Commun. Math. Phys. 126, 167–199 (1989)

Communications in
Mathematical
Physics

The Chern-Simons Theory and Knot Polynomials

J. Fröhlich[1] and C. King[2]

[1] Theoretical Physics, ETH-Hönggerberg, CH-8093 Zürich, Switzerland
[2] Department of Mathematics, Cornell University, Ithaca, NY 14853, USA

Abstract. The Chern-Simons gauge theory is studied using a functional integral quantization. This leads to a differential equation for expectations of Wilson lines. The solution of this differential equation is shown to be simply related to the two-variable Jones polynomial of the corresponding link, in the case where the gauge group is $SU(N)$. A similar equation has been used before to get the Jones polynomial from a braid representation of the link. The main novelty of our approach is that we get the Jones polynomial from a plat representation of the link.

1. Introduction

The purpose of this paper is to explore the connection between the non-abelian Chern-Simons gauge theory and knot polynomials. The motivation for this work comes, in part, from Witten's beautiful paper [1] in which he shows that the expectation of a collection of Wilson loops in the Chern-Simons theory is related to the Jones polynomial of the corresponding link.

Our approach is different from that in [1]. We start from the classical action of Chern-Simons theory in Minkowski space and impose a particular gauge condition which we call *light-cone gauge*. In this gauge the action is quadratic in the gauge potential. The Feynman path integral then formally yields the propagator for the gauge potential in light-cone gauge, up to an overall normalization constant, denoted λ. Next, we complexify space-time and analytically continue the propagator from the Minkowski space to the Euclidean region. In the Euclidean region we may then derive simple differential equations for the expectations of Wilson lines in light-cone gauge with coefficients depending on λ. These equations are similar to equations studied previously in connection with representations of the braid groups [2, 3] and the Wess-Zumino-Witten models of two-dimensional conformal field theory, [4]. If we now require that the solutions of our differential equations for expectations of Wilson lines are compatible with unitarity (reflection positivity) of the theory then the parameter λ is constrained to take the values $(k+c_2(G)/2)^{-1}$, $k=1,2,3,\ldots$, where G is the gauge group, and $c_2(G)$ is the eigenvalue of the quadratic Casimir operator in the adjoint representation. This follows from the results in [2] and will be discussed in more detail elsewhere. We discuss the equation more explicitly in the case where the gauge group is $SU(N)$ and the Wilson loops are chosen in the fundamental representation. Not surprisingly, the solutions are simply related to the two-variable generalization of the Jones polynomial [5]. We show that the solution satisfies the skein relations of this polynomial, and we show that it is a link invariant. These two properties characterize it uniquely, up to an overall multiplicative constant. In order to prove that it is a link invariant, we find what Birman calls a plat representation for the Wilson loops [6]. Birman found a collection of plat moves which serve the same purpose as the Markov moves for a braid representation. We prove invariance of the solution under these moves, which implies that it is a link invariant.

For groups other than $SU(N)$ or for higher dimensional representations of $SU(N)$, we believe that the solution of our equation is again a link invariant. However the skein relations for such an invariant are much more complicated than for the Jones polynomial, involving crossings with several twists (see [7] for a discussion of this). It is not clear that these invariants can be computed in any simple way from our differential equations.[1]

The paper is organized as follows. In Sect. 2 we relate the Chern-Simons theory to the differential equations for Wilson line expectations. In Sect. 3 this equation is completely solved in a simple but non-trivial case. Sections 4 and 5 contain our results about the general case, and establish the skein relations which define the Jones polynomial. In the appendix we prove the main result about the regularity of solutions of our differential equations.

From the point of view of this paper, it is the differential equations for Wilson line expectations, the so-called Knizhnik-Zamolodchikov equations [4], that are fundamentally related to link-invariants. The circumstance that these equations can be derived from a gauge theory, the Chern-Simons theory, is interesting mainly because it provides an intrinsically three-dimensional interpretation of link polynomials, [1].

2. The Differential Equation for Wilson Line Expectations

The action of Chern-Simons theory for a gauge group G is given by

$$S[A]=\frac{k}{4\pi}\int \operatorname{tr}\left(A\wedge dA+\frac{2}{3}A\wedge A\wedge A\right).$$

The connection A is a Lie G-valued 1-form on three-dimensional space-time $\mathbb{M}^3$ (or Euclidean space-time $\mathbb{E}^3$). For $G=U(1)$, the cubic term in $S[A]$ is absent, and the path integral for the abelian Chern-Simons theory is given by the Gaussian measure

$$Z^{-1}\exp(iS[A])\,[DA],$$

[1] For alternative approaches, where equations between so-called braid- and fusion matrices are used, see, however [8]

where the partition function, Z, is formally defined by

$$Z = \int \exp(iS[A])\,[DA]\,.$$

Let $\sigma(t)$ be a closed parametrized curve in $\mathbb{E}^3$. The Wilson loop observable corresponding to σ is

$$W(\sigma; A) = \exp(i \oint \mathbf{A} \cdot d\boldsymbol{\sigma})\,.$$

The expectation $\langle W(\sigma)\rangle$ of the Wilson loop observable is given by the Gaussian integral

$$\langle W(\sigma)\rangle = Z^{-1} \int W(\sigma; A) \exp(iS[A])\,[DA]$$

and yields the Gauss "self-linking number" of σ; see e.g. [1]. In order to evaluate $\langle W(\sigma)\rangle$, it is necessary to regulate $W(\sigma; A)$. One may, for example, replace $W(\sigma; A)$ by

$$W(j, A) = \exp(i \int \mathbf{j}(x) \cdot \mathbf{A}(x) d^3x)\,, \qquad \nabla \cdot \mathbf{j} = 0\,,$$

where the current $\mathbf{j}$ is bounded, with support in a tubular neighbourhood of σ. Then

$$\langle W(j)\rangle = Z^{-1} \int W(j; A) \exp(iS[A])\,[DA]$$

can be evaluated unambiguously. If the support of $\mathbf{j}$ consists of several disjoint components then $\langle W(j)\rangle$ can be evaluated in terms of the Gauss linking numbers of those components and of "self-linking numbers" of the individual components which, however, depend on the precise choice of $\mathbf{j}$, a dependence that persists when $\mathbf{j}$ tends to the distributional current with support on σ. Thus, the expectation $\langle W(\sigma)\rangle$ is a topological invariant, albeit one which depends on the regularization scheme.

Note that, in the abelian Chern-Simons theory the "coupling constant" k in the action may be an arbitrary real number.

The method of calculating $\langle W(\sigma)\rangle$ outlined above for the abelian Chern-Simons theory does *not* generalize to the non-abelian theory, both because the action is no longer quadratic and because the Wilson loop observable is not the exponential of a linear functional in A. Let G be a non-abelian, compact simple Lie group. Under a gauge transformation $g: \mathbb{M}^3 \to G$ which tends to the identity at infinity, continuously, the connection A transforms according to

$$A^g = g^{-1} A g + g^{-1} dg\,,$$

and the action evaluated at A^g is

$$S[A^g] = S[A] + 2\pi k n_g\,,$$

where n_g is the winding number of the map g. [Note that for any non-abelian, compact simple Lie group G, $\pi_3(G) = \mathbb{Z}$.] Furthermore, under a diffeomorphism Φ of $\mathbb{M}^3$ which tends to the identity at infinity, the connection becomes $\Phi^* A$, and

$$S[\Phi^* A] = S[A]\,.$$

Similarly, the Wilson loop observables, i.e. the path ordered exponentials of A corresponding to loops, are gauge- and diffeomorphism-invariant. For these reasons, the Chern-Simons theory is called a topological quantum field theory.

We are interested in the Feynman path integral defined by $S[A]$, whose partition function is formally given by

$$Z = \int \exp(iS[A])\,[DA].$$

From the transformation properties of $S[A]$ under gauge transformations of A it follows that $\exp(iS[A])$ is gauge-invariant, *provided* the "coupling constant" k is an *integer*. Moreover, $\exp(iS[A])$ is diffeomorphism invariant. It might therefore appear that Z is independent of the metric on space-time. However, the regularizations needed to define the functional integral break general covariance; in particular, the definition of $[DA]$ requires choosing a metric on space-time.

In order to evaluate the path integrals of Chern-Simons theory, such as its partition function Z, gauge fixing is necessary. [For normalized expectations of the *abelian* theory, gauge fixing can be avoided, because the integrals are Gaussian.] We choose a gauge that renders the action *quadratic, even for the non-abelian theories.* We call it light-cone gauge, for reasons that will be evident.

A point in $\mathbb{M}^3$ is denoted by $x=(x^0, x^1, x^2)$, where x^2 is the time-coordinate of x. The 0-direction is called "transfer direction."

We introduce light-cone coordinates

$$x^+ = x^1 + x^2, \qquad x^- = x^1 - x^2.$$

The corresponding components of the connection A are denoted by A_+ and A_-, with

$$A_\pm = A_1 \pm A_2.$$

Here A_i can be written as

$$A_i = \sum_{a=1}^{r} A_i^a \frac{T_a}{\sqrt{2}},$$

where A^a is a real-valued 1-form on space-time and $\{T_a\}_{a=1}^r$ are an orthonormal set of generators of the Lie algebra of G, so that

$$\mathrm{tr}(T_a T_b) = -\delta_{ab}.$$

The light-cone gauge is defined by setting A_-^a to zero, for all $a=1,\ldots,r$. This is not a complete gauge fixing, since gauge transformations which only depend on x^0 and x^+, i.e. are independent of x^-, preserve the light-cone gauge condition.

The Chern-Simons action in light-cone gauge is given by

$$S_{\mathrm{l.c.}}[A] = \frac{k}{4\pi} \int \mathrm{tr}(A_+ \partial_- A_0 - A_0 \partial_- A_+),$$

where $2\partial_- = \partial_1 - \partial_2$. Notice that the action has no derivatives in the transfer direction, and that the cubic term vanishes in this gauge, so the action is quadratic.

In order to complete our definition of the light-cone gauge, we must specify the correct interpretation of the formal "integration measure" $[DA]$. Let $\mathfrak{A}$ denote the affine space of all connections $A=(A_0, A_+, A_-)$. This space carries a metric given by

$$\langle \delta A, \delta A \rangle = \langle \delta A_0, \delta A_0 \rangle + \langle \delta A_+, \delta A_- \rangle,$$

where

$$\langle \delta a, \delta a \rangle = \int_{\mathbb{M}^3} \mathrm{tr}(\delta a \cdot \delta a)\, d\,\mathrm{vol},$$

for a Lie G-valued function δa on space-time. To define $d\,\mathrm{vol}$, we choose the usual Lorentz metric on $\mathbb{M}^3$.

The integration measure $[DA]$ is then defined to be the formal volume form associated with the metric on $\mathfrak{A}$ defined above. The equation $A_-=0$ determines a surface, $\mathscr{S}_{\mathrm{l.c.}}$, in $\mathfrak{A}$ which intersects all gauge orbits. A general connection in $\mathfrak{A}$ can be represented as a gauge transformation of a connection in $\mathscr{S}_{\mathrm{l.c.}}$. Thus we may introduce coordinates (A_0, A_+, g) on $\mathfrak{A}$, with $A_-=0$. Then $[DA]$ is proportional to

$$[DA_0]\,[DA_+]\,[Dg],$$

where $[DA_0]$ and $[DA_+]$ are formal Lebesgue measures, and $[Dg]$ is the volume form on the group of gauge transformations. The Jacobian, i.e. the Faddeev-Popov determinant, turns out to be independent of A and g. After these preparations, we can proceed to calculate the propagators of Chern-Simons theory in the light-cone gauge.

Let D be the Feynman propagator of the two-dimensional d'Alembertian

$$\Box=\partial_2^2-\partial_1^2=-4\partial_+\partial_-,$$

with $2\partial_+=\partial_1+\partial_2$, $2\partial_-=\partial_1-\partial_2$. Then

$$\langle T(A_+^a(x)A_+^b(y))\rangle=0,$$
$$\langle T(A_0^a(x)A_0^b(y))\rangle=0,$$
$$\langle T(A_+^a(x^0,x^+,x^-)A_0^b(y^0,y^+,y^-))\rangle=\kappa\delta^{ab}\delta(x^0-y^0)\partial_+D(x_\perp-y_\perp),$$

with $x_\perp=(x^1,x^2)$. The possible values of κ are determined by the requirement that expectation values of gauge-invariant operators satisfy unitarity. Since the Chern-Simons action in light-cone gauge is quadratic, all Green functions are completely determined by the two-point functions.

