

---

# Gauge Theory and Integrability, I

by Kevin Costello<sup>\*</sup>, Edward Witten<sup>†</sup>, and Masahito Yamazaki<sup>‡</sup>

**Abstract.** Several years ago, it was proposed that the usual solutions of the Yang-Baxter equation associated to Lie groups can be deduced in a systematic way from four-dimensional gauge theory. In the present paper, we extend this picture, fill in many details, and present the arguments in a concrete and down-to-earth way. Many interesting effects, including the leading nontrivial contributions to the  $R$ -matrix, the operator product expansion of line operators, the framing anomaly, and the quantum deformation that leads from  $\mathfrak{g}[[z]]$  to the Yangian, are computed explicitly via Feynman diagrams. We explain how rational, trigonometric, and elliptic solutions of the Yang-Baxter equation arise in this framework, along with a generalization that is known as the dynamical Yang-Baxter equation.

## Contents

1	Introduction . . . . .	47
2	Review of Integrability . . . . .	48
	2.1 The Yang-Baxter Equation . . . . .	48
	2.2 Quasi-Classical $R$ -Matrix . . . . .	49
3	Four-Dimensional Gauge Theory . . . . .	51
	3.1 The Starting Point . . . . .	51
	3.2 Generalization . . . . .	53
	3.3 Wilson Lines . . . . .	54
	3.4 The Yang-Baxter Equation and Unitarity . . . . .	55
	3.5 Elementary Examples . . . . .	58

	3.6 First Look at the Framing Anomaly . . . . .	59
4	$R$ -Matrix from Crossing Wilson Lines . . . . .	61
5	OPE of Parallel Wilson Lines . . . . .	63
	5.1 Overview . . . . .	63
	5.2 Lowest Order Computation . . . . .	64
	5.3 Relation to the $R$ -Matrix . . . . .	66
	5.4 Interpretation of the Result . . . . .	66
	5.5 The Yangian . . . . .	68
6	The Framing Anomaly and Curved Wilson Lines . . . . .	69
7	Networks of Wilson Lines . . . . .	73
	7.1 Overview . . . . .	73
	7.2 Vanishing of Higher Order Anomalies . . . . .	74
	7.3 Calculating the Anomaly . . . . .	75
	7.4 Cancelling the Anomaly . . . . .	76
	7.5 Indecomposable Vertices . . . . .	76
	7.6 Vertices Constrained by Symmetries . . . . .	77
	7.7 A General Formula for the Angles at a Trivalent Vertex . . . . .	78
	7.8 Trivalent Vertices Linking Fundamental Representations of $\mathfrak{sl}_n$ . . . . .	79
	7.9 Vertices Related to $\Gamma$ -Matrices . . . . .	80
	7.10 A Vertex Connecting Representations of $E_6$ . . . . .	81
	7.11 A 4-Valent Vertex for the 56 of $E_7$ . . . . .	81
	7.11.1 Anomalies . . . . .	82
8	Two-Loop Correction to Gauge Invariance . . . . .	83
	8.1 Preliminaries . . . . .	83
	8.2 Anomalies on Wilson Lines . . . . .	84
	8.2.1 Cancellation by Counter-Terms . . . . .	85
	8.3 Enumerating Feynman Diagrams . . . . .	87
	8.3.1 Excluding Diagrams (A2), (C1), (C2) and (D2) . . . . .	87
	8.3.2 Excluding Diagrams of Type (C3) . . . . .	88
	8.3.3 Excluding Diagrams (C4) and (C5) . . . . .	89
	8.3.4 Excluding Diagrams (B) and (D1) . . . . .	89
	8.4 The Anomaly Associated to Diagram (A1) . . . . .	89

<sup>\*</sup> Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada

<sup>†</sup> School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey  
E-mail: witten@ias.edu

<sup>‡</sup> Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo

8.4.1	Color Factor . . . . .	89
8.5	Numerical Factor . . . . .	90
8.5.1	A Comment on Signs . . . . .	91
8.5.2	Regularizing the Integral . . . . .	91
8.5.3	The Failure of the Amplitude to be Gauge Invariant . . . . .	92
8.5.4	Calculating the Anomaly Integral . . . . .	93
8.6	The Anomaly as a Correction to the Al- gebra . . . . .	93
8.6.1	Matching with Standard Relations in the Yangian . . . . .	94
8.7	An Example of an Anomalous Wilson Line . . . . .	95
9	Trigonometric Solutions of the Yang-Baxter Equations . . . . .	97
9.1	Preliminaries . . . . .	97
9.2	Example . . . . .	99
9.3	The $r$ -Matrix . . . . .	99
9.4	Specializing to $\mathfrak{sl}_2$ . . . . .	100
10	Elliptic Solutions of the Yang-Baxter Equation . . . . .	100
10.1	Preliminaries . . . . .	100
10.2	Rigid Holomorphic Bundles . . . . .	101
10.3	Specializing to $N = 2$ . . . . .	101
11	The Dynamical Yang-Baxter Equation . . . . .	102
11.1	Reduction to an Abelian Subgroup . . . . .	102
11.2	$BF$ Theory and the Dynamical Yang- Baxter Equation . . . . .	104
11.3	Quantum Treatment . . . . .	105
	Acknowledgments . . . . .	107
A	Rational $R$ -Matrix for $SO_N$ . . . . .	107
B	Examples of Trigonometric and Elliptic $R$ -Ma- trix . . . . .	110
B.1	Rational . . . . .	110
B.2	Trigonometric Solution . . . . .	110
B.3	Elliptic Solution . . . . .	111
B.4	Relations . . . . .	111
C	Anomalies to Quantizing Wilson Lines . . . . .	111
D	Derivation of Eqn. (6.13) . . . . .	116
E	Evaluation of Eqn. (8.34) . . . . .	116
	References . . . . .	118

## 1. Introduction

Integrable systems of  $1+1$ -dimensional many-body physics and two-dimensional statistical mechanics first emerged in Bethe's discovery of the Bethe Ansatz [1] and Onsager's solution of the two-dimensional Ising model [2], respectively. Subsequent study led to remarkable generalizations and new discoveries, continuing to the present day.

Much of the wisdom about integrable models can be distilled into the Yang-Baxter equation [3, 4, 5, 6] and its interpretation via quantum groups [7, 8]. Many of the classic papers on this subject are reprinted in [9]. See also, for example, [10] for an introduction.

In the present paper, we will aim to explain, simplify, and further develop a new approach to the Yang-Baxter equation and related integrable systems of statistical mechanics that was proposed several years ago by one of us [11, 12]. In this approach, the solutions of the Yang-Baxter equation and their properties are deduced from a four-dimensional gauge theory that can be regarded as a  $T$ -dual version of three-dimensional Chern-Simons gauge theory. An informal introduction to this approach can be found in [13]. In spirit, this approach is in keeping with a vision that was proposed long ago by Atiyah [14]: the Yang-Baxter equation in two dimensions is deduced by starting with a theory in higher dimensions.

The purpose of the present paper is to further develop this approach, filling in many details and hopefully presenting the arguments in a concrete and down-to-earth way. In section 2, we review the basic facts about integrability and the Yang-Baxter equation that will be needed. In section 3, we introduce the relevant four-dimensional gauge theory. We explain why it leads automatically to solutions of the Yang-Baxter equation and more specifically why possible choices of compactification to two dimensions lead to rational, trigonometric, and elliptic solutions of that equation. We explain why in this theory, one can at the classical level introduce Wilson line operators associated not just to representations of the gauge group  $G$  but to representations of an infinite-dimensional algebra  $\mathfrak{g}[[z]]$ . This generalization turns out to be crucial in understanding the theory. Quantum mechanically, as we learn later,  $\mathfrak{g}[[z]]$  will be promoted to the Yangian deformation of  $\mathfrak{g}[[z]]$  (or its trigonometric or elliptic generalization). In section 3, we also review some simple examples of solutions of the Yang-Baxter equation and see how one can deduce from these simple examples that there must be a framing anomaly for Wilson line operators.

In sections 4–8, we study some increasingly subtle quantum effects in this theory. In section 4, we compute directly the first nontrivial term in the quantum  $R$ -matrix. General theorems [7, 8] actually determine the whole structure in terms of this lowest order term together with formal properties of the theory. However, our goal in the present paper is to see everything as explicitly as possible rather than relying on abstract arguments.

In section 5, we study the first nontrivial quantum correction to the operator product expansion (OPE) of Wilson line operators. We show that to get a closed OPE, one has to consider line operators associated to representations of  $\mathfrak{g}[[z]]$ , not just representations of the underlying finite-dimensional gauge group  $G$ . We also explain that this first quantum correction to the classical OPE implies that in yet higher orders, there

will have to be further deformations, which will deform  $\mathfrak{g}[[z]]$  to the Yangian (or one of its generalizations).

In section 6, we compute the framing anomaly for Wilson line operators in this theory, recovering from a Feynman diagram calculation the result that was predicted on more abstract grounds in section 3. The framing anomaly found here is somewhat analogous to the framing anomaly for Wilson operators in three-dimensional Chern-Simons theory, but its consequences are more far-reaching.

In section 7, we generalize the analysis from Wilson line operators to networks of Wilson lines – graphs that can be drawn in the plane in which the line segments are Wilson line operators and the “vertices” are invariant couplings that describe (for example) the “fusion” of two Wilson line operators to make a single one. In this context, there is a quantum anomaly that generalizes and can largely be deduced from the framing anomaly.

In section 8, we reconsider, following the elementary considerations in section 5, the deformation from  $\mathfrak{g}[[z]]$  to the Yangian. At the two-loop level, that is, in order<sup>1</sup>  $\hbar^2$ , there is a potential anomaly in the coupling of two gauge bosons to a Wilson line operator. To avoid or cancel the anomaly, Wilson lines must be associated (in the rational case) to representations of a quantum deformation of  $\mathfrak{g}[[z]]$  known as the Yangian. A surprising consequence of this is that an ordinary Wilson operator associated to a finite-dimensional representation of  $G$  may be anomalous and hence absent in the quantum theory. For example, for  $G = SO(N)$  (or any simple Lie group other than  $SU(N)$ ), there is no Wilson line operator associated to the adjoint representation.

In sections 9 and 10, we analyze the variants of the construction that lead to trigonometric and elliptic solutions of the Yang-Baxter equation, respectively. In section 11, we explain how a generalization of the Yang-Baxter equation known as the “dynamical Yang-Baxter equation” [21, 22, 23, 24] fits in this framework. In brief, one finds an ordinary Yang-Baxter equation when one expands around a classical gauge theory solution that has no moduli; moduli lead to a dynamical Yang-Baxter equation.

All of our explicit computations in the present paper are in the lowest nontrivial order in  $\hbar$  in which some quantum effect occurs. In a companion paper [25], we will explain how to construct the Yangian algebra, and its trigonometric and elliptic generalizations, “exactly” in the present framework, not just in lowest order of perturbation theory. We have put the

<sup>1</sup> There does not seem to be a standard terminology for counting loops in Feynman diagrams that contain Wilson operators. We refer to a contribution that is of order  $\hbar^n$  relative to a leading order contribution as an  $n$ -loop effect.

word “exactly” in quotes because the theory, in the form in which it has been developed so far, is a perturbative theory, so “exactly” really means “to all orders in perturbation theory.” It is anticipated that the D4-NS5 system of string theory would provide the framework for a nonperturbative description, along the lines of the study of the D3-NS5 system in [26], but this has not yet been developed.

Finally, we recall the existence of another and superficially quite different relationship between integrable spin systems and four-dimensional gauge theory [27, 28]. A connection between the two approaches is not yet known.

## 2. Review of Integrability

In this section, we review some standard facts about integrable models, aiming just to explain what is needed for the purposes of this paper. The goal of the rest of the paper will be to explain these facts from the standpoint of four-dimensional gauge theory.

### 2.1 The Yang-Baxter Equation

We consider a system of particles whose internal quantum numbers take values in some vector space  $V$ . It is often convenient to pick a basis  $\{e_i\}_{i=1}^{\dim V}$  of  $V$ . A particle is also characterized by a complex parameter  $z$  that in the context of integrable systems is known as the spectral parameter. It will play a crucial role in what follows. The particles live in a two-dimensional spacetime and travel on (possibly curved) one-dimensional worldlines in this spacetime. When two worldlines cross (Fig. 1), their spectral parameters are assumed to be unchanged, but their internal state is transformed by a matrix that in general depends on the spectral parameters.

