by a set of measure zero), their images $\psi(\alpha)$ in the loop representation are genuine functions of homotopy classes of loops.

2. The transform is *faithful*. To see this, note first that for any choice of a measure M, $L^2(\mathcal{T}, dV)$ provides an irreducible representation of the algebra of operators $\hat{T}^I$, defined by

$$\begin{aligned} \hat{T}^0(\vec{n}) &= \cosh(\vec{n} \cdot \vec{a}) \\ \hat{T}^1(\vec{n}) &= -i\hbar \sinh(\vec{n} \cdot \vec{a}) \; \vec{n} \times \Big(\frac{\partial}{\partial \vec{a}} + \frac{1}{2}\vec{\nabla} M\Big). \end{aligned} \tag{4.7}$$

   In particular, this is true for our choice (4.6). Now, suppose the transform has a kernel $\mathcal{K}$. Then, K is a closed subspace of $L^2(\mathcal{T}, dV)$ and, since (4.6) is of the type (3.9), it follows that $\mathcal{K}$ remains invariant under the action of the $\hat{T}^I$. Hence, $\mathcal{K}$ must be either the zero subspace or the full Hilbert space. It cannot be the full Hilbert space because, in particular, the elements $T^0(\alpha)$ of $L^2(\mathcal{T}, dV)$ cannot lie in the kernel for any α since its norm is positive definite. Hence $\mathcal{K}$ must contain only the zero vector.

3. It follows from our discussion of the connection representation that for any positive M, the representation (4.7) of the $\hat{T}^I$-algebra on $L^2(\mathcal{T}, dV = (\exp -M)dV_o)$ is unitarily equivalent to the representation on $L^2(\mathcal{T}, dV_o)$. Hence, in particular, it follows that the loop representations of the $\hat{T}^I$-algebra obtained by using different $c \geq 0$, and independent $\vec{q}_1$, $\vec{q}_2$ are also unitarily equivalent. In particular then, these "genuine" loop representations are unitarily equivalent to the one constructed by Marolf [7].

The transform enables us to endow the loop states with an inner product and to represent the observables $\hat{T}^I$ directly on the loop states. The action (4.7) of the $\hat{T}^I(\alpha)$-operators translates to the loop representation yielding

$$\begin{aligned}
&\Big(\hat{T}^0(\vec{k})\psi\Big)(\vec{n}) = \frac{1}{2}\Big(\psi(\vec{n}+\vec{k}) + \psi(\vec{n}-\vec{k})\Big)\\
&\Big(\hat{T}^1(\vec{k})\psi\Big)(\vec{n}) = -\frac{i\hbar}{2}(\vec{k}\times\vec{n})\Big(\psi(\vec{k}+\vec{n}) - \psi(\vec{k}-\vec{n})\Big)+\\
&+\frac{ic\hbar}{4}\sum_{i=1,2}(\vec{k}\times\vec{q}_i)\Big(\psi(\vec{k}+\vec{n}+\vec{q}_i) - \psi(\vec{k}+\vec{n}-\vec{q}_i) + \psi(\vec{k}-\vec{n}+\vec{q}_i) - \psi(\vec{k}-\vec{n}-\vec{q}_i)\Big).
\end{aligned} \tag{4.8}$$

Using these expressions, one may check explicitly that the commutation relations that result from (2.10) and the algebraic conditions that arise from (2.11) are satisfied by the $\hat{T}^I$ in this loop representation.

