THE GEOMETRY AND PHYSICS OF KNOTS.

M.F. Atiyah

1. LINKING NUMBERS AND FUNCTIONAL INTEGRALS

1.1 INTRODUCTION

The aim of these lectures is to present a new approach to the Jones polynomial invariants of knots (Annals of Math. 1988) due to Witten ("Jones polynomial and quantum field theory" to appear in Proceedings IAMP Swansea 1988). They represent a very abbreviated version in which many subtle points have been omitted or only alluded to.

1.2 KNOTS AND LINKS IN $\mathbf{R}^3$

A knot is just an oriented closed connected smooth curve in $\mathbf{R}^3$. A general curve with possibly many components is referred to as a link. Knots may also be considered as embedded in S^3 or more generally in an arbitrary (compact, oriented) three dimensional manifold M^3.

The main problem is to classify knots by suitable invariants. The earliest attempt was the introduction by Alexander (1928) of a one variable polynomial knot invariant with integral coefficients. The Alexander polynomial is not a complete invariant for knots but is useful and readily computable. Moreover it can be constructed from standard techniques of algebraic topology (homology of a covering branched over the knot). One defect of the Alexander polynomial is that it fails to distinguish 'chirality', that is a knot and its mirror image have the same polynomial.

The Jones polynomial (1984) $V(q)$ is a finite Laurent series in q with the following properties.

1) It is chiral giving different values for example to the left handed and right handed trefoil knots.
2) It is associated with the Lie group $SU(2)$ and there are other polynomial invariants

associated with other Lie groups e.g. $V_N(q)$ for $SU(N)$.

3) $V(q)$ is related to integrable systems in $1+1$ dimensions regarded either from the view point of statistical mechanics or conformal field theory.

4) As yet it does not appear to be related to standard algebraic topology.

These properties pose the question of why integrable systems in 2 dimensions produce topological invariants in 3 dimensions. In 3 and 4 dimensions we have non-Abelian gauge theories which are known to be related to the topology of 3 and 4 manifolds and we might anticipate that they are also related to the Jones polynomial. I made this conjecture at the Hermann Weyl Symposium in 1987 and it was answered by Witten at the Swansea conference. We can also turn the question around and ask what is the relationship between solvable 2 dimensional models (conformal field theories) and topological gauge theories in 3 dimensions. Witten's work sheds some light on this.

1.3 WITTEN THEORY

Witten considers a special quantum field theory in $2+1$ dimensions. This quantum field theory produces expectation values of observables which are equal to the values

$$V_N(\exp \frac{2\pi i}{k+N})$$

of the Jones polynomial where k and N are integers. Given these values for general k the Jones polynomials $V_N(q)$ are determined.

The Witten field theory has a number of general features.

(i) It is almost a standard quantum field theory, i.e. the Lagrangian is basically one of the standard theories previously considered by physicists with a slight twist which we will come to later.

(ii) Witten's approach allows generalisations to all Lie groups and to all 3 manifolds. Hence it can be used to generate new mathematics.

(iii) The price for all this beauty is that the theory is not rigorous. However it is very computable. So we can calculate and check that the computed answers are consistent. It is enough to check the calculated values and how they change under certain elementary transformations. This is essentially what Jones did, the difference being that Witten's theory assigns a meaning to these rules. Consistency has not however been checked yet for all three manifolds.

(iv) A useful analogy in thinking of the relationship between Jones and Witten is to recall the Betti numbers of a manifold. Originally these were calculated via a triangulation of the manifold. A satisfactory understanding of their meaning however had to await the development of the general machinery of homology groups. Similarly one should think of Witten theory as providing a non-abelian quantum homology theory. The numerical invariants are set in a more general conceptual context which incorporates machinery for their computation.

(v) Witten's theory is an example of a topological quantum field theory (TQFT). There are now others, for example one explaining the Donaldson invariants of a 4 manifold. The precise description of TQFT's will not be given here, however they share a number of common features

a) they are related to non-abelian gauge theories,

b) the invariants appear as expectation values and,

c) they are tied to certain low dimensions.