We write this matrix as  $R(z_1, z_2) : V \otimes V \rightarrow V \otimes V$ , or in more detail, in the chosen basis, as  $R_{ij}^{kl}(z_1, z_2)$ . However, although there are interesting  $R$ -matrices that lack this property,<sup>2</sup> we will be concerned in this paper with the case that the  $R$ -matrix depends only on the difference  $z = z_1 - z_2$  of the two spectral parameters. (In some applications of the Yang-Baxter equation, the spectral parameter is interpreted as a particle momentum or rapidity, and the fact that the  $R$ -matrix depends only on the difference of spectral parameters is interpreted as a consequence of Galilean invariance or Lorentz invariance.)

<sup>2</sup> The basic example is the chiral Potts model [15, 16], in which the spectral parameter takes values in a curve of genus greater than 1. Interestingly this model arises as a root-of-unity degeneration of the model of [17], which in turn arises from supersymmetric indices of four-dimensional  $\mathcal{N} = 1$  quiver gauge theories [18, 19, 20].

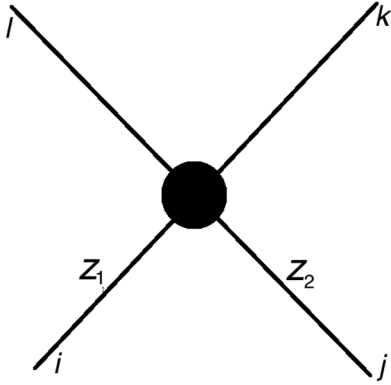


Figure 1. Crossing of two worldlines in a two-dimensional spacetime. The “blob” indicates a scattering process the amplitude for which will (in the context of the present paper) ultimately be computed via gauge theory. When two particles cross, their spectral parameters  $z_1$  and  $z_2$  are unchanged but their “internal” state is transformed.

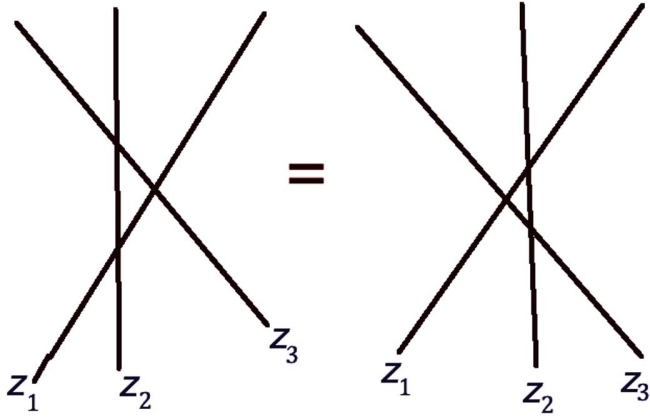


Figure 2. The Yang-Baxter equation asserts the equivalence between these two pictures.

The Yang-Baxter equation says that when three worldlines cross in a pairwise fashion, the arrangement in which they cross does not matter (Fig. 2). We denote the three particles as  $a, b, c$ , and write, for example,  $V_a$  for the vector space of internal states of particle  $a$ ,  $z_a$  for its spectral parameter, and  $R_{ab}(z_a - z_b) : V_a \otimes V_b \rightarrow V_a \otimes V_b$  for the corresponding  $R$ -matrix.<sup>3</sup> Then the Yang-Baxter equation reads<sup>4</sup>

$$(2.1) \quad R_{12}(z_1 - z_2)R_{13}(z_1 - z_3)R_{23}(z_2 - z_3) \\ = R_{23}(z_2 - z_3)R_{13}(z_1 - z_3)R_{12}(z_1 - z_2).$$

<sup>3</sup> We also denote  $R_{ab} \otimes 1 : V_a \otimes V_b \otimes V_c \rightarrow V_a \otimes V_b \otimes V_c$  simply as  $R_{ab}$ .

<sup>4</sup> The general form of this equation without assuming that the spectral parameter depends only on the difference of rapidities is simply

$$R_{12}(z_1, z_2)R_{13}(z_1, z_3)R_{23}(z_2, z_3) = R_{23}(z_2, z_3)R_{13}(z_1, z_3)R_{12}(z_1, z_2).$$

In terms of the basis  $\{e_i\}$  of  $V$ , the equation takes the imposing form

$$(2.2) \quad \sum_{o,p,q} R_{12}(z_1 - z_2)^{nm} R_{13}(z_1 - z_3)^{ql} R_{23}(z_2 - z_3)^{op}_{jk} \\ = \sum_{r,s,t} R_{23}(z_2 - z_3)^{ml}_{rt} R_{13}(z_1 - z_3)^{nt}_{sk} R_{12}(z_1 - z_2)^{sr}_{ij},$$

where the meaning of the indices is more clear in a picture (Fig. 3).

The Yang-Baxter equation can be naturally supplemented with a condition sometimes called “unitarity,” which asserts that a picture in which two worldlines cross and then cross back is equivalent to one in which they do not cross at all (Fig. 4). In formulas, the relation is

$$(2.3) \quad R_{21}(z_2 - z_1)R_{12}(z_1 - z_2) = 1.$$

All solutions of the Yang-Baxter equation studied in this paper satisfy such a unitarity condition. There is also a crossing relation, which we will come to in sections 3.5 and 3.6.

The Yang-Baxter equation is highly over-constrained, especially if the dimension of  $V$  is large: the  $R$ -matrix has  $\mathcal{O}((\dim V)^4)$  coefficients, while the Yang-Baxter equation has  $\mathcal{O}((\dim V)^6)$  components. Moreover, the presence of the spectral parameter further constrains the possible solutions to the Yang-Baxter equation. Nevertheless, the Yang-Baxter equations does have solutions and these lead to a remarkably rich theory.

## 2.2 Quasi-Classical $R$ -Matrix

While no complete classification is known of the general solution of the Yang-Baxter equation, there are more complete results for the case of a so-called quasi-classical  $R$ -matrix,<sup>5</sup> a concept that we now explain.

A quasi-classical  $R$ -matrix is a solution  $R_{\hbar}(z)$  of the Yang-Baxter equation that depends on another continuous parameter  $\hbar$  as well as on the spectral parameter  $z$ , and that is holomorphic near  $\hbar = 0$  with  $R_0(z) = 1$ . Thus  $R_{\hbar}(z)$  has an expansion near  $\hbar = 0$  that begins

$$(2.4) \quad R_{\hbar}(z) = I + \hbar r(z) + \mathcal{O}(\hbar^2).$$

Here  $r(z)$  is called the classical  $r$ -matrix.

By considering the  $\mathcal{O}(\hbar^2)$  term of the Yang-Baxter equation (2.1), one learns that the classical  $r$ -matrix obeys an equation that is known as the classical Yang-Baxter equation:

<sup>5</sup> There are known solutions of Yang-Baxter equations which are not quasi-classical. The chiral Potts model [15, 16] is again a basic example.

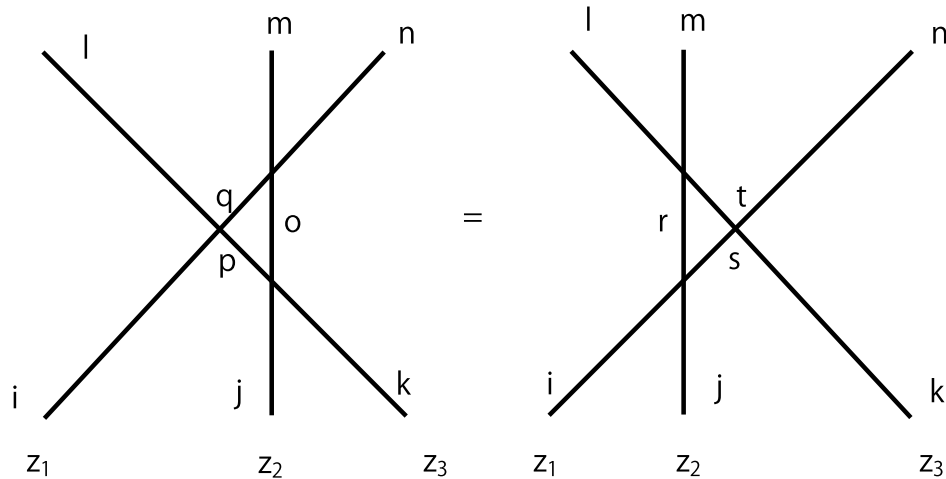


Figure 3. A graphical representation of the Yang-Baxter equation. On the left, one sums over labels  $p, q, o$ , and on the right one sums over  $r, s, t$ . An appropriate  $R$ -matrix element is attached to each vertex.

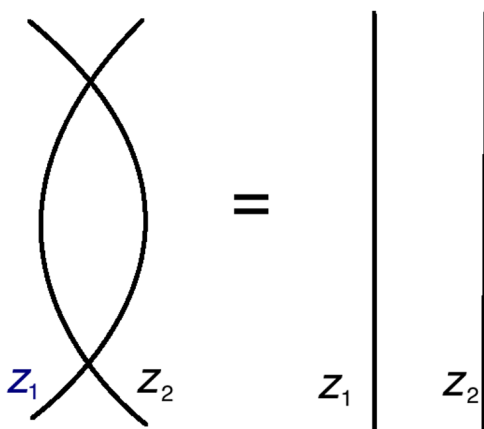


Figure 4. In the context of the Yang-Baxter equation, “unitarity” is the equivalence of these two pictures.

$$(2.5) \quad [r_{12}(z_1 - z_2), r_{13}(z_1 - z_3) + r_{23}(z_2 - z_3)] \\ + [r_{13}(z_1 - z_3), r_{23}(z_2 - z_3)] = 0.$$

Note that this equation is quadratic in the classical  $r$ -matrix  $r$ , whereas the original Yang-Baxter equation was cubic in the  $R$ -matrix  $R$ .

Now, Belavin and Drinfeld [29] classified solutions of the classical Yang-Baxter equation (2.5), modulo trivial equivalences,<sup>6</sup> under certain assumptions. The assumptions were motivated by the examples which were known at that time. The solution is assumed to be associated to the Lie algebra  $\mathfrak{g}$  of a semi-simple<sup>7</sup> Lie group  $G$ . No reality condition will be important in the present paper, so we consider  $G$  to be

<sup>6</sup> The classical Yang-Baxter equation is invariant under conjugation and under adding to  $r$  a multiple of the identity. The latter possibility reflects the fact that the Yang-Baxter equation is invariant under multiplying  $R(z)$  by a function of  $z$ .

<sup>7</sup> In gauge theory, it is natural to consider the somewhat larger class of Lie groups consisting of those whose Lie algebra

is a complex semi-simple Lie group with complex Lie algebra  $\mathfrak{g}$ . The classical  $r$ -matrix  $r$  is assumed to be an element of  $\mathfrak{g} \otimes \mathfrak{g}$ :

$$(2.6) \quad r(z) = \sum_{a,b} r_{ab}(z)(t^a \otimes t^b),$$

where  $t^a$  are a basis of  $\mathfrak{g}$ . The classical  $r$ -matrix is assumed to be non-degenerate, namely  $\det_{a,b}(r_{ab}(z)) \neq 0$ .

Then the result shows that the poles of  $r(z)$  in the complex plane spans a lattice, which is either of rank 0, 1 or 2. Solutions of the classical Yang-Baxter equation of any of the three types are almost uniquely determined<sup>8</sup> by the choice of  $\mathfrak{g}$  and the rank of the lattice. The solutions for rank 0, 1, and 2 can be written explicitly in terms of rational, trigonometric, and elliptic functions and are known as rational, trigonometric, and elliptic solutions of the Yang-Baxter equation.

In the language of representation theory, rational, trigonometric and elliptic solutions have their algebraic counterparts, namely the Yangian  $Y_{\hbar}(\mathfrak{g})$ , the quantum affine algebra  $U_{q,\hbar}(\mathfrak{g})$  and the elliptic algebra  $E_{q,\tau,\hbar}(\mathfrak{g})$ . Solutions of the full (rather than classical)

algebra admits an invariant, non degenerate bilinear form. An important example is a reductive group, which is locally the product of a semi-simple group and a torus (an abelian group). We will find at least two reasons to consider reductive groups in this paper. One reason is that the simplest example for some purposes is actually the case  $G = GL_N(\mathbb{C})$ , which is reductive but not semi-simple. Another reason is that in the framework we will follow in this paper, trigonometric solutions of the Yang-Baxter equation are most easily understood starting with a gauge group that is reductive but not semi-simple.