In order to calculate the expectations of Wilson loop observables, it is convenient to perform a Wick rotation. The propagators are analytically continued in the time variables to the imaginary axis. A point in the Euclidean region is written as (t,z), with $t\equiv x^0$, $z\equiv x^1+ix^2$, x^0, x^1, x^2 real. The *formal* Euclidean versions of A_+ and A_- are denoted by α and $\bar\alpha$, respectively. Then we find that

$$\left.\begin{aligned}&\bar\alpha\equiv 0\\&\langle\alpha^a(t,z)\alpha^b(s,w)\rangle=0\\&\langle A_0^a(t,z)A_0^b(s,w)\rangle=0\\&\langle\alpha^a(t,z)A_0^b(s,w)\rangle=4\lambda\delta^{ab}\delta(t-s)\frac{1}{z-w},\end{aligned}\right\}\tag{2.1}$$

where $\lambda=-\dfrac{\kappa}{16\pi}$. All n-point Euclidean Green functions are determined by (2.1). It turns out that unitarity of the theory requires that

$$\lambda=-(k+c_2(G)/2)^{-1},\qquad k=1,2,3,\ldots.$$

This follows from (2.11) below and the results in [2] and will be discussed in more detail elsewhere. Since the action of Chern-Simons theory in light-cone gauge is quadratic, one might think that the overall scale λ of the two-point Green functions is a free parameter. As a trace of the non-linearity of the theory one finds that unitarity imposes the quantization of λ described above. It would be desirable to derive this fact directly from a more careful study of the path integrals defining Chern-Simons theory (rather than by using the results in [2]).

Now we will use the two-point functions in (2.1) to compute the Wilson loop expectations. The method we use was developed in [9] and [10] for the computation of loop expectations in two-dimensional Yang-Mills theory, in axial gauge. We recall the method below.

In light-cone gauge, the equation for parallel transport along a curve $\sigma(t) = (t, z(t))$ is

$$du(t) = (\tfrac{1}{2}\alpha dz + A_0 dt)u(t), \tag{2.2}$$

where we assume $0 \leqq t \leqq 1$, and $u(0)$ is the identity on $\mathbb{C}^N$. In order to compute expectations involving these parallel transport operators, we make an assumption about the meaning of the differential $du(t)$ in (2.2). Specifically, we assume that the increment $(\frac{1}{2}\alpha dz + A_0 dt)$ depends only on the fields supported in the region $\{(s,z)|s>t\}$, and that $u(t)$ depends only on the fields supported in the region $\{(s,z)|s \leqq t\}$. Since the Chern-Simons action in light-cone gauge has no time derivatives, the fields in these two regions are independent, and hence the increment $(\frac{1}{2}\alpha dz + A_0 dt)$ is independent of $u(t)$. This is the crucial fact that allows us to derive our differential equation; [if (2.2) were a genuine stochastic differential equation, our assumption would mean that $u(t)$ is given by an Ito integral].[2]

First we define new variables for the curve $\sigma(t)$:

$$L(t) = \int_0^t \alpha(s, z(s))dz(s),$$

$$M(t) = \int_0^t A_0(s, z(s))ds.$$

Then (2.2) becomes

$$du(t) = (\tfrac{1}{2}dL(t) + dM(t))u(t). \tag{2.3}$$

Now suppose that we have n curves $\sigma_1(t), \ldots, \sigma_n(t)$ each given in the form

$$\sigma_i(t) = (t, z_i(t)) \qquad i = 1, \ldots, n,\ 0 \leqq t \leqq 1.$$

We assume that the curves never intersect so that $z_i(t) \neq z_j(t)$ for different i and j, and all $0 \leqq t \leqq 1$; see Fig. 2.1. The parallel transport operators along $\sigma_1, \ldots, \sigma_n$ will be denoted $u_1, \ldots, u_n$, respectively, and are defined by (2.3) with new variables $\{L_i(t), M_i(t)\}$, $i = 1, \ldots, n$. We will now compute a differential equation for the operator

$$\varphi_n(t) = \langle u_1(t) \otimes \ldots \otimes u_n(t) \rangle, \tag{2.4}$$

where the expectation is taken with respect to the Chern-Simons measure in the light-cone gauge. Since $u_i(t)$ is an operator on $\mathbb{C}^N$, $\varphi_n(t)$ maps the tensor product $(\mathbb{C}^N)^n$ into itself.

---

[2] There are other interpretations of the meaning of $du(t)$. However, they change our equations for the expectation values of Wilson lines, see (2.7), only by terms that yield a different dependence on the "framing" of the Wilson lines and drop out in (2.15)

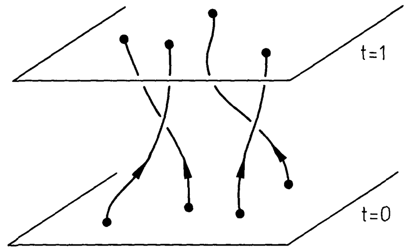


Fig. 2.1. Wilson lines in $\mathbb{R} \times \mathbb{C}$

We use (2.3) to get the differential equation for $\varphi_n(t)$. We take the differential of (2.4); in order to include all terms of order dt, we must include the second order differentials of the right-hand side. This gives

$$d\varphi_n(t) = \sum_{i=1}^{n} \langle u_1(t) \otimes \ldots \otimes du_i(t) \otimes \ldots \otimes u_n(t) \rangle$$
$$+ \sum_{1 \leq i < j \leq n} \langle u_1(t) \otimes \ldots \otimes du_i(t) \otimes \ldots \otimes du_j(t) \otimes \ldots \otimes u_n(t) \rangle .$$

By our assumption regarding the meaning of $du(t)$, these expectations factorize, and we get

$$d\varphi_n(t) = \sum_{i=1}^{n} \langle I \otimes \ldots \otimes (\tfrac{1}{2} dL_i(t) + dM_i(t)) \otimes \ldots \otimes I \rangle \varphi_n(t)$$
$$+ \sum_{1 \leq i < j \leq n} \langle I \otimes \ldots \otimes (\tfrac{1}{2} dL_i(t) + dM_i(t)) \otimes \ldots \otimes (\tfrac{1}{2} dL_j(t)$$
$$+ dM_j(t)) \otimes \ldots \otimes I \rangle \varphi_n(t) , \tag{2.5}$$

where I is the identity on $\mathbb{C}^N$. We evaluate the expectations of the increments as follows. First, from (2.1) we have

$$\langle L_i^a(t) M_j^b(t) \rangle = 4\lambda \delta^{ab} \int_0^t \int_0^t dz_i(s) ds' \delta(s-s') \frac{1}{z_i(s) - z_j(s')}$$
$$= 4\lambda \delta^{ab} \int_0^t \frac{z_i'(s)}{z_i(s) - z_j(s)} ds .$$

Taking the differential of both sides gives

$$\langle dL_i^a(t) dM_j^b(t) \rangle = 4\lambda \delta^{ab} \frac{z_i'(t)}{z_i(t) - z_j(t)} dt .$$

Inserting the generators $\{T_a\}$ of the Lie algebra gives

$$\langle dL_i(t) \otimes dM_j(t) \rangle = 2\lambda \frac{z_i'}{z_i - z_j} \sum_{a=1}^{r} T_a \otimes T_a dt . \tag{2.6}$$

In order to use this in (2.5) we define the following operators on $(\mathbb{C}^N)^n$:

$$\Omega_{ij} = \sum_{a=1}^{r} I \otimes \ldots \otimes T_a \otimes \ldots \otimes T_a \otimes \ldots \otimes I ,$$

where the generators T_a occur in the i and j positions. The first expectation on the right-hand side of (2.5) is zero, and the second term gives

$$d\varphi_n = \lambda \sum_{1 \leqq i < j \leqq n} \frac{z_i' - z_j'}{z_i - z_j} \Omega_{ij} \varphi_n dt .$$

So we finally get the promised differential equation:

$$\frac{d\varphi_n}{dt} = \lambda \sum_{1 \leqq i < j \leqq n} \frac{z_i' - z_j'}{z_i - z_j} \Omega_{ij} \varphi_n . \tag{2.7}$$

This equation has been studied before in connection with linear representations of the braid group [2], statistics of quantum fields [3] and monodromy equations in two-dimensional conformal field theory [4]. The relevance of the equations to these topics is due to the following observations.

Let M_n be the subset of $\mathbb{C}^n$ consisting of points ζ whose coordinates $(z_1, \ldots, z_n)$ are all distinct. We define an operator-valued connection on M_n by

$$\omega = \lambda \sum_{1 \leqq i < j \leqq n} \Omega_{ij} d \log(z_i - z_j) , \tag{2.8}$$

where, here, λ is any complex number. Then it is easy to show that ω is flat, so that

$$d\omega + \omega \wedge \omega = 0 .$$

This follows from the infinitesimal braid relations for the operators $\{\Omega_{ij}\}$ (see [10] for a simple proof):

$$[\Omega_{ij}, \Omega_{kl}] = 0 , \qquad [\Omega_{ij}, \Omega_{jk} + \Omega_{ki}] = 0 \tag{2.9}$$

for all i, j, k, l distinct. Therefore the holonomy of ω gives a linear representation of the fundamental group of M_n, which is the pure braid group on n strings [6]. Also the symmetry of (2.8) under permutations of the coordinates means that ω induces a flat connection on the quotient space $\tilde{M}_n$ obtained by identifying points in M_n whose coordinates are equal up to a permutation. Since the fundamental group of $\tilde{M}_n$ is the braid group on n strings [6], we also get a linear representation of the full braid group. It is known how to obtain a link invariant from this representation, by defining a suitable trace on $(\mathbb{C}^N)^n$ ([2, 7]). In the case where G is $SU(N)$, this link invariant yields the two-variable generalization of the Jones polynomial [5]. Therefore the differential equation (2.7), which is the equation of parallel transport for the connection ω, can be used to obtain knot polynomials. This already indicates a connection between the Chern-Simons theory and knot polynomials.

We now wish to explore more fully the statement that the expectation of a Wilson loop in the Chern-Simons theory yields a knot polynomial. We will do this by using a differential equation for the loop expectation which is similar to (2.7). Our approach is different to that described in the previous paragraph. We do not want to use a braid representation; rather we want to use the original loop itself. The equation we use is again suggested by the Chern-Simons theory.

If we take a cross-section of the Wilson loop, we will intersect an even number of Wilson lines, half of which "go up" and half of which "go down"; see Fig. 2.2. The upward moving lines carry a representation R of the group G, while the downward

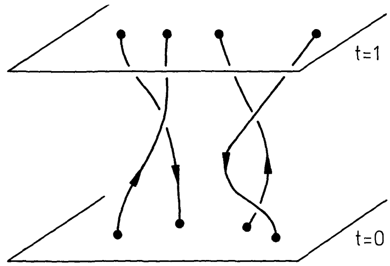


Fig. 2.2. Cross section of a Wilson loop

moving lines carry the representation $\bar{R}$. Our previous equation (2.7) applied to the case where the n lines carry the same representation of G, so we generalize it in the following way.

Let $\sigma_1(t), \ldots, \sigma_{2n}(t)$ be parametrized curves in $\mathbb{R} \times \mathbb{C}$, given by

$$\sigma_i(t) = (t, z_i(t)), \quad i = 1, \ldots, 2n, \quad 0 \leqq t \leqq 1.$$

Again we assume that $z_i(t) \neq z_j(t)$ for any t, if i and j are different. The parallel transport operators on the curves $\sigma_1, \ldots, \sigma_n$ will be written $u_1, \ldots, u_n$, and these are defined by (2.2). On the other n lines we use the complex conjugate representation, and we define operators $\bar{u}_{n+1}, \ldots, \bar{u}_{2n}$ by the equation,

$$d\bar{u}(t) = \sum_{a=1}^{r} \left(\frac{1}{2} \alpha^a dz + A_0^a dt \right) \frac{\bar{T}_a}{\sqrt{2}} \bar{u}(t), \tag{2.10}$$

where $\bar{T}_a$ is the complex conjugate of T_a. The operator we consider is

$$\psi_n(t) = \langle u_1(t) \otimes \ldots \otimes u_n(t) \otimes \bar{u}_{n+1}(t) \otimes \ldots \otimes \bar{u}_{2n}(t) \rangle$$

which maps $(\mathbb{C}^N)^{2n}$ into itself. We can view $\psi_n(t)$ as a linear operator taking states in $(\mathbb{C}^N)^{2n}$ at time $t=0$ to states in $(\mathbb{C}^N)^{2n}$ at time t. We now compute an equation of motion for $\psi_n(t)$ using the same method as before. This produces the following operators on $(\mathbb{C}^N)^{2n}$:

$$S_{ij} = \begin{cases} \sum_{a=1}^{r} I \otimes \ldots \otimes T_a \otimes \ldots \otimes T_a \otimes \ldots \otimes I, & 1 \leqq i < j \leqq n \\ \sum_{a=1}^{r} I \otimes \ldots \otimes T_a \otimes \ldots \otimes \bar{T}_a \otimes \ldots \otimes I, & 1 \leqq i \leqq n,\ n+1 \leqq j \leqq 2n, \\ \sum_{a=1}^{r} I \otimes \ldots \otimes \bar{T}_a \otimes \ldots \otimes \bar{T}_a \otimes \ldots \otimes I, & n+1 \leqq i < j \leqq 2n \end{cases}$$

where again the non-identity factors occur in the i and j positions. We also define $S_{ji} = S_{ij}$ for $i<j$. These operators also satisfy the infinitesimal braid relations.