(vi) TQFT's in 3 dimensions are related to rational conformal field theories in 2 dimensions.

1.4 A 5 MINUTE REVIEW OF QUANTUM FIELD THEORY

A relativistic quantum field theory in $d+1$ dimensions consists of a $d+1$ dimensional manifold M (space-time), some fields $\varphi(x)$ which depend on the points $x \in M$, a Lagrangian density $L(\varphi)$ and a Lagrangian

$$\mathcal{L} = \int_M L(\varphi)dx$$

which is a functional of φ.

The quantities of interest are calculated using the Feynman path integral (which is of course not rigorous) for example the partition function

$$Z = \int \exp(\frac{i}{\hbar}\mathcal{L}(\varphi))\mathcal{D}\varphi.$$

and the vacuum expectation values of an observable $W(\varphi)$

$$< W >= Z^{-1} \int \exp(\frac{i}{\hbar}\mathcal{L}(\varphi))W(\varphi)\mathcal{D}\varphi.$$

1.5 GAUGE THEORIES

A gauge theory in 3 dimensions depends on a compact Lie group G. The fields are connections, or gauge potentials which are one forms

$$A(x) = \sum_1^3 A_\mu(x)dx^\mu$$

where $A_\mu(x) \in LG$ the Lie algebra of G. The covariant derivative is defined by $D_\mu = \partial_\mu + A_\mu$ and the curvature is a two form

$$F = \sum F_{\mu\nu}dx^\mu dx^\nu$$

where $F_{\mu\nu} = [D_\mu, D_\nu]$.

It is important to remember that the infinite dimensional gauge group $\mathcal{G}$ consisting of maps from M into G acts naturally by conjugating the covariant derivatives. All interesting physics is meant to be invariant under this gauge action.

The most familiar Lagrangian for a gauge theory is the Yang-Mills (Y-M) Lagrangian which is the square of the L^2 norm of the curvature. This is quadratic in derivatives of the fields (connections) and therefore plays the role of a "kinetic energy term". However it is metric dependent whereas we are interested in Lagrangians which are metric independent in order to obtain solely topological information. To avoid using the volume form defined by the metric we look for a 3 form which is itself independent of the metric. There is essentially only one, the Chern-Simons form,

$$cs(A) = \mathrm{Tr}(A \wedge dA + \frac{2}{3}A \wedge A \wedge A)$$

well known to both mathematicians and physicists. The Chern-Simons Lagrangian is

$$\mathcal{L}_k(A) = \frac{k}{4\pi}\int_M \mathrm{Tr}(A \wedge dA + \frac{2}{3}A \wedge A \wedge A).$$

In physics a combination of the Yang-Mills and Chern-Simons Lagrangians are used but in Witten's theory we drop the Yang-Mills term.

Before considering the gauge invariance of this Lagrangian recall that the space of maps from a 3 manifold into a compact Lie group G is disconnected. The connected component of a map g is determined (for simply connected G) by an integer $\deg(g)$ called the degree of the map. We find that the Chern-Simons Lagrangian is invariant under the subgroup $\mathcal{G}_0$ of maps of degree zero and more generally if A^g denotes the connection A transformed by g then

$$\mathcal{L}_k(A^g) = \mathcal{L}_k(A) + 2\pi k \deg(g).$$

As the quantities of interest involve $\exp(i\mathcal{L}_k(A))$ they can only be gauge invariant if k is an integer.

The Chern-Simons form can be understood as follows. Let $\mathcal{A}$ be the affine space of all connections. The tangent space to $\mathcal{A}$ at a point A consists of 1 forms on M with values in the Lie algebra. The curvature F_A of A defines a linear map on such 1 forms η by

$$\eta \mapsto \int_M \mathrm{Tr}(\eta \wedge F_A)$$

and this defines a 1 form on $\mathcal{A}$. The Bianchi identity implies that this is a closed one form and hence it is the differential of a function on $\mathcal{A}$. This function is

$$A \mapsto \int_M cs(A).$$

Notice that all the preceding discussion depends critically on the dimension of M being 3.