<sup>8</sup> In the elliptic case, there is a discrete choice to be made. For  $\mathfrak{g} = \mathfrak{sl}_N$ , one is free to pick a generator of the finite group  $\mathbb{Z}_N$ . There also are some subtleties in the trigonometric case, involving the possibility of an “external field.” We explain in sections 9 and 10 what these issues mean from the point of view of four-dimensional gauge theory.



Yang-Baxter equation depend on the choice of a representation of one of these algebras and the  $R$ -matrix  $R_h(z) \in \text{End}(V \otimes V)$  is then an intertwiner for the tensor products of these representations. In the algebraic approach, the spectral parameter enters as part of the data needed to specify a representation.

Elliptic solutions of the classical Yang-Baxter equation exist only for  $\mathfrak{g} = \mathfrak{sl}_N$ , whereas trigonometric and rational solutions exist for any semisimple  $\mathfrak{g}$ . Rational solutions of the classical Yang-Baxter equation have  $G$  as a group of symmetries, while trigonometric solutions admit only the maximal torus of  $G$  as a symmetry group and elliptic solutions for  $\mathfrak{sl}_N$  have only a finite group  $\mathbb{Z}_N \times \mathbb{Z}_N$  of automorphisms.

In this paper, we will see how quasi-classical  $R$ -matrices with these properties can emerge from four-dimensional gauge theory.

### 3. Four-Dimensional Gauge Theory

#### 3.1 The Starting Point

The four-dimensional gauge theory that is relevant to our subject [11, 12] may be described as follows.

The theory in question is only defined on a four-manifold with some additional structure. We start with the basic case, which is a product 4-manifold  $\mathbb{R}^2 \times \mathbb{C}$  with real coordinates<sup>9</sup>  $x, y$  on  $\mathbb{R}^2$  and a holomorphic coordinate  $z$  on  $\mathbb{C}$ .

The fundamental field of our theory is a 3-component partial connection

$$(3.1) \quad A = A_x dx + A_y dy + A_{\bar{z}} d\bar{z},$$

where we did not include the component  $A_z dz$  that is of type (1,0) along  $\mathbb{C}$ . The fields  $A_x, A_y$ , and  $A_{\bar{z}}$  all depend nontrivially on  $z$  and  $\bar{z}$  as well as  $x$  or  $y$ ; that is, they are not constrained to vary holomorphically or antiholomorphically on  $\mathbb{C}$ . Since  $A_z$  is missing, it would not be possible to place a reasonable reality condition on this space of fields. Instead we take the gauge group to be a complex Lie group  $G$  with complex Lie algebra  $\mathfrak{g}$ , and view  $A_x, A_y$ , and  $A_{\bar{z}}$  as independent complex fields. The construction will make use of an invariant and nondegenerate bilinear form on  $\mathfrak{g}$ , which we will denote as  $\text{Tr}$ . The notation is motivated by the fact that if  $\mathfrak{g}$  is semisimple or more generally if it is reductive (the direct sum of a semi simple Lie algebra with an abelian one), then an invariant quadratic form can be defined as the trace in a suitable representation. However, our discussion in this paper applies whether  $\text{Tr}$  actually has this interpretation or not. For a simple summand of  $\mathfrak{g}$ , the Killing form of  $\mathfrak{g}$  gives an invariant nondegenerate quadratic

form. We choose an orthonormal basis  $t_a$  of  $\mathfrak{g}$  with respect to this Killing form and normalize  $\text{Tr}$  by

$$(3.2) \quad \text{Tr}(t_a t_b) = \delta_{ab}.$$

The action of our theory is given by

$$(3.3) \quad S = \frac{1}{2\pi} \int_{\mathbb{R}^2 \times \mathbb{C}} dz \wedge \text{CS}(A),$$

where  $\text{CS}(A)$  is the Chern-Simons three-form

$$(3.4) \quad \begin{aligned} \text{CS}(A) &:= \text{Tr} \left( A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right) \\ &= \varepsilon^{ijk} \text{Tr} \left( A_i \partial_j A_k + \frac{2}{3} A_i A_j A_k \right). \end{aligned}$$

Here and afterwards the indices  $i, j, \dots$  run over  $x, y$  and  $\bar{z}$  ( $\varepsilon$  is a totally antisymmetric tensor with  $\varepsilon^{xy\bar{z}} = 1$ ). The VEV (vacuum expectation value) of an observable  $\mathcal{O}$  is given by the path-integral

$$(3.5) \quad \langle \mathcal{O} \rangle = \frac{\int \mathcal{D}A \mathcal{O} \exp\left(\frac{iS}{\hbar}\right)}{\int \mathcal{D}A \exp\left(\frac{iS}{\hbar}\right)}.$$

The action  $S$  is obviously not invariant under four-dimensional diffeomorphisms, because the use of the 1-form  $dz$  spoils the four-dimensional symmetry. Nor does it have the three-dimensional diffeomorphism symmetry of three-dimensional Chern-Simons theory; this is the symmetry that enables one to define quantum invariants of knots. But we still have two-dimensional diffeomorphism symmetry - invariance under orientation-preserving diffeomorphisms of  $\mathbb{R}^2$  (or of its generalization  $\Sigma$  that will be introduced later). This will ultimately lead to the Yang-Baxter equation and the unitarity relation.

We understand the action  $S$  as a holomorphic function of complex variables  $A_x, A_y, A_{\bar{z}}$ , and this implies that the construction that we will be describing is somewhat formal. There is no difficulty in formally carrying out perturbation theory in such a holomorphic theory. That approach was taken in [11, 12] and it is the approach that we will follow here. (We expect that a nonperturbative definition of the theory can be given by considering the D4-NS5 system of string theory, along the lines of the study of the D3-NS5 system in [26], but we will not pursue this in the present paper.) The parameter  $\hbar$  that appears in the action is, at the quantum level, the loop-counting parameter. In the semi-classical limit  $\hbar \rightarrow 0$ , this parameter will be identified with the parameter of the same name that appears in the quasi-classical  $R$ -matrix (2.4). The parameter  $\hbar$  has dimensions of length, in the sense that for  $C = \mathbb{C}$ , the theory is invariant under a common rescaling of  $z$  and  $\hbar$ . The factor of  $1/(2\pi)$  in the action is included here to match with the literature on integrable models.

<sup>9</sup> When convenient, we will denote  $x$  and  $y$  as  $x^1$  and  $x^2$ .

A reflection of the fact that the construction is formal and leads (in the form we present here) only to a perturbative theory is the following. There is no quantization condition for  $\hbar$  that will ensure that the action is gauge-invariant mod  $2\pi\mathbb{Z}$ . This contrasts with three-dimensional Chern-Simons theory, which is defined with such a condition.

The action is invariant, modulo surface terms that are irrelevant in perturbation theory, under gauge transformations acting in the usual way.

$$(3.6) \quad A_i \mapsto g^{-1}A_i g + g^{-1}\partial_i g, \quad (i = x, y, \bar{z}).$$

This is true because the Chern-Simons three-form is gauge-invariant modulo an exact form. Alternatively, we can integrate by parts to put the action in a manifestly gauge-invariant form, after discarding surface terms that are irrelevant in perturbation theory:

$$(3.7) \quad S = -\frac{1}{2\pi} \int_{\mathbb{R}^2 \times \mathbb{C}} z \text{Tr} F \wedge F.$$

This is the standard topological term of the Yang-Mills theory, where the  $\theta$ -angle now depends linearly on  $z$ .

Some readers might be more comfortable starting with a standard 4-component connection

$$(3.8) \quad A = A_x dx + A_y dy + A_{\bar{z}} d\bar{z} + A_z dz,$$

with gauge transformations acting in the usual way on all four components, and again with the action (3.3). In this case, one finds that due to the presence of the differential form  $dz$  in the action (3.3), the  $A_z$  component drops out from the action, and hence we have an extra gauge symmetry

$$(3.9) \quad A \mapsto A + \chi dz.$$

We can then fix this extra gauge symmetry by choosing a gauge  $A_z = 0$ . The 4-component gauge transformation for the 4-component gauge field (3.9) is not consistent with this gauge since it will in general generate a non-trivial  $A_z$  component. However, a combination of the 4-component gauge transformation, with the extra gauge symmetry (3.9) with  $\chi = -A_z$ , remains as a residual gauge symmetry. This is the 3-component gauge transformation (3.6).

In the following, we will always choose  $A_z = 0$ , so that  $A$  is the 3-component connection and the only remaining gauge symmetry is the conventional gauge transformation (3.6).

A possibly more familiar theory that is defined in a similar way with a partial connection is holomorphic Chern-Simons theory. This theory is defined on a Calabi-Yau threefold  $X$  with holomorphic 3-form  $\Omega$ . The dynamical variable is a  $(0,1)$  connection  $A =$

$\sum_{i=1}^3 A_i d\bar{z}^i$  and the action is the integral of the Chern-Simons  $(0,3)$ -form, wedged with  $\Omega$ :

$$(3.10) \quad S = \int_X \Omega \wedge \text{CS}(A^{(0,1)}).$$

The definition of this action depends only on the complex structure and holomorphic volume form  $\Omega$  of the 3-fold  $X$ .

In this light, the four-dimensional theory of (3.3) is intermediate between ordinary Chern-Simons theory in three dimensions and holomorphic Chern-Simons theory on a Calabi-Yau threefold. These theories arise as effective theories of branes in the topological A-model and B-model respectively [30] and are related by mirror symmetry. The four-dimensional theory that we will be studying here is intermediate between the two cases and on an appropriate four-manifold can be related by  $T$ -duality - mirror symmetry in some but not all dimensions of spacetime - to either one of them.

The classical equations of motion of the theory read

$$(3.11) \quad F_{xy} = 0, \quad F_{x\bar{z}} = F_{y\bar{z}} = 0.$$

This means that the gauge field defines a flat bundle on  $\mathbb{R}^2$ , which then varies holomorphically as we move along  $\mathbb{C}$ .

The equations (3.11) imply that all local gauge-invariant quantities that can be constructed from the field  $A$  actually vanish. This is the reason that the theory works at the quantum level. Because the loop-counting parameter  $\hbar$  has dimensions of length or inverse mass, the theory is unrenormalizable by power-counting. But this does not cause difficulty because all conceivable counterterms actually vanish by the equations of motion. The theory thus can be quantized in perturbation theory [11, 12]. However, it is affected by framing anomalies somewhat similar to those of three-dimensional Chern-Simons theory, but more subtle.

The fact that the theory is unrenormalizable by power counting actually leads to a very important simplification. After gauge-fixing, when one concretely constructs the theory in perturbation theory, it is infrared-free. The fact that the theory is infrared-free makes it straightforward, once one introduces Wilson line operators, to deduce a local procedure to compute their expectation values. From this local procedure, one then can immediately recover the Yang-Baxter equation of an integrable system. (This will be explained in detail in section 3.4.) By contrast, three-dimensional Chern-Simons theory is renormalizable by power counting and does not lead as directly to a local picture.

### 3.2 Generalization

We comment next on replacing  $\mathbb{R}^2 \times \mathbb{C}$  by a more general 4-manifold.

In this paper, we will exclusively study the special case that the 4-manifold is a product of two Riemann surfaces,

$$(3.12) \quad M = \Sigma \times C,$$

where  $\Sigma$  is a smooth oriented 2-manifold, and  $C$  is a complex manifold endowed with a holomorphic (or sometimes meromorphic, as discussed shortly) 1-form  $\omega$ , which plays a role similar to  $\Omega$  in the holomorphic Chern-Simons theory of eqn. (3.10). We will sometimes refer to  $\Sigma$  as the “topological plane” and  $C$  as the “holomorphic plane,” though in general neither one of them is really a plane.

We can then define a natural generalization of the action (3.3) by

$$(3.13) \quad I = \frac{1}{2\pi} \int_M \omega \wedge \text{CS}(A).$$

As long as  $\omega$  is closed, this action is gauge-invariant modulo total derivatives that do not affect perturbation theory.

