COHERENT STATES, YANG-MILLS THEORY, AND REDUCTION

BRIAN C. HALL

ABSTRACT. This paper explains the ideas behind a joint work with Bruce Driver on the quantization of Yang-Mills theory on a spacetime cylinder, with an emphasis on the notions of coherent states and reduction.

CONTENTS

1. Introduction 1
2. Classical Yang-Mills theory on a spacetime cylinder 3
3. Formal and semiformal quantization 5
4. The Segal–Bargmann transform to the rescue 7
5. Coherent states: from $\mathcal{A}_{\mathbb{C}}$ to $K_{\mathbb{C}}$ 10
6. Identification of $T^*(K)$ with $K_{\mathbb{C}}$ 13
7. Reduction of the Laplacian 14
8. Does quantization commute with reduction? 15
9. Notes 17
References 19

1. INTRODUCTION

The quantization of Yang-Mills theory is an important example of the quantization of reduced Hamiltonian systems. This paper concerns the simplest non-trivial case of quantized Yang-Mills theory, namely, pure Yang-Mills on a spacetime cylinder. Most of this paper is an exposition of a joint work [DH1] with Bruce Driver, with an emphasis on the concepts rather than the mathematical technicalities.

Driver and I use as our main tool the Segal–Bargmann transform, or equivalently, coherent states. We reach two main conclusions. First, upon reduction the *ordinary* coherent states on the space of connections become the *generalized* coherent states in the sense of [H1] on the finite-dimensional compact structure group. Second, coherent states provide a way to make rigorous the generally accepted idea that upon reduction the Laplacian for the infinite-dimensional space of connections becomes the Laplacian on the structure group. In the rest of the introduction I give a schematic description of the the paper. More details are found in the body of the paper and in [DH1].

Driver and I use the canonical quantization approach rather than the path-integral approach, and we work in the temporal gauge. As stated, we assume that spacetime is a cylinder, namely, $S^1 \times \mathbb{R}$. We fix a compact connected structure

Date: Summer 1999.

group K, which I will assume here is simple connected, with Lie algebra $\mathfrak{k}$. The configuration space for the classical theory is the space $\mathcal{A}$ of $\mathfrak{k}$-valued connection 1-forms over the spatial circle. The gauge group $\mathcal{G}$, consisting of maps of the spatial circle into K, acts naturally on $\mathcal{A}$. The based gauge group $\mathcal{G}_0$, consisting of gauge transformations that equal the identity at one fixed point in the spatial circle, acts freely on $\mathcal{A}$, and the quotient $\mathcal{A}/\mathcal{G}_0$ is simply the compact structure group K. This reflects that in this simple case the only gauge-invariant quantity is the holonomy of a connection around the spatial circle.

Meanwhile we have the complexification of $\mathcal{A}$, namely, $\mathcal{A}_{\mathbb{C}} = \mathcal{A} + i\mathcal{A}$, which is identifiable with the cotangent bundle of $\mathcal{A}$ and is the phase space for the unreduced system. We have also the complexification $K_{\mathbb{C}}$ of the structure group K, which is identifiable with the cotangent bundle of K. Here $K_{\mathbb{C}}$ is the unique simply connected complex Lie group whose Lie algebra is $\mathfrak{k} + i\mathfrak{k}$. One defines in the obvious way the (based) complexified gauge group $\mathcal{G}_{0,\mathbb{C}}$, which acts holomorphically on $\mathcal{A}_{\mathbb{C}}$. The quotient $\mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}}$ is $K_{\mathbb{C}}$. This is the reduced phase space for the theory.

Now we have the ordinary Segal–Bargmann tranform for $\mathcal{A}$, which maps from an L^2 space of functions on $\mathcal{A}$ to an L^2 space of *holomorphic* functions on $\mathcal{A}_{\mathbb{C}}$. Much more recently there is a generalized Segal–Bargmann transform for K [H1], which maps from an L^2 space of functions on K to an L^2 space of holomorphic functions on $K_{\mathbb{C}}$. The gist of [DH1] is that the *ordinary* Segal–Bargmann transform for $\mathcal{A}$, when restricted to the gauge-invariant subspace is precisely the *generalized* Segal-Bargann transform for $\mathcal{A}/\mathcal{G}_0 = K$. To say the same thing in the language of coherent states, taking the ordinary coherent states for $\mathcal{A}$ and projecting them onto the gauge-invariant subspace gives the generalized coherent states for K (in the sense of [H1]). So [DH1] gives a new way of understanding the generalized Segal–Bargmann transform (or generalized coherent states) for a compact Lie group K.

There is another purpose to the paper [DH1]. The Segal–Bargmann transform for $\mathcal{A}$ may be expressed in terms of the Laplacian $\Delta_{\mathcal{A}}$ for $\mathcal{A}$. The generalized Segal–Bargmann transform for K is expressed in a precisely parallel way in terms of the Laplacian for K. Thus the result that the Segal–Bargmann transform for $\mathcal{A}$ becomes the generalized Segal–Bargmann transform for K when restricted to the gauge-invariant subspace, this result gives a mathematically precise meaning to the following generally accepted principle.

(1) On the gauge-invariant subspace, $\Delta_{\mathcal{A}}$ reduces toΔ_K.

Driver and I wish to interpret Theorem 5.2 of [DH1] as a mathematically rigorous version of this principle, which, as explained below, does not really make mathematical sense as written. Since there is no spatial curvature when space is one-dimensional, the Hamiltonian for our theory is just a multiple of $\Delta_{\mathcal{A}}$. So we may say that upon restriction to the gauge-invariant subspace the Hamiltonian becomes a multiple of Δ_K.

I discuss two additional points. First, I discuss why, even at a formal level, $\Delta_{\mathcal{A}}$ *should* go to Δ_K on the invariant subspace. For the general situation of a manifold modulo an isometric group action, even in the finite-dimensional case, the Laplacians before and after reduction do not agree. So beyond the technicalities associated to the infinite-dimensionality, something special is happening in this case. Second, I consider the possibility of doing things in the opposite order, namely, first passing to the reduced phase space $K_{\mathbb{C}}$, and then constructing coherent states by

means of geometric quantization. It turns out that the two procedures give the same answer, *provided* that on includes as part of the geometric quantization the mysterious "half-form correction."

Acknowledgments. The idea of deriving the generalized Segal–Bargmann transform from the infinite-dimensional ordinary Segal–Bargmann transform is due to L. Gross and P. Malliavin [GM]. However, [GM] was not intended to be about Yang-Mills theory. What I am here calling the gauge group $\mathcal{G}_0$ they call the loop group, and its action in [GM] is not unitary. To get the Yang-Mills interpretation that we were striving for, Driver and I modified the approach of Gross and Malliavin so as to make the action of $\mathcal{G}_0$ unitary. (More precisely, we take a certain limit under which the action of $\mathcal{G}_0$ becomes formally unitary.)

The idea that the generalized coherent states for K could be obtained from the ordinary coherent states for $\mathcal{A}$ by reduction is due to K. Wren [W]. Wren uses the "Rieffel induction" approach proposed by Landsman [L1] and carried out in the abelian case by Landsman and Wren [LW]. See also the exposition in the book of Landsman [L2, IV.3.7]. I describe in Section 5 the relationship of our results to those of Wren.

I am endebted to Bruce Driver for clarifying to me many aspects of what is discussed here. I also acknowledge valuable discussions with Andrew Dancer, and I thank Dan Freed for a valuable suggestion regarding the half-form correction.

2. Classical Yang-Mills theory on a spacetime cylinder

Yang-Mills theory on a spacetime cylinder is an exactly solvable model [R]. Nevertheless, I believe that there are things to learn here, both classically and quantum mechanically, by comparing what happens before gauge symmetry is imposed to what happens afterward. I begin with the classical theory, borrowing heavily from the treatment of Landsman [L2].

We work on the spacetime manifold $S^1 \times \mathbb{R}$, with S^1 being space and $\mathbb{R}$ time. Fix a connected compact Lie group K, the structure group, which for simplicity I take to be simply connected, and fix an Ad-invariant inner product on the Lie algebra $\mathfrak{k}$ of K. We work in the temporal gauge, which has the advantage of allowing the classical Yang-Mills equations to be put into Hamiltonian form. The temporal gauge is only a partial gauge-fixing, leaving still a large gauge group $\mathcal{G}$, namely the group of mappings of the space manifold S^1 into the structure group K. Note that the gauge group is just a loop group in this case. I will concern myself only with the based gauge group $\mathcal{G}_0$, consisting of maps of S^1 into K which equal the identity at one fixed point in S^1. The remaining gauge symmetry can easily be added later.