The reader familiar with the theory of connections may wonder why the connection is a 1 form on M not on the total space of a principal bundle. However over a three manifold all G bundles (for G simply connected) are trivial and therefore we can choose a section and pull the connection form back to the base.

If we fix a $k \geq 1$ the Lagrangian $\mathcal{L}_k$ defines a quantum field theory and we want to consider the expectation values of observables. As there is no dependence on the metric we expect these to be topological invariants.

The partition function

$$Z(M) = \int \exp(i\mathcal{L}_k(A))\mathcal{D}A$$

is a complex number which is an invariant of M. For simple manifolds such as S^3 this will not be interesting but for more general manifolds it will be. Notice that k plays the role of $1/\hbar$.

To define observables let $K \subset M$ be a knot. Then given a connection A we can consider parallel transport, or monodromy around the knot which defines an element of G (up to conjugacy). If, in addition we specify a representation λ of G then we can define

$$W_K(A) = \mathrm{Tr}_\lambda(\mathrm{Mon}_K(A))$$

where the trace is taken after we represent the monodromy element using λ. Physicists call this a Wilson loop.

Taking the vacuum expectation gives

$$\begin{aligned} Z(M,K) &=< W_K(A) > Z(M) \\ &= \int \exp(i\mathcal{L}_k(A))W_K(A)\mathcal{D}A. \end{aligned}$$

More generally for a link with connected components $K_1, \ldots, K_p$ we form the product $\prod_1^p W_{K_i}(A)$ where we can choose different representations for each component.

Finally if we take $G = SU(N)$ and the standard representation then we find that

$$Z(M,K) = V_N(\exp(\frac{2\pi i}{k+N})).$$

2. STATIONARY PHASE APPROXIMATION.

2.1 INTRODUCTION

The partition function we are interested in has the form

$$Z(M) = \int \exp(ik\mathcal{L}(A))\mathcal{D}A$$

and we want to consider the expansion of this as $k \to \infty$.

2.2 FINITE DIMENSIONS

In finite dimensions we can consider an oscillatory integral of the form

$$\int_M \exp(itf(x))dx.$$

The idea is that for large t the integrand is oscillating so wildly that it cancels itself except where f has critical points. For simplicity assume that f has non-degenerate critical points $p_1, \dots, p_q$ then the more precise statement is that there is an asymptotic expansion

$$\int_M \exp(itf(x))dx \sim \sum_{i=1}^{q} \left(\sum_{-\infty}^{\dim M} a_n^i t^{n/2} \right)$$

where the sum is over the critical points. The leading term in this expansion involves the values of f and the matrix of second derivatives of f at the critical points. As we will consider only the leading term here it is enough to consider the case where f is a quadratic form. In the one variable case we have

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{\pi}} \exp(i\lambda x^2) = |\lambda|^{-1/2} \exp(\frac{\pi i}{4} \mathrm{sign}\lambda).$$

In the case of n variables where Q denotes the matrix of a non-degenerate quadratic form this becomes

$$\int_{-\infty}^{\infty} \frac{dx^1 \dots dx^n}{\pi^{n/2}} \exp(ix^t Qx) = |\det Q|^{-1/2} \exp(\frac{\pi i}{4} \mathrm{sign} Q)$$

where $\mathrm{sign}Q$ is the signature of Q.