Now we should discuss the possible role of zeroes and poles of  $\omega$ . Naively, since the action involves only the ratio  $\omega/\hbar$ , a zero of  $\omega$  corresponds to a point at which  $\hbar \rightarrow \infty$ . Thus, in a theory that one only knows how to define perturbatively, it should not be straightforward to make sense of the behavior near a zero of  $\omega$ . We expect that essentially new ingredients are needed to make sense of that behavior. We will not explore this issue in the present paper.

Conversely, near a pole of  $\omega$ ,  $\hbar$  is effectively going to zero and perturbation theory should be within reach. However, poles of  $\omega$  are still subtle for the following reason. If  $\omega$  has a pole at a point  $p \in C$ , then  $d\omega$  does not vanish near  $p$  but is a distribution supported at  $p$ . Accordingly, gauge-invariance will fail unless we place some suitable conditions near  $p$  on the gauge field  $A$  and the gauge parameter  $g$ . If  $\omega$  has a double pole at  $p$ , one can restore gauge invariance by asking that  $A = 0$  at  $p$  and  $g = 1$  at  $p$ . What one has to do if  $\omega$  has a simple pole at  $p$  is more subtle and will be described in section 9.

Only simple and double poles are relevant, as one sees if one considers the possibilities for a complex Riemann surface  $C$  with a holomorphic one-form  $\omega$  that is allowed to have poles but not zeroes. By the Riemann-Roch theorem, the number of zeroes of any meromorphic differential  $\omega$  minus the number of its poles is  $2g - 2$ , where  $g$  is the genus of  $C$ . Thus if  $\omega$  has no zeroes,  $C$  must have genus 0 or 1. Moreover, for  $g = 0$  we have either (1) a single pole with multiplicity 2, which corresponds to  $C = \mathbb{C}$  with differential  $dz$

(which has a double pole at  $\infty$ ) or (2) two simple poles, in which case we can take  $C = \mathbb{C}^\times = \mathbb{C}/\mathbb{Z}$  with differential  $\omega = \frac{dz}{z}$ , which has simple poles at 0 and  $\infty$ . For  $g = 1$ , there are no poles at all;  $C$  is a complex torus or elliptic curve  $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$  (with modulus  $\tau$ ) with the holomorphic differential  $dz$ . In each case, the choice of  $\omega$  is unique up to a normalization constant that can be absorbed in rescaling  $\hbar$ .

Summarizing, we have the following three possibilities for  $C$ :

$$(3.14) \quad \begin{aligned} C = \mathbb{C}, \quad \omega = dz, \quad &\text{double pole at } \{\infty\}, \quad (\text{rational}), \\ C = \mathbb{C}^\times, \quad \omega = \frac{dz}{z}, \quad &\text{poles at } \{0, \infty\}, \quad (\text{trigonometric}), \\ C = E = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z}), \quad \omega = dz, \quad &\text{no poles,} \quad (\text{elliptic}). \end{aligned}$$

As indicated, the three choices of  $C$  match the three broad classes of quasi-classical R-matrices that were summarized in section 2.2, if we assume that  $C$  parametrizes the spectral parameter of the classical  $r$ -matrix. Developing this relationship is the purpose of the present paper.

A notable fact is that the three examples are all abelian groups. This is no coincidence, of course. Since the holomorphic differential  $\omega$  on  $C$  has no zeroes, its inverse is a holomorphic vector field  $\zeta = \omega^{-1}$  that generates an abelian group symmetry. For the three cases, in the coordinates used in eqn. (3.14), the group action is  $z \rightarrow z + a$  in the case that  $C$  is the complex plane or an elliptic curve, or  $z \rightarrow \lambda z$  in the case of  $\mathbb{C}^\times$ . It is because of this group action, which is a symmetry of the action (3.3) and the theory constructed from it, that the  $R$ -matrix  $R(z_1, z_2)$  that we eventually construct is a function only of the difference  $z_1 - z_2$  or the ratio  $z_1/z_2$ , as the case may be.

Though this will not be developed in the rest of the paper, we will briefly describe a more general possible choice of 4-manifold. Suppose that the 4-manifold  $M$  admits a complex-valued closed 1-form  $\omega$ . We require that  $\text{Re}\omega$  and  $\text{Im}\omega$  are everywhere linearly independent. This means that locally  $\omega = df + ig$ , where  $f$  and  $g$  are real-valued functions and  $df$  and  $dg$  are linearly independent.  $M$  can then locally be foliated by the smooth two-manifolds that are defined by setting  $f$  and  $g$  to constants. Thus  $M$  has a two-dimensional integrable foliation. The gauge field  $A$  is a 3-component partial  $\mathfrak{g}$ -valued connection, or alternatively it is an ordinary connection with the extra gauge symmetry

$$(3.15) \quad A \mapsto A + \chi \omega.$$

The action is still given by (3.13).

The considerations to this point have been purely classical, but there are important quantum corrections. As we discuss briefly in section 3.6 (see [11, 12])



for a detailed account), at the quantum level there is a framing anomaly which means that we can only define the theory on some, but not all, 4-manifolds  $M$  of the type mentioned above. For the product manifold of (3.12), the framing anomaly implies that  $\Sigma$  must be equipped with a framing. In particular, if compact,  $\Sigma$  must be a two-torus. For integrable lattice models associated to solutions of the Yang-Baxter equation, the most important examples are that  $\Sigma$  is  $\mathbb{R}^2$  or a two-torus.

### 3.3 Wilson Lines

Now let us consider the gauge-invariant operators of the theory. There are no local ones because they all vanish by the equations of motion. The simplest gauge-invariant operators – and the only ones that we will study in the present paper – are Wilson line operators.

In ordinary gauge theory, a natural gauge-invariant quantity is the trace, in some representation  $\rho$  of the gauge group, of the holonomy of the connection around a closed loop  $K$ . Quantum field theorists usually write this quantity as

$$(3.16) \quad W_\rho(K) = \text{Tr}_\rho P \exp \left( \oint_K A_i(x^1, x^2, z, \bar{z}) dx^i \right),$$

where  $P$  denotes path-ordering along the loop  $K$ , and  $\text{Tr}_\rho$  the trace in the representation  $\rho$ . In the present context,  $K$  cannot be an arbitrary loop in the four-manifold  $\Sigma \times C$ . On the contrary, because we only have a partial connection with no  $dz$  term, there is no notion of parallel transport in the  $C$  direction.<sup>10</sup> Accordingly, we are restricted to the case that  $K$  is a loop in the topological plane  $\Sigma$ , at a specified point<sup>11</sup>  $z = z_0$  in  $C$ . This already makes contact in a preliminary way with some aspects of the standard Yang-Baxter picture that we reviewed in section 2.1. A Wilson operator is supported on a 1-manifold  $K$  in the two-manifold  $\Sigma$  (which one can think of as the worldline of a particle in a two-dimensional spacetime), and it is labeled by a spectral parameter, that is, by a point in  $C$ , and by a choice of a representation  $\rho$  of  $G$ . Here  $\rho$  will play the role of the vector space  $V$  of internal states of a particle, introduced at the beginning of section 2.1.

We can also introduce more general Wilson operators, which do not have analogs in standard gauge theories. The existence of these operators is related

<sup>10</sup> Either there is no  $A_z$  and no way to define parallel transport along a path on which  $z$  is not constant, or there is an  $A_z$  but also an extended gauge invariance (3.15), and parallel transport in the  $z$  direction is not gauge-invariant.

<sup>11</sup> This classical statement will later be subject to some revision because of the framing anomaly.

to the fact that the loop  $K$  is highly restricted, as described in the last paragraph. At the classical level, these Wilson loops are labeled by a representation  $\hat{\rho}$  of the infinite-dimensional Lie algebra  $\mathfrak{g}[[z]] = \prod_{n \geq 0} (\mathfrak{g} \otimes z^n)$  of series in  $z$  whose coefficients are in the finite-dimensional Lie algebra  $\mathfrak{g}$ . (The same algebra  $\mathfrak{g}[[z]]$  will appear regardless of the choice of  $C$  because the considerations will be local along  $C$ .) To be more exact, Wilson operators supported at  $z = 0$  will be associated to representations of  $\mathfrak{g}[[z]]$ . Wilson operators supported at  $z = z_0$  are similarly associated to representations of  $\mathfrak{g}[[z - z_0]]$  (which is obtained from  $\mathfrak{g}[[z]]$  by  $z \rightarrow z - z_0$ ).

Since the relevant concepts may be unfamiliar, we pause for an explanation. Roughly speaking, an element of  $\mathfrak{g}[[z]]$  is a  $\mathfrak{g}$ -valued function of  $z$ . If  $\mathfrak{g}$  has a basis  $t_a$ ,  $a = 1, \dots, \dim \mathfrak{g}$ , then a basis, in the relevant sense, of the space of  $\mathfrak{g}$ -valued functions of  $z$  is provided by

$$(3.17) \quad t_{a,n}(z) = t_a z^n, \quad n \geq 0.$$

So to define a representation of  $\mathfrak{g}[[z]]$ , we need to give, for every  $a$  and  $n$ , a matrix (or operator)  $t_{a,n}$  that represents the action of  $t_{a,n}(z)$ . Concretely, if  $\mathfrak{g}$  has a basis with  $[t_a, t_b] = f_{ab}^c t_c$ , then the natural commutation relation for  $\mathfrak{g}$ -valued functions of  $z$  is  $[t_a z^n, t_b z^m] = f_{ab}^c t_c z^{n+m}$ . Therefore, the corresponding representation matrices should obey

$$(3.18) \quad [t_{a,n}, t_{b,m}] = f_{ab}^c t_{c,n+m}.$$

It is important that there is no central extension here and that this algebra has finite-dimensional representations. We will be primarily interested in finite-dimensional representations, and more specifically representations with the property that there is some  $n_0$  such that  $t_{a,n} = 0$  for  $n \geq n_0$ . To orient the reader, we consider the first nontrivial example, which arises for  $n_0 = 2$ . The nonzero generators are just  $t_{a,0}$ , which generates the finite-dimensional algebra  $\mathfrak{g}$ , and  $t_{a,1}$ , which commutes with itself and transforms in the adjoint representation of  $\mathfrak{g}$ . To construct a representation  $\hat{\rho}$  of this algebra, we can take a direct sum  $\rho \oplus \rho$  of two copies of any representation  $\rho$  of  $\mathfrak{g}$ . If  $t_a$  are the representation matrices of  $\rho$ , then a representation of  $\mathfrak{g}[[z]]$  is given by

$$(3.19) \quad t_{a,0} = \begin{pmatrix} t_a & 0 \\ 0 & t_a \end{pmatrix}, \quad t_{a,1} = \begin{pmatrix} 0 & t_a \\ 0 & 0 \end{pmatrix}.$$

This representation is indecomposable as a representation of  $\mathfrak{g}[[z]]$ , though it is decomposable as a representation of  $\mathfrak{g}$ . There are many elaborations on this theme with  $n_0 \geq 2$ .

Because the line  $K$  in (3.16) is supported at a point in  $C$ , its definition depends only on the components  $A_i(x^1, x^2, z, \bar{z})$  of the gauge field, with  $i = 1, 2$ . We will

preserve this fact in defining (at the classical level) more general Wilson operators. For fixed  $x^1$  and  $x^2$ ,  $A_i(x^1, x^2, z, \bar{z})$  is a  $\mathfrak{g}$ -valued function of  $z$  and  $\bar{z}$ . Formally setting  $\bar{z} = 0$ , we get (for each point in the topological plane and each  $i$ ) a  $\mathfrak{g}$ -valued function of  $z$ , namely  $\widehat{A}_i(x^1, x^2, z) = A_i(x^1, x^2, z, 0)$ . A precise definition of what we mean by setting  $\bar{z}$  to 0 with  $z$  fixed is that we define

$$(3.20) \quad \widehat{A}_i(x^1, x^2, z) := \sum_{k \geq 0} \frac{z^k}{k!} \frac{\partial^k}{\partial z^k} A_i(x^1, x^2, z, \bar{z}) \Big|_{z=\bar{z}=0}.$$

We do not need to worry about convergence of this series, because we consider representations that are annihilated by a sufficiently high power of  $z$ . Thus for any given representation, we can terminate the series after finitely many terms and consider  $\widehat{A}_i$  to have a polynomial dependence on  $z$ .