In the temporal gauge, the Yang-Mills equations have a configuration space $\mathcal{A}$ consisting of connections on the space manifold. The connections are 1-forms with values in the Lie algebra $\mathfrak{k}$. Since our space manifold is one-dimensional, we may think of the connections as $\mathfrak{k}$-valued functions. There is a natural norm on $\mathcal{A}$ given by

$$\|A\|^2 = \int_0^1 |A(\tau)|^2 \, d\tau, \quad A \in \mathcal{A}.$$

Here S^1 is the interval $[0,1]$ with ends identified, and $|A(\tau)|^2$ is computed using the inner product on $\mathfrak{k}$. The norm allows us to define a distance function

$$d(A,B) := \|A - B\|.$$

The gauge group $\mathcal{G}_0$ acts on $\mathcal{A}$ by

$$(g \cdot A)_\tau = g_\tau A_\tau g_\tau^{-1} - g_\tau^{-1} \frac{dg}{d\tau}. \tag{2}$$

The map $A \to gAg^{-1}$ is linear, invertible, and norm-preserving, hence a "rotation" of $\mathcal{A}$. So the action of each $g \in \mathcal{G}_0$ is distance-preserving, a combination of a rotation and a translation in $\mathcal{A}$.

The phase space of the theory is the cotangent bundle of $\mathcal{A}$, $T^*(\mathcal{A}) \cong \mathcal{A} + \mathcal{A}$. The action of $\mathcal{G}_0$ on $\mathcal{A}$ extends in a natural way to an action on $\mathcal{A} + \mathcal{A}$ given by

$$g \cdot (A, P) = \left(g \cdot A, gPg^{-1}\right).$$

So the translation part of (2) affects only the "position" A and not the "momentum" P.

The Yang-Mills equations take place in the phase space $\mathcal{A} + \mathcal{A}$ and have three parts. First we have a dynamical part. The equations of motion are just Hamilton's equations, for the Hamiltonian function

$$H\left(A + iP\right) = \frac{1}{2} \left\|P\right\|^2.$$

Normally there would be another term involving the curvature of A, but that term is necessarily zero in this case, since S^1 is one-dimensional. Thus the solutions of Hamilton's equations are embarassingly easy to write down: the general solution is

$$\left(A\left(t\right), P\left(t\right)\right) = \left(A_0 + tP_0, P_0\right).$$

This is just free motion in $\mathcal{A}$. Observe that the Hamiltonian H is invariant under the action of $\mathcal{G}_0$ on $\mathcal{A} + \mathcal{A}$.

Second we have a constraint part. This says that the solutions (trajectories in $\mathcal{A}+\mathcal{A}$) have to lie in a certain set, which I will denote $J^{-1}\left(0\right)$, which is "the zero set of the moment mapping for the action of $\mathcal{G}_0$." I will not repeat here the formulas, which may be found for example in [DH1, Sect. 2]. This constraint is of a simple sort, in that $J^{-1}\left(0\right)$ is invariant under the dynamics and under the action of $\mathcal{G}_0$ on $\mathcal{A} + \mathcal{A}$. So the constraint does not alter the dynamics, it merely restricts us to certain special solutions of the original equations of motion.

Third we have a philosophical part. This says that the only functions on phase space that are physically observable are ones that are gauge-invariant.

The last two points together say that we may as well think of the dynamics as taking place in $J^{-1}\left(0\right)/\mathcal{G}_0$, which is the same as $T^*(\mathcal{A}/\mathcal{G}_0)$. This is Marsden–Weinstein or symplectic reduction. Since the Hamiltonian function H is $\mathcal{G}_0$-invariant, it makes sense as a function on $J^{-1}\left(0\right)/\mathcal{G}_0$.

Now, we are in a very simple situation, with the space manifold being just a circle. In this case two connections are gauge-equivalent if and only if they have the same holonomy around the spatial circle. So the orbits of $\mathcal{G}_0$ are labeled by the holonomy $h\left(A\right)$ of a connection A around the circle, where for $A \in \mathcal{A}$, $h\left(A\right)$ is an element of the structure group K. It is easily seen that any $x \in K$ can be the holonomy of some A, and so we have

$$\mathcal{A}/\mathcal{G}_0 \cong K.$$

Thus

$$J^{-1}\left(0\right)/\mathcal{G}_0 \cong T^*(\mathcal{A}/\mathcal{G}_0) \cong T^*(K).$$

After this reduction, the dynamics become geodesic motion in K. The geodesics may be written explicitly as xe^{tX} with $x \in K$ and $X \in \mathfrak{k}$.

We require one last discussion before turning to the quantum theory. We may think of $\mathcal{A}+\mathcal{A}$ as the complex vector space $\mathcal{A}_{\mathbb{C}} = \mathcal{A} + i\mathcal{A}$, in the same way that we think of $T^*(\mathbb{R}) \cong \mathbb{R}+\mathbb{R}$ as $\mathbb{C}$. We may then think of elements of $\mathcal{A}_{\mathbb{C}}$ as functions (or 1-forms) with values in the complexified Lie algebra $\mathfrak{k}_{\mathbb{C}} = \mathfrak{k} + i\mathfrak{k}$. The action of $\mathcal{G}_0$ extends to an action on $\mathcal{A}_{\mathbb{C}}$ by

$$(g \cdot Z)_\tau = g_\tau Z_\tau g_\tau^{-1} - g_\tau^{-1} \frac{dg}{d\tau},$$

where $Z : [0,1] \to \mathfrak{k}_{\mathbb{C}}$. Note that the translation part in in the real direction; that is, $g_\tau^{-1} \frac{dg}{d\tau}$ is in $\mathcal{A}$. One can think of elements of $\mathcal{A}_{\mathbb{C}}$ as complex connections and thus define their holonomy. But the holonomy now takes values in the *complexified* group $K_{\mathbb{C}}$, where $K_{\mathbb{C}}$ is the unique simply connected complex Lie group with Lie algebra $\mathfrak{k} + i\mathfrak{k}$. For example, if $K = SU(n)$ then $K_{\mathbb{C}} = SL(n;\mathbb{C})$. The complexified (based) gauge group $\mathcal{G}_{0,\mathbb{C}}$ is then the group of based loops with values in $K_{\mathbb{C}}$. The same reasoning as on $\mathcal{A}$ shows that the only $\mathcal{G}_{0,\mathbb{C}}$-invariant quantity on $\mathcal{A}_{\mathbb{C}}$ is the holonomy; so $\mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}} = K_{\mathbb{C}}$.

It turns out that restricting to the zero set of the moment mapping and then dividing out by the action of $\mathcal{G}_0$ gives the same result as working on the whole phase space and then dividing out by the action of $\mathcal{G}_{0,\mathbb{C}}$. Thus

$$J^{-1}(0)/\mathcal{G}_0 = \mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}} = K_{\mathbb{C}}.$$

On the other hand, we have already said that $J^{-1}(0)/\mathcal{G}_0$ is identifiable with $T^*(K)$. So we have a natural identification

$$K_{\mathbb{C}} \cong T^*(K).$$

This is explained in detail in Section 6 and the resulting identification is given there explicitly.

3. Formal and semiformal quantization

In this section we will see what is involved in trying to quantize this system. This discussion will set the stage for the entrance of the Segal–Bargmann transform and the coherent states in the next two sections.

Let us first try to quantize our Yang-Mills example at a purely formal level, that is, without worrying too much whether our formulas make sense. I want to do the quantization *before* the reduction by $\mathcal{G}_0$. If we did the reduction before the quantization, then we would have a finite-dimensional system, which is easily quantized. So it is of interest to do the quantization first and see if this gives the same result. See [R], where quantization is done after the reduction, and [Di], where quantization is done before the reduction.

Since our system has a configuration space $\mathcal{A}$, we may formally take our unreduced quantum Hilbert space to be

$$L^2(\mathcal{A}, \mathcal{D}A),$$

where $\mathcal{D}A$ is the fictitious Lebesgue measure on $\mathcal{A}$. The quantization of the constraint equation (see [DH1, Sect. 2]) then imposes the condition that our wave functions be $\mathcal{G}_0$-invariant. Note that the quantization of the second part of the classical theory (the constraint) automatically incorporates the third part as well

(the $\mathcal{G}_0$-invariance). So we want the reduced (physical) quantum Hilbert space to be

$$L^2(\mathcal{A}, \mathcal{D}A)^{\mathcal{G}_0} := \left\{ f \in L^2(\mathcal{A}, \mathcal{D}A) \,|\, f(g \cdot A) = f(A), \forall g, A \right\}.$$

Recall that in our example, in which space is a circle, two connections are $\mathcal{G}_0$-equivalent if and only if they have the same holonomy around the spatial circle. That means that a $\mathcal{G}_0$-invariant function must be of the form

$$f(A) = \phi(h(A)), \tag{3}$$

where $h(A) \in K$ is the holonomy of A and where ϕ is a function on the structure group K. Furthermore, as we shall see more clearly in the next section, it is reasonable to think that for a function of the form (3), integrating $|f(A)|^2$ with respect to $\mathcal{D}A$ is the same as integrating $|\phi(g)|^2$ with respect to a multiple of the Haar measure on K. Thus

$$L^2(\mathcal{A}, \mathcal{D}A)^{\mathcal{G}_0} \cong L^2(K, C \cdot dg), \tag{4}$$

for some (probably infinite) constant C. This is our physical Hilbert space.