Finally we need to consider the case where a group G acts on the space. For example consider $U(1)$ acting on $\mathbf{R}^2 - \{0\}$. Then if f is invariant under G (so is a function of the radial variable in the case of the plane) the matrix Q has zero eigenvalues along the G-orbit directions. Choose a transverse slice and write the integral in terms of a 'product' of the measure over the orbit times a measure over the transversal. This produces a Jacobian-like factor which when integrated gives the correct volume of the orbit. This contribution arises from a map B from LG to the tangent space to the orbit and gives the orbit volume as $(\det R)^{1/2} = |\det B|$ where $R = B^*B$. Thus in general the leading term in the expansion is

$$\frac{(\det R)^{1/2}}{|\det Q_0|^{1/2}} \exp(\frac{\pi i}{4} \mathrm{sign} Q_0)$$

where Q_0 denotes the part of the quadratic form which is non-degenerate i.e. transverse to the orbit.

2.3 FIELD THEORY

Now we do the stationary phase approximation for a field theory. As a simple first case consider the functional integral

$$\int \exp(i < \Delta\varphi, \varphi >)\mathcal{D}\varphi$$

where Δ is a positive Laplace type operator on the compact manifold M. By the above we expect an answer of the form $(\det \Delta)^{-1/2}$. There is a standard method of regularising such determinants due to Ray and Singer which defines them in terms of the zeta function:

$$\zeta_\Delta(s) = \sum \lambda^{-s}.$$

This function is clearly analytic for the real part of s large and possesses an analytic continuation to zero which is a regular point. Then we define

$$\det \Delta = \exp(-\zeta'_\Delta(0)).$$

Note that for a constant k

$$\det k\Delta = k^{\zeta_\Delta(0)} \det \Delta$$

gives the scaling behaviour. For odd-dimensional manifolds, $\zeta_\Delta(0) = 0$ so that $\det k\Delta = \det \Delta$.

Next we need to consider the case where the operator in the exponent is self adjoint but is not positive definite, for example, if it is a Dirac operator D (which has both positive and negative eigenvalues). We can certainly define the absolute value as

$$|\det D| = (\det D^* D)^{1/2}.$$

The phase is defined by considering the η-function introduced by Atiyah, Patodi and Singer

$$\eta_D(s) = \sum |\lambda|^{-s} \text{sign}\lambda.$$

As with the zeta function the η-function possesses an analytic continuation to the regular point $s = 0$ and we define

$$\text{sign}D = \eta_D(0).$$

Now consider a gauge theory. We have to take account of the contribution of the gauge orbits. The Faddeev-Popov prescription is to define the volume of a gauge orbit as $(\det R)^{1/2}$ where R is determined as follows.

The Lie algebra of the gauge group is the space $\Omega^0(LG)$ of functions on the manifold with values in the Lie algebra of G. The tangent space to $\mathcal{A}$ at a connnection A is the space $\Omega^1(LG)$ of one forms with values on LG. We want the map B which maps an element of the Lie algebra of the gauge group to the tangent space. By considering an infinitesimal gauge transformation it is easy to check that B is the covariant derivative

$$d_A : \Omega^0(LG) \to \Omega^1(LG).$$

Hence the volume of the $\mathcal{G}$ orbit through A is

$$\det R = \det d_A^* d_A = \det \Delta_A^0.$$

(We denote the Laplacian on r-forms by Δ^r).

Now we have regularised all the terms and hence the leading term in the asymptotic expansion. Next we turn to Witten's theory.

2.4 APPLICATION TO THE CHERN-SIMONS LAGRANGIAN

To do the stationary phase approximation to $\mathcal{L}_k(A)$ we need first to find the critical points. But it is easy to see that these are precisely the flat connections A i.e. those for which $F_A = 0$. Any flat connection determines a representation of the fundamental group in G and because of the $\mathcal{G}$ action we need consider only equivalence classes of such representations. Suppose that $A_{\alpha_j}, j = 0, \ldots, m$ are the flat connections and denote by $\alpha_0, \ldots, \alpha_m$ the corresponding representations. Assume that α_0 is the trivial representation and that the other critical points for $j \neq 0$ are all non-degenerate in a sense that we shall define later. The stationary phase approximation then gives us a sum of terms

$$\int \exp(i\mathcal{L}_k(A))\mathcal{D}A \sim b_{\alpha_0} + \sum_{i=1}^{m} b_{\alpha_i}$$

where we isolate the trivial connection as it poses some problems that we shall not deal with here.