Next we consider gauge transformations. The generator of such a gauge transformation is a  $\mathfrak{g}$ -valued function  $u(x^1, x^2, z, \bar{z})$ . We can restrict such a function to  $\bar{z} = 0$ , in the same sense described in the last paragraph, and extract a  $\mathfrak{g}$ -valued function  $\widehat{u}(x^1, x^2, z) = u(x^1, x^2, z, 0)$ , which we can interpret as a  $\mathfrak{g}[[z]]$ -valued function of  $x^1, x^2$ . The theory we are studying is invariant under a gauge transformation of  $A_i$  generated by  $u(x^1, x^2, z, \bar{z})$ . When we restrict to  $\bar{z} = 0$ , the action of  $u(x^1, x^2, z, \bar{z})$  on  $A_i$  becomes an action of  $\widehat{u}(x^1, x^2, z)$  on  $\widehat{A}_i(x^1, x^2, z)$ . Here we can view  $\widehat{u}(x^1, x^2, z)$  as a  $\mathfrak{g}[[z]]$ -valued function on  $\Sigma$ , and its action of  $\widehat{A}_i$  is the natural action of such a function on  $\widehat{A}_i$ , viewed as a  $\mathfrak{g}[[z]]$ -valued gauge field on  $\Sigma$ .

So finally if  $\widehat{\rho}$  is a representation of  $\mathfrak{g}[[z]]$  of the allowed class, we can define a corresponding Wilson operator by modifying eqn. (3.16) in an almost trivial way:

$$(3.21) \quad W_{\widehat{\rho}}(K) = \text{Tr}_{\widehat{\rho}} P \exp \left( \oint_K A_i(x^1, x^2, z, 0) dx^i \right).$$

Here  $A_i(x^1, x^2, z, 0)$  is expanded around  $z = 0$  with fixed  $\bar{z} = 0$ . Wilson lines supported at some other point  $z = z_0$  in  $C$  are similarly defined by expanding around  $z = z_0$  with fixed  $\bar{z} = \bar{z}_0$ . (In this case, one considers representations of  $\mathfrak{g}[[z - z_0]]$ .)

However, here we should point out a crucial subtlety that is important for applications of the extension to  $\mathfrak{g}[[z]]$ . It is very undesirable to take a trace in eqn. (3.21) because this causes much of the interesting structure to disappear. We can see that by going back to eqn. (3.19). This representation is indecomposable as a representation of  $\mathfrak{g}[[z]]$ , and it does not just come from a representation of the finite-dimensional algebra  $\mathfrak{g}$ . The holonomy operator along a given path in this representation is different from what it would be if one sets  $t_{a,1} = 0$ , which would give a decomposable representation of  $\mathfrak{g}[[z]]$ . But if we take the *trace* of the holonomy, then  $t_{a,1}$  will play no role

because it is strictly upper triangular, and we cannot distinguish the given representation from its decomposable cousin.

Because the theory is infrared-free, as explained at the end of section 3.1, the holonomy itself rather than its trace is a meaningful observable. To see this, we take  $\Sigma = \mathbb{R}^2$  and we consider a 1-manifold  $K \subset \Sigma$  (supported at a point in  $C$ ) that is not compact but has its ends at infinity along  $\Sigma$ . Because the theory is infrared-free, the gauge field can be considered to vanish at infinity along  $\Sigma$  and then the holonomy along  $K$  is a gauge-invariant observable, with no need to take its trace. It is really this that gives power to the fact that the theory has Wilson operators associated to a large class of representations of  $\mathfrak{g}[[z]]$ . The importance of the extension to  $\mathfrak{g}[[z]]$  will not be fully clear until we analyze the operator product expansion of Wilson operators in section 5.

### 3.4 The Yang-Baxter Equation and Unitarity

Having defined Wilson operators, we can now return to Figs. 2 and 4 in which the Yang-Baxter equation and the associated unitarity relation are illustrated. We now interpret these figures as representing configurations of Wilson lines  $K_i \in \Sigma$  (supported at the indicated points  $z_i \in C$ ). It is now not hard to argue for the equivalence of the left and right hand sides of these pictures. Two-dimensional diffeomorphism symmetry means that we are free to move the  $K_i$  around as long as we do not change the topology of the situation. However, this alone is not quite enough to prove the equivalences suggested in the pictures. For example, in Fig. 2, we are free to move the “middle” Wilson line to the left or right, as long as we do not try to pass through a configuration (Fig. 5) in which the three lines all meet at a point; such a configuration is not equivalent by a diffeomorphism of  $\Sigma$  to a configuration without a triple intersection. In trying

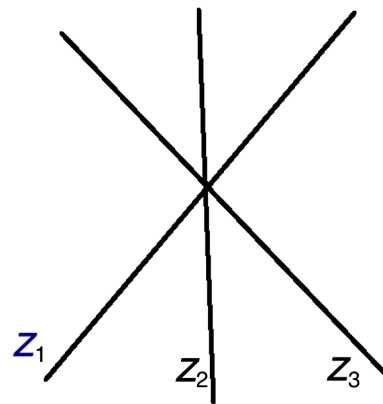


Figure 5. Three lines meeting at a point in  $\Sigma$ , but with distinct values of the spectral parameters  $z_i$ .

to prove the equivalences between the left and right of Fig. 2 by moving the middle Wilson line from left to right, we have to ask whether there is a discontinuity in the path integral at the moment that a triple intersection occurs. However, as long as the spectral parameters  $z_1, z_2, z_3$  are all distinct, none of the lines are meeting in four dimensions and it is manifest that the configuration that has a triple intersection when projected to  $\Sigma$  is not associated to any singularity.<sup>12</sup> In particular there is no discontinuity and the pictures on the left and right of Fig. 2 are equivalent as long as the  $z_i$  are distinct (in fact, it is enough that they are not all equal). Likewise the pictures on the left and right of Fig. 4 are equivalent as long as  $z_1 \neq z_2$ .

It takes more than this to argue that the theory has an  $R$ -matrix that satisfies the Yang-Baxter equation and the unitarity relation. The usual  $R$ -matrix formalism, as summarized in section 2.1, involves a much more specific interpretation of the pictures. Each line in Fig. 1 is supposed to have associated to it a space  $V$  of “internal states” accessible to a particle. In the present framework, the meaning of this is clear: a line is a Wilson line associated to some representation  $\rho$  of  $\mathfrak{g}$  (or more generally of  $\mathfrak{g}[[z]]$ ), and  $\rho$  corresponds to  $V$ . But the usual  $R$ -matrix picture is much simpler than one would expect in quantum field theory in general. In the usual  $R$ -matrix picture, each line segment between two crossings is labeled by a basis vector  $e_i$  of  $V$ , and to a crossing one associates a local factor, the  $R$ -matrix element  $R_{ij}^{kl}(z_1 - z_2)$ , which depends only on the data at a particular crossing and not on any other details in which the local picture is embedded. Moreover, this  $R$ -matrix element depends only on the difference  $z_1 - z_2$ . In quantum field theory in general, one would not expect a local picture like this. Finally, though in standard presentations of  $R$ -matrix theory one might take this for granted and skip it over, it is noteworthy that in  $R$ -matrix theory, the two-dimensional regions bounded by the lines do not carry any labels. This is a nontrivial point and in fact there is a generalization of the Yang-Baxter equation (the dynamical Yang-Baxter equation [21, 22, 23, 24]) in which the bulk regions do carry labels, above and beyond the labels carried by the line segments.

In trying to explain these facts in the present context, the most basic question is why there is a local picture of any sort. The reason for this is that the theory is infrared-free, as was noted at the end of section 3.1. Concretely, in constructing perturbation theory, as we will do starting in section 4, one picks a Riemannian metric on  $\Sigma \times C$ . If one scales up the metric on  $\Sigma$  by a large factor, so that different crossings are very far apart (compared to the distances between

the points in  $C$  at which a given set of Wilson line operators are supported), then the infrared-free nature of the theory guarantees that some kind of local picture will be possible.

To explain more, let us first ask what would happen in the absence of any line operators. The theory under study is topological in the  $\Sigma$  direction, so (ignoring further subtleties that arise because we are dealing with a theory whose action is a holomorphic function of complex variables) in general we would expect the theory to have a space  $\mathcal{H}$  of quantum states. These would roughly correspond to vacua of a standard quantum field theory. In general, one would expect to label the regions between the lines – that is, any region of  $\Sigma$  that is not near one of the Wilson operators – by a basis vector of  $\mathcal{H}$ . Accordingly, if  $\mathcal{H}$  has dimension bigger than 1, we would get something like the dynamical Yang-Baxter equation [21, 22, 23, 24], with labels for regions as well as line segments, rather than the standard Yang-Baxter equation in which regions between the lines are unlabeled. We discuss this situation in section 11.

To get something as simple as the standard Yang-Baxter equation, we want  $\mathcal{H}$  to be one-dimensional, which will happen if the space of classical solutions of the theory, modulo gauge transformations, is a point. This is also the condition that eliminates the subtleties associated with having a holomorphic action; perturbation theory is straightforward in principle if there is only one classical solution to expand around, and it has only a finite group of automorphisms.<sup>13</sup> The simplest case is that  $C = \mathbb{C}$ . In quantizing the theory on  $\Sigma \times \mathbb{C}$  for any  $\Sigma$ , we require that the gauge field  $A$  and the generator of a gauge transformation both vanish at infinity. With this choice, the only (stable<sup>14</sup>) classical solution, up to a gauge transformation, is  $A = 0$ . So we are in the situation in which the bulk regions do not carry labels and perturbation theory is straightforward in principle.

It is likewise possible when  $C$  is  $\mathbb{C}^\times$  or an elliptic curve to ensure that the classical phase space is a point, leading to straightforward perturbation theory

<sup>13</sup> If there is a unique classical solution up to gauge transformation, but it has a nontrivial automorphism group  $H$ , then in developing perturbation theory one wants to divide by the volume of  $H$ . If  $H$  is not a finite group, this volume might be hard to interpret. However, this issue involves only an overall constant factor in the path integral, independent of what collection of Wilson lines one considers.

<sup>14</sup> There are many classical solutions on  $\Sigma \times \mathbb{C}$  that correspond to bundles on  $\mathbb{C}$  (trivialized at infinity) that are unstable in the sense of algebraic geometry. This likely makes them unsuitable as a starting point for perturbation theory. At any rate, the fact that is really important for us is that the trivial connection on  $\Sigma \times \mathbb{C}$  is a classical solution that has no infinitesimal deformations or gauge automorphisms. In perturbation theory around this solution, we do not meet unstable bundles.

<sup>12</sup> An interpretation of the Yang-Baxter equation and the spectral parameter somewhat along these lines was conjectured by M. F. Atiyah in the 1980’s [14].

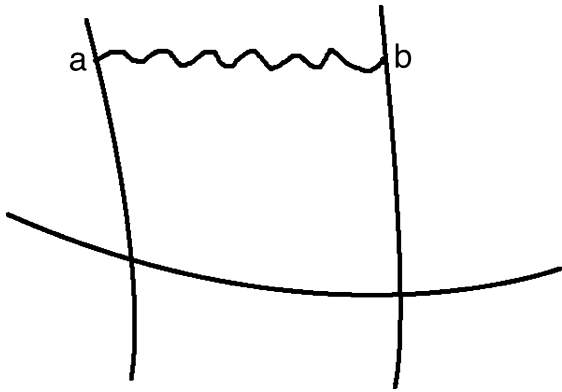


Figure 6. The infrared-free nature of the theory under study means that when we scale up the metric of  $\Sigma$  after gauge-fixing, gluon exchange between Wilson lines that are not crossing becomes irrelevant.

and (as we argue shortly) a conventional Yang-Baxter equation. The details are more involved and we defer a discussion to sections 9 and 10.

Let us now imagine doing perturbation theory in an infrared-free theory in the presence of a configuration of Wilson lines. What sort of perturbative corrections are significant? A typical example of an effect that is not significant is gauge boson exchange between two Wilson lines that are not crossing (Fig. 6). By scaling up the metric of  $\Sigma$ , the points  $a$  and  $b$  in the figure can be made arbitrarily far apart, regardless of where they lie on the Wilson lines in question, and the contribution of gluon exchange between them goes to zero. So this can be ignored.