Finally, let us exhibit the inner products between loop states. For simplicity, let us consider the subset of loop states $\{T^0_A(\vec{k}),\, \vec{k} \in \mathbb{R}^2/\mathbb{Z}_2\} \subset L^2(\mathcal{T}, m\, dV_o)$, and, in the measure, set $\vec{q}_1 = (1,0)$ and $\vec{q}_2 = (0,1)$. The general expression for the scalar product is then

$$\begin{aligned}
< T^0(\vec{k}), T^0(\vec{n}) >=& 2\int_0^\infty da_1 \int_0^\infty da_2\, e^{-c(\cosh a_1 + \cosh a_2)} T^0(\vec{k}) T^0(\vec{n})\\
&= F(k_1+n_1)F(k_2+n_2) + F(k_1-n_1)F(k_2-n_2)+\\
&+G(k_1+n_1)G(k_2+n_2) + G(k_1-n_1)G(k_2-n_2),
\end{aligned} \tag{4.9}$$

where

$$F(0) = K_0(c)$$
$$F(1) = K_1(c)$$
$$F(n) = K_0(c)\left(2^{n-1} + n\sum_{k=1}^{\frac{n}{2}}(-1)^k\frac{1}{k}\binom{n-k-1}{k-1}2^{n-2k-1}\right) + \sum_{m=1}^{\frac{n}{2}}(2m-1)!!\frac{1}{c^m}K_m(c)\times$$
$$\times\left(2^{n-1}\binom{\frac{n}{2}}{\frac{n}{2}-m} + n\sum_{k=1}^{\frac{n}{2}-m}(-1)^k\frac{1}{k}\binom{n-k-1}{k-1}2^{n-2k-1}\binom{\frac{n}{2}-k}{\frac{n}{2}-k-m}\right),\ n\,\text{even}$$

$$F(n) = K_1(c)\left(2^{n-1} + n\sum_{k=1}^{\frac{n-1}{2}}(-1)^k\frac{1}{k}\binom{n-k-1}{k-1}2^{n-2k-1}\right) + \sum_{m=1}^{\frac{n-1}{2}}(2m-1)!!\frac{1}{c^m}K_{m+1}(c)\times$$
$$\times\left(2^{n-1}\binom{\frac{n-1}{2}}{\frac{n-1}{2}-m} + n\sum_{k=1}^{\frac{n-1}{2}-m}(-1)^k\frac{1}{k}\binom{n-k-1}{k-1}2^{n-2k-1}\binom{\frac{n-1}{2}-k}{\frac{n-1}{2}-k-m}\right),\ n\,\text{odd}$$

$$G(0) = 0$$
$$G(n) = e^{-c}\sum_{m=1}^{n}\frac{1}{c^m}\sum_{k=0}^{E(\frac{n-m}{2})}(-1)^k\binom{n-k-1}{k}2^{n-2k-1}\frac{(n-2k-1)!}{(n-2k-m)!}. \tag{4.10}$$

Here, the functions K_n are the modified Bessel functions of integer order [16]. It is important to notice that, unlike in the case of the "time-like sector" [6], the scalar product is *not* proportional to $\delta_{\vec{k},\vec{n}}$. Note also that, as expected, the above expressions diverge as $c \to 0$ (corresponding to the trivial measure $M = 0$), due to the divergence of the Bessel functions and of the terms proportional to c^{-m} in (4.14). Let us give a few special cases of (4.13) for illustration:

$$\begin{aligned}
&< T^0(0,0), T^0(0,0) >= 2K_0(c)^2\\
&< T^0(1,0), T^0(1,0) >= < T^0(0,1), T^0(0,1) >= 2K_0(c)(\frac{1}{c}K_1(c) + K_0(c))\\
&< T^0(1,0), T^0(0,0) >= < T^0(0,1), T^0(0,0) >= 2K_0(c)K_1(c)\\
&< T^0(1,0), T^0(0,1) >= 2K_1(c)^2\\
&\qquad\qquad\vdots
\end{aligned} \tag{4.11}$$

Finally we note that, although the wave functions $T^0(k,n)$ do not form an orthonormal set, since the weight function $e^{-M(\vec{a})}$ is everywhere positive on $\mathcal{T}$, there exists a sequence of polynomials $P_m(a_1)Q_n(a_2)$, where

$$
\begin{aligned}
P_m(a_1) &= \sum_{k=1}^{m} P_{mk} T^0(k,0) \\
Q_m(a_2) &= \sum_{k=1}^{m} Q_{mk} T^0(0,k),
\end{aligned}
\tag{4.12}
$$

with constants P_{mk}, Q_{mk}, such that

$$
< P_k Q_l, P_m Q_n >= \delta_{km}\delta_{ln}. \tag{4.13}
$$

Such a basis would still have a discrete labelling, but the direct geometric interpretation of the wave functions in terms of loops on T^2 would be lost.