Repeating the derivation of (2.7), we get the following differential equation for ψ_n:

$$\frac{d\psi_n}{dt} = \lambda \sum_{1 \leqq i < j \leqq 2n} \frac{z_i' - z_j'}{z_i - z_j} S_{ij} \psi_n, \tag{2.11}$$

with $\lambda = -(k + c_2(G)/2)^{-1}$.

In order to apply this to the computation of Wilson loop expectations, we can restrict the operators to a subspace of $(\mathbb{C}^N)^{2n}$, corresponding to gauge-invariant states. Let R be the representation of G on $\mathbb{C}^N$, and let $\bar{R}$ be its complex conjugate. Then we define the gauge invariant subspace to be

$$V_n = \{v \in (\mathbb{C}^N)^{2n} | R(g) \otimes \ldots \otimes R(g) \otimes \bar{R}(g) \otimes \ldots \otimes \bar{R}(g) v = v, \text{ all } g \in G\}.$$

As an example, suppose $\{e_\alpha\}$ is an orthonormal basis of $\mathbb{C}^N$, and consider the state

$$v_0 = N^{-n} \sum_{\alpha_1, \ldots, \alpha_n = 1}^{N} e_{\alpha_1} \otimes \ldots \otimes e_{\alpha_n} \otimes e_{\alpha_1} \otimes \ldots \otimes e_{\alpha_n}.$$

Then v_0 is in V_n, and the matrix element of $\psi_n(t)$ in this state is

$$(v_0, \psi_n(t) v_0) = \langle \operatorname{tr}(u_1(t) u_{n+1}^+(t)) \ldots \operatorname{tr}(u_n(t) u_{2n}^+(t)) \rangle.$$

where the trace is normalized to give $\operatorname{tr} I = 1$. Suppose now that the curves $\sigma_1, \ldots, \sigma_{2n}$ satisfy $\sigma_i(0) = \sigma_{n+i}(0)$ and $\sigma_i(1) = \sigma_{n+i}(1)$ for all $i = 1, \ldots, n$. Then each term $\operatorname{tr}(u_i(1) u_{n+i}^+(1))$ is the trace of the holonomy around a closed curve, and so $(v_0, \psi_n(1) v_0)$ is the expectation of a product of n Wilson loops. Similarly, let P be any permutation of the integers $\{1, \ldots, n\}$ and consider the state

$$v(P) = N^{-n} \sum_{\alpha_1, \ldots, \alpha_n = 1} e_{\alpha_{P(1)}} \otimes \ldots \otimes e_{\alpha_{P(n)}} \otimes e_{\alpha_1} \otimes \ldots \otimes e_{\alpha_n}.$$

Then $v(P)$ belongs to V_n for any permutation P. The inner product $(v(P), \psi_n(1) v_0)$ is again the expectation of a product of Wilson loops, if the curves satisfy

$$\sigma_i(0) = \sigma_{n+i}(0), \quad \sigma_i(1) = \sigma_{n+P(i)}(1) \quad \text{for all} \quad i = 1, \ldots, n.$$

The number of Wilson loops is the number of cycles in P.

In [10] it was proved that the operators $\{S_{ij}\}$ map V_n into itself, for any group G. Therefore the solution of (2.11) restricts to this subspace also, assuming that $\psi_n(0)$ leaves V_n invariant. In fact we will always take $\psi_n(0)$ to be the identity operator.

We are now faced with two problems when we try to use (2.11) to compute expectations of Wilson loops. The first problem is that a Wilson loop is obtained by joining the "up" curves and the "down" curves at their endpoints. However the right-hand side of (2.11) is singular at these joining points, and the solution itself may be singular there also. Indeed the solution of (2.11) in the case $G = U(1)$ is given by

$$\psi_n(t) = \prod_{\substack{1 \leqq i < j \leqq n \\ n+1 \leqq i < j \leqq 2n}} (z_i - z_j)^\lambda \prod_{\substack{1 \leqq i \leqq n \\ n+1 \leqq j \leqq 2n}} (z_i - z_j)^{-\lambda} \psi_n(0),$$

which is singular when $z_i = z_j$, being either zero or infinity.

The second problem is the framing problem pointed out by Witten [1]. In our approach this shows up as the dependence of the solution of (2.11) on quantities which are not topological invariants. This problem occurs already in the case $n = 1$, where there is one up curve and one down curve. The equation is

$$\frac{d\psi_1}{dt} = \lambda \frac{z_1' - z_2'}{z_1 - z_2} \sum_{a=1}^{r} T_a \otimes \bar{T}_a \psi_1.$$

The gauge invariant subspace V_1 is one-dimensional, containing just the state v_0, and

$$\sum_a T_a \otimes \bar{T}_a v_0 = -cv_0, \tag{2.12}$$

where c is the Casimir operator, which we assume to be a multiple of the identity. So the solution is

$$\psi_1(t) = (z_1 - z_2)^{-c\lambda}\psi_1(0), \qquad z_i \equiv z_i(t).$$

This result depends on the winding number of $z_1 - z_2$ about the origin (in general $c\lambda$ is not an integer). This winding number has no topological significance and should not appear.

It turns out that we can solve both these problems by considering a ratio of expectations. Let $W(\sigma)$ be the Wilson loop observable for a closed curve σ, computed for the gauge group G in the representation R. We denote by $\langle W(\sigma)\rangle_G$ the expectation of this in the Chern-Simons theory with gauge group G, evaluated in light-cone gauge. In addition we consider the Wilson loop observable $\tilde{W}(\sigma)$ computed for the group $U(1)$ around the same curve σ. We denote by $\langle \tilde{W}(\sigma)\rangle_{U(1)}$ the expectation taken with respect to the $U(1)$ Chern-Simons action in light-cone gauge, where the square of the $U(1)$ charge is adjusted to be the Casimir of G in the representation R, as in (2.12). Then we consider the ratio

$$\langle W(\sigma)\rangle_G \langle \tilde{W}(\sigma)\rangle_{U(1)}^{-1}. \tag{2.13}$$

In [1] it was shown that by choosing a framing for the curve σ both the numerator and denominator in (2.13) can be computed and yield topological invariants. In our approach we consider only this ratio of expectations, but we do not require a framing on the curve σ. Once again, we consider $2n$ curves $\sigma_1, \ldots, \sigma_{2n}$ which never intersect, and we define parallel transport operators both for the group G and for $U(1)$. If we write $\chi_n(t)$ for the operator $\psi_n(t)$ in the $U(1)$ case, we get the equation of motion

$$\frac{d\chi_n}{dt} = c\lambda \sum_{1 \leqq i < j \leqq 2n} \frac{z_i' - z_j'}{z_i - z_j}\theta_{ij}\chi_n,$$

where $\theta_{ij} = \theta_{ji}$ is given by

$$\theta_{ij} = \begin{cases} 1 & 1 \leqq i < j \leqq n,\ n+1 \leqq i < j \leqq 2n \\ -1 & 1 \leqq i \leqq n,\ n+1 \leqq j \leqq 2n \end{cases}.$$

The operator which leads to the ratio of expectations (2.13) is

$$\tilde{\psi}_n(t) = \psi_n(t)\chi_n(t)^{-1},$$

and its equation of motion is

$$\frac{d\tilde{\psi}_n}{dt} = \lambda \sum_{1 \leqq i < j \leqq 2n} \frac{z_i' - z_j'}{z_i - z_j}\tilde{S}_{ij}\tilde{\psi}_n.$$

The new operators are

$$\tilde{S}_{ij} = S_{ij} - (c\theta_{ij} I \otimes \ldots \otimes I), \tag{2.14}$$

which also satisfy the infinitesimal braid relations (2.9). Deriving this equation was the goal of this section, and in the remaining sections we will ignore its origins in the Chern-Simons theory. For this reason, the parameter λ is no longer constrained to be $-(k+c/2)^{-1}$, and for convenience we drop the tilde on ψ_n. So we consider the equation

$$\frac{d\psi_n}{dt}=\lambda \sum_{1\leqq i<j\leqq 2n} \frac{z_i'-z_j'}{z_i-z_j}\tilde{S}_{ij}\psi_n\,. \tag{2.15}$$

In the remainder of the paper we will study (2.15) in the case where $G=SU(N)$, and R is the fundamental representation. In the next section we will consider the case of four curves $\sigma_1, \sigma_2, \sigma_3, \sigma_4$, and we will solve (2.15) in terms of hypergeometric functions. In order to do this, we will need to restrict λ in (2.15) by the condition

$$0<-N\lambda<1\,, \tag{2.16}$$

which is automatically satisfied if $\lambda=-(k+N)^{-1}$, $k=1,2,3,\ldots$.

We will see later in Sect. 4 that this restriction is also necessary in the general case, so henceforth we will assume that λ satisfies (2.16).

3. The Solution for Four Curves

We consider the case where $n=2$, and G is $SU(N)$, so that there are four curves $\sigma_1, \sigma_2, \sigma_3, \sigma_4$. The gauge invariant subspace V_2 is two-dimensional. If we let $\{e_\alpha\}$ be an orthonormal basis of $\mathbb{C}^N$, then V_2 is spanned by the states

$$v_0=N^{-2}\sum_{\alpha,\beta=1}^{N} e_\alpha\otimes e_\beta\otimes e_\alpha\otimes e_\beta\,,$$

$$v_1=N^{-2}\sum_{\alpha,\beta=1}^{N} e_\alpha\otimes e_\beta\otimes e_\beta\otimes e_\alpha\,.$$

The operators $\{\tilde{S}_{ij}\}$ are given by 2×2 matrices on this subspace (see [9, 10]):

$$\tilde{S}_{12}=\tilde{S}_{34}=\begin{pmatrix} N & -1\\ -1 & N\end{pmatrix},$$

$$\tilde{S}_{13}=\tilde{S}_{24}=\begin{pmatrix} 0 & 1\\ 0 & -N\end{pmatrix},$$

$$\tilde{S}_{14}=\tilde{S}_{23}=\begin{pmatrix} -N & 0\\ 1 & 0\end{pmatrix}.$$

We define a new variable

$$x=\frac{(z_1-z_3)(z_2-z_4)}{(z_1-z_2)(z_3-z_4)},$$

in terms of which (2.15) becomes,

$$\frac{d\psi_2}{dx}=\lambda\left\{\frac{1}{x}\begin{pmatrix} 0 & 1\\ 0 & -N\end{pmatrix}+\frac{1}{1-x}\begin{pmatrix} N & 0\\ -1 & 0\end{pmatrix}\right\}\psi_2\,. \tag{3.1}$$

An equation very similar to (3.1) was studied in [4]; the solutions can be expressed in terms of hypergeometric functions. As t increases from zero to one, $x(t)$ describes a curve in the complex plane, which never passes through $x=0$ or $x=1$. Equation (3.1) describes the analytic continuation of a matrix-valued function along this curve. It is clear that the solution has monodromy about $x=0$ and $x=1$ (indeed this is the characteristic property of the hypergeometric function).