Next we consider the Hamiltonian. Formally quantizing the function $\frac{1}{2}\|P\|^2$ in the usual way gives

$$\hat{H} = -\frac{\hbar^2}{2}\Delta_{\mathcal{A}} = -\frac{\hbar^2}{2}\sum_{k=1}^{\infty}\frac{\partial^2}{\partial x_k^2}, \tag{5}$$

where the x_k's are coordinates with respect to an orthonormal basis of $\mathcal{A}$. We must now try to determine how $\hat{H}$ acts on the $\mathcal{G}_0$-invariant subspace. In light of what happens when performing the reduction before the quantization, it is reasonable to guess that on the invariant subspace $\Delta_{\mathcal{A}}$ reduces to Δ_K, that is,

$$\Delta_{\mathcal{A}}[\phi(h(A))] = (\Delta_K \phi)(h(A)). \tag{6}$$

(See also [Di, W].) If we accept this and if we ignore the infinite constant C in (4) then we conclude that our quantum Hilbert space is

$$L^2(K, dx)$$

and our Hamiltonian is

$$\hat{H} = -\frac{\hbar^2}{2}\Delta_K.$$

This concludes the formal quantization of our system.

We now begin to consider how to make this mathematically precise. One approach is to forget about the measure theory (i.e. the Hilbert space) and to try to prove (6). As it turns out, the answer is basis-dependent–choosing different bases in (5) will give different answers. Another way of saying this is that the matrix of second derivatives of a function f of the form (3) is in general non-trace-class. However, if one uses the most obvious sort of basis, then indeed it turns out that (5) is true. See the appendix of [DH1].

Even without the problem of basis-dependence, the above approach is unsatisfying because we would like to define $\hat{H}$ as an operator in some Hilbert space. Since Lebesgue measure $\mathcal{D}A$ does not actually exist, one reasonable procedure is to "approximate" $\mathcal{D}A$ by a Gaussian measure $dP_s(A)$ with large variance s. This means that P_s is formally given by the expression

$$dP_s(A) = c_s e^{-\|A\|^2/2s} \mathcal{D}A,$$

where c_s is supposed to be a normalization constant that makes the total integral one. Formally as $s \to \infty$ we get back a multiple of Lebesgue measure $\mathcal{D}A$. The measure P_s does exist rigorously, provided that one allows sufficiently non-smooth connections.

There is good news and bad news with this approach. First the good news. 1) Even though the connections in the support of P_s are not smooth, the holonomy of such a connection makes sense, as the solution to a *stochastic* differential equation. 2) If we define the gauge-invariant subspace to be

$$L^2(\mathcal{A}, P_s)^{\mathcal{G}_0} = \{f \mid \text{for all } g \in \mathcal{G}_0,\ f(g \cdot A) = f(A) \text{ for } P_s\text{-almost every } A\},$$

then the Gross ergodicity theorem [G2] asserts that $L^2(\mathcal{A}, P_s)^{\mathcal{G}_0}$ is precisely what we expect, namely, the space of functions of the form $f(A) = \phi(h(A))$, with ϕ a function on K. 3) There is a natural dense subspace of $L^2(\mathcal{A}, P_s)$ on which $\Delta_{\mathcal{A}}$ is unambiguously defined, consisting of smooth cylinder functions. Here a cylinder function is one which depends on only finitely many of the infinitely many variables in $\mathcal{A}$. See [DH1, Defn. 4.2].

Note that the map which takes $f(A)$ to $f(g \cdot A)$ is not unitary, because the measure P_s is not invariant under the action of $\mathcal{G}_0$. Driver and I wish to avoid "unitarizing" the action of $\mathcal{G}_0$, because if we did unitarize then there would be no gauge-invariant subspace. See [DH2]. Instead of unitarizing the action for a fixed value of s, we will eventually let s tend to infinity, at which point unitarity is formally recovered.

The bad news about this approach is that $\Delta_{\mathcal{A}}$ is not a closable operator, and that functions of the holonomy are not cylinder functions. This means that if we approximate $\phi(h(A))$ by cylinder functions, then the value of $\Delta_{\mathcal{A}}\phi(h(A))$ depends on the choice of approximating sequence. So we still have a major problem in making mathematical sense out of the quantization.

4. The Segal–Bargmann transform to the rescue

In this section I will explain how the Segal–Bargmann transform can be used to make sense out of the quantization. At the same time, we will see how the generalized Segal–Bargmann transform for the structure group K arises from the restriction of the ordinary Segal–Bargmann transform for the gauge-invariant subspace. Although it is technically easier to describe the quantization in terms of the Segal–Bargmann transform, there is a formally equivalent description in terms of coherent states, as I will explain in the next section. See [B, S1, S2, S3] and also [BSZ, H6] for results on the ordinary Segal–Bargmann transform.

Let me explain the normalization of the Segal–Bargmann transform that I wish to use, first for the finite-dimensional space $\mathbb{R}^d$. Let $\mathcal{H}(\mathbb{C}^d)$ denote the space of holomorphic (complex analytic) functions on $\mathbb{C}^d$. For any positive constant $\hbar$, define

$$C_\hbar : L^2(\mathbb{R}^d, dx) \to \mathcal{H}(\mathbb{C}^d)$$

by the formula

$$C_\hbar f(z) = (2\pi\hbar)^{-d/2} \int_{\mathbb{R}^d} e^{-(z-q)^2/2\hbar} f(q)\, dx, \quad z \in \mathbb{C}^d. \tag{7}$$

Here $(z-q)^2$ means $\Sigma (z_k - q_k)^2$. If we restrict attention to $z \in \mathbb{R}^d$, then this is the standard expression for the solution of the heat equation $\partial u/\partial\hbar = \frac{1}{2}\Delta u$, at time $\hbar$

with initial condition f. Thus we may write

$$C_\hbar f = \text{ analytic continuation of } e^{\hbar\Delta/2} f.$$

Here $e^{\hbar\Delta/2} f$ is just a mnemonic for the solution of the heat equation with initial condition f, and the analytic continuation is in the space variable (analytic continuation from $\mathbb{R}^d$ to $\mathbb{C}^d$). Because $\hbar$ is playing the role of time in the heat equation, it is tempting call this parameter t instead of $\hbar$; this is what we do in [DH1].

Now let $\nu_\hbar$ be the measure on $\mathbb{C}^d$ given by

$$d\nu_\hbar(z) = (\pi\hbar)^{-d/2} e^{-(\operatorname{Im} z)^2/\hbar} dz,$$

where dz refers to the $2d$-dimensional Lebesgue measure on $\mathbb{C}^d$.

Theorem 1 (Segal-Bargmann transform)**.** *For each positive value of $\hbar$, $C_\hbar$ is a unitary map of $L^2(\mathbb{R}^d, dq)$ onto $\mathcal{H}L^2(\mathbb{C}^d, \nu_\hbar)$, where $\mathcal{H}L^2$ denotes the space of entire holomorphic functions on $\mathbb{C}^d$ which are square-integrable with respect to $\nu_\hbar$.*

This is not quite the form of the transform given by either Segal or Bargmann. Comparing to Bargmann's map A (and taking $\hbar = 1$ since that is what Bargmann does) we have

$$C_1 f(z) = (4\pi)^{-d/4} e^{-z^2/4} Af\left(\frac{z}{\sqrt{2}}\right).$$

The factor in front of Af converts from the measure in [B] to the measure $\nu_\hbar$ that I am using, and also emphasizes the role of the heat equation. The factor of $\sqrt{2}$ accounts for the difference between Bargmann's convention that $z = (q+ip)/\sqrt{2}$ and my convention that $z = q + ip$, which is preferable for me because on a more general manifold, the map $z \to z/\sqrt{2}$ does not make sense.

The $C_\hbar$ form of the Segal–Bargmann transform has the advantage of making explicit the symmetries of position-space. The measure dq on $\mathbb{R}^d$ and the measure $\nu_\hbar$ on $\mathbb{C}^d$ are both invariant under rotations and translations of q-space, and the transform commutes with rotations and translations of q-space. Since a gauge transformation is just a combination of a rotation and a translation, this property of $C_\hbar$ will be useful.

On the other hand, as it stands this form of the Segal–Bargmann transform does not permit taking the infinite-dimensional limit, as we must do if we want to quantize $\mathcal{A}$, since neither dq nor $\nu_\hbar$ makes sense when d tends to infinity. Fortunately, it is not too hard to fix this problem by adding a little bit of Gaussian-ness to our measures in the q-directions. It turns out that if we do this correctly, then we can keep the same formula for the Segal–Bargmann transform while making a small change in the measures, and still have a unitary map.

Theorem 2. *For all $s > \hbar/2$, let P_s denote the measure on $\mathbb{R}^d$ given by*

$$dP_s(q) = (2\pi s)^{-d/2} e^{-q^2/2s} dq$$

and let $M_{s,\hbar}$ denote the measure on $\mathbb{C}^d$ given by

$$dM_{s,\hbar}(q+ip) = (\pi\hbar)^{-d/2} (\pi r)^{-d/2} e^{-q^2/r} e^{-p^2/\hbar},$$

where $r = 2s - \hbar$. Then the map $S_{s,\hbar} : L^2(\mathbb{R}^d, P_s) \to \mathcal{H}(\mathbb{C}^d)$ given by

$$S_{s,\hbar} f = \text{ analytic continuation of } e^{\hbar\Delta/2} f$$

is a unitary map of $L^2(\mathbb{R}^d, P_s)$ onto $\mathcal{H}L^2(\mathbb{C}^d, M_{s,\hbar})$.