One may object to the introduction of a non-trivial measure factor $\exp -M$ on the grounds that the trivial measure $da_1 \wedge da_2$ is distinguished by its modular invariance, i.e. invariance under the action of the modular group, whose generators act on the connection variables (a_1, a_2) according to

$$
\begin{aligned}
(a_1, a_2) &\to (a_2, -a_1) \\
(a_1, a_2) &\to (a_1, a_1 + a_2).
\end{aligned}
\tag{4.14}
$$

Modular invariance (i.e. invariance under large diffeomorphisms) however is not a physical requirement of the 2+1-theory, and its imposition leads to orbifold singularities in the reduced configuration space [17]. Although our modified measures are not modular invariant, their corresponding quantum representations still allow for a unitary implementation of the modular group. The action of the generators on wave functions is given by

$$
\begin{aligned}
\Psi(a_1, a_2) &\to e^{(M(a_1,a_2) - M(a_2,-a_1))/2}\, \Psi(a_2, -a_1) \\
\Psi(a_1, a_2) &\to e^{(M(a_1,a_2) - M(a_1,a_1+a_2))/2}\, \Psi(a_1, a_1 + a_2).
\end{aligned}
\tag{4.15}
$$

For our particular choice $M = c(T^0(1,0) + T^0(0,1))$ one obtains

$$
\begin{aligned}
\Psi(a_1, a_2) &\to \Psi(a_2, -a_1) \\
\Psi(a_1, a_2) &\to e^{\frac{c}{2}(T^0(0,1) - T^0(1,1))}\, \Psi(a_1, a_1 + a_2).
\end{aligned}
\tag{4.16}
$$

This action takes on a simple form in the loop representation only if one makes a change of basis (to one in which the operators $\hat{T}^0(\vec{k})$ are diagonal).

5 Conclusion

Let us begin with a summary of the main results. While the connection representation can be constructed without a volume element or a measure on $\mathcal{T}$, the loop transform does require this additional structure. The Teichmüller space $\mathcal{T}$ is equipped with a natural, Liouville volume element dV_o. Unfortunately, because the traces of holonomies grow unboundedly in the coordinates which are canonical for the Liouville form, the integral in the loop transform defined using dV_o is in general ill-defined. Our strategy was to exploit the freedom available in the choice of the volume element to introduce a damping factor $\exp -M$ $(M \geq 0)$ and define the transform using $dV = (\exp -M)dV_o$ instead. The transform is then well-defined if M is chosen so that all the traces of holonomies T^0_α belong to $L^2(\mathcal{T}, dV)$. The requirement that the observables $\hat{T}^I$ have manageable expressions in the loop representation further restricts the damping factor M: it has to be of the type $\sum b_i T(\alpha_i^o)$ for some real numbers b_i and homotopy generators $\{\alpha_i^o\}$. The key question then is whether the two requirements on the damping factor can be met simultaneously. In the genus $g = 1$ case, we saw that it was quite straightforward to achieve this. In the more general case, the issue remains open although the available freedom in the choice of constants and homotopy generators seems large enough to meet these conditions.

The solution we propose here does have an inelegant feature: the expressions of the operators $\hat{T}^1(\alpha)$ in the loop representation now involve not only the homotopy generator α labelling the operator but also the fiducial loops α_i^o we fixed to define the measure. Could we have avoided this by modifying the strategy slightly? For example, in the definition of the transform, $\psi(\alpha) = \int(\exp -M)dV_o T^0(\alpha)\Psi$, could we not have constructed the damping factor M from the homotopy generator α itself, without introducing any fiducial generators α_i^o? This is a tempting strategy since it avoids all references to fiducial loops. However, it does not work, essentially because the transform no longer has the form of an inner product of Ψ with $T^0(\alpha)$. More specifically, in the resulting loop representation, it is not possible to express the action of even the $\hat{T}^0$ operators in a manageable way!