In order to solve (3.1), we consider the associated vector equation

$$w'(x)=\lambda\left\{\frac{1}{x}\begin{pmatrix}0 & 1\\ 0 & -N\end{pmatrix}+\frac{1}{1-x}\begin{pmatrix}N & 0\\ -1 & 0\end{pmatrix}\right\}w(x), \tag{3.2}$$

where $w(x)$ is a two-component vector. This equation has two linearly independent solutions. There are two natural bases to choose in the solution space; the one which diagonalizes the monodromy about $x=0$, and the one which diagonalizes the monodromy about $x=1$. The first basis is $(w_1(x), w_2(x))$, where

$$w_1(x)=\begin{pmatrix}(1-x)^{-N\lambda}F(\lambda,-\lambda;1+N\lambda;x)\\ \dfrac{-\lambda}{1+N\lambda}xF(1+N\lambda-\lambda,1+N\lambda+\lambda;\,2+N\lambda;\,x)\end{pmatrix},$$

$$w_2(x)=x^{-N\lambda}\begin{pmatrix}(1-x)^{-N\lambda}F(\lambda-N\lambda,-\lambda-N\lambda;\,1-N\lambda;\,x)\\ -NF(\lambda,-\lambda;\,-N\lambda;\,x)\end{pmatrix}.$$

Here $F(a,b;c;x)$ is the hypergeometric function, which is analytic and single-valued for $|x|<1$. The second basis, which diagonalizes the monodromy about $x=1$, is $(w_3(x), w_4(x))$, where

$$w_3(x)=\begin{pmatrix}\dfrac{-\lambda}{1+N\lambda}(1-x)F(1+N\lambda-\lambda,1+N\lambda+\lambda;\,2+N\lambda;\,1-x)\\ xF(1+N\lambda-\lambda,1+N\lambda+\lambda;\,1+N\lambda;\,1-x)\end{pmatrix},$$

$$w_4(x)=(1-x)^{-N\lambda}\begin{pmatrix}-NF(\lambda,-\lambda;\,-N\lambda;\,1-x)\\ xF(1+\lambda,1-\lambda;\,1-N\lambda;\,1-x)\end{pmatrix}.$$

These bases are related by a constant transition matrix, given by

$$\begin{pmatrix}w_1(x)\\ w_2(x)\end{pmatrix}=M\begin{pmatrix}w_3(x)\\ w_4(x)\end{pmatrix},$$

$$M=\begin{bmatrix}N\dfrac{\Gamma(N\lambda)\Gamma(-N\lambda)}{\Gamma(\lambda)\Gamma(-\lambda)} & \dfrac{-N}{N^2-1}\dfrac{\Gamma(N\lambda)\Gamma(N\lambda)}{\Gamma(\lambda+N\lambda)\Gamma(-\lambda+N\lambda)}\\ -N\dfrac{\Gamma(-N\lambda)\Gamma(-N\lambda)}{\Gamma(\lambda-N\lambda)\Gamma(-\lambda-N\lambda)} & -N\dfrac{\Gamma(N\lambda)\Gamma(-N\lambda)}{\Gamma(\lambda)\Gamma(-\lambda)}\end{bmatrix}.$$

Note that by our assumption (2.16) we avoid any poles of the gamma function. We can now compute the solution of (3.1) around any closed curve in the plane which avoids the two singular points $x=0$ and $x=1$. For example, suppose γ is a closed curve which encircles the point $x=0$ once in the positive direction, but does not encircle $x=1$. Then the solution of (3.1) in the basis (w_1, w_2) is

$$\psi_2(\gamma)=\begin{pmatrix}1 & 0\\ 0 & \exp(-2\pi i N\lambda)\end{pmatrix}.$$

This is also the solution of (3.1) for a curve which encircles $x=1$ once in the positive direction, but does not encircle $x=0$, given in the basis (w_3, w_4). Similarly, if γ encircles both $x=0$ and $x=1$ once in the positive direction, the solution in the (w_1, w_2) basis is

$$\psi_2(\gamma)=\begin{pmatrix}1 & 0\\ 0 & \exp(-2\pi i N\lambda)\end{pmatrix} M \begin{pmatrix}1 & 0\\ 0 & \exp(-2\pi i N\lambda)\end{pmatrix} M^{-1}.$$

Every closed curve can be built out of these elementary curves, and so the solution can be computed. We can also solve (3.1) along a curve γ which is not closed, and we will denote the solution by $\psi_2(\gamma)$ also. In this case a basis for the solution spaces at the initial and final points of γ must be specified in order to write down the solution.

It is straightforward to continue from this point and obtain a representation of the braid group on four strings using the monodromy of the solution of (3.1) (see [4]). However we want to consider what happens when we apply (3.1) to a Wilson loop which is constructed by joining together the ends of the curves $\sigma_1, \sigma_2, \sigma_3, \sigma_4$. There are two ways of doing this; we can join σ_1 to σ_3 and σ_2 to σ_4, or we can join σ_1 to σ_4 and σ_2 to σ_3. Either of these two ways can be used at the beginning and the end of the curves, as we illustrate with the example in Fig. 3.1a. In terms of the variable x, these are the points $x=0$ and $x=1$ respectively, which are precisely the singular points of the solutions of (3.1). So the Wilson loop in Fig. 3.1a corresponds to the curve γ in Fig. 3.1b, which begins at 0 and ends at 1.

If we now consider the behavior of the solutions w_1, w_2, w_3, w_4 at $x=0$ and $x=1$, we see

$$\begin{aligned} w_1(0)&=\begin{pmatrix}1\\0\end{pmatrix}, \quad w_2(0)=\begin{pmatrix}0\\0\end{pmatrix},\\ w_3(1)&=\begin{pmatrix}0\\1\end{pmatrix}, \quad w_4(1)=\begin{pmatrix}0\\0\end{pmatrix}. \end{aligned} \tag{3.3}$$

We have used the fact that λ satisfies the bound (2.16). From (3.3), we see that any solution $w(x)$ has a limit as x approaches 0 or 1. At $x=0$, the limit is

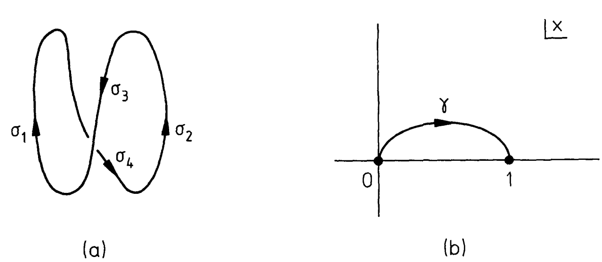


Fig. 3.1. a A single Wilson loop and **b** the corresponding path in the x-plane

proportional to v_0, and at $x=1$ it is proportional to v_1. So this defines the solution $w(x)$ along a curve which terminates at $x=0$ or $x=1$. If the curve begins at these singular points, we must make a choice about how to define the solution. Firstly the initial value must be proportional to the appropriate vector, i.e. at $x=0$ it must be proportional to v_0 and at $x=1$ it must be proportional to v_1. There are many solutions $w(x)$ which converge to a given vector av_0 at $x=0$, where a is a constant. However there is a unique solution which is single-valued in the disc $|x|<1$, namely $aw_1(x)$. This is the only solution which does not depend on the behavior of γ in the vicinity of $x=0$. Therefore we define the solution along a curve γ which begins at $x=0$ as follows. The initial state is av_0, and the solution for $0<|x|<1$ is $aw_1(x)$. This solution can now be analytically continued along the rest of γ, giving a solution $\psi_2(\gamma)$ for the whole curve. Similarly if γ begins at $x=1$, the initial state must be bv_1, for some constant b, and then the solution for $0<|1-x|<1$ is $bw_3(x)$.

Notice that the only allowed initial state at $x=0$, namely v_0, is the state which contracts the indices on the curves σ_1 and σ_3, and also on σ_2 and σ_4. Since the point $x=0$ is reached when the curves join together in these pairings, this means that the initial state is precisely the one required to compute the corresponding Wilson loop expectation. Similarly the state v_1 at $x=1$ contracts the indices on σ_1 and σ_4 and on σ_2 and σ_3.

If a curve passes through the point $x=0$, the solution is continuous but not analytic in general. For example consider the set of curves in $\mathbb{R}\times\mathbb{C}$ shown in Fig. 3.2b. The corresponding curve γ in the x-plane is shown in Fig. 3.2a; it begins at x and ends at y, and passes through 0. Suppose $|x|<1$, $|y|<1$, and that the solution at x is

$$w(x)=aw_1(x)+bw_2(x).$$

Then the solution along γ is

$$\psi_2(\gamma)w(x)=aw_1(y).$$

Returning to the example shown in Fig. 3.1, we have defined $\psi_2(\gamma)$ for the curve γ. The matrix element of $\psi_2(\gamma)$ which computes the Wilson loop expectation is

$$(v_1,\psi_2(\gamma)v_0)=N\frac{\Gamma(N\lambda)\Gamma(-N\lambda)}{\Gamma(\lambda)\Gamma(-\lambda)},$$

where we used the explicit expressions for the matrix M. Recalling the identity

$$\Gamma(z)\Gamma(1-z)=\frac{\pi}{\sin\pi z},$$

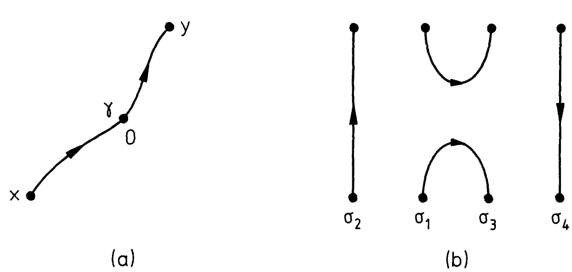


Fig. 3.2. A curve which passes through $x=0$

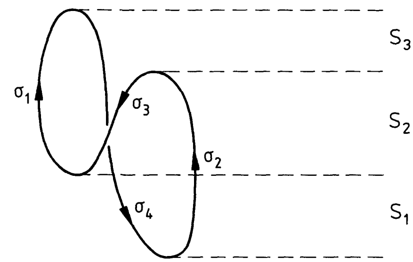


Fig. 3.3. A deformed version of Fig. 3.1a

this reduces to

$$(v_1, \psi_2(\gamma)v_0) = \frac{\sin \pi\lambda}{\sin N\pi\lambda}. \tag{3.4}$$

This is equal to the expectation of an unknotted Wilson loop in the $SU(N)$ Chern-Simons theory calculated by Witten [1], if we take λ to be $-(k+N)^{-1}$.

Obviously the loop in Fig. 3.1a is untypical, since all four curves begin and end at the same horizontal levels. A more typical loop is pictured in Fig. 3.3. The solution of (3.1) for this loop is obtained by solving the equation separately in each of the horizontal slices where the number of curves is constant, and multiplying the solutions in the appropriate order. In the slice labelled S_1 in Fig. 3.3, there are only two curves σ_2 and σ_4, so we could evaluate this by using (2.15) with $n=1$. However in order to be compatible with the solution in the next slice S_2, we use (2.15) with $n=2$ in the slice S_1. The functions $z_1(t)$ and $z_3(t)$ no longer appear, and (2.15) becomes

$$\frac{d\psi_2}{dt} = \lambda \frac{z_2' - z_4'}{z_2 - z_4} \tilde{S}_{24} \psi_2. \tag{3.5}$$

Referring back to the Chern-Simons theory, the solution of (3.5) corresponds to the operator $\langle I \otimes u_2 \otimes I \otimes \bar{u}_4 \rangle$, i.e. the parallel transport operators on the missing curves are replaced by the identity. The right-hand side of this equation annihilates the state v_0. Since this is the initial state which we want to use in the slice S_1, this means that the state is unchanged and ψ_2 is just the identity. So the final state for S_1, which is the initial state for S_2, is again v_0, as we had previously. Similarly the final state for S_2 is proportional to v_1. The equation in the slice S_3 corresponding to (3.5) has the operator $\tilde{S}_{13}$ on the right-hand side, which annihilates v_1. So the solution ψ_2 in the slice S_3 is again the identity. Therefore the solution obtained for the loop in Fig. 3.3 is the same as for Fig. 3.1, and the matrix element of this operator between the initial state v_0 and the final state v_1 is equal to (3.4).

Our equation (3.1) can be used to obtain a solution for any combination of Wilson loops whose intersection with a horizontal plane contains no more than four points. An example is shown in Fig. 3.4. Again the reader can check that the final state for each slice is an allowed initial state for the succeeding slice. So the operators in the slices can be multiplied and the resulting operator has a non-zero matrix element which gives the loop expectation.

We now turn to the problem of identifying this solution in terms of knot polynomials. To do this we need the skein relations which serve to define the knot

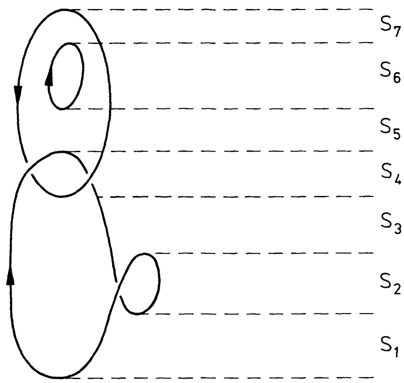


Fig. 3.4. A collection of loops with the slices indicated

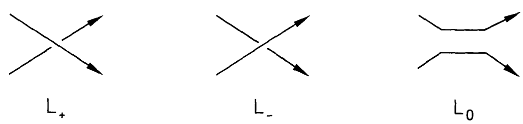


Fig. 3.5. Crossing for the skein relation

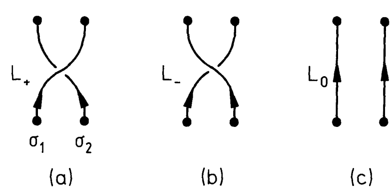


Fig. 3.6. Crossing of two up curves

polynomial. We will write $P_L(t,x)$ for the two-variable generalization of the Jones polynomial. This polynomial satisfies the following recursion relation [5]:

$$t^{-1}P_{L_+} - tP_{L_-} = xP_{L_0},$$

where the links L_+, L_-, and L_0 differ at only one crossing, in the manner shown in Fig. 3.5. Notice that $P_L(t,x)$ is defined for oriented links, and hence for Wilson loops. Furthermore $P_L(t,x)$ is invariant under ambient isotopies of the link L. In order to relate our solution to this knot polynomial, we shall study the skein relation.