If we multiply the measures on both sides by $(2\pi s)^{d/2}$ and then let $s \to \infty$ we recover the $C_\hbar$ version of the transform. On the other hand, for any finite value of s it is possible to let $d \to \infty$ to get a transform that is applicable to our gauge-theory example. So we consider $L^2(\mathcal{A}, P_s)$, where P_s is the Gaussian measure on $\mathcal{A}$ described in Section 3, which is just the infinite-dimensional limit of the measures P_s on $\mathbb{R}^d$. We consider also the Gaussian measure $M_{s,\hbar}$ on $\mathcal{A}_\mathbb{C}$ that is the infinite-dimensional limit of the corresponding measures on $\mathbb{C}^d$.

We then work with **cylinder functions** in $L^2(\mathcal{A}, P_s)$, that is, functions that depend on only finitely many of the infinitely many variables in $\mathcal{A}$. (See [DH1, Defn. 4.2].) On such functions the Segal–Bargmann transform makes sense, since on such functions $\Delta_\mathcal{A}$ reduces to the Laplacian for some finite-dimensional space. It then follows from Theorem 2 that the Segal–Bargmann transform $S_{s,\hbar}$ is an isometric map of the space of cylinder functions in $L^2(\mathcal{A}, P_s)$ into $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$. This transform extends by continuity to a unitary map of $L^2(\mathcal{A}, P_s)$ onto $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$. Recall that $\Delta_\mathcal{A}$ by itself is a non-closable operator as a map of $L^2(\mathcal{A}, P_s)$ to itself. Considering $e^{\hbar\Delta_\mathcal{A}/2}$ as a map from $L^2(\mathcal{A}, P_s)$ to itself will not help matters. But by considering $e^{\hbar\Delta_\mathcal{A}/2}$ followed by analytic continuation, as a map from $L^2(\mathcal{A}, P_s)$ to $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$, we get a map which is not only closable but continuous (even isometric). It then makes perfect sense to apply this operator (the Segal–Bargmann transform) to functions of the holonomy.

The following theorem summarizes the above discussion.

Theorem 3. *For all $s > \hbar/2$ the map $S_{s,\hbar}$ given by*

$$S_{s,\hbar} f = \text{ analytic continuation of } e^{\hbar\Delta_\mathcal{A}/2} f$$

makes sense and is isometric on cylinder functions, and extends by continuity to a unitary map of $L^2(\mathcal{A}, P_s)$ onto $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$.

We are now ready to state the main result (Theorem 5.2) of [DH1].

Theorem 4. *Suppose $f \in L^2(\mathcal{A}, \tilde{P}_s)$ is of the form*

$$f(A) = \phi(h(A))$$

where ϕ is a function on K. Then there exists a unique holomorphic function Φ on $K_\mathbb{C}$ such that

$$S_{s,\hbar} f(C) = \Phi(h_\mathbb{C}(C)).$$

The function Φ is given by

$$\Phi = \text{analytic continuation } e^{\hbar\Delta_K/2}\phi.$$

Note that in light of the definition of $S_{s,\hbar}$, this result says that on the gauge-invariant subspace, $e^{\hbar\Delta_\mathcal{A}/2}$ (followed by analytic continuation) reduces to $e^{\hbar\Delta_K/2}$ (followed by analytic continuation). Thus Theorem 4 is a formally equivalent to the principle (1) with which we started.

Now, the gauge-invariant subspace $L^2(\mathcal{A}, P_s)^{\mathcal{G}_0}$ consists of functions of the form $f(A) = \phi(h(A))$, with ϕ a function on K. It may be shown that

$$\int_\mathcal{A} |\phi(h(A))|^2 \, dP_s(A) = \int_K |\phi(x)|^2 \rho_s(x) \, dx,$$

where ρ_s is the heat kernel at the identity on K at time s. Similarly,

$$\int_{\mathcal{A}_\mathbb{C}} |\Phi(h_\mathbb{C}(Z))|^2 \, dM_{s,\hbar}(Z) = \int_{K_\mathbb{C}} |\Phi(g)|^2 \mu_{s,\hbar}(g) \, dg,$$

where $\mu_{s,\hbar}$ is a suitable heat kernel on $K_{\mathbb{C}}$ and dg is Haar measure on $K_{\mathbb{C}}$. So the gauge-invariant subspace on the real side is identifiable with $L^2(K, \rho_s(x)\, dx)$ and on the complex side with $\mathcal{H}L^2(K_{\mathbb{C}}, \mu_{s,\hbar}(g)\, dg)$. So we have the following commutative diagram in which all maps are unitary.

$$\begin{array}{ccc} L^2(\mathcal{A}, P_s)^{\mathcal{G}_0} & \xrightarrow{e^{\hbar\Delta_{\mathcal{A}}/2}} & \mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})^{\mathcal{G}_0} \\ \updownarrow & & \updownarrow \\ L^2(K, \rho_s(x)\, dx) & \xrightarrow{e^{\hbar\Delta_K/2}} & \mathcal{H}L^2(K_{\mathbb{C}}, \mu_{s,\hbar}(g)\, dg) \end{array} \tag{8}$$

The horizontal maps contain an implicit analytic continuation.

This result embodies a rigorous version of the principle (1) and also shows that a form of the Segal–Bargmann transform for $\mathcal{A}$ can descend to a Segal–Bargmann transform for $\mathcal{A}/\mathcal{G}_0 = K$. But so far we still have the regularization parameter s, which we are supposed to remove by letting it tend to infinity. On the full space $L^2(\mathcal{A}, P_s)$ or $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$ the limit $s \to \infty$ does not exist; this was the point of putting in the s in the first place. But on the gauge-invariant subspaces, identified with functions on K or $K_{\mathbb{C}}$, the limit does exist. As $s \to \infty$, the heat kernel measure ρ_s on K converges to normalized Haar measure on K. This confirms our earlier conjecture that the fictitious Lebesgue measure on $\mathcal{A}$ (formally the $s \to \infty$ limit of P_s) pushes forward to the Haar measure on K. Meanwhile, the measure $\mu_{s,\hbar}$ converges as $s \to \infty$ to a certain measure I call $\nu_\hbar$, which coincides with the "K-averaged heat kernel measure" of [H1]. So taking the limit in the bottom line of (8) gives

$$L^2(K, dx) \quad \xrightarrow{e^{\hbar\Delta_K/2}} \quad \mathcal{H}L^2(K_{\mathbb{C}}, \mu_{s,\hbar}) \tag{9}$$

This supports the expected conclusion that our reduced quantum Hilbert space is $L^2(K, dx)$ and that the quantum Hamiltonian is $\left(-\hbar^2/2\right)\Delta_K$. It further shows that the generalized Segal–Bargmann transform for K, as given in (9), arises naturally from the ordinary Segal–Bargmann transform for the space of connections, upon restriction to the gauge-invariant subspace. The transform in (9) is precisely the K-invariant form of the transform which was previously constructed in [H1] from a purely finite-dimensional point of view.

5. Coherent states: from $\mathcal{A}_{\mathbb{C}}$ to $K_{\mathbb{C}}$

Let us now reformulate the results of the last section in terms of coherent states. Klauder and Skagerstam [KS] think of coherent states as a collection of states ψ_α in some Hilbert space, labeled by points some parameter space X such that there is a resolution of the identity

$$I = \int_X |\psi_\alpha\rangle \langle\psi_\alpha|\, d\nu(\alpha) \tag{10}$$

for some measure ν on X. One may then define a "coherent state transform," that is, a linear map $C : H \to L^2(X, \nu)$ given by taking the inner product of a vector in H with each of the coherent states:

$$C(v)(\alpha) = \langle\psi_\alpha | v\rangle .$$

The resolution of the identity implies that

$$\begin{aligned}\int_X |\langle \psi_\alpha \,|v\rangle|^2 \, d\nu\,(\alpha) &= \int_X \langle v\,|\psi_\alpha\rangle \,\langle \psi_\alpha\,|v\rangle \, d\nu\,(\alpha) \\ &= \langle v| \int_X |\psi_\alpha\rangle \,\langle \psi_\alpha| \, d\nu\,(\alpha)\,|v\rangle \\ &= \langle v\,|v\,\rangle\,.\end{aligned}$$

Thus the resolution of the identity (10) is at least formally equivalent to the statement that C is an isometric linear map. Note that C is only isometric but not unitary; in all the interesting cases the image of C is a proper subspace of $L^2(X,\nu)$, which may be characterized by a certain reproducing kernel condition. Although the resolution of the identity looks on the surface like an orthonormal basis expansion, it is in fact quite different. The coherent states are typically non-orthogonal and "overcomplete." The overcompleteness is reflected in the fact that C does not map onto $L^2(X,\nu)$.