A second strategy [5] would be to avoid the loop transform altogether and introduce the loop representation ab-initio. Thus, one may begin with the quantum algebra of the $\hat{T}^I$-operators and attempt to find a representation directly on a vector space of suitable functions $\psi(\alpha)$ of (homotopy classes of) loops. Unfortunately, any ansatz which avoids the introduction of the transform and reference to the connection representation faces two important problems. First, in such representations, it seems difficult to incorporate the numerous identities and inequalities satisfied by the loop variables [13]. Second, it is difficult to simply guess the class of "suitable" functions $\psi(\alpha)$ one has to begin with. In the case of 2+1 gravity on a torus, for example, it would a priori seem natural to begin with functions $\psi(\vec{n}) = \sum c_i \delta_{\vec{n},\vec{n}_i}$, obtained by taking linear combinations of characteristic functions of homotopy classes. When the quantization program [3] is completed, however, one finds that the spectrum of all the $\hat{T}^0(\vec{n})$-operators is bounded between $(-1, 1)$ while classically, on the geometrodynamic sector, the classical $T^0(\alpha)(\vec{n})$ take values precisely in the complement of this interval! That is, harmless assumptions on the initial "regularity" conditions end up having unexpected, physically important and often undesired consequences. In the example just described, the quantum theory can be constructed but it corresponds to the "time-like sector" of the moduli space of flat connections which has no geometrodynamic analog. More importantly, the "correct" regularity conditions that will finally lead one to the desired sector may be quite involved and difficult to guess because one's favourite loop states – such as the characteristic functions of homotopy classes – may not belong to the physical Hilbert space. This is the case both for the representations presented in the last section and the ones found by Marolf [7]. All these problems are avoided if one constructs the loop representation via the loop transform.

2+1 gravity is a "toy model" for the physical 3+1 theory. What lessons can one learn from its analysis? First, we found that, once appropriate care is taken in defining measures, "old fashioned" loop representations *do* exist for the 2+1 theory. Results of [7] had been used by some to question the existence and utility of the loop representation in cases when the gauge group is non-compact. Our analysis removes these objections. It does point out, however, that even in absence of infinite-dimensional, field-theoretic problems the task of choosing physically interesting sectors can be a rather delicate issue in loop representations. In the connection representation of the 2+1 theory, restricting oneself to the geometrodynamic sector was straightforward: we simply restricted the wave functions to have support in this sector. In the loop representation, by contrast, the restriction is implemented by imposing different "regularity conditions" on loop states which, in turn, lead to quite different inner products. Without recourse to the transform, it would have been hopeless to unravel this subtle intertwining between the physics of the representation and the mathematical regularity conditions. The second lesson therefore is that it would be "safer" to construct the loop representation via the transform also in the 3+1 theory. The relation between the mathematical assumptions and their physical implications would then be more transparent and the numerous identities and inequalities [13] between traces of holonomies automatically incorporated in the loop representation. This in turn suggests that the construction of suitable measures on $\mathcal{A}/\mathcal{G}$ would be a central problem in the 3+1 theory as well. This is the third lesson. The details of the required strategies in 3+1 dimensions will, however, be quite different from those that have been successful in the present analysis. The 3+1-problem is both more difficult and easier. It is more difficult because the space $\mathcal{A}/\mathcal{G}$ is now *infinite*-dimensional. However, it is also easier because the gauge group $\mathbb{C}SU(2)$ of the 3+1 theory is the complexification of a *compact* group, and states in the connection representation are *holomorphic* functions of connections. It therefore seems possible that the integration theory on $\mathcal{A}/\mathcal{G}$ developed for compact groups [4] would admit a natural generalization to this case. This approach would lead to measures of a quite different sort than the ones introduced in this paper and in particular would not refer to any fiducial loops.