There is a basic asymmetry in Eq. (3.1), namely the transfer direction is singled out. Therefore we must consider two different skein relations, depending on whether the crossing curves in Fig. 3.5 carry the same representation or conjugate representations. Suppose first that σ_1 and σ_2 cross, as in Fig. 3.6.

If σ_3 and σ_4 are vertical during the crossing, then the variable x follows the paths γ_+ and γ_- shown in Fig. 3.7, corresponding to the crossings L_+ and L_-. Notice that interchanging z_1 and z_2 replaces x by $1-x$.

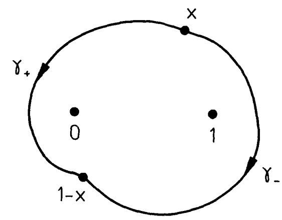


Fig. 3.7. Curves in the x-plane for Fig. 3.6

We will write $\psi_\pm$ for the solutions of (3.1) along the paths $\gamma_\pm$. For convenience we assume that $|\gamma_+(z)|<1$ and $|1-\gamma_-(z)|<1$ for all z on the curves $\gamma_\pm$. Suppose we have an initial state $w(x)$ given by

$$w(x)=c_1w_1(x)+c_2w_2(x).$$

Then from our explicit expressions for w_1, w_2 we get

$$\psi_+w(x)=c_1w_1(1-x)+c_2e^{-2\pi iN\lambda}w_2(1-x).$$

We can also re-express $w(x)$ in the other basis

$$w(x)=c_3w_3(x)+c_4w_4(x),$$

and then we have

$$\psi_-w(x)=c_3w_3(1-x)+c_4e^{2\pi iN\lambda}w_4(1-x).$$

We now claim that there is a simple relationship between these solutions. If we write M_{ij} for the entries of the transition matrix M, we get

$$c_3=c_1M_{11}+c_2M_{21},$$
$$c_4=c_1M_{12}+c_2M_{22},$$
$$w_1=M_{11}w_3+M_{12}w_4,$$
$$w_2=M_{21}w_3+M_{22}w_4.$$

Therefore

$$(e^{\pi iN\lambda}\psi_+-e^{-\pi iN\lambda}\psi_-)w(x)=c_1M_{11}(e^{\pi iN\lambda}-e^{-\pi iN\lambda})w_3(1-x)$$
$$+c_2M_{22}(e^{-\pi iN\lambda}-e^{\pi iN\lambda})w_4(1-x). \tag{3.6}$$

Referring back to the matrix M, we see that

$$M_{11}=-M_{22}=\frac{\sin\pi\lambda}{\sin N\pi\lambda}.$$

Hence

$$(3.6)=(e^{\pi i\lambda}-e^{-\pi i\lambda})(c_1w_3(1-x)+c_2w_4(1-x)).$$

Finally from our explicit solutions we can simplify this further. Define a transposition operator acting on V_2 (our gauge invariant subspace) by

$$\hat{t}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

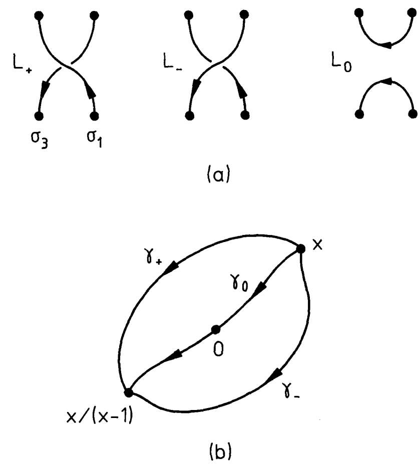


Fig. 3.8. Crossing of an up curve and a down curve

Then a straightforward computation using the properties of hypergeometric functions gives

$$\hat{t}w_1(x)=w_3(1-x),$$

$$\hat{t}w_2(x)=w_4(1-x).$$

Therefore we get

$$e^{\pi iN\lambda}\psi_+ - e^{-\pi iN\lambda}\psi_- = (e^{\pi i\lambda}-e^{-\pi i\lambda})\hat{t}. \tag{3.7}$$

This relation corresponds to the skein relation for $P_L(t,x)$, with $t=\exp(-\pi iN\lambda)$ and $x=2i\sin\pi\lambda$. The operator $\hat{t}$ in (3.7) interchanges the labellings on the curves σ_1 and σ_2, so that the three terms in (3.7) can be fitted into a product of operators at the same place.

Now consider the case where σ_1 and σ_3 cross, as in Fig. 3.8a. The corresponding curves $\gamma_+,\gamma_-,\gamma_0$ in the x-plane are drawn in Fig. 3.8b. Notice that interchanging z_1 and z_3 replaces x by $x(x-1)^{-1}$. Again we assume the curves lie inside the disc $|z|<1$.

Let $w(x)$ be an initial state, given by

$$w(x)=c_1w_1(x)+c_2w_2(x).$$

We will denote by ψ_+, ψ_-, and ψ_0 the solutions of (3.1) along the curves γ_+,γ_-, and γ_0. We have

$$\psi_+w_1(x)=\psi_-w_1(x)=w_1\left(\frac{x}{x-1}\right),$$

$$\psi_+w_2(x)=e^{-2\pi iN\lambda}\psi_-w_2(x).$$

Therefore

$$(e^{\pi iN\lambda}\psi_+ - e^{-\pi iN\lambda}\psi_-)w(x)=(e^{\pi iN\lambda}-e^{-\pi iN\lambda})c_1w_1\left(\frac{x}{x-1}\right). \tag{3.8}$$

As indicated in Fig. 3.2, the solution of (3.1) along the curve γ_0 in Fig. 3.8b picks out the analytic part of $w(x)$, which in this case is $c_1w_1(x)$, and evaluates it at

$x(x-1)^{-1}$. So (3.8) implies

$$e^{\pi i N\lambda}\psi_+ - e^{-\pi i N\lambda}\psi_- = (e^{\pi i N\lambda} - e^{-\pi i N\lambda})\psi_0 . \tag{3.9}$$

Unfortunately this is not the skein relation we want: the coefficient on the right-hand side of (3.9) is wrong. Therefore the solution of (3.1) is not the knot polynomial.

However it is very easy to manufacture a knot polynomial from our solution. Notice that in Fig. 3.8b the curve γ_0 passes through $x=0$; this corresponds to the fact that in L_0 two curves join and separate again. Suppose that we have a set of Wilson loops W of the kind shown in Fig. 3.4. Let γ be the corresponding curve in the x-plane. Then γ is a union of curves $\{\gamma_i\}$, $i=1,\ldots,n$, each of which begins and ends at either $x=0$ or $x=1$, and otherwise avoids these points. We have defined the solution of (3.1) for γ by

$$\psi_2(\gamma) = \prod_{i=1}^{n} \psi_2(\gamma_i) .$$

From this operator we get the matrix element $(v_f, \psi_2(\gamma)v_i)$, where v_i and v_f are, respectively, the initial and final states for W in V_2. We now define the following number, which we call the *loop polynomial* for W:

$$P(W;\lambda,N) = \left(\frac{\sin N\pi\lambda}{\sin\pi\lambda}\right)^n (v_f, \psi_2(\gamma)v_i) . \tag{3.10}$$

If we denote by W_+, W_-, and W_0 the collections of Wilson loops which differ only at one crossing in the way shown in Fig. 3.6 or Fig. 3.8, then from (3.7) and (3.9) we deduce

$$e^{\pi i N\lambda}P(W_+;\lambda,N) - e^{-\pi i N\lambda}P(W_-;\lambda,N) = (e^{\pi i\lambda} - e^{-\pi i\lambda})P(W_0;\lambda,N) ,$$

which is exactly the skein relation for the knot polynomial. In the next sections we will extend the definition of $P(W;\lambda,N)$ to a general collection of Wilson loops and we will show that it also satisfies the skein relation. We will then show that $P(W;\lambda,N)$ is invariant under ambient isotopies of the Wilson loops. These results together imply that it is exactly the link polynomial $P_L(t,x)$.

4. The General Case

We now consider a general collection of Wilson loops W, whose cross-section can contain more than four points. As illustrated in Fig. 4.1, the number of intersection points of W with a plane perpendicular to the transfer axis changes as the plane moves up or down.

In Fig. 4.1 the number of curves intersected by a horizontal plane is either zero, two, four, or six, depending on its position. This number changes by two at the points where the tangent vector to W is horizontal (and the curvature is non-zero). We call these the *turning points* of W. There are an even number of turning points; half are "minima" and half are "maxima," as in Fig. 4.1. (We restrict attention to "generic" sets of curves for which these assertions are true.)

In order to apply Eq. (2.15), we want to describe W by a set of curves $\{\sigma_i\}$, each of the form $\sigma(t) = (t, z(t))$ for some function $z(t)$. Suppose W has $2n$ turning points.

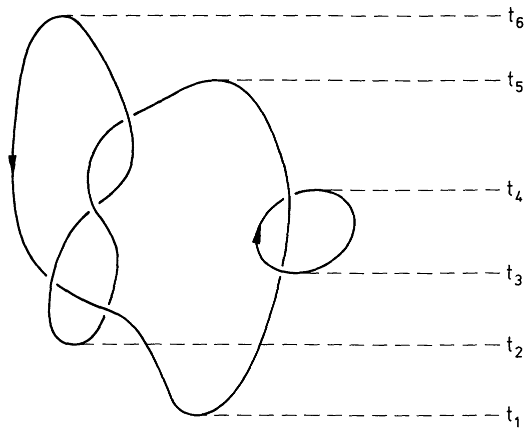


Fig. 4.1. A general collection of loops

We can view each turning point as a point where two curves join, one oriented upwards and one oriented downwards. In terms of the operator $\psi_n(t)$ in Sect. 2, the up curve carries the representation N of $SU(N)$ and the down curves carries the representation $\bar{N}$. Since at each turning point two of these curves join, it follows that W can be described by $2n$ curves $\sigma_1, \ldots, \sigma_{2n}$. We assume that $\sigma_1, \ldots, \sigma_n$ are the up curves, and that $\sigma_{n+1}, \ldots, \sigma_{2n}$ are the down curves. For example, we can describe W in Fig. 4.1 by six curves $\sigma_1, \ldots, \sigma_6$. Each curve begins at some "time" t_i and ends at another "time" t_j, where $1 \leqq i < j \leqq 6$, and $\{t_i\}$ are the coordinates of the turning points in the transfer direction.

Each up curve σ_i begins at some turning point simultaneously with a down curve. This down curve can be written $\sigma_{\pi(i)+n}$ for some permutation π in S_n. Similarly the up curve σ_i ends with another down curve $\sigma_{\varrho(i)+n}$, where ϱ is another permutation in S_n. We will see that π and ϱ determine the initial and final states for the operator $\psi(W)$ which we will define for W using Eq. (2.15), in the same way that v_0 and v_1 were the initial and final states for the solution of Fig. 3.3.

The "time" coordinates of the turning points of W can be listed as $t_1 < t_2 < \ldots < t_{2n}$. We will call the set $[t_p, t_{p+1}] \times \mathbb{C}$ the p^{th} *slice*, and our first concern is to solve Eq. (2.15) in each slice. In the p^{th} slice, some subset of the curves $\sigma_1, \ldots, \sigma_{2n}$ are present, corresponding to a subset K_p of the integers $\{1, \ldots, 2n\}$. Therefore the differential equation in this slice corresponding to (2.15) is

$$\frac{d\psi_p}{dt} = \lambda \sum_{\substack{i,j \in K_p \\ i<j}} \frac{z_i' - z_j'}{z_i - z_j} \tilde{S}_{ij} \psi_p . \tag{4.1}$$

The solution of (4.1) is a function of the coordinates $\{z_i\}$, for $i \in K_p$. Referring back to the Chern-Simons interpretation the operator ψ_p should be interpreted as the expectation of the tensor product of parallel transport operators along the curves $\{\sigma_i\}$ for $i \in K_p$, with the missing parallel transport operators replaced by the identity.