As an example, consider the finite-dimensional Segal–Bargmann transform, in my normalization. The coherent states are then the states $\psi_z \in L^2(\mathbb{R}^n, dx)$ given by

$$\psi_z(x) = (2\pi\hbar)^{-n/2}\, e^{-(\bar{z}-x)^2/2\hbar}, \quad z \in \mathbb{C}^n\,.$$

This means that the coherent state transform is given by

$$(C_\hbar f)(z) = \langle \psi_z \,|f\,\rangle_{L^2(\mathbb{R}^n,dx)} = \int_{\mathbb{R}^n} (2\pi\hbar)^{-n/2}\, e^{-(z-x)^2/2\hbar} f(x)\, dx,$$

as above. In this case the parameter space X is $\mathbb{C}^n$ and the measure on X is the measure $\nu_\hbar$ of the last section. The overcompleteness of the coherent states means here that the image of $C_\hbar$ is not all of $L^2(\mathbb{C}^n,\nu)$, but only the holomorphic subspace.

Next consider what happens to a set of coherent states under reduction. Suppose we have a set of coherent states in a Hilbert space H, satisfying a resolution of the identity (10). Then suppose that V is a closed subspace of H and that P is the orthogonal projection onto V. Since $P^2 = P^* = P$, (10) gives

$$P = PIP = \int_X |P\psi_\alpha\rangle \,\langle P\psi_\alpha| \, d\nu\,(\alpha)\,.$$

Thus by projecting each coherent state into V we get a resolution of the identity (and hence a coherent state transform) for the subspace V. Note that at the moment the parameter space for the coherent states, and the measure on it, are unchanged by the projection. However, it may happen that certain sets of distinct coherent states become the same after the projection is applied. In that case we may reduce (or "collapse") the parameter space X by identifying any two parameters α and β for which $P\psi_\alpha = P\psi_\beta$. The measure ν then pushes forward to a measure $\tilde{\nu}$ on the reduced parameter space $\tilde{X}$.

This is what happens in our Yang-Mills case. We have states $\psi_Z^{(s)} \in L^2(\mathcal{A}, P_s)$ defined by the condition that

$$S_{s,\hbar} f(z) = \left\langle \psi_Z^{(s)} \,\middle|\, f \right\rangle_{L^2(\mathcal{A},P_s)}, \tag{11}$$

where now the parameter space, that is, the set of Z's, is the space $\mathcal{A}_{\mathbb{C}}$ of complexified connections. I suppress the dependence of $\psi_Z^{(s)}$ on $\hbar$. We want to project the

ψ_Z's onto the gauge-invariant subspace, that is, onto the space of functions of the form $\phi(h(A))$. The projection amounts to the same thing as restricting attention in (11) to f's of the form $f(A) = \phi(h(A))$. For such f's, Theorem 4 tells us that

$$\begin{aligned}\left\langle \psi_Z^{(s)} \middle| f \right\rangle &= \left[S_{s,\hbar}(\phi \circ h)\right](Z) \\ &= \Phi(h_{\mathbb{C}}(Z)),\end{aligned}$$

where Φ is the analytic continuation to $K_{\mathbb{C}}$ of $e^{\hbar\Delta_K/2}\phi$. We see then that for f in the invariant subspace, the right side of (11) depends only on the holonomy of Z. Thus upon projection into the gauge-invariant subspace the parameter space for the coherent states collapses from $\mathcal{A}_{\mathbb{C}}$ to $\mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}} = K_{\mathbb{C}}$.

If we identify the gauge-invariant subspace with $L^2(K, \rho_s)$ as in the previous section, then the reduced coherent states are the vectors $\tilde{\psi}_g^{(s)} \in L^2(K, \rho_s)$, with $g \in K_{\mathbb{C}}$, given by

$$\tilde{\psi}_g^{(s)}(x) = \frac{\overline{\rho_\hbar(gx^{-1})}}{\rho_s(x)}, \quad g \in K_{\mathbb{C}},$$

so that, as required, we have

$$\begin{aligned}\left\langle \tilde{\psi}_g^{(s)} \middle| \phi \right\rangle_{L^2(K,\rho_s)} &= \int_K \frac{\rho_\hbar(gx^{-1})}{\rho_s(x)} \phi(x)\ \rho_s(x)\ dx \\ &= \Phi(g).\end{aligned}$$

Here $\rho_\hbar(gx^{-1})$ refers to the analytic continuation of the heat kernel from K to $K_{\mathbb{C}}$, and for $g \in K$, the convolution $\int_K \rho_\hbar(gx^{-1})\phi(x)\,dx$ is nothing but $\left(e^{\hbar\Delta_K/2}\phi\right)(g)$.

The $\tilde{\psi}_g^{(s)}$ satisfy a resolution of the identity with respect to the measure $\mu_{s,\hbar}$ on $K_{\mathbb{C}}$. This measure is the one which is naturally induced from the measure $M_{s,\hbar}$ on $\mathcal{A}_{\mathbb{C}}$, upon reduction from $\mathcal{A}_{\mathbb{C}}$ to $K_{\mathbb{C}}$. That is, $\mu_{s,\hbar}$ is the "push-forward" of $M_{s,\hbar}$ from $\mathcal{A}_{\mathbb{C}}$ to $K_{\mathbb{C}}$, under the map $h_{\mathbb{C}}$. Now, as $s \to \infty$, $\rho_s(x)$ converges to the constant function $\mathbf{1}$. Thus we obtain in the limit coherent states $\tilde{\psi}_g \in L^2(K, dx)$ given by

$$\tilde{\psi}_g(x) := \lim_{s\to\infty} \tilde{\psi}_g^{(s)}(x) = \overline{\rho_\hbar(gx^{-1})}, \quad g \in K_{\mathbb{C}}. \tag{12}$$

These satisfy the following resolution of the identity:

$$I = \int_{K_{\mathbb{C}}} \left|\tilde{\psi}_g\right\rangle \left\langle \tilde{\psi}_g\right| d\nu_\hbar(g),$$

where $\nu_\hbar = \lim_{s\to\infty} \mu_{s,\hbar}$. The measure $\nu_\hbar$ coincides with the "K-averaged heat kernel measure" of [H1].

Although we are "supposed to" let $s \to \infty$, we get a well-defined coherent state theory for any $s > \hbar/2$. The case $s = \hbar$, as well as the limiting case $s \to \infty$, had previously been described in [H1]. For other values of s we get something new, which I investigate from a finite-dimensional point of view in [H5].

Let me compare the above results to those in the paper of Wren [W], which motivated Driver and me to develop our paper [DH1]. Wren uses the "Rieffel induction" method proposed by Landsman [L1], applied to this same problem of Yang-Mills theory on a spacetime cylinder. The commutative case was considered previously by Landsman and Wren in [LW]. Wren uses a fixed Gaussian measure and a "unitarized" action of the gauge group. In this approach there is no gauge-invariant subspace (see [DH2]) and so an integration over the gauge group is used to define a reduced Hilbert space, which substitutes for the gauge-invariant subspace. Wren

shows that the reduced Hilbert space can be identified with $L^2(K, dx)$ and further shows that under the reduction map the ordinary coherent states map precisely to the coherent states $\tilde{\psi}_g$ in (12). So the appearance of these coherent states in [DH1] was expected on the basis of Wren's results.

The paper [DH1] set out to understand better two issues raised by [W]. First, because in [W] there is no true gauge-invariant subspace to project onto, the resolution of the identity for the classical coherent states does not survive the reduction. That is, Rieffel induction does not tell you what the right measure is to get a resolution of the identity. Of course, the relevant measure had already been described in [H1], but it would be nice not to have to know this ahead of time. By contrast, in our approach the measure $\nu_\hbar$ arises naturally by pushing forward the Gaussian measure $M_{s,\hbar}$ to $K_{\mathbb{C}}$ and then letting s tend to infinity. Second, the calculation in [W] concerning the reduction of the Hamiltonian is non-rigorous, mainly because the *unconstrained* Hamiltonian is not well-defined. Driver and I used the Segal–Bargmann transform in order to get some form of the Hamiltonian $\Delta_{\mathcal{A}}$ to make rigorous sense.

Finally, let me mention that the generalized coherent states on K are do not fall into the framework of Perelomov [P], because there does not seem to be in the compact group case anything analogous to the irreducible unitary representation of the Heisenberg group on $L^2(\mathbb{R}^n)$.

6. Identification of $T^*(K)$ with $K_{\mathbb{C}}$

I am thinking of K as the configuration space for the reduced classical Yang-Mills theory, and of $K_{\mathbb{C}}$ as the corresponding phase space. For this to be sensible, there should be an identification of $K_{\mathbb{C}}$ with the standard phase space over K, namely the cotangent bundle $T^*(K)$. In this section I will explain how such an identification comes out of the reduction process. The resulting identification coincides with the one described in [H3, H4] from an intrinsic point of view.