A typical contribution that cannot be ignored is a gauge boson exchange between two lines that are crossing (Fig. 7). In this case, the points  $a$  and  $b$  can be near the crossing point, so they cannot be assumed to be far away in  $\Sigma$ . We can still scale up the metric in the picture in order to exploit the infrared-free nature of the theory. But all that happens when we do this is that the lines that are crossing turn into straight lines near the point in  $\Sigma$  where they cross, and they become very widely separated from any other crossing points. So the diagram of Fig. 7 may be nontrivial – and it is nontrivial, as we will calculate in section 4 – but it will be local: it will not depend on the details of a larger picture in which the crossing of Fig. 7 might be embedded.

Now we can put the pieces together and explain why something along the lines of standard Yang-Baxter theory will emerge. We assume a situation with a unique classical solution<sup>15</sup>  $A = 0$ . Moreover, we know

<sup>15</sup> In the trigonometric and elliptic cases, this argument needs to be stated a little more carefully. There is a unique classical solution, and although it is not gauge-equivalent globally to  $A = 0$ , this is true locally. Given this, the argument proceeds as in the text.

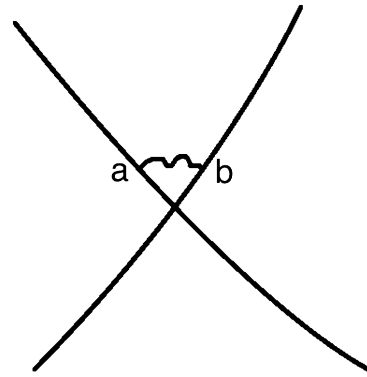


Figure 7. The infrared-free nature of the theory under study means that when we scale up the metric of  $\Sigma$  after gauge-fixing, gluon exchange between Wilson lines that are not crossing becomes irrelevant.

that quantum effects are negligible except near crossings. Thus away from crossings, we can assume that  $A = 0$  everywhere. This means that a line segment between crossings just describes a free particle in the relevant representation of  $\mathfrak{g}$  (or of  $\mathfrak{g}[[z]]$ ) and can be labeled by a basis vector in that representation. And crucially, the amplitude associated to a given crossing can only depend on the local data at that crossing – the representations and labels of the lines that are crossing. Thus the equivalence of the two pictures of Fig. 2, which follows from rather general arguments that were given above, turns into the more precise numerical equivalence of Fig. 3, with a local  $R$ -matrix at each crossing. The same reasoning applies to the unitarity relation of Fig. 4. Because of the infrared-free nature of the theory, it turns into the concrete unitarity relation  $R_{21}R_{12} = 1$  of Yang-Baxter theory.

For the case that  $C$  is  $\mathbb{C}$  or an elliptic curve, the local  $R$ -matrix  $R_{ij}^k(z_1, z_2)$  is actually a function only of the difference  $z_1 - z_2$ , because the classical action (3.3) is invariant under shifting  $z$  by a constant. For the case of  $C = \mathbb{C}^\times$  with differential  $dz/z$ , the equivalent statement is that the  $R$ -matrix (written in these multiplicative coordinates) is a function of the ratio  $z_1/z_2$ .

For the case that  $C = \mathbb{C}$ , related to the Yangian, a few further nice things happen which make this case particularly simple and elegant. First of all, the action is invariant under a common rescaling of  $z$  and  $\hbar$ , so actually the  $R$ -matrix is a function only of a single variable  $(z_1 - z_2)/\hbar$ . Second, in quantizing the theory with  $C = \mathbb{C}$ , we divide only by gauge transformations that are 1 at infinity along  $C$ . But we are left with gauge transformations that are constant at infinity along  $C$ , and these behave as global symmetries. Thus the  $R$ -matrix for  $C = \mathbb{C}$  has  $G$  as an automorphism group. (This is not true for the other choices of  $C$ , as we will see in sections 9 and 10.) The properties stated in this paragraph make it straightforward to



understand some simple examples. We present some of these elementary examples in the next section. We present them because they are fun – though probably well-known to many readers – and also because they enable one to see in a completely direct and elementary way why the theory must have a framing anomaly.

### 3.5 Elementary Examples

For some elementary examples, we take  $G = GL_N$  (or  $SL_N$ , which would be equivalent for the purposes of this analysis), and we will consider the case that  $\rho$  is the fundamental representation of  $G$  or its dual. We denote these representations as  $V$  and  $V^*$ , respectively. In all cases, we will take  $C = \mathbb{C}$ , so that the  $R$ -matrix has  $G$  symmetry.

First we consider the  $R$ -matrix for crossing of two copies of  $V$ . It will be a  $G$ -invariant linear map  $R(z_1 - z_2) : V \otimes V \rightarrow V \otimes V$ . Such an operator is a linear combination of the identity and the operator  $P : V \otimes V \rightarrow V \otimes V$  that exchanges the two factors:  $R(z) = A(z) + PB(z)$ , with  $z = z_1 - z_2$ . In this particular case,  $R(z)$  can also be written fairly conveniently as a matrix with all its indices:

$$(3.22) \quad R_{ij}^{i'j'}(z) = \delta_i^{i'} \delta_j^{j'} A(z) + \delta_j^{i'} \delta_i^{j'} B(z).$$

Here  $i$  and  $j$  refer to “incoming” lines and  $i'$  and  $j'$  to outgoing ones. The  $A(z)$  term describes two lines crossing without “charge exchange,” while  $B(z)$  describes crossing with charge exchange. The non-zero matrix elements of  $R$  are depicted in Fig. 8.

The Yang-Baxter equation is in general invariant under multiplying the  $R$ -matrix by a scalar function – a  $z$ -dependent multiple of the identity. So it is only sensitive to the ratio  $U(z) = B(z)/A(z)$ . It is not diffi-

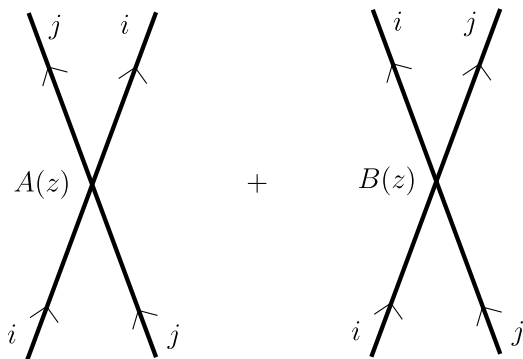


Figure 8. The  $R$ -matrix for crossing of two Wilson lines in the fundamental representation  $V$ . The non-zero matrix elements of  $R$  correspond to processes in which the final states equal the initial states, possibly modulo a permutation.

cult to work out the Yang-Baxter equation in this case (Fig. 9) and to learn that it is equivalent to

$$(3.23) \quad U(z_1 - z_3)U(z_2 - z_3) + U(z_1 - z_2)U(z_1 - z_3) = U(z_1 - z_2)U(z_2 - z_3).$$

After dividing by the product  $U(z_1 - z_2)U(z_1 - z_3)U(z_2 - z_3)$ , we learn that  $1/U(z)$  is a multiple of  $z$ , so (remembering that the  $R$ -matrix for  $C = \mathbb{C}$  is a function of  $\hbar/z$ )  $U$  must be a constant multiple of  $\hbar/z$ . Determining the constant from the  $\mathcal{O}(\hbar)$  contribution to the  $R$ -matrix (see section 4), we find

$$(3.24) \quad U(z) = \frac{\hbar}{z}.$$

Now we consider the  $R$ -matrix for crossing of a copy of  $V$  with a copy of the dual representation  $V^*$ . The  $R$ -matrix is now a linear map  $R(z) : V \otimes V^* \rightarrow V \otimes V^*$ . Again  $R(z)$  is determined by the  $G$  symmetry in terms of two functions:  $R(z) = C(z) + QD(z)$ . Here  $Q$  is the  $G$ -invariant projection operator from  $V \otimes V^*$  to its  $G$ -invariant subspace. A picture is rather clear (Fig. 10), but now a formula analogous to eqn. (3.22) is less transparent:

$$(3.25) \quad R_{ij}^{i'j'}(z) = \delta_i^{i'} \delta_j^{j'} C(z) + \delta_i^{j'} \delta_j^{i'} D(z).$$

(As before, lower indices refer to incoming lines and upper indices to outgoing ones.) The Yang-Baxter equation will involve only the ratio  $W(z) = D(z)/C(z)$ .

It is again not difficult to write down the Yang-Baxter equation. We learn (Fig. 11) that

$$(3.26) \quad U(z_1 - z_2)W(z_2 - z_3) + W(z_1 - z_3)W(z_2 - z_3) = U(z_1 - z_2)W(z_1 - z_3).$$

This is equivalent to  $1/W(z_2 - z_3) - 1/W(z_1 - z_3) = 1/U(z_1 - z_2)$ , with the general solution

$$(3.27) \quad W(z) = -\frac{\hbar}{z - \hbar b},$$

with a constant  $b$ . This constant  $b$  can actually be determined by the unitarity relation  $R_{21}(-z)R_{12}(z) = 1$ . A specific matrix element of this relation, for the case of crossing of  $V$  and  $V^*$  (see Fig. 12) gives

$$(3.28) \quad NW(z)W(-z) + W(z) + W(-z) = 0,$$

leading to

$$(3.29) \quad W(z) = -\frac{\hbar}{z + \frac{\hbar N}{2}}.$$

We need not separately consider the  $R$ -matrix for crossing of two copies of  $V^*$ , because it simply equals the  $R$ -matrix for crossing of two copies of  $V$ . The reason is that the outer automorphism of  $GL_N$  or  $SL_N$  that exchanges  $V$  and  $V^*$  is a symmetry of the action (3.3), and of the theory derived from it.

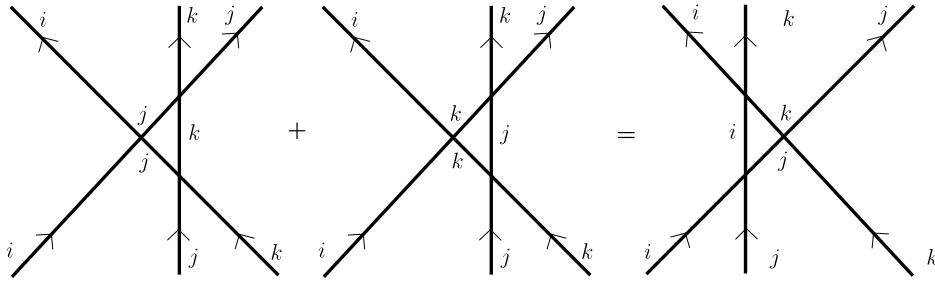


Figure 9. The Yang-Baxter equation for crossing of three Wilson lines in the fundamental representation  $V$ .

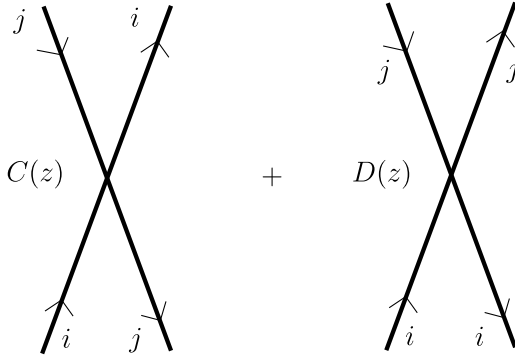


Figure 10. The R-matrix for crossing of two Wilson lines in the dual representations  $V$  and  $V^*$ .

Not determined by these arguments are overall scalar functions in the R-matrices for  $V \otimes V \rightarrow V \otimes V$  and  $V \otimes V^* \rightarrow V \otimes V^*$ . These can be partly but not entirely determined by the unitarity relation; to some extent these overall scalar functions depend on arbitrary choices in quantizing the theory.