The operators $\{\tilde{S}_{ij}\}$ appearing in (4.1) act only on the factors of the tensor product $(\mathbb{C}^N)^{2n}$ which correspond to the integers in K_p. Acting on a vector $v(P)$ in

V_n, they generate a subspace of V_n, which we will describe below. First of all, we separate the lines missing from this slice into two sets, which are labelled by subsets I_p and J_p of the integers $\{1,\ldots,2n\}$. An integer is in I_p if the corresponding curve begins at a "time" $t \geqq t_{p+1}$. Similarly an integer is in J_p if the corresponding curve ends at a "time" $t \leqq t_p$. All the missing curves are included in these two possibilities. We now define the subspace of V_n on which we want the solution of (4.1) to act. The subspace is called $V_n(I_p;J_p)$. A vector $v(P)$ in V_n belongs to $V_n(I_p;J_p)$ if the permutation P satisfies

$$\begin{aligned} P(j) &= \pi(j) \quad \text{all } j \in I_p,\ 1 \leqq j \leqq n, \\ P(j) &= \varrho(j) \quad \text{all } j \in J_p,\ 1 \leqq j \leqq n. \end{aligned} \tag{4.2}$$

The subspace $V_n(I_p;J_p)$ is then the linear span of the vectors $v(P)$ satisfying (4.2). In order to see that the solution of (4.1) leaves this subspace invariant, we use the following explicit expressions for the operators $\{\tilde{S}_{ij}\}$. These can all be deduced from the following result. Let $\{T_a\}$ $(a=1,\ldots,N^2)$ be a linearly independent set of $N \times N$ skew-hermitian matrices satisfying

$$\mathrm{tr}(T_a T_b) = -\delta_{ab}.$$

Then as an operator on $\mathbb{C}^N \otimes \mathbb{C}^N$,

$$\sum_{a=1}^{N^2} T_a \otimes T_a = -\hat{\tau},$$

where $\hat{\tau}$ is the transposition operator given by

$$\hat{\tau} u \otimes v = v \otimes u.$$

Recall that V_n is the linear span of the vectors $\{v(P)\}$, where

$$v(P) = N^{-n} \sum_{\alpha_1,\ldots,\alpha_n=1}^{N} e_{\alpha_{P(1)}} \otimes \ldots \otimes e_{\alpha_{P(n)}} \otimes e_{\alpha_1} \otimes \ldots \otimes e_{\alpha_n}.$$

We write τ_{ij} for the transposition in S_n which interchanges i and j. Then we have

$$\tilde{S}_{ij} v(P) = \begin{cases} Nv(P) - v(P\tau_{ij}) & 1 \leqq i < j \leqq n \\ Nv(P) - v(\tau_{ij}P) & n+1 \leqq i < j \leqq 2n. \end{cases} \tag{4.3}$$

Furthermore if $1 \leqq i, j \leqq n$, then

$$\tilde{S}_{i,j+n} v(P) = \begin{cases} 0 & \text{if } P(i) = j \\ v(P\tau_{ij}) - Nv(P) & \text{if } P(i) \neq j. \end{cases} \tag{4.4}$$

It is clear that the operators on the right-hand side of (4.1) do not affect the conditions (4.2), and so ψ_p leaves $V_n(I_p;J_p)$ invariant. Furthermore we can read off the eigenvalues of $\tilde{S}_{ij}$ now. From (4.3) we see that for $1 \leqq i < j \leqq n$, or $n+1 \leqq i < j \leqq 2n$, $\tilde{S}_{ij}$ has two eigenvalues, namely $N \pm 1$, and they occur with equal multiplicities. From (4.4), we see that for $1 \leqq i, j \leqq n$, $\tilde{S}_{i,j+n}$ has eigenvalues 0 and $-N$, the former occurring with multiplicity $(n-1)!$. In fact the subspace of V_n annihilated by $\tilde{S}_{i,j+n}$ is isomorphic to V_{n-1}.

In order to solve (4.1), we again look at the corresponding vector equation

$$\frac{dw}{dt}=\lambda \sum_{\substack{i,j\in K_p\\ i<j}} \frac{z_i'-z_j'}{z_i-z_j}\tilde{S}_{ij}w. \tag{4.5}$$

The solution of (4.5) is a multi-valued analytic function of the coordinates $\{z_i\}$, for $i\in K_p$. It has non-trivial monodromy about the singular subspaces where any two coordinates are equal. Although we cannot solve (4.5) explicitly, we can compute the monodromy of the solution. Suppose that $|z_k-z_l|$ is very small, for some $k,l\in K_p$. Then we can find the monodromy of (4.5) by solving the reduced equation

$$\frac{du}{dt}=\lambda\frac{z_k'-z_l'}{z_k-z_l}\tilde{S}_{kl}u. \tag{4.6}$$

Suppose first that $1\leqq k<l\leqq n$. Then the solution of (4.6) is

$$u=(z_k-z_l)^{\lambda(N+1)}u_+ +(z_k-z_l)^{\lambda(N-1)}u_-, \tag{4.7}$$

where $u_\pm$ are eigenvectors of $\tilde{S}_{kl}$, with eigenvalues $N\pm 1$. From (4.7) we can read off the monodromy, which is given by a 2×2 matrix, as in Sect. 3. The solution of the full equation (4.5) has the same monodromy, and so it can be written as

$$w(\zeta)=(z_k-z_l)^{\lambda(N+1)}w_1(\zeta)+(z_k-z_l)^{\lambda(N-1)}w_2(\zeta), \tag{4.8}$$

where now $w_1(\zeta)$ and $w_2(\zeta)$ are single-valued analytic functions of z_k and z_l, for $|z_k-z_l|$ small. By this we mean that the other coordinates lie outside some disc in the complex plane containing z_k and z_l. We have denoted by ζ the point in $\mathbb{C}^{|K_p|}$ with coordinates $\{z_i\}$ $(i\in K_p)$, where $|K_p|$ is the cardinality of K_p. In the appendix we prove that the solution $w(\zeta)$ has the form given in Eq. (4.8), and we prove that $w_1(\zeta)$ and $w_2(\zeta)$ are analytic.

We will write $\tau_{kl}(\zeta)$ for the point obtained from ζ by interchanging the coordinates z_k and z_l. Let $\hat{\tau}_{kl}$ be the operator on V_n given by

$$\hat{\tau}_{kl}v(P)=v(P\tau_{kl}).$$

Then $\hat{\tau}_{kl}$ satisfies

$$\begin{aligned}
&[\hat{\tau}_{kl},\tilde{S}_{kl}]=0,\\
&\hat{\tau}_{kl}\tilde{S}_{ki}\hat{\tau}_{kl}=\tilde{S}_{li}, \quad i\neq k,l,\\
&\hat{\tau}_{kl}\tilde{S}_{il}\hat{\tau}_{kl}=\tilde{S}_{ik}, \quad i\neq k,l,\\
&[\hat{\tau}_{kl},\tilde{S}_{ij}]=0, \quad i,j\neq k,l.
\end{aligned}$$

Recall that

$$\tilde{S}_{kl}=N-\hat{\tau}_{kl},$$

so from (4.7) we have

$$\hat{\tau}_{kl}u_\pm=\mp u_\pm.$$

Therefore from (A10) in the appendix we deduce

$$\begin{aligned}
&\hat{\tau}_{kl}w_1(\zeta)=-w_1(\tau_{kl}(\zeta)),\\
&\hat{\tau}_{kl}w_2(\zeta)=w_2(\tau_{kl}(\zeta)).
\end{aligned} \tag{4.9}$$

If we take $n+1 \leqq k < l \leqq 2n$, the same result holds. In this case $\hat{\tau}_{kl}$ is a different operator, given by

$$\hat{\tau}_{kl} v(P) = v(\tau_{kl} P) \qquad n+1 \leqq k < l \leqq 2n .$$

Now suppose that $1 \leqq k \leqq n$ and $n+1 \leqq l \leqq 2n$. Then the solution of (4.6) is

$$u = u_0 + (z_k - z_l)^{-\lambda N} u_1 , \tag{4.10}$$

where u_0 is annihilated by $\tilde{S}_{kl}$ and u_1 has eigenvalue $-N$. Notice that our condition (2.16) on λ implies regularity of (4.10) as $|z_k - z_l|$ approaches zero. Once again the solution of the full equation (4.5) can be written

$$w(\zeta) = w_3(\zeta) + (z_k - z_l)^{-\lambda N} w_4(\zeta) , \tag{4.11}$$

where $w_3(\zeta)$ and $w_4(\zeta)$ are single-valued analytic functions of z_k and z_l, for $|z_k - z_l|$ small as described above. Again this decomposition follows from the results proved in the appendix.

When $z_k = z_l$, we will write η in place of ζ. The solution $w(\eta) = w_3(\eta)$ is annihilated by $\tilde{S}_{kl}$, so it satisfies the equation

$$\begin{aligned} \frac{dw_3}{dt} = \lambda & \sum_{\substack{i,j \in K_p \setminus \{k,l\} \\ i<j}} \frac{z_i' - z_j'}{z_i - z_j} \tilde{S}_{ij} w_3 \\ & + \lambda \sum_{\substack{i \in K_p \\ i \neq k,l}} \frac{z_i' - z_k'}{z_i - z_k} (\tilde{S}_{ik} + \tilde{S}_{il}) w_3 . \end{aligned}$$

For each $i \neq k, l$, the operator $(\tilde{S}_{ik} + \tilde{S}_{il})$ annihilates the kernel of $\tilde{S}_{kl}$, so it annihilates $w_3(\eta)$. Therefore when $z_k = z_l$, $w_3(\eta)$ satisfies the equation derived from (4.5) by omitting all terms involving the curves σ_k and σ_l. Conversely, suppose $u(\eta)$ is a solution of this reduced equation with the curves σ_k and σ_l missing. Then there is a solution $w(\zeta)$ of (4.5), given by (4.11) for $|z_k - z_l|$ small, which is equal to $u(\eta)$ when $z_k = z_l$. In fact there are many solutions, since the other function $w_4(\zeta)$ in (4.11) can be chosen arbitrarily. We will pick out the analytic solution for which $w_4(\zeta) = 0$, and we will call this the *extension* of $u(\eta)$ for $|z_k - z_l|$ small. This is the analog of our definition in Sect. 3 of the solution of (3.1) along a curve which begins at either $x = 0$ or $x = 1$.

We can now define the solution of (4.1) in the p^{th} slice. This slice is defined by $t_p \leqq t \leqq t_{p+1}$, where t_p and t_{p+1} are the time coordinates of two turning points of W. For $t_p < t < t_{p+1}$, (4.1) is solved by analytically continuing a solution of (4.5) along the curve $\gamma(t) = \{\sigma_i(t)\}$ for $i \in K_p$ in the space $\mathbb{C}^{|K_p|}$. This leaves the subspace $V_n(I_p; J_p)$ invariant. So the only question is what happens at the endpoints where $t = t_p$ and $t = t_{p+1}$. From (4.11) we see that nothing goes wrong at these points. If two curves σ_k and σ_l meet at t_{p+1}, the solution $w(\zeta)$ converges to a solution $w_3(\eta)$ of the equation in the succeeding slice. Furthermore, if $w(\zeta)$ is in $V_n(I_p; J_p)$, then $w_3(\eta)$ is in $V_n(I_{p+1}; J_{p+1})$, since by assumption we have

$$I_{p+1} = I_p , \qquad J_{p+1} = J_p \cup \{k, l\} .$$

Similarly if two curves σ_k and σ_l begin at t_p, then the solution from the previous slice is in $V_n(I_{p-1}; J_{p-1})$ and by assumption

$$I_{p-1} = I_p \cup \{k, l\}, \qquad J_{p-1} = J_p.$$

This means that the solution from the previous slice is annihilated by $\tilde{S}_{kl}$, and so its extension in the p^{th} slice is defined.

Therefore the solution of (4.1) extends to the whole interval $t_p \leqq t \leqq t_{p+1}$, where it defines a continuous mapping from the solution space $V_n(I_{p-1}; J_{p-1})$ of the previous slice to the solution space $V_n(I_{p+1}; J_{p+1})$ of the succeeding slice. We call this operator ψ_p.

The solution for W is now obtained by composing these operators in the correct order. We define

$$\psi(W) = \psi_{2n-1}\psi_{2n-2} \cdots \psi_2\psi_1$$

and call this the solution of (2.15) for W. We evaluate it on the initial state $v(\pi)$, and it maps this to a final state proportional to $v(\varrho)$. To get a number out we compute the matrix element $(v(\varrho), \psi(W)v(\pi))$.