To leading order in the stationary phase approximation only quadratic terms are important and we get

$$\int \mathrm{Tr} A \wedge d_\alpha A =< A, *d_\alpha A >$$

where d_α is d twisted by the flat connection α. Hence the self adjoint operator is $*d_\alpha : \Omega^1(LG) \to \Omega^1(LG)$ so that $Q = *d_\alpha$. Using the formula of Section 2.3 a little calculation enables us to identify the terms for a general, non-trivial representation (connection) α as

$$\frac{(\det \Delta^0_\alpha)^{3/4}}{(\det \Delta^1_\alpha)^{1/4}} \exp(\frac{\pi i}{4}\eta_L(0))$$

where L is the operator $*d + d*$ on odd forms and therefore L^2 is $\Delta^1 + \Delta^3$. To see this it is best to decompose $\Omega^1(LG)$ as the sum of $d^*_\alpha(\Omega(LG))$ and $\mathrm{Ker} d_\alpha$ (the non-degeneracy of α means that there is no cohomology for d_α).

2.5 TOPOLOGICAL INVARIANCE

The Lagrangian $\mathcal{L}_k$ has been chosen to be independent of the metric. However to perform the calculations above we have used a metric in many places. As in other problems in physics where special choices are made to perform a calculation it is not clear that the end result is metric invariant.

If we square the first piece of the expression we obtain

$$\frac{(\det \Delta^0_\alpha)^{3/2}}{(\det \Delta^1_\alpha)^{1/2}} = T_\alpha$$

which has been proved by Ray and Singer to be independent of the metric. The proof consists of calculating the variation of T_α under an infinitesimal change in the metric and showing that this is zero. They conjectured that this was the same as the Reidemeister torsion which is constructed from a ratio of determinants of combinatorial Laplacians obtained from a triangulation. This was proved by Cheeger and Müller. The Reidemeister torsion is intimately related to the Alexander polynomial.

More correctly this was shown for a non-degenerate connection α , that is, one for which the complex

$$\Omega^0(LG) \xrightarrow{d_\alpha} \Omega^1(LG) \xrightarrow{d_\alpha} \Omega^2(LG) \xrightarrow{d_\alpha} \Omega^3(LG)$$

has no cohomology. Note that this is a sensible thing to ask because, for a flat connection α, $d_\alpha^2 = 0$. This explains what we meant above by a non-degenerate critical point of $\mathcal{L}_k$.

It follows that $T_\alpha^{1/2}$ is a topological invariant.

For the term $\exp(\frac{\pi i}{4}\eta_L(0))$ we have to use the Atiyah-Patodi-Singer index theorem. This says that $\eta_{\alpha_j} - \eta_{\alpha_0}$ is a topological invariant (mod $\mathbb{Z}$) independent of the metric, in fact (for $G = SU(N)$)

$$\eta_{\alpha_j} - \eta_{\alpha_0} = \frac{N}{\pi} I_{\alpha_j}$$

where $\mathcal{L}_k(\alpha) = \frac{k}{4\pi} I_\alpha$.

So we have

$$Z(M) \sim \exp(\frac{\pi i}{4}\eta_{\alpha_0}) \left(\sum_j \exp i(k+N) I_{\alpha_j} T_{\alpha_j}^{1/2} \right).$$

This leaves the term $\exp(i\pi\eta_0/4)$. We deal with this by adding a counterterm to the original Lagrangian (this is a standard trick in field theory). This term is chosen to be of the same general form as the other terms in the Lagrangian and almost cancels the term above after applying the stationary phase approximation. However at the end there is still a finite discrete dependence on the metric. This is removed by choosing a homotopy class of framing F for the manifold M. We will not go into the details of this but it is an important subtlety in the Witten theory.