In the theory as we have developed it so far, the angles at which two Wilson lines cross are of no consequence. We can use the above formulas to test this expectation in an interesting way. We consider (see the left of Fig. 13) two Wilson lines in the representation  $V$ , with spectral parameters  $z_1$  and  $z_2$ , and differing from the vertical by small angles  $\pm\alpha$ . Rotating one line clockwise by a small angle  $2\alpha$  and the other one clockwise by a larger angle  $\pi - 2\alpha$ , we arrive at the right hand side of the figure, which depicts the crossing of a pair of nearly vertical Wilson lines in the representations  $V$  and  $V^*$ . The rotation converts a charge exchange for  $VV$  crossing to an “annihilation” process for  $VV^*$ . Naively, the two parts of Fig. 13 should be equivalent. This would imply

$$(3.30) \quad U(z) \stackrel{?}{=} W(-z).$$

A look back at our previous formulas shows, however, that this is false. What is true instead is that

$$(3.31) \quad U\left(z - \frac{\hbar N}{2}\right) = W(-z).$$

### 3.6 First Look at the Framing Anomaly

What accounts for this discrepancy? The answer is that the theory has a framing anomaly for Wilson operators, which will be explored from another point of view in section 6. The framing anomaly is analogous to the perhaps familiar framing anomaly for Wilson operators in Chern-Simons theory, but more subtle. It can be formulated in different but topologically equivalent ways. However, in an approach natural in perturbation theory, the framing anomaly can be formulated as follows. To begin with we take the topological plane to really be a plane  $\Sigma = \mathbb{R}^2$ , and we quantize with a gauge choice that uses a flat metric on  $\mathbb{R}^2$ . (See section 4.) We consider a Wilson operator supported on a general curve  $K \subset \Sigma$ , and we let  $\varphi(p)$  be the angle between the tangent vector to  $K$  at a given point  $p \in K$  and some chosen direction in  $\mathbb{R}^2$  (e.g. the vertical). (We define this angle to increase if  $K$  bends in a clockwise direction.) Thus  $\varphi$  is not quite well-defined as a function on  $K$ , but it is well-defined up to an additive constant, and its differential  $d\varphi$  is well-defined.

The framing anomaly means that what is constant along  $K$  is not the spectral parameter  $z$ , as one would expect from a classical analysis, but  $z - \hbar h^\vee \varphi / (2\pi)$ , where  $h^\vee$  is the dual Coxeter number of the gauge group. As a perhaps surprising example of the implications of this statement, a Wilson operator whose support is a simple closed curve in  $\Sigma$  is anomalous and does not exist in the quantum theory, because in going around a simple closed loop,  $\varphi$  increases by  $2\pi$ .

Now in our problem, we can formulate the comparison between the two parts of Fig. 13 in a slightly different way. To go from the left to the right of the figure, we make the  $z_1$  Wilson line bend in the plane by an angle  $\Delta\varphi = \pi$  before crossing the  $z_2$  Wilson line, as in Fig. 14. But because of the framing anomaly, when we do this  $z_1$  is shifted to  $z_1 - \hbar h^\vee \Delta\varphi / (2\pi) = z_1 - \hbar N / 2$ , where  $N$  is the dual Coxeter number of  $SL_N$  (or  $GL_N$ ). Thus what should coincide with  $W(-z)$  is not  $U(z)$  but  $U(z - \hbar N / 2)$ , and this is precisely what we found in eqn. (3.31).

In light of the framing anomaly, one might ask the following question. In the usual formulation of the

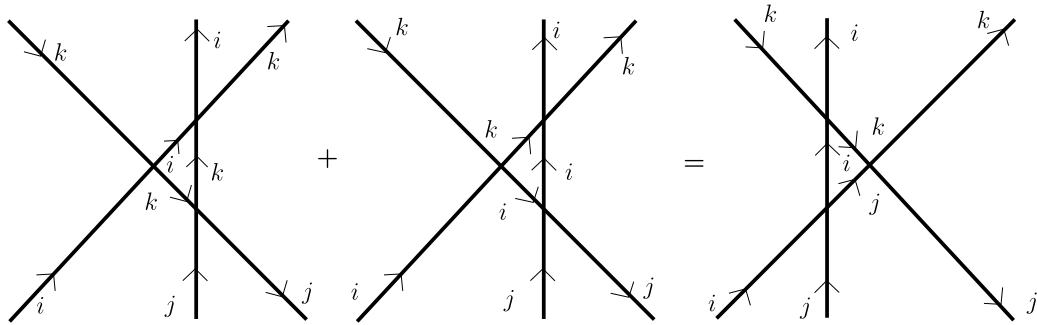


Figure 11. The Yang-Baxter equation for three lines labeled by representations  $V$ ,  $V$ , and  $V^*$ .

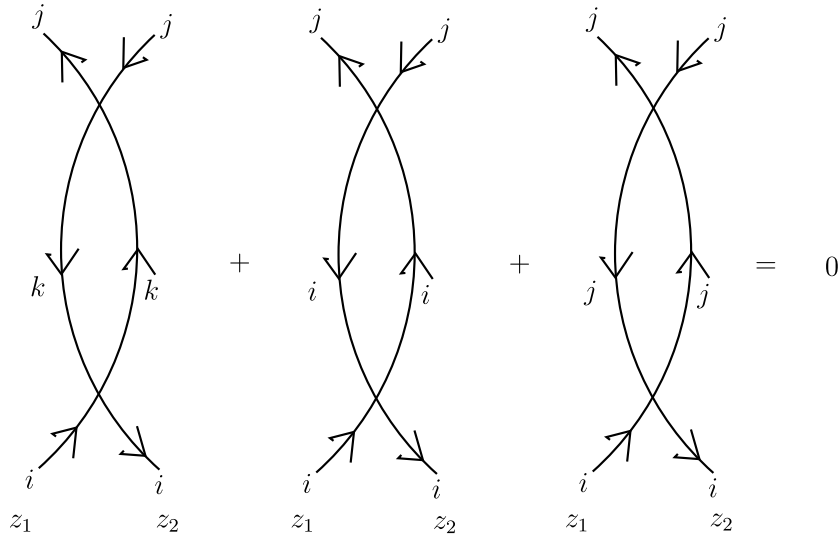


Figure 12. The “unitarity” relation for lines labeled by  $V$  and  $V^*$  implies that the given sum vanishes for  $i \neq j$ .

Yang-Baxter equation for crossing of three lines, what are the angles at which the lines cross? The answer is clear if one considers the case that the three representations involved are all the same. It is usually then assumed that the three  $R$ -matrices  $R_{12}$ ,  $R_{23}$ , and  $R_{13}$  are given by the same matrix-valued function of  $z$ . For this to be true, the relative angles must be the same at all three crossings. Since the actual relation is that the 13 crossing angle is the sum of the 12 and 23 crossing angles, the three angles are equal only in the limit that they all go to zero. So the usual formulation of the Yang-Baxter equation refers to the case of nearly parallel lines in the limit that the crossing angles vanish. That is actually why Fig. 13 has been drawn with lines at small angles  $\pm\alpha$  to the vertical, and it was implicitly assumed in our discussion of Fig. 14.

The reader might be slightly perplexed that we began our explanation of the framing anomaly by restricting to the special case  $\Sigma = \mathbb{R}^2$ . To understand this point properly, one has to analyze the framing anomaly for four-manifolds as well as the framing anomaly for Wilson operators. This is somewhat beyond the scope of the present paper. However, the

upshot is that to avoid an anomaly, the two-manifold  $\Sigma$  must be “framed,” meaning that its tangent bundle must be trivialized. (This is actually analogous to the framing anomaly of three-dimensional Chern-Simons theory, which is defined on a framed three-manifold.) On a framed two-manifold  $\Sigma$ , one can define along any embedded oriented one-manifold  $K \subset \Sigma$  a function with the properties of the  $\varphi$  used above. For this, recall that a “framing” is a pair of everywhere linearly independent vector fields  $v_1$  and  $v_2$  on  $\Sigma$ . Given a framing, one can pick a metric such that  $v_1$  and  $v_2$  are everywhere orthonormal, and then one can define  $\varphi$  as the angle between the tangent vector to  $K$  and the  $v_1$  direction. Note that the condition that  $\Sigma$  should be framed is very restrictive; for example, a compact framed two-manifold must have zero Euler characteristic and hence must have genus 1. Thus – as we also saw from the anomaly for a closed loop in the plane – the framing anomaly in the four-dimensional theory discussed here is much more restrictive than its three-dimensional cousin.

If we think of the vertical direction in the above figures as the Euclidean “time,” then the reader will

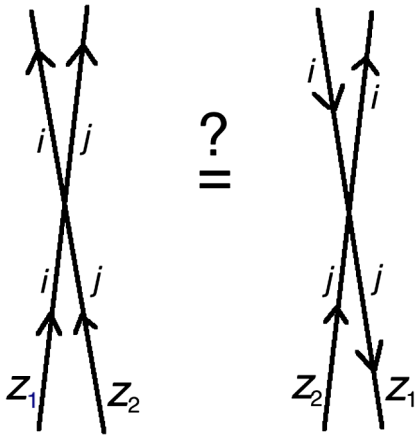


Figure 13. On the left are two Wilson lines in the fundamental representation of  $GL_N$  or  $SL_N$ ; they differ from the vertical by small angles  $\pm\alpha$ . The one on the left has spectral parameter  $z_1$  and the one on the right has spectral parameter  $z_2$ . To get the picture on the right, the  $z_2$  Wilson line has been rotated clockwise by the small angle  $2\alpha$ , and the  $z_1$  Wilson line has been rotated clockwise by the much larger angle  $\pi - 2\alpha$ . Naively the two pictures should be equivalent, implying that  $U(z_1 - z_2) = W(z_2 - z_1)$ , a claim that turns out to be false.

note that we have interpreted a  $V^*$ -valued particle moving forward in time as a  $V$ -valued particle moving backward in time. This is reminiscent of the relation between particles and antiparticles in relativistic quantum field theory, and indeed in the application of  $R$ -matrix theory to integrable models of relativistic quantum field theory, this operation becomes crossing symmetry whereby an  $S$ -matrix element with a particle in the initial state, after analytic continuation to negative energy, is interpreted as an  $S$ -matrix element with an antiparticle in the final state. That is why in  $R$ -matrix theory, the relation between  $R$ -matrix elements associated to a pair of dual representations  $\rho$  and  $\rho^*$  is often called “crossing.”

#### 4. $R$ -Matrix from Crossing Wilson Lines

Here and in the next two sections, we will perform concrete Feynman diagram calculations to compute (1) the  $\mathcal{O}(\hbar)$  term in the  $R$ -matrix; (2) the  $\mathcal{O}(\hbar)$  quantum correction to the operator product expansion (OPE) for Wilson line operators; (3) the framing anomaly. In fact, in the case of the framing anomaly, the  $\mathcal{O}(\hbar)$  term that we compute gives the complete answer; in the other cases, there are higher order contributions to the effects that we calculate, although they can be determined by general principles (such as the Yang-Baxter equation and associativity of the

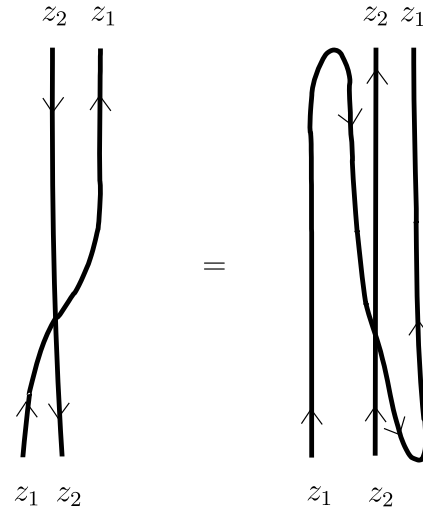


Figure 14. An improved view of the unitarity relation that was explored in Fig. 13, formulated to take the framing anomaly into account. Two lines both labeled by  $V$  enter at the bottom and emerge at the top at nearly vertical angles. If they cross at nearly vertical angles, we get an  $R$ -matrix element for a  $VV$  crossing, while if one of them turns by an angle very close to  $\pi$  before the crossing, then we get a  $VV^*$  crossing, but with  $z_1$  replaced by  $z_1 - N\hbar/2$  because of the framing anomaly. Thus the naive relation  $U(z) = W(-z)$  suggested by Fig. 13 is corrected to  $U(z - N\hbar/2) = W(-z)$ .

operator product expansion) once the lowest order terms are known. We will perform independent Feynman diagram calculations of the three effects, but actually the three effects can be deduced from each other to a large extent. We already deduced the framing anomaly from the  $\mathcal{O}(\hbar)$  term in the  $R$ -matrix in section 3.6, and we will explain in section 5 why the quantum correction to the OPE is inevitable given the quantum correction to the  $R$ -matrix.