We will see in Sect. 5 that in order to get a link invariant from the solution, it will be necessary to use the loop polynomial introduced in Sect. 3. Our general definition of the loop polynomial is

$$P(W; \lambda, N) = \left(\frac{\sin N\pi\lambda}{\sin \pi\lambda}\right)^n (v(\varrho), \psi(W)v(\pi)). \tag{4.12}$$

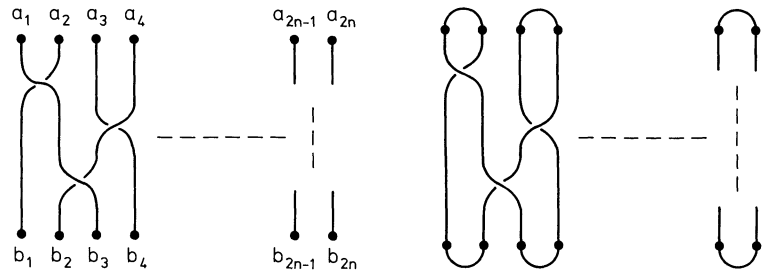
This definition depends explicitly on the number of turning points in W, but of course the link invariant does not.

In the next section we will prove that (4.12) is the knot polynomial for the collection of Wilson loops W. We do this by establishing the skein relations and by showing that it is invariant under ambient isotopies (these are the deformations of a set of closed curves which preserve its link equivalence class). When a link is given as the closure of a braid, invariance under ambient isotopies follows from invariance under the Markov moves (see [6]). However in our case we do not have a braid representation. Fortunately there is another way of representing a link called a *plat representation*. This is defined in [6] and [5], and we recall the definition below.

Consider a braid on $2n$ strings as in Fig. 4.2a. We label the top endpoints $a_1, \ldots, a_{2n}$ and the bottom endpoints $b_1, \ldots, b_{2n}$. To obtain the plat corresponding to b, we put caps on the top and bottom of the braid, making it into a link. We join a_{2k-1} to a_{2k} by an arc lying above the braid, for each $k = 1, \ldots, n$. Similarly b_{2k-1} is joined to b_{2k} by an arc lying below the braid, as shown in Fig. 4.2b. The resulting set of loops is called a *plat*.

Obviously a plat is determined by a braid, and vice versa. Birman [6] showed that every link is ambient isotopic to a plat. She also showed that two plats are equivalent if and only if one can be reached from the other by a sequence of *plat moves*. These plat moves are shown in Fig. 4.3.

The first move twists a cap; the second and third moves interchange two neighboring caps in the ways indicated. The fourth move adds on an extra cap at top and bottom; this is the only move which changes the number of caps.



(a) (b)

Fig. 4.2. **a** A braid and **b** its closure as a plat

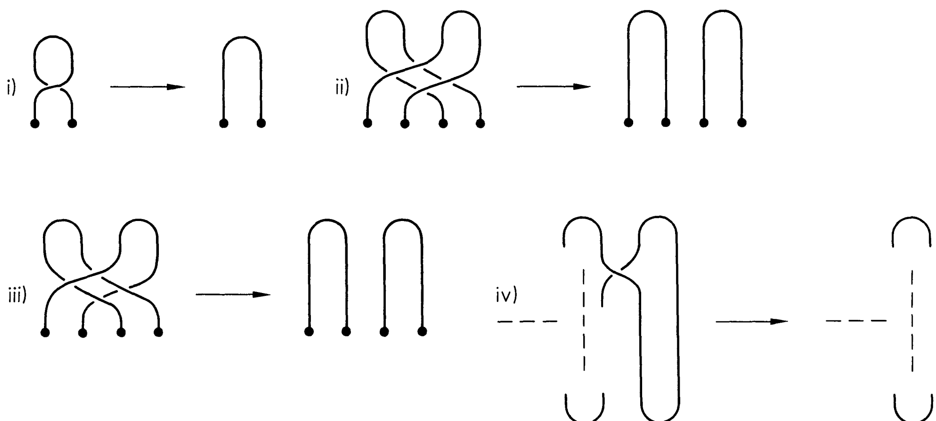
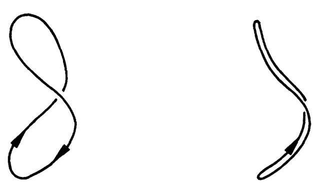


Fig. 4.3. The plat moves



(a) (b)

Fig. 4.4. Adding an unlinked loop

In the next section we will show how a collection of Wilson loops can be made into a plat without changing the loop polynomial. Then we will prove invariance under the plat moves and that will complete the proof that we have a link invariant.

We will need a result in the next section concerning the expectation of an unknotted, unlinked Wilson loop. In fact we just need to consider a loop of the type shown in Fig. 4.4a, which is described by two curves σ_k and σ_l, going up and down respectively. This loop is part of a larger collection W, but is not linked to any other loops in W.

First of all, the integrability conditions for Eq. (2.15) allow us to deform the curves σ_k and σ_l into the shape in Fig. 4.4b without changing the solution (notice that the endpoints are fixed). This can be done so that $|z_k - z_l|$ is always sufficiently small that the solution defined by (4.11) is always analytic in z_k and z_l [i.e. $w_4(\zeta)$ is always zero]. The curves begin and end with $z_k = z_l$, so by analyticity we get the same solution by fixing $z_k = z_l$ throughout the interval. The corresponding collection W' with this loop removed has two turning points less than W, so the solution space V_{n-1} for W' is smaller than V_n, the solution space for W. However the matrix elements of $\psi(W)$ and $\psi(W')$ are the same because we normalized the states $\{v(P)\}$. Therefore from (4.12) we get

$$P(W; \lambda, N) = \frac{\sin N\pi\lambda}{\sin \pi\lambda} P(W'; \lambda, N). \tag{4.13}$$

5. The Skein Relations and the Link Invariant

We will now establish the skein relations for the loop polynomial $P(W; \lambda, N)$, for a general collection of Wilson loops. Consider first the case where two up curves σ_k and σ_l cross in the p^{th} slice in the ways illustrated in Fig. 3.6a and b. The curves not shown are assumed to be constant in this interval. In Fig. 3.6c all the curves are constant. We will write $\psi_\pm$ for the solutions of (4.1) along the curves $L_\pm$ in this interval. Let ζ be the initial point in $\mathbb{C}^{|K_p|}$, where the curves begin; again we denote by $\tau_{kl}(\zeta)$ the point with z_k and z_l interchanged. Suppose that $w(\zeta)$ is an initial state. Then from (4.8) we have

$$w(\zeta) = (z_k - z_l)^{\lambda(N+1)} w_1(\zeta) + (z_k - z_l)^{\lambda(N-1)} w_2(\zeta).$$

Therefore

$$\begin{aligned}\psi_+ w(\zeta) &= e^{-i\pi\lambda(N+1)} (z_k - z_l)^{\lambda(N+1)} w_1(\tau_{kl}(\zeta)) \\ &\quad + e^{-i\pi\lambda(N-1)} (z_k - z_l)^{\lambda(N-1)} w_2(\tau_{kl}(\zeta)), \\ \psi_- w(\zeta) &= e^{i\pi\lambda(N+1)} (z_k - z_l)^{\lambda(N+1)} w_1(\tau_{kl}(\zeta)) \\ &\quad + e^{i\pi\lambda(N-1)} (z_k - z_l)^{\lambda(N-1)} w_2(\tau_{kl}(\zeta)).\end{aligned}$$

We have used the analyticity of $w_1(\zeta)$ and $w_2(\zeta)$ here. Using (4.9) now gives

$$\begin{aligned}&(e^{i\pi\lambda N}\psi_+ - e^{-i\pi\lambda N}\psi_-) w(\zeta) \\ &= e^{i\pi\lambda}\{(z_k - z_l)^{\lambda(N+1)} \hat{\tau}_{kl} w_1(\zeta) + (z_k - z_l)^{\lambda(N-1)} \hat{\tau}_{kl} w_2(\zeta)\} \\ &\quad - e^{-i\pi\lambda}\{(z_k - z_l)^{\lambda(N+1)} \hat{\tau}_{kl} w_1(\zeta) + (z_k - z_l)^{\lambda(N-1)} \hat{\tau}_{kl} w_2(\zeta)\}.\end{aligned}$$

Therefore

$$e^{i\pi\lambda N}\psi_+ - e^{-i\pi\lambda N}\psi_- = (e^{i\pi\lambda} - e^{-i\pi\lambda})\hat{\tau}_{kl} \tag{5.1}$$

which is the required skein relation. The same relation holds when two down curves cross.

As in Sect. 3 the other skein relation corresponding to the crossing of an up curve and a down curve behaves differently. The situation is illustrated in Fig. 3.8a. Once again the other curves are assumed to be constant, and we write $\psi_\pm$ for the solutions of (4.1) along the curves $L_\pm$ in Fig. 3.8a. If $w(\zeta)$ is an initial state, then

(4.11) gives

$$w(\zeta)=w_3(\zeta)+(z_k-z_l)^{-\lambda N}w_4(\zeta),$$

and again using the analyticity of w_3, w_4 we get

$$\psi_+ w(\zeta)=w_3(\tau_{kl}(\zeta))+e^{-i\pi\lambda N}(z_k-z_l)^{-\lambda N}w_4(\tau_{kl}(\zeta)),$$
$$\psi_- w(\zeta)=w_3(\tau_{kl}(\zeta))+e^{i\pi\lambda N}(z_k-z_l)^{-\lambda N}w_4(\tau_{kl}(\zeta)).$$

Therefore

$$(e^{i\pi\lambda N}\psi_+ - e^{-i\pi\lambda N}\psi_-)w(\zeta)=(e^{i\pi\lambda N}-e^{i\pi\lambda N})w_3(\tau_{kl}(\zeta)). \tag{5.2}$$

The right-hand side of (5.2) is the solution for the curve L_0 in Fig. 3.8a. If we consider the collection of loops W' in which $L_\pm$ is replaced by L_0, we see that W' has two more turning points than W. Therefore when we apply (5.2) to the evaluation of the loop polynomials for W and W', the change in n (the number of turning points) changes the coefficient on the right-hand side of (5.2). Let W_+, W_-, and W_0 be three collections of loops which differ only at one crossing in the way shown in Fig. 3.5. Then (5.1) and (5.2) imply that

$$e^{i\pi\lambda N}P(W_+;\lambda,N)-e^{-i\pi\lambda N}P(W_-;\lambda,N)=(e^{i\pi\lambda}-e^{-i\pi\lambda})P(W_0;\lambda,N). \tag{5.3}$$

Our final task is to prove that $P(W;\lambda,N)$ is a link invariant. As explained in Sect. 4, we do this by obtaining a plat representation for W. It should be noted that our method for doing this relies on the skein relation (5.3).

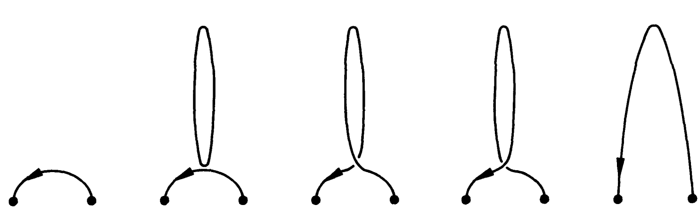
First of all, in Sect. 4 we derived (4.13), which shows how $P(W;\lambda,N)$ changes when a single unknotted, unlinked loop is removed from W. Consider a turning point in W, as illustrated in Fig. 5.1a.

We will introduce a single unknotted link above this turning point, as in (b). Using the skein relations, we can compare this with the situations in (c) and (d) where the loop is connected at the turning point. We will write $W_a, W_b, \ldots$ for the corresponding collections of loops. Then from (4.13) and (5.3) we have

$$\begin{aligned} e^{i\pi\lambda N}P(W_c;\lambda,N)-e^{-i\pi\lambda N}P(W_d;\lambda,N)&=(e^{i\pi\lambda}-e^{-i\pi\lambda})P(W_b;\lambda,N)\\ &=(e^{i\pi N\lambda}-e^{-i\pi N\lambda})P(W_a;\lambda,N). \end{aligned}$$

But since the solutions of (c) and (d) are the same, and are equal to (e), we get

$$P(W_e;\lambda,N)=P(W_a;\lambda,N). \tag{5.4}$$



(a) (b) (c) (d) (e)

Fig. 5.1. Moving up a turning point

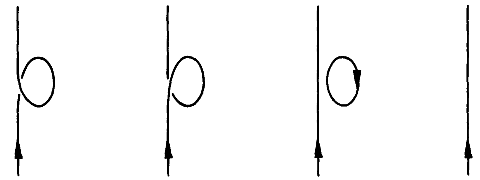


Fig. 5.2. The last plat move

Comparing (a) and (e), we see that the turning point has been pulled up without changing $P(W;\lambda,N)$. We get a plat representation for W by doing this for each turning point which looks like (a). The new turning points can be arranged in a horizontal line as in Fig. 4.2b. Similarly the turning points which are inversions of (a) can be pulled down to the same horizontal level, without changing $P(W;\lambda,N)$.