To define the counterterm consider an oriented 4-manifold X. Then the intersection form on two dimensional homology has a signature which is referred to as the signature of X and is a topological invariant equal to one third of the Pontrjagin class of X. If X is a manifold with boundary M and we choose a framing F for the boundary there is a relative Pontrjagin class $p_1(X, F)$, and

$$\sigma(M, F) = \operatorname{sign} X - \frac{1}{3} p_1(X, F)$$

is easily seen to be independent of X (given two choices glue along M) and is an invariant of (M, F). The corrected topologically invariant formula for the large k limit of $Z(M, F)$ is then

$$Z(M) \sim \exp(\frac{\pi i}{4}\sigma(M, F)) \left(\sum_j \exp i(k+N) I_{\alpha_j} T_{\alpha_j}^{1/2} \right).$$

Finally let us make two remarks. Firstly if the orientation of M is reversed $Z(M, F)$ is complex conjugated. Thus it is essential that $Z(M, F)$ is not real in order that changes in chirality are detected. Note that the Reidemeister torsion piece of $Z(M, F)$ is not sensitive to orientation but the η invariant changes sign because the positive and negative eigenvalues of a Dirac type operator are interchanged if the orientation is reversed.

Secondly we are of course interested in $Z(M, K)$ where K is a knot. We have been looking at one extreme case $Z(M, \emptyset)$. If we consider the other extreme case $Z(S^3, K)$, for example for $G = U(1)$ then we get the classical Gauss formula for the linking number of two curves expressed as a double integral of a Green's function over the product of the curves. To regularize the self-linking number of a knot a normal framing is needed, so as to push the knot away from itself (a process referred to by physicists as point-splitting).

3. HAMILTONIAN APPROACH

3.1 RELATION BETWEEN THE LAGRANGIAN AND HAMILTONIAN FORMULATIONS

As before we start with a Lagrangian $\mathcal{L}(\varphi)$. Now we think of our manifold M as a product of a manifold X representing the space directions and the interval $[0, T]$ representing the time variable. In the context of the Witten theory this situation arises from cutting M along a Riemann surface X. Then $X \times [0, T]$ is an approximation to M near the cut. In the functional integral approach the Hamiltonian enters when we consider the transition amplitude between initial and final states φ_0 and φ_T:

$$\int_{\varphi_0}^{\varphi_T} \exp(\frac{i}{\hbar}\mathcal{L}(\varphi))\mathcal{D}\varphi.$$

If we introduce the Hilbert space of states $\mathcal{H}$ and the Hamiltonian operator H on $\mathcal{H}$ which is the generator of time translations this functional integral is given by

$$< \exp(iTH)\varphi_0, \varphi_T > .$$

If one now imposes periodic boundary conditions in the time variable and sums over all states one obtains the relation:

$$\int \exp(\frac{i}{\hbar}\mathcal{L}(\varphi))\mathcal{D}\varphi = \mathrm{Trace}(\exp(iTH))$$

where the integral is over all fields φ on the manifold $X \times S^1$ where S^1 is $[0, T]$ with the endpoints identified. This expression can be made more plausible by the procedure of going to imaginary time (i.e. replacing a Minkowskian field theory by a Euclidean one) in which case the right hand side becomes the partition function (of statistical mechanics) $\mathrm{Trace}\exp(-TH)$.