So let us see what comes out of the reduction process. From the symplectic point of view we have the Marsden–Weinstein symplectic quotient $J^{-1}(0)/\mathcal{G}_0$. Since the action of $\mathcal{G}_0$ on $\mathcal{A}_{\mathbb{C}} = T^*(\mathcal{A})$ arises from an action of $\mathcal{G}_0$ on the configuration space $\mathcal{A}$, general principles tell us that $J^{-1}(0)/\mathcal{G}_0$ coincides with $T^*(\mathcal{A}/\mathcal{G}_0) = T^*(K)$. On the other hand, from the complex point of view we may analytically continue the action of $\mathcal{G}_0$ on $\mathcal{A}_{\mathbb{C}}$ to get an action of $\mathcal{G}_{0,\mathbb{C}}$ on $\mathcal{A}_{\mathbb{C}}$. Dividing out by this action gives $\mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}} = K_{\mathbb{C}}$. But in this case there is a natural identification of $J^{-1}(0)/\mathcal{G}_0$ with $\mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}}$: each orbit of $\mathcal{G}_{0,\mathbb{C}}$ intersects $J^{-1}(0)$ in precisely one $\mathcal{G}_0$-orbit. This may be seen from [L2].

This result is not a coincidence. In general, given a Kähler manifold M (in our example $\mathcal{A}_{\mathbb{C}}$) and an action of a group G that preserves both the complex and symplectic structure of M, we may analytically continue to get an action of $G_{\mathbb{C}}$ on M, an action which preserves the complex but not the symplectic structure of M. Then if $J^{-1}(0)$ is the moment mapping for the action of G, one expects that

$$J^{-1}(0)/G = M/G_{\mathbb{C}}. \tag{13}$$

This would mean that for each orbit O of $G_{\mathbb{C}}$ in M the intersection of O with $J^{-1}(0)$ is precisely a single G-orbit. Now, (13) is not actually true in general, but only with various provisos and qualifications. (See [Ki, MFK].) Still, this is an important idea and in our case it works out exactly.

So putting everything together we have the following identifications.

$$\begin{array}{ccccc} \mathcal{A}_{\mathbb{C}}/\mathcal{G}_{0,\mathbb{C}} & = & J^{-1}(0)/\mathcal{G}_0 & = & T^*(\mathcal{A}/\mathcal{G}_0) \\ \updownarrow & & & & \updownarrow \\ K_{\mathbb{C}} & & & & T^*(K) \end{array}$$

If one does the calculations, one obtains the following explicit identification of $T^*(K)$ with $K_{\mathbb{C}}$. First, use left-translation to trivialize the cotangent bundle, so that $T^*(K) \cong K \times \mathfrak{k}^*$. Then use the inner product on $\mathfrak{k}$ to identify $K \times \mathfrak{k}^*$ with $K \times \mathfrak{k}$. Finally map from $K \times \mathfrak{k}$ to $K_{\mathbb{C}}$ by the map

$$\Phi(x, Y) = xe^{iY}, \quad x \in K,\ Y \in \mathfrak{k}. \tag{14}$$

The map Φ is a diffeomorphism of $K \times \mathfrak{k}$ onto $K_{\mathbb{C}}$, and Φ^{-1} is called the polar decomposition of $K_{\mathbb{C}}$.

For example, suppose $K = SU(n)$ so that $K_{\mathbb{C}} = SL(n;\mathbb{C})$. Then given $g \in SL(n;\mathbb{C})$ we may use the standard polar decomposition for matrices to write

$$g = xp$$

with x unitary and p positive. Since $\det g = 1$ it follows that $\det x = \det p = 1$. Then p has a unique self-adjoint logarithm ξ, which will have trace zero. Letting $Y = \xi/i$ we have

$$g = xe^{iY}$$

with Y skew and trace zero, so $Y \in su(n)$.

Now in [H3] (see also [H4]) I argued from an intrinsic, finite-dimensional point of view that the above identification of $T^*(K)$ with $K_{\mathbb{C}}$ was natural. The argument was based on the notion of "adapted complex structures." There is a good reason that the reduction argument gives the same identification as the adapted complex structures do. Suppose X is a finite-dimensional compact Riemannian manifold such that $T^*(X)$ has a global adapted complex structure, and suppose G is a compact Lie group which acts freely and isometrically on X. Then a result of R. Aguilar [A] says that $T^*(X/G)$ has a global adapted complex structure and that this complex structure coincides with the one inherited from $T^*(X)$ by means of reduction. We have the same sort of situation here, with $X = \mathcal{A}$ and $G = \mathcal{G}_0$. Of course, $\mathcal{G}_0$ is not compact and $\mathcal{A}$ is neither comact nor finite-dimensional, but nevertheless what happens is reasonable in light of Aguilar's result.

7. Reduction of the Laplacian

Why *should* $\Delta_{\mathcal{A}}$ correspond to Δ_K on gauge-invariant functions? Let us strip away the infinite-dimensional technicalities and consider the analogous question in finitely many dimensions. Suppose X is a finite-dimensional connected Riemannian manifold and suppose G is a Lie group that acts by isometries on X. For simplicity I will assume that G is compact and that G acts freely on X. Then X/G is again a manifold, which has a unique Riemannian metric such that the quotient map $q : X \to X/G$ is a Riemannian submersion. This means that at each point $x \in X$, the differential of q is an isometry when restricted to the orthogonal complement of the tangent space to the G-orbit through x.

Given this metric on X/G we may consider the Laplace-Beltrami operator $\Delta_{X/G}$. For a smooth function f on X/G we may ask whether $\left(\Delta_{X/G} f\right) \circ q$ coincides with $\Delta_X (f \circ q)$. This amounts to asking whether Δ_X and $\Delta_{X/G}$ agree on the G-invariant

subspace of $C^\infty(X)$. Since Δ_X commutes with isometries, it will at least preserve the G-invariant subspace.

The answer in general is no, Δ_X and $\Delta_{X/G}$ do not agree on $C^\infty(X)^G$. For example, consider $SO(2)$ acting on $\mathbb{R}^2 \setminus \{0\}$ by rotations. The quotient manifold is diffeomorphic to $(0,\infty)$, with the point $r \in (0,\infty)$ corresponding to the orbit $x^2+y^2=r^2$ in $\mathbb{R}^2 \setminus \{0\}$. The induced metric on $(0,\infty)$ is the usual metric on $(0,\infty)$ as a subset of $\mathbb{R}$. So the Laplace-Beltrami operator on $(0,\infty)$ is just d^2/dr^2. On the other hand, the formula for the two-dimensional Laplacian on radial functions is

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right) f\left(\sqrt{x^2+y^2}\right) = \left[\frac{d^2 f(r)}{dr^2}+\frac{1}{r}\frac{df(r)}{dr}\right]\Bigg|_{r=\sqrt{x^2+y^2}}. \tag{15}$$

The source of the trouble is the discrepancy between the intrinsic volume measure dr on $(0,\infty)$ and the push-forward of the volume measure from $\mathbb{R}^2 \setminus \{0\}$, which is $2\pi r\, dr$.

In general, each G-orbit in X inherits a natural Riemannian metric from X, and we may compute the total volume of this orbit with respect to the associated Riemannian volume measure. The function $\mathrm{Vol}(G\cdot x)$ on X/G measures the discrepancy between the intrinsic volume measure on X/G and the push-forward of the volume measure on X. The two Laplacians on $C^\infty(X)^G$ will be related by the formula

$$\Delta_X = \Delta_{X/G} + \nabla\left(\log \mathrm{Vol}(G\cdot x)\right)\cdot\nabla. \tag{16}$$

(The gradient may be thought of as that for X/G, although this coincides in a natural sense with that for X, on G-invariant functions.) Formula (15) is a special case of (16) with volume factor $2\pi r$. So the two Laplacians agree if and only if the G-orbits all have the same volume.

Let us return, then, to the case of $\mathcal{A}/\mathcal{G}_0$. By considering the appendix of [DH1] it is easily seen that the metric on K that makes the map $h: \mathcal{A} \to K$ a Riemannian submersion is simply the bi-invariant metric on K induced by the inner product on $\mathfrak{k}$. (We use on $\mathcal{A}$ the metric coming from the L^2 norm as in Section 2.) So in light of (16) the statement that $\Delta_{\mathcal{A}}$ and Δ_K agree on the $\mathcal{G}_0$-invariant subspace is formally equivalent to the statement that all the $\mathcal{G}_0$-orbits have the same volume. But again from [DH1] it may be seen that there exist isometries of $\mathcal{A}$ that map any $\mathcal{G}_0$-orbit to any other, so formally all should have the same volume.

To look at it another way, we need to see that the (fictitious) volume measure $\mathcal{D}A$ on $\mathcal{A}$ pushes forward to a multiple of the Haar measure on K. Accepting P_s as an approximation to $\mathcal{D}A$, this pushes forward to the measure $\rho_s(x)\,dx$, which indeed converges to dx as s tends to infinity. It should be noted, however, that nothing so simple is likely to happen in higher-dimensional Yang-Mills theory. See for example [Ga].