Acknowledgements: We would like to thank Jorma Louko and Don Marolf for discussions and Steve Carlip for pointing out reference [14]. This work was supported in part by the NSF grant PHY93-96246, the Eberly research fund of Penn State University, and a DFG fellowship to RL.

Appendix

In this appendix we present some details on the genus-2 case. In particular, we will demonstrate that the problems encountered in the genus-1 case persist and therefore again a measure with an appropriate damping factor is needed.

The reduced configuration space for the genus-2 case is the six-dimensional Teichmüller space and may be parametrized globally by the Fenchel-Nielsen coordinates [14]. They are a set of length and angle coordinates of a pants decomposition of the genus-g surface. The surface is cut along a set of $3g$ simple geodesic loops and, at the i'th cut, $l_i \in \mathbb{R}^+$ measures the intrinsic length of the border curve and $\tau_i \in \mathbb{R}$ the relative twisting angle of the opposite sides of the cut. As was shown by Wolpert [18], the Weil-Petersson symplectic form ω in these coordinates is simply given by $\omega = \sum_{i=1}^{3g-3} dl_i \wedge d\tau_i$.

To find a suitable damping factor, we need estimates on how fast the traces of holonomies diverge on $\mathcal{T}$. Thus, we have to express the loop holonomies as functions of the $[l_i, \tau_i]$. For $g \geq 2$ this task has been carried out by Okai [19].

We will adopt the notation of [19] and denote the six Fenchel-Nielsen parameters by $(l_{-\infty}, l_0, l_\infty, \tau_{-\infty}, \tau_0, \tau_\infty)$. Next one has to choose a set of six loops on Σ such that the traces of their holonomies are independent, i.e. parametrize the Teichmüller space locally. There are clearly many different ways of doing this. The simplest choice we have found is illustrated in Fig.1:

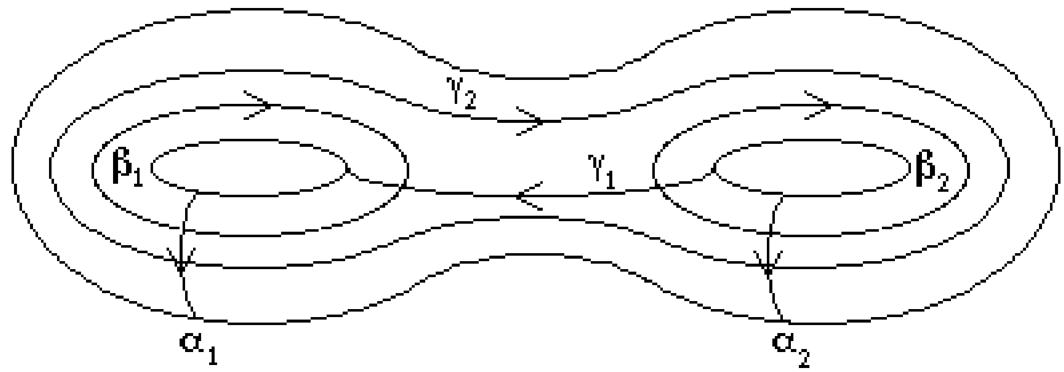

Fig.1

The loops $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ are the usual homotopy generators, and in addition we have the two loops (γ_1, γ_2). The normalized traced holonomies of these loops are the following functions of (l_i, τ_i):

$$
\begin{aligned}
L_1(\alpha_1) &= \cosh \frac{l_{-\infty}}{2} \\
L_1(\alpha_2) &= \cosh \frac{l_{\infty}}{2} \\
L_1(\gamma_1) &= \cosh \frac{l_0}{2} \\
L_1(\beta_1) &= \sinh \frac{\tau_{-\infty}}{2} \sinh \frac{\tau_0}{2} + \cosh s_{\infty} \cosh \frac{\tau_{-\infty}}{2} \cosh \frac{\tau_0}{2} \\
L_1(\beta_2) &= \sinh \frac{\tau_{\infty}}{2} \sinh \frac{\tau_0}{2} + \cosh s_{-\infty} \cosh \frac{\tau_{\infty}}{2} \cosh \frac{\tau_0}{2} \\
L_1(\gamma_2) &= \sinh \frac{\tau_{-\infty}}{2} \sinh \frac{\tau_{\infty}}{2} + \cosh s_0 \cosh \frac{\tau_{-\infty}}{2} \cosh \frac{\tau_{\infty}}{2},
\end{aligned}
\tag{A.1}
$$