On ground states φ the Hamiltonian H is zero and T is therefore irrelevant. A topologically invariant theory is independent of the size of the circle and therefore also time independent. In TQFT's then we expect that the Hamiltonian is zero, that is, there is no dynamics. However there is still something interesting in the theory. Associated to the manifold X is a Hilbert space of the theory $\mathcal{H}_X$ and as H is trivial the trace gives us

$$Z(X \times S^1) = \dim \mathcal{H}_X.$$

This indicates that $\mathcal{H}_X$ should be finite dimensional. If we take a diffeomorphism f of X then it should act also on the Hilbert space $\mathcal{H}_X$ and we are considering the case where we take $X \times [0, 1]$ and identify the X's at the endpoint using f (denote this space by X_f). This means we have periodic boundary conditions twisted by f and the partition function is

$$Z(X_f) = \mathrm{Trace}(f|_{\mathcal{H}_X}).$$

This depends only on the isotopy class of f. If for instance we consider $S^1 \times S^1$ and let f be given by an element of $SL(2, \mathbf{Z})$ then the partition function Z defines a character of $SL(2, \mathbf{Z})$. This gives rise to character formulae.

So there are interesting things happening even when the dynamics are trivial.

3.2 THE HILBERT SPACE OF WITTEN'S THEORY

Let us fix a G and for convenience take it to be $SU(N)$ and also fix k. Given a closed surface X we want to get a finite dimensional Hilbert space $\mathcal{H}_X$ and in particular to determine its dimension.

Consider connections on X. We will use the same notation as in section 2 but now everything is defined over a two dimensional X rather than a three dimensional M. The space $\mathcal{A}$ of all connections has a natural symplectic form. If α and β are two tangent vectors at A then they are one forms on X with values on LG and we define

$$(\alpha, \beta)_k = \frac{k}{4\pi^2} \int_X \mathrm{Tr}(\alpha \wedge \beta).$$

This is the natural object in two dimensions, just as in three dimensions there was the one form given by the curvature. Note that the group $Diff^+(X)$ of orientation preserving diffeomorphisms preserves this symplectic form. (This holds for $G = U(1)$ directly, but for non-abelian G the relevant group is a semidirect product of $\mathcal{G}$ and $Diff^+(X)$.)

Recall that in finite dimensional classical mechanics we have a phase space $\mathbf{R}^{2n}$ with co-ordinates $q_1, \ldots, q_n$ and $p_1, \ldots, p_n$, the positions and conjugate momenta and the quantization of this is the Hilbert space $L^2(\mathbf{R}^n)$. On this the observables q_i are represented by multiplication and the p_i by differentiation. The special thing about the p, q co-ordinates is that the symplectic form is

$$\sum dq_i \wedge dp_i.$$

In principal we can apply this procedure to $\mathcal{A}$ with the symplectic form we have defined and get a big Hilbert space $\mathbf{H}_k$. The gauge group $\mathcal{G}$ is meant to preserve all the physically interesting things so that it acts projectively on the Hilbert space $\mathbf{H}_k$. The physical "part"of this Hilbert space is the subspace of vectors left invariant under the action of $\mathcal{G}$. This defines a finite dimensional Hilbert space $\mathcal{H}_{X,k}$.

There is a more direct way to determine $\mathcal{H}_{X,k}$. There is a smaller phase space, the reduced phace space, which when quantized gives rise to $\mathcal{H}_{X,k}$. If we consider the moment map

$$\mu : \mathcal{A} \to L(\mathcal{G})^*$$

the reduced phase space is defined to be the quotient

$$\frac{\mu^{-1}(0)}{\mathcal{G}}.$$

In finite dimensional examples this reduces the dimension by twice the dimension of the group. In the infinite dimensional case we are considering the reduced phase space is a finite dimensional, compact, symplectic manifold (possibly with singularities that we will ignore.) For example if $G = U(1)$ then $\mathcal{A} = \Omega^1$ and $L\mathcal{G} = \Omega^0$. The dual of $L\mathcal{G}$ is naturally Ω^2 and the moment map is the exterior derivative. The reduced phase space is the first de Rham cohomology space of X. In the Witten case the moment map is (up to a constant)

$$\mu(A)(\xi) = \int \mathrm{Tr} F_A \xi.$$

The reduced phase space $\mathcal{M}$ is therefore the set of isomorphism classes of flat connections on the surface X. In the usual way these correspond to the space of representations of the fundamental group of X. Recall that for a Riemann surface such as X the fundamental group has a nice finite presentation and therefore the space of representations is finite dimensional and compact. The singular points of $\mathcal{M}$ are the reducible connections.