8. Does quantization commute with reduction?

When quantizing a reduced Hamiltonian system such as Yang-Mills theory, one may ask whether the quantization should be done before or after the reduction. If we were very optimistic, we might hope that it doesn't matter, that one gets the same answer either way. If this were so, we could say that *quantization commutes with reduction*. Of course the question of whether quantization commutes with reduction may well depend on the system being quantized and on how one interprets

the question. I want to consider this question from the point of view of geometric quantization and I want specifically to compare the Segal–Bargmann space obtained by first quantizing $\mathcal{A}_{\mathbb{C}}$ and then reducing by $\mathcal{G}_0$ to the one obtained by directly quantizing $K_{\mathbb{C}}$.

In geometric quantization [Wo] one begins with a symplectic manifold (M, ω) and constructs over M a Hermitian complex line bundle L with connection, whose curvature form is $i\omega/\hbar$. If M is a cotangent bundle then such a bundle exists and may be taken to be topologically and Hermitianly trivial (though the connection is necessarily non-trivial). The "prequantum Hilbert space" is then the space of sections of L which are square-integrable with respect to the symplectic volume measure on M. To obtain the "quantum Hilbert space" one picks a "polarization" and restricts to the space of square-integrable polarized sections of L. If M is a Kähler manifold, i.e. it has a complex structure which is compatible in a natural sense with ω, then there is a natural Kähler polarization. In that case L may be given the structure of a holomorphic line bundle and the quantum Hilbert space becomes the space of square-integrable holomorphic sections of L.

In the case $M = \mathbb{C}^n$ the resulting bundle is holomorphically trivial. So by choosing a nowhere vanishing holomorphic section, the space of holomorphic sections of L may be identified with the space of holomorphic functions on $\mathbb{C}^n$. This nowhere vanishing section will not, however, have constant norm. This means that the inner product on the space of holomorphic functions will be an L^2 inner product with respect to a measure which is Lebesgue measure times the norm-squared of the trivializing section. Working this out we get simply the Segal–Bargmann space, with different normalizations of the space coming from different possible choices of the trivializing section. The construction depends on Planck's constant $\hbar$. In summary: applying geometric quantization to $\mathbb{C}^n$, using a Kähler polarization, yields the Segal–Bargmann space.

To apply geometric quantization to the infinite-dimensional space $\mathcal{A}_{\mathbb{C}}$ we may try to quantize $\mathbb{C}^n$ and then let n tend to infinity. For this to make sense with my normalization, we need to add the additional parameter s. So we obtain the Segal–Bargmann space $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$. We then want to reduce by the action of $\mathcal{G}_0$, which amounts to restricting to the space of functions in $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$ which are $\mathcal{G}_0$-invariant, and thus by analyticity, $\mathcal{G}_{0,\mathbb{C}}$-invariant. The resulting space is identifiable with $\mathcal{H}L^2(K_{\mathbb{C}}, \mu_{s,\hbar})$. Finally, letting s tend to infinity we obtain $\mathcal{H}L^2\left(K_{\mathbb{C}}, \nu_\hbar\right)$. It is therefore reasonable to say that $\mathcal{H}L^2\left(K_{\mathbb{C}}, \nu_\hbar\right)$ is the space obtained by quantizing $\mathcal{A}_{\mathbb{C}}$ and then reducing by $\mathcal{G}_0$.

Meanwhile, we may apply geometric quantization directly to $K_{\mathbb{C}}$. I do this calculation in [H4] and find that geometric quantization yields the space $\mathcal{H}L^2\left(K_{\mathbb{C}}, \gamma_\hbar\right)$, where $\gamma_\hbar$ and $\nu_\hbar$ are related by the formula

$$d\nu_\hbar\left(g\right) = a_\hbar\, u\left(g\right)\, d\gamma_\hbar\left(g\right). \tag{17}$$

Here $a_\hbar$ is an irrelevant constant and u is a function which is non-constant except when K is commutative. So it seems that quantizing $K_{\mathbb{C}}$ directly does *not* yield the same answer. However, this is not the end of the story. One can quantize $K_{\mathbb{C}}$ taking into account the "half-form correction" (also known as the "metaplectic correction"). This "corrected" quantization yields an extra factor in the measure, a factor that coincides precisely with the factor $u\left(g\right)$ in (17)! On the other hand, in the $\mathbb{C}^n$ case the half-form correction does not affect the answer, so even with the half-form correction we would get $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$ and then ultimately $\mathcal{H}L^2\left(K_{\mathbb{C}}, \nu_\hbar\right)$. So

our conclusion is the following: In this example, if we use geometric quantization with a Kähler polarization and the half-form correction, quantization does in fact commute with reduction.

Let me conclude by mentioning a related setting in which one can ask whether quantization commutes with reduction. In an influential paper [GS], Guillemin and Sternberg consider the geometric quantization of a compact Kähler manifold M. They assume then that there is an action of a compact group G on M that preserves the complex structure and the symplectic structure of M and they consider as well the Marsden–Weinstein quotient $M^G := J^{-1}(0)/G$, where J is the moment mapping for the action of G. Under certain conditions they show that there is a natural invertible linear map between on the one hand the G-invariant subspace of the quantum Hilbert space over M and on the other hand the quantum Hilbert space over M^G. They interpret this result as a form of quantization commuting with reduction.

However, Guillemin and Sternberg do not show that this invertible linear map is unitary, and indeed there seems to be no reason that it should be in general. So in their setting we may say that quantization fails to commute *unitarily* with reduction. Dan Freed [F] has suggested to me that inclusion of the half-form correction in the quantization might the map unitary, and indeed our Yang-Mills example seems to confirm this. (It was Freed's suggestion that led me to work out that u is just the half-form correction.) After all, upon inclusion of the half-form correction we get the same measure (except for an irrelevant overall constant) and therefore the same inner product whether quantizing before or after the reduction. Nevertheless, I do not believe that one will get a unitary correspondence in general. So we are left with the following open question.

> Given a Kähler manifold M with an action of a group G, under what conditions on M and G will quantization commute *unitarily* with reduction?

Although the question may be considered with or without the half-form correction, what little evidence there is so far suggests that the answer is more likely to be yes if the half-form correction is included.

9. Notes

Section 2. One should say something about the degree of smoothness assumed on the connections and gauge transformations. Although it does not matter so much at the classical level, it seems natural to take the space of connections to be the Hilbert space of square-integrable connections. This amounts to completing $\mathcal{A}$ with respect to the natural norm, the one which appears in the formula for the classical Hamiltonian. We may then take the gauge group to be the largest group whose action on $\mathcal{A}$ makes sense. This is the group of "finite energy" gauge transformations, namely, the ones for which $\|g^{-1}\,dg\|$ is finite. It is easily shown that in our example of a spatial circle, two square-integrable connections are related by a finite energy gauge (based) gauge transformation if and only if they have the same holonomy. In the quantized theory we will be forced to consider a larger space of connections.

Section 3. The measure P_s is a Gaussian measure, about which there is an extensive theory. For example, see [G1, K, GJ]. The distinctive feature of Gaussian measures on infinite-dimensional spaces is the presence of two different spaces, a

Hilbert space H whose norm enters the formal expression for the measure, and a larger topological vector space B on which the measure lives. Although one should think of the Gaussian measure as being canonically associated to H, the measure lives on B, and H is a measure-zero subspace. In our example H is the space of square-integrable connections and B is a suitable space of distributional connections. Since the elements of B are highly non-smooth, the holonomy must be defined as the solution of a stochastic differential equation.

If one glosses over questions of smoothness, the Gross ergodicity theorem [G2] sounds as if it ought to be trivial. But we have just said that we must enlarge the space of connections in order for the measure P_s to exist. Unfortunately, we may not correspondingly enlarge the gauge group without losing the quasi-invariance of the measure P_s under the action of $\mathcal{G}_0$, without which the definition of $L^2(\mathcal{A}, P_s)^{\mathcal{G}_0}$ does not make sense. So we end up unavoidably in a situation in which two connections with the same holonomy are not necessarily $\mathcal{G}_0$-equivalent, because the would-be gauge transformation is not smooth enough to be in $\mathcal{G}_0$. It was the "J-perp" theorem, which arose as a corollary of Gross's proof of the ergodicity theorem, which led him to suggest to me to look for an analog of the Segal–Bargmann transform on K.

Section 4. Driver and I define the holomorphic subspace of $L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$ to be the L^2 closure of the space of holomorphic cylinder functions. An important question then is whether a function of the form $F(Z) = \Phi(h_{\mathbb{C}}(Z))$, with $\Phi \in \mathcal{H}L^2(K_{\mathbb{C}}, \mu_{s,\hbar})$, is in this holomorphic subspace. The answer is yes, but the proof that we give is indirect.

I am *defining* $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})^{\mathcal{G}_0}$ to be the image of $L^2(\mathcal{A}, P_s)^{\mathcal{G}_0}$ under the Segal–Bargmann transform. Certainly every element of $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})^{\mathcal{G}_0}$ is actual invariant under the action of $\mathcal{G}_0$ on $\mathcal{A}_{\mathbb{C}}$. The converse is probably true as well, namely that every element of $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$ which is $\mathcal{G}_0$-invariant is in the image of $L^2(\mathcal{A}, P_s)^{\mathcal{G}_0}$, but we have not proved this.