Now we just need to check invariance under the plat moves shown in Fig. 4.3. The first three plat moves can be implemented by making different choices for the loops which are added on when W is put in a plat representation, as in Fig. 5.1b. The result is independent of this choice, and so is invariant under these moves.

To check plat move (iv), we again use the skein relations. Firstly, by (5.4) we can change the new caps at top and bottom so that the plat looks like Fig. 5.2a.

The solutions for Fig. 5.2a and b are equal, and are related to (c) by the skein relation:

$$e^{i\pi\lambda N}P(W_a;\lambda,N)-e^{-i\pi\lambda N}P(W_b;\lambda,N)=(e^{i\pi\lambda}-e^{-i\pi\lambda})P(W_c;\lambda,N).$$

But from (4.13) we have

$$P(W_c;\lambda,N)=\frac{\sin N\pi\lambda}{\sin\pi\lambda}P(W_d;\lambda,N).$$

Combining these gives

$$P(W_a;\lambda,N)=P(W_d;\lambda,N).$$

Therefore $P(W;\lambda,N)$ is invariant under all the plat moves, and so is a link invariant. Since it satisfies the skein relations (5.3), it is equal to the two-variable Jones polynomial $P_L(\exp(-\pi i\lambda N),2i\sin\pi\lambda)$.

Appendix

The principal result in this appendix is contained in the following lemma.

Lemma 1. *Let $M(z)$ be a $n\times n$ matrix which is an analytic function of z in a simply connected domain D containing* 0. *Suppose that there is a vector u satisfying*

$$M(0)u=0,$$

and suppose also that $\|M(0)\|=c<1$. Then the differential equation

$$\frac{d\psi}{dz}=\frac{1}{z}M(z)\psi$$

with the initial condition

$$\psi(0)=u$$

has a unique analytic solution in D.

Proof. We will construct the solution in a neighborhood of $z=0$. Analytic continuation then extends the solution to the whole domain D.

Consider the modified equation

$$\frac{d\psi}{dz}=\frac{1}{z}M(0)\psi+\lambda Q(z)\psi, \tag{A1}$$

where $Q(z)=z^{-1}(M(z)-M(0))$ is analytic in D, and λ is a parameter. Setting $\lambda=1$ recovers the original equation. When $\lambda=0$, the solution of (A1) is

$$\psi(z)=u.$$

We construct the full solution as a power series in λ:

$$\psi(z)=u+\lambda\psi_1(z)+\lambda^2\psi_2(z)+\lambda^3\psi_3(z)+\ldots. \tag{A2}$$

Substituting (A2) into (A1) and equating powers of λ gives

$$\frac{d\psi_n}{dz}=\frac{1}{z}M(0)\psi_n(z)+Q(z)\psi_{n-1}(z). \tag{A3}$$

We can solve Eq. (A3) inductively in n. We assume that

$$\psi_n(0)=0 \quad \text{all} \quad n\geqq 1.$$

Our induction hypothesis is that $\psi_{n-1}(z)$ is analytic in a neighborhood $|z|<2R$, so that we can write

$$Q(z)\psi_{n-1}(z)=\sum_{k=1}^{\infty} a_k z^k, \qquad |z|<2R.$$

The corresponding expansion for $\psi_n(z)$ is

$$\psi_n(z)=\sum_{k=1}^{\infty} b_k z^k. \tag{A4}$$

Equation (A3) implies that

$$b_k=(k-M(0))^{-1}a_{k-1}, \quad \text{all} \quad k\geqq 1.$$

Using our assumption that $\|M(0)\|=c<1$, it follows that the series (A4) converges in the disc $|z|<R$. By analytic continuation with (A3), this extends to the original disc $|z|<2R$.

So our induction hypothesis is verified. It is interesting to note that Eq. (A3) implies that the first non-zero coefficient in (A4) is

$$b_n=(n-M(0))^{-1}Q(0)^n u.$$

Finally we need a bound on $\psi_n(z)$ which is uniform in n. From (A3) we have

$$\psi_n(z)=\int\frac{1}{\zeta}M(0)\psi_n(\zeta)d\zeta+\int Q(\zeta)\psi_{n-1}(\zeta)d\zeta, \tag{A5}$$

where the line integrals begin at 0 and end at z. When $f(z)$ is analytic we will write

$$\| f \|_R = \sup_{|z| \leqq R} |f(z)| = \sup_{|z| = R} |f(z)|,$$

where we used the maximum modulus principle. Returning to (A5) we get

$$\begin{aligned}\|\psi_n\|_R &\leqq cR \left\| \frac{1}{z} \psi_n \right\|_R + \|Q\|_R \|\psi_{n-1}\|_R R \\ &= c\|\psi_n\|_R + R\|Q\|_R \|\psi_{n-1}\|_R .\end{aligned}$$

Hence

$$\|\psi_n\|_R \leqq \frac{\|Q\|_R R}{1-c} \|\psi_{n-1}\|_R .$$

Iterating this bound gives

$$\|\psi_n\|_R \leqq \left(\frac{\|Q\|_R R}{1-c}\right)^n \|u\| .$$

Therefore by Weierstrass' Theorem the series (A2) converges to an analytic function $\psi(z)$ in the disc $|z| \leqq R$, where R satisfies the bound

$$R < \frac{1-c}{\lambda \|Q\|_R} .$$

Setting $\lambda = 1$ gives the desired result. □

We now exploit Lemma 1 to derive our results concerning the singularities of our differential equation (2.15). We single out any two coordinates z_k and z_l, and any eigenvalue μ of $\tilde{S}_{kl}$. We define

$$\varphi = (z_k - z_l)^{-\lambda\mu} \psi ,$$

and then Eq. (2.15) becomes

$$\begin{aligned}\frac{d\varphi}{dt} &= \lambda(\tilde{S}_{kl} - \mu) \frac{1}{z_k - z_l}\left(\frac{dz_k}{dt} - \frac{dz_l}{dt}\right)\varphi + \lambda \sum_{j \neq k,l} \tilde{S}_{jk} \frac{1}{z_j - z_k}\left(\frac{dz_j}{dt} - \frac{dz_k}{dt}\right)\varphi \\ &\quad + \lambda \sum_{j \neq k,l} \tilde{S}_{jl} \frac{1}{z_j - z_l}\left(\frac{dz_j}{dt} - \frac{dz_l}{dt}\right)\varphi \\ &\quad + \lambda \sum_{i,j \neq k,l} \tilde{S}_{ij} \frac{1}{z_i - z_j}\left(\frac{dz_i}{dt} - \frac{dz_j}{dt}\right)\varphi . \end{aligned} \tag{A6}$$

At first, we will keep fixed the remaining $2n-2$ variables. Let

$$M_0 = \lambda(\tilde{S}_{kl} - \mu),$$

$$Q(z) = \lambda \sum_{j \neq k,l} \tilde{S}_{jk} \frac{1}{z - z_j},$$

$$P(z) = \lambda \sum_{j \neq k,l} \tilde{S}_{jl} \frac{1}{z - z_j}.$$

Then Eq. (A 6) becomes

$$\frac{d\varphi}{dt}=M_0\frac{1}{z-w}\left(\frac{dz}{dt}-\frac{dw}{dt}\right)\varphi+Q(z)\frac{dz}{dt}\varphi+P(w)\frac{dw}{dt}\varphi.$$

The remaining $2n-2$ coordinates appear only in $Q(z)$ and $P(w)$, as external parameters. By the integrability conditions (2.9) this equation is equivalent to the pair of equations

$$\frac{\partial\varphi}{\partial z}=M_0\frac{1}{z-w}\varphi+Q(z)\varphi, \tag{A 7}$$

$$\frac{\partial\varphi}{\partial w}=M_0\frac{1}{w-z}\varphi+P(w)\varphi. \tag{A 8}$$

From our assumption (2.16) we deduce that

$$\|M_0\|<1.$$

So suppose that the vector u satisfies

$$\tilde{S}_{kl}u=\mu u. \tag{A 9}$$

The matrices $Q(z)$, $P(z)$ have simple poles at the $2n-2$ coordinates being held fixed. So let D be a simply connected domain in $\mathbb{C}$ which excludes these points. Then $Q(z)$, $P(z)$ are analytic in D. By Lemma 1, Eq. (A 7) has a solution $\varphi(z,w)$ which is analytic for z in D and satisfies

$$\varphi(w,w)=u.$$

Equation (A 8) guarantees that $\varphi(z,w)$ is analytic for w in D also.

We can now use the full equation (A 6) to analytically continue this solution into $\mathbb{C}^{2n}$. In particular, the vector u which satisfies (A 9) can be an analytic function of the other $2n-2$ variables. So we can extend $\varphi(z,w)$ to an analytic function in a simply connected domain in $\mathbb{C}^{2n}$, where we restrict the variables so that $z_i \neq z_j$ unless $i=k$ and $j=l$.

Each eigenvector u of $\tilde{S}_{kl}$ provides a solution of Eq. (2.15) of the form

$$\psi=(z_k-z_l)^{\lambda\mu}\varphi,$$

where $\varphi(z_k=z_l)=u$ and

$$\tilde{S}_{kl}u=\mu u.$$

By taking a basis of these eigenvectors we generate the full solution space of (2.15), and so any solution can be written as a linear combination of these solutions.

Finally, we need a further property of the solutions of (A 7) and (A 8). Suppose there is a matrix A which satisfies

$$[A,M_0]=0,$$

$$AQ(z)A^{-1}=P(z),$$

and suppose also that

$$Au=au.$$

Then it follows from (A 7), (A 8) that

$$A\varphi(z,w)=a\varphi(w,z) \tag{A10}$$

for all w, z in D. This property extends to the solution of the full equation (A 6).

In our case the matrix A is $\hat{t}_{kl}$, which interchanges the coordinates k and l, when either $1 \leqq k < l \leqq n$, or $n+1 \leqq k < l \leqq 2n$. The eigenvalues are ± 1.

Acknowledgements. C. K. would like to thank the ETH-Zürich for their kind hospitality, and M. Loss for useful discussions. We thank C. Wieczerkowski for his interest in our work and very helpful conversations, and K. Gawedzki and R. Stora for their encouraging interest in our approach.

References

1. Witten, E.: Quantum field theory and the Jones polynomial. Commun. Math. Phys. **121**, 351–399 (1989)
2. Tsuchiya, A., Kanie, Y.: Vertex operators in the conformal field theory of P_1, and monodromy representations of the braid group. Lett. Math. Phys. **13**, 303–312 (1987); Kohno, T.: Monodromy representations of braid groups and Yang-Baxter equations. Ann. de l'Inst. Fourier de l'Univ. de Grenoble **37**, 139–160 (1987); Quantized universal enveloping algebras and monodromy of braid groups. Nagoya preprint No. 3, 1989
3. Fröhlich, J.: Statistics of fields, the Yang-Baxter equation, and the theory of knots and links, 1987 Cargèse lectures. In: Non-perturbative quantum field theory. 't Hooft G. et al. (eds.). New York: Plenum Press 1988
4. Knizhnik, V.G., Zamolodchikov, A.B.: Current algebra and Wess-Zumino model in two dimensions. Nucl. Phys. B **247**, 83–103 (1984)
5. Jones, V.F.R.: A polynomial invariant for links via von Neumann algebras. Bull. AMS **12**, 103–111 (1985); Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335–388 (1987)
6. Birman, J.S.: Braids, links, and mapping class groups. Princeton University Press 1974
7. Turaev, V.G.: The Yang-Baxter equation and invariants of links. LOMI preprint E-3-87
8. Reshetikhin, N.Yu.: Quantized universal enveloping algebras, the Yang-Baxter equation and invariants of links, I, II, LOMI preprints E-4-87, E-17-87; Fröhlich, J., King, C.: Two-dimensional conformal field theory, and three-dimensional topology, ETH-TH/89-9, preprint, February 1989, to appear in J. Mod. Phys. A (1989)
9. Bralic, N.: Exact computation of loop averages in two-dimensional Yang-Mills theory. Phys. Rev. D **22**, 3090–3103 (1980)
10. Gross, L., King, C., Sengupta, A.: Two dimensional Yang-Mills theory via stochastic differential equations. Cornell University preprint (1989), to appear in Ann. Phys. (N.Y.)

Communicated by K. Gawedzki

Received March 28, 1989; in revised form April 25, 1989