The Hilbert space $\mathcal{H}_{X,k}$ is therefore the quantization of $\mathcal{M}$ using the symplectic form induced by the symplectic form $(\,,)_k$ on $\mathcal{A}$.

3.3 QUANTIZING $\mathcal{M}$ VIA ALGEBRAIC GEOMETRY

One way of quantizing $\mathcal{M}$ is to choose a complex structure on X. This induces a complex structure on $\mathcal{M}$. For example if $G = U(1)$ then $\mathcal{M}$ is the space of all topologically trivial holomorphic line bundles or the Jacobian of X. In the general case (for $G = U(N)$) $\mathcal{M}$ is the moduli space of rank N vector bundles on X.

The symplectic form when properly normalised is an integral class and therefore represents the first chern class of a line bundle $\mathcal{L}$. This line bundle is holomorphic and the quantization of $\mathcal{M}$ is the space of all holomorphic sections, that is

$$\mathcal{H}_{X,k} = H^0(\mathcal{M}, \mathcal{L}^k).$$

As we want topologically invariant objects we have to examine how these constructions depend on the choice of complex structure. The space of all complex structures on X is the Teichmuller space and the spaces of holomorphic sections define a bundle of Hilbert spaces over this. These are all expected to be projectively isomorphic that is the corresponding projective bundle is trivial. This Hilbert bundle has a naturally defined connection whose curvature is a scalar multiple of the Kahler form on Teichmuller space and therefore gives a natural trivialization of the projective bundle.

This completes our discussion of the general methods.

3.4 FURTHER DEVELOPMENTS

a) If we want to insert a knot into the theory then it intersects the surface X in some points $p_1, \ldots, p_r$ and associated to these we have representations $\lambda_1, \ldots, \lambda_r$ and the Hilbert space should depend on these. By evaluating an element of $\mathcal{G}$ at each of these points and applying the appropriate representation and taking a tensor product this extra data defines a representation of $\mathcal{G}$. The Hilbert space of interest then should be the subspace of the big Hilbert space $\mathbf{H}$ which transforms according to this representation.

From the viewpoint of algebraic geometry we obtain generalised moduli spaces that have only recently been described. There we look at representations of the fundamental group of X with these points deleted which applied to loops around the points give particular conjugacy classes of order k defined by the λ_i.

To take the extreme case let $X = S^2$ and λ_i the standard representation of $G = SU(N)$. The group of orientation preserving diffeomorphisms that fix the set of points acts on the big Hilbert space and this is related to the braid group.

b) To define the bundle of Hilbert spaces over the moduli space of Riemann surfaces we need to look at the boundary which is made up of degenerate curves of lower genus. Much of the detail of this has been worked out.

c) If the marked points are expanded into holes ands we take the boundary values of functions defined on X then we get representations of the Virasoro algebra. This path leads back to conformal field theory.

d) Some work of N. Hitchin seems to be closely related to all these things. He considers the moduli space $\mathcal{M}$ as a fiber in a fibering over a vector space in which the generic fibres are abelian varieties. $\mathcal{M}$ is then a sort of limit of abelian varieties. This should lead to a relationship with abelian theory except that we have to deal with the monodromy around the point of the base corresponding to $\mathcal{M}$. The conjecture is that $\mathcal{H}_{X,k}$ is the subset of the Hilbert space of an abelian theory stable under the monodromy action. This is analogous to considering representations of a compact group as those of the abelian maximal torus invariant under the Weyl group.

(Notes taken by A. L. Carey and M. K. Murray.)

Mathematical Institute
24-29 St. Giles
Oxford, OX13LB U.K.