where the length parameters $s_{-\infty}$, s_0 and s_{∞} are functions of the l_i alone,

$$
\begin{aligned}
\cosh s_{-\infty} &= \frac{\cosh \frac{l_{\infty}}{2} \cosh \frac{l_0}{2} + \cosh \frac{l_{-\infty}}{2}}{\sinh \frac{l_{\infty}}{2} \sinh \frac{l_0}{2}} \\
\cosh s_0 &= \frac{\cosh \frac{l_{\infty}}{2} \cosh \frac{l_{-\infty}}{2} + \cosh \frac{l_0}{2}}{\sinh \frac{l_{\infty}}{2} \sinh \frac{l_{-\infty}}{2}} \\
\cosh s_{\infty} &= \frac{\cosh \frac{l_{-\infty}}{2} \cosh \frac{l_0}{2} + \cosh \frac{l_{\infty}}{2}}{\sinh \frac{l_{-\infty}}{2} \sinh \frac{l_0}{2}}.
\end{aligned}
\tag{A.2}
$$

Note that the particular L_1's chosen in (A.1), unlike the Fenchel-Nielsen variables, are not good global coordinates on Teichmüller space, since a simultaneous sign change of the

τ_i leads to the same values for the independent L_1-variables. It is obvious from (A.1) and (A.2) that a problem similar to that encountered in the genus-1 case arises here too (and, in fact, for any higher genus), since none of the loop variables in (A.1) are square-integrable with respect to the measure $\Pi_i \, dl_i d\tau_i$. Moreover, there is now an additional problem for small l, namely, another divergence in the loop transform coming from terms like $\cosh s_\infty$. This problem also occurs in some calculations in string theory (see, for example, [20]), and may be dealt with by introducing a cut-off for small lengths.

Unfortunately, it appears difficult to extract from (A.1) and (A.2) estimates for the asymptotic growth of traces of holonomies around arbitrary elements of the homotopy group. This happens because – although possible in principle – it is in practice difficult to express arbitrary $T^0(\alpha)$ as functions of the independent set. Thus, while it is tempting to choose the damping factor simply as

$$\exp - \sum_{i=1}^{2} [L_1(\alpha_i) + L_1(\beta_i) + L_1(\gamma_i)], \qquad (A.3)$$

there is no guarantee that this damping will suffice to make the loop transform well-defined for any α in the homotopy group.

Even if one restricts the quantum wave functions to sums of tensor products of the six basic loop functions $L_1(\alpha_1), ..., L_1(\gamma_2)$, the situation is still non-trivial. Since some of those functions themselves have a complicated, coupled dependence on the Fenchel-Nielsen parameters, it is not immediately clear whether (A.3) gives a sufficient damping in all of the asymptotic regions. One could probably get more control over the measure by re-expressing it in terms of the independent variables L_1. The Jacobian of this transformation can be readily expressed as a simple function of L_2-variables (c.f. (2.13)), but again there is no straightforward way of rewriting it in terms of the independent set $\{L_1\}$. The algebraic problems encountered here are not insurmountable, however, a detailed case by case analysis appears to be necessary to obtain a complete control on the asymptotic behaviour of the traces of holonomies around arbitrary loops. Only then will one be able to construct explicit measures that make the loop transforms for higher-genus surfaces well-defined.

References

[1] Gambini, R. and Trias, A.: Gauge dynamics in the C-representation, *Nucl. Phys.* B278 (1986) 436-448

[2] Goldberg, J., Lewandowski, J. and Stornaiolo, C.: Degeneracy in loop variables, *Commun. Math. Phys.* (1992) 148, 377; Ashtekar, A. and Lewandowski, J.: Completeness of Wilson loop functionals on the moduli space of $SL(2,\mathbb{C})$ and $SU(1,1)$ connections, *Class. Quant. Grav.* (1993) L69-74.