Section 5. Except when $s = \hbar$ the coherent states $\psi_Z^{(s)}$ in $L^2(\mathcal{A}, P_s)$ are non-normalizable states. When $s = \hbar$, the coherent states $\psi_Z^{(s)}$ are normalizable states provided that Z is a square-integrable (complex) connection [HS, Sect. 2.3]. But even then the measure $M_{\hbar,\hbar}$ does not live on the space of square-integrable connections, and so it is a bit delicate to formulate the resolution of the identity. This shows that it is technically easier to formulate things in terms of the Segal–Bargmann transform instead of the coherent states. Nevertheless, we may think continue to think of unitarity for the Segal–Bargmann transform as formally equivalent to a resolution of the identity for the coherent states.

Section 6. There are several obstructions to (13) holding in general. One needs some condition to guarantee that the analytic continuation of the G-action exists globally. Even when it does, one needs to worry about the possibility of "unstable points," that is points whose $G_{\mathbb{C}}$-orbit does not intersect the zero set of the moment mapping, and about the possibility that the $G_{\mathbb{C}}$-orbits may not be closed. In the case of a cotangent bundle of a compact Riemannian manifold whose cotangent bundle admits a global adapted complex structure, none of these problems actually arises. See [A, Sect. 7].

Section 8. I jumping to conclusions about the correct action of the gauge group $\mathcal{G}_0$ on the Segal–Bargmann space $\mathcal{H}L^2(\mathcal{A}_{\mathbb{C}}, M_{s,\hbar})$. One should properly use geometric quantization to determine this action. To do this, we restrict first to the finite-dimensional space $\mathcal{H}L^2(\mathbb{C}^n, \nu_\hbar)$ and then consider the action of the group of

rotations and translations of $\mathbb{R}^n$ on this space. Going through the calculations, on finds that with my normalization these rotations and translations act in the obvious way, namely, by composing a function in $\mathcal{H}L^2(\mathbb{C}^n, \nu_\hbar)$ with the rotation or translation. Note that this holds only for rotations and translations in the x-directions. Now we have said that the action of $\mathcal{G}_0$ on $\mathcal{A}$ consists just of a rotation and a translation. So, taking $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$ as the best approximation of $\mathcal{H}L^2(\mathbb{C}^n, \nu_\hbar)$ when $n = \infty$, it is reasonable to say that the action of $\mathcal{G}_0$ on $\mathcal{H}L^2(\mathcal{A}_\mathbb{C}, M_{s,\hbar})$ should be just $F(Z) \to F\left(g^{-1} \cdot Z\right)$.

There is a large body of work extending the results of [GS]; see for example the survey article of Sjamaar [Sj].

References

[A] R. M. Aguilar, Symplectic reduction and new global unbounded solutions of the homogeneous complex Monge–Ampère equation, preprint.

[BSZ] J. Baez, I. Segal, and Z. Zhou, "Introduction to Algebraic and Constructive Quantum Field Theory," Princeton Univ. Press, Princeton, NJ, 1992.

[B] V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, Part I, *Comm. Pure Appl. Math.* **14** (1961), 187-214.

[Di] J. Dimock, Canonical quantization of Yang-Mills on a circle, *Rev. Math. Phys.* **8** (1996), 85-102.

[D] B. Driver, On the Kakutani-Itô-Segal-Gross and Segal–Bargmann-Hall isomorphisms, *J. Funct. Anal.* **133** (1995), 69-128.

[DG] B. Driver and L. Gross, Hilbert spaces of holomorphic functions on complex Lie groups, *in* "New trends in stochastic analysis. Proceedings of a Taniguchi international workshop" (K. Elworthy, S. Kusuoka, and I. Shigekawa, Eds.) Singapore: World Scientific, Singapore, 1997, pp. 76-106.

[DH1] B. K. Driver and B. C. Hall, Yang–Mills theory and the Segal–Bargmann transform, *Comm. Math. Phys.* **201** (1999), 249-290.

[DH2] B. K. Driver and B. C. Hall, The energy representation has no non-zero fixed vectors, preprint.

[F] D. Freed, personal communication.

[Ga] K. Gawedzki, Yang-Mills theory as Schrödinger quantum mechanics on the space of gauge-group orbits, *Phys. Rev. D*, **26** (1982) 3593-3610.

[GJ] J. Glimm and A. Jaffe, "Quantum physics. A functional integral point of view," Second edition, Springer-Verlag, New York/Berlin, 1987.

[G1] L. Gross, Abstract Wiener spaces, *in* "Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probablility," Vol. II, Univ. of California Press, 1967.

[G2] L. Gross, Uniqueness of ground states for Schrödinger operators over loop groups, *J. Funct. Anal.* **112** (1993), 373-441.

[GM] L. Gross, and P. Malliavin, Hall's transform and the Segal–Bargmann map, *in* "Itô's stochastic calculus and probability theory"(M. Fukushima, N. Ikeda, H. Kunita, and S. Watanabe, Eds.), Springer-Verlag, New York/Berlin, 1996, pp. 73-116.

[GS] V. Guillemin and S. Sternberg, Geometric quantization and multiplicities of group representations, *Invent. Math.* **67** (1982), 515-538.

[H1] B. C. Hall, The Segal–Bargmann "coherent state" transform for compact Lie groups, *J. Funct. Anal.* **122** (1994), 103-151.

[H2] B. C. Hall, The inverse Segal–Bargmann transform for compact Lie groups, *J. Funct. Anal.* **143** (1997), 98-116.

[H3] B. C. Hall, Phase space bounds for quantum mechanics on a compact Lie group, *Comm. Math. Phys.* **184** (1997), 233-250.

[H4] B. C. Hall, Quantum mechanics in phase space, *in* "Perspectives on quantization" (L. Coburn and M. Rieffel, Eds.), Contemp. Math., Vol. 214, Amer. Math.Soc., Providence, RI, 1998, 47–62.

[H5] B. C. Hall, A new form of the Segal–Bargmann transform for Lie groups of compact type, *Can. J. Math.*, to appear.

[H6] B. C. Hall, Holomorphic methods in analysis and mathematical physics, lecture notes to be published by the Mexican Mathematical Society.

[HS] B. C. Hall and A. N. Sengupta, The Segal–Bargmann transform for path-groups, *J. Funct. Anal.* **152** (1998), 220-254.

[Ki] F. Kirwan, "Cohomology of quotients in symplectic and algebraic geometry," Mathematical Notes 31, Princeton Univ. Press, Princeton, N.J., 1984.

[KS] J. R. Klauder and B.-S. Skagerstam (Eds.), "Coherent states. Applications in physics and mathematical physics," World Scientific, Singapore, 1985.

[K] H.-H. Kuo, "Guassian measures in Banach spaces," Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, New York/Berlin, 1975.

[L1] N. P. Landsman, Rieffel induction as generalized quantum Marsden–Weinstein reduction, *J. Geom. Phys.* **15** (1995), 285–319.

[L2] N. P. Landsman, "Mathematical topics between classical and quantum mechanics," Springer-Verlag, New York/Berlin, 1998.

[LW] N. P. Landsman and K. K. Wren, Constrained quantization and θ-angles, *Nucl. Phys. B* **502** (1997), 537-560.

[MFK] D. Mumford, J. Fogarty, and F. Kirwan, "Geometric invariant theory," Third edition, Ergebnisse der Mathematik und ihrer Grenzgebiete 34, Springer-Verlag, New York/Berlin, 1994.

[P] A. Perelomov, "Generalized coherent states and their applications," Springer-Verlag, New York/Berlin, 1986.

[R] S. Rajeev, Yang-Mills theory on a cylinder, *Phys. Lett. B* **212** (1988), 203-205.

[S1] I. E. Segal, Mathematical problems of relativistic physics, Chap. VI. *in* "Lectures in applied mathematics. Proceedings of the summer seminar, Boulder, Colorado, 1960," Vol. II, (M. Kac, Ed) American Math. Soc, Providence, R.I., 1963, pp. 73-84.

[S2] I. E. Segal, Mathematical characterization of the physical vacuum for a linear Bose-Einstein field, *Illinois J. Math.* **6** (1962), 500-523.

[S3] I. E. Segal, The complex wave representation of the free Boson field, *in* "Topics in functional analysis. Essays dedicated to M.G. Krein on the occasion of his 70th birthday," (I. Gohberg and M. Kac, Eds.) Advances in Mathematics Supplementary Studies, Vol. 3, Academic Press, New York, 1978, pp. 321-343.

[Sen] A. Sengupta, The two-parameter Segal–Bargmann transform, preprint.

[Sj] R. Sjamaar, Symplectic reduction and Riemann–Roch formulas for multiplicities, *Bull. (N.S.) Amer. Math. Soc.* **33** (1996), 327-338.

[Wo] N. J. M. Woodhouse, "Geometric quantization," Second edition, Oxford Univ. Press, New York, 1992.

[W] K. K. Wren, Constrained quantization and θ-angles II, *Nucl. Phys. B* **521** (1998), 472-502.

Department of Mathematics, University of Notre Dame, Computing Center and Mathematics Building, Notre Dame, IN 46556 U.S.A.,

E-mail address: bhall@nd.edu