[3] Ashtekar, A. *Lectures on Non-Perturbative Canonical Gravity*, World Scientific, Singapore, 1991

[4] Ashtekar, A., Marolf, D. and Mourão, J.: Integration on the space of connections modulo gauge transformations, in: *Proceedings of the Lanczos centenary conference*, ed. D. Brown et al., SIAM

[5] Rovelli, C. and Smolin, L.: Loop space representation of quantum general relativity, *Nucl. Phys.* B331 (1990) 80-152

[6] Ashtekar, A., Husain, V., Rovelli, C., Samuel, J., and Smolin, L.: 2+1 gravity as a toy model for the 3+1 theory, *Class. Quant. Grav.* 6 (1989) L185-193; Smolin, L.: Loop representation for quantum gravity in 2+1 dimensions, in: *Knots, topology and quantum field theory*, Proceedings of the 12th Johns Hopkins Workshop; Ashtekar, A.: Lessons from 2+1 dimensional quantum gravity, in: *Strings 90*, ed. R. Arnowitt et al., World Scientific, Singapore

[7] Marolf, D.M.: Loop representations for 2+1 gravity on a torus, *Class. Quant. Grav.* 10 (1993) 2625-2647

[8] Witten, E.: 2+1 dimensional gravity as an exactly soluble system, *Nucl. Phys.* B311 (1989) 46-78

[9] Louko, J. and Marolf, D.M.: The solution space of 2+1-gravity on $\mathbb{R} \times T^2$ in Witten's connection formulation, *Class. Quant. Grav.* 11 (1994) 311-330; Carlip, S.: Six ways to quantize (2+1)-dimensional gravity, to appear in *Proc. of the Fifth Canadian Conference on General Relativity and Relativistic Astrophysics*, Waterloo, Ontario, May 1993

[10] Moncrief, V.: Reduction of the Einstein equations in 2+1 dimensions to a Hamiltonian system over Teichmüller space, *J. Math. Phys.* 30 (1989) 2907-2914; Mess, G.: Lorentz spacetimes of constant curvature, preprint Institut des Hautes Études Scientifiques IHES/M/90/28

[11] Loll, R.: Independent SU(2)-loop variables and the reduced configuration space of SU(2)-lattice gauge theory, *Nucl. Phys.* B368 (1992) 121-142; Yang-Mills theory without Mandelstam constraints, *Nucl. Phys.* B400 (1993) 126-144

[12] Giles, R.: Reconstruction of gauge potentials from Wilson loops, *Phys. Rev.* D24 (1981) 2160-2168

[13] Loll, R.: Loop variable inequalities in gravity and gauge theory, *Class. Quant. Grav.* 10 (1993) 1471-1476

[14] Abikoff, W.: The real analytic theory of Teichmüller space, Lecture Notes in Mathematics, vol.820, Springer, Berlin, 1980

[15] Ashtekar, A. and Romano, J.: Chern-Simons and Palatini actions and 2+1 gravity, *Phys. Lett.* B229 (1989) 56-60

[16] Gradsteyn, I.S. and Ryzhik, I.M.: *Table of integrals, series, and products*, corrected and enlarged edition, Academic Press, San Diego, 1980

[17] Moncrief, V.: Recent advances in ADM reduction, in: *Directions in General Relativity*, vol.1, ed. B.L. Hu et al., Cambridge University Press, 1993, 231-243

[18] Wolpert, S.: On the Weil-Petersson geometry of the moduli space of curves, Ann. of Math. 115 (1982) 501-528

[19] Okai, T.: An explicit description of the Teichmüller space as holonomy representations and its applications, Hiroshima Math. J. 22 (1992) 259-271

[20] D'Hoker, E. and Phong, D.H.: The geometry of string perturbation theory, *Rev. Mod. Phys.* 60 (1988) 917-1065