In all cases, we take  $C = \mathbb{C}$ , corresponding to a rational solution of the Yang-Baxter equation. In the case of the framing anomaly, the considerations are manifestly local, so the choice of  $C$  does not matter. For the  $R$ -matrix and the OPE, once the result is known for  $\mathbb{C}$ , it can be deduced from global considerations for the other choices of  $C$ . We leave this for sections 9 and 10.

We will compute in a way that involves a choice of metric on  $\Sigma \times C$ . As explained in section 3.4, for the output of Feynman diagrams to have a straightforward interpretation in the usual language of  $R$ -matrix theory, we have to scale up the metric on  $\Sigma$  by a large factor. In the limit,  $\Sigma$  becomes  $\mathbb{R}^2$  near the crossing and the supports of the two Wilson lines that are crossing become straight lines in  $\mathbb{R}^2$ . In  $\mathcal{O}(\hbar)$ , the angle at which the lines cross does not matter (the reader



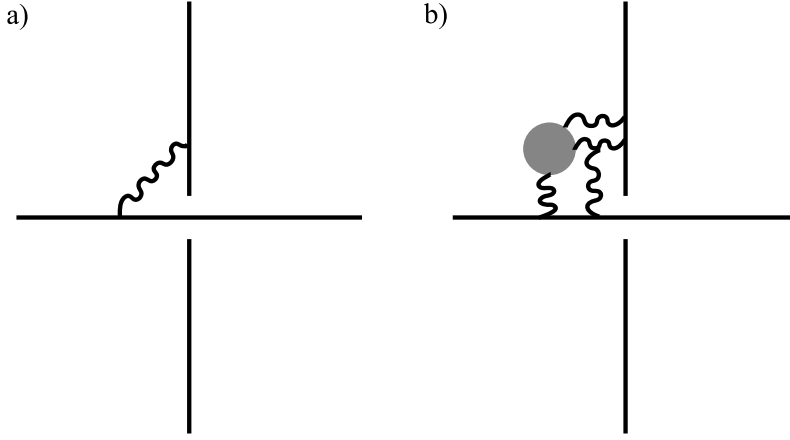


Figure 15. (a) The leading order Feynman diagram for gluon exchange between crossing Wilson lines. (b) A higher order diagram, with further complications hidden in the “blob.”

can verify this by a slight generalization of the calculation that we will describe), so we can take the two lines to be the  $x$ -axis and the  $y$ -axis in the  $xy$  plane. In higher orders, the crossing angle would matter via the framing anomaly.

The  $\mathcal{O}(\hbar)$  contribution to the  $R$ -matrix involves one gluon exchange between the two lines, as sketched in Fig. 15. We evaluate the contribution of this diagram for the case that the Wilson lines are associated to representations  $\rho$  and  $\rho'$  of  $\mathfrak{g}$  supported respectively at  $z = z_1$  and  $z = z_2$ .

The metric that we will use on  $\mathbb{R}^2 \times C$  is  $ds^2 = dx^2 + dy^2 + dzd\bar{z} = g_{\mu\nu}dx^\mu dx^\nu$ , or in another language

$$(4.1) \quad g_{xx} = g_{yy} = 1, \quad g_{z\bar{z}} = g_{\bar{z}z} = \frac{1}{2}, \quad \text{other components zero.}$$

The corresponding inverse metric is

$$(4.2) \quad g^{xx} = g^{yy} = 1, \quad g^{z\bar{z}} = g^{\bar{z}z} = 2, \quad \text{other components zero.}$$

For a gauge-fixing condition, we pick

$$(4.3) \quad 0 = \frac{\partial}{\partial x} A_x + \frac{\partial}{\partial y} A_y + 4 \frac{\partial}{\partial z} A_{\bar{z}}.$$

This is the closest analog of the usual Lorentz gauge for this theory with a partial gauge connection. The factor of four is explained by noting that if the gauge field  $A$  satisfies this equation and also the linearized equations of motion  $dz \wedge dA = 0$ , then each component of  $A$  is harmonic for the metric we have chosen.

In this gauge, the four-dimensional propagator for the gauge field is then given by

$$(4.4) \quad \begin{aligned} & \langle A_x^a(x, y, z, \bar{z}) A_y^b(x', y', z', \bar{z}') \rangle \\ &= -\delta^{ab} \frac{4}{4\pi} \frac{\partial}{\partial z} \frac{1}{(x-x')^2 + (y-y')^2 + |z-z'|^2} \\ &= \delta^{ab} \frac{1}{2\pi} \frac{2(\bar{z}-\bar{z}')}{((x-x')^2 + (y-y')^2 + |z-z'|^2)^2}, \\ & \langle A_{\bar{z}}^a(x, y, z, \bar{z}) A_x^b(x', y', z', \bar{z}') \rangle \\ &= -\delta^{ab} \frac{1}{4\pi} \frac{\partial}{\partial y} \frac{1}{(x-x')^2 + (y-y')^2 + |z-z'|^2} \\ &= \delta^{ab} \frac{1}{2\pi} \frac{y-y'}{((x-x')^2 + (y-y')^2 + |z-z'|^2)^2}, \\ & \langle A_y^a(x, y, z, \bar{z}) A_{\bar{z}}^b(x', y', z', \bar{z}') \rangle \\ &= -\delta^{ab} \frac{1}{4\pi} \frac{\partial}{\partial x} \frac{1}{(x-x')^2 + (y-y')^2 + |z-z'|^2} \\ &= \delta^{ab} \frac{1}{2\pi} \frac{x-x'}{((x-x')^2 + (y-y')^2 + |z-z'|^2)^2}. \end{aligned}$$

For later purposes, we can reinterpret the propagator as a two-form on two copies of  $\mathbb{R}^2 \times C$ . Setting  $x = x' - x''$ ,  $y = y' - y''$ ,  $z = z' - z''$ ,  $\bar{z} = \bar{z}' - \bar{z}''$ , we define

$$(4.5) \quad \begin{aligned} P^{ab}(x, y, z, \bar{z}) &:= \frac{1}{2} \sum_{i,j=z,y,\bar{z}} \langle A_i^a(x', y', z', \bar{z}') A_j^b(x'', y'', z'', \bar{z}'') \rangle dx^i \wedge dx^j \\ &= -\frac{\delta^{ab}}{4\pi} \left( dy \wedge d\bar{z} \frac{\partial}{\partial x} + d\bar{z} \wedge dx \frac{\partial}{\partial y} + 4dx \wedge dy \frac{\partial}{\partial z} \right) \\ &\quad \times \frac{1}{(x^2 + y^2 + z\bar{z})} \\ &= \frac{\delta^{ab}}{2\pi} (xdy \wedge d\bar{z} + yd\bar{z} \wedge dx + 2\bar{z}dx \wedge dy) \frac{1}{(x^2 + y^2 + z\bar{z})^2}. \end{aligned}$$

In the following, we often use the propagator two-form with adjoint indices stripped off:

$$(4.6) \quad P^{ab}(x, y, z, \bar{z}) = \delta^{ab} P(x, y, z, \bar{z}).$$

The defining equations of the 2-form  $P(x, y, z, \bar{z})$  are

$$(4.7) \quad \frac{i}{2\pi} dz \wedge dP(x, y, z, \bar{z}) = \delta_{x,y,z,\bar{z}=0},$$

$$(4.8) \quad (\partial_x t_{\partial_x} + \partial_y t_{\partial_y} + 4\partial_z t_{\partial_z})P(x, y, z, \bar{z}) = 0,$$

where  $\delta_{x,y,z,\bar{z}=0}$  is a delta-function distribution localized at  $x = y = z = \bar{z} = 0$ , and  $t_V$  indicates contraction with a vector field  $V$ .

To verify the normalizations, we can check these equations explicitly. It is easy to verify that  $dz \wedge dP = 0$  away from the origin  $x = y = z = \bar{z} = 0$ , and that eqn. (4.8) is satisfied.

Note that, when restricted to the unit three-sphere,

$$(4.9) \quad P(x, y, z, \bar{z}) = \frac{1}{2\pi} (x dy \wedge d\bar{z} - y dx \wedge d\bar{z} + 2\bar{z} dx \wedge dy).$$

Since the forms on the two sides of this equation are the same on the unit three-sphere, Stokes' theorem tell us that applying the operator  $dz \wedge d$  to both sides and integrating over the ball of radius 1 will give the same answer. Therefore we have the identity (using the coordinates  $u, v$  with  $z = u + iv$ )

$$(4.10) \quad \begin{aligned} & \frac{i}{2\pi} \int_{x^2+y^2+z\bar{z} \leq 1} dz \wedge dP(x, y, z, \bar{z}) \\ &= \frac{i}{2\pi} \frac{1}{2\pi} \int_{x^2+y^2+z\bar{z} \leq 1} 4 dx dy dz d\bar{z} \\ &= \frac{i}{2\pi} \frac{1}{2\pi} \int_{x^2+y^2+z\bar{z} \leq 1} (-8i) dx dy du dv \\ &= \frac{i}{2\pi} \frac{1}{2\pi} (-8i) \frac{\pi^2}{2} = 1. \end{aligned}$$

The four-dimensional bulk gauge field  $A^a$  couples to a Wilson line in the representation  $\rho$  by a factor  $t_{a,\rho}$ , which is the matrix by which the Lie algebra element  $t_a$  acts in the representation  $\rho$ . For the case of a single gluon coupling to a Wilson line, this factor does not depend on where on the line the gluon is inserted. For gluon exchange between two Wilson lines, as in Fig. 15, we simply get such a factor on each line. The propagator between a Wilson line supported on the  $x$  axis at  $z = z_1$  and one supported on the  $y$  axis at  $z = z_2$  then evaluates to

$$(4.11) \quad \begin{aligned} I_1 &= \hbar((t_{a,\rho} \otimes t_{b,\rho'}) \int dx dy' P^{ab}(x-x', y-y', z_1-z_2, \bar{z}_1-\bar{z}_2) \\ &= \hbar c_{\rho,\rho'} \int dx dy' P(x-x', y-y', z_1-z_2, \bar{z}_1-\bar{z}_2), \end{aligned}$$

where the color factor reads

$$(4.12) \quad c_{\rho,\rho'} = \sum_a t_{a,\rho} \otimes t_{a,\rho'}.$$

Here  $c_{\rho,\rho'}$  can be viewed as the image of an element  $c = \sum_a t_a \otimes t_a$  of  $\mathfrak{g} \otimes \mathfrak{g}$  in the representation  $\rho \otimes \rho'$  of  $\mathfrak{g} \otimes \mathfrak{g}$ . The factor of  $\hbar$  is the loop counting parameter.

It is straightforward to evaluate the integral in (4.11), with the result

$$(4.13) \quad I_1 = \hbar c_{\rho,\rho'} \frac{1}{2\pi} \int dx dy \frac{2(\bar{z}_1 - \bar{z}_2)}{(x^2 + y^2 + |z_1 - z_2|^2)^2} = \frac{\hbar c_{\rho,\rho'}}{z_1 - z_2}.$$

This reproduces the standard semi-classical expansion of the rational  $R$ -matrix

$$(4.14) \quad R = I + \hbar r + \mathcal{O}(\hbar^2) = I + \frac{\hbar c_{\rho,\rho'}}{z_1 - z_2} + \mathcal{O}(\hbar^2).$$

General theorems (see [7, p. 814] or [8, p. 418]) imply that the full rational  $R$ -matrix is determined up to a prefactor by the general conditions that it obeys together with the leading order term that we have just computed. Some interesting special cases of this statement are rather easy, as we have reviewed in section 3.5.

## 5. OPE of Parallel Wilson Lines

### 5.1 Overview

The next topic that we will consider is the operator product expansion (OPE) of Wilson lines. We begin with generalities about line operators in diffeomorphism-invariant theories.

Consider parallel line operators  $K$  and  $K'$  in a theory with diffeomorphism invariance in any dimension  $\geq 2$  (Fig. 16(a)). Diffeomorphism invariance means that there is no natural notion of whether  $K$  and  $K'$  are “near” or “far” and therefore that we can think of them as being arbitrarily near. This implies that it must be possible in any diffeomorphism invariant theory to interpret a product  $KK'$  of two parallel line operators as a single line operator  $K''$ . This is the operator product expansion for line operators.

Although there is in a diffeomorphism invariant theory no natural notion of  $K$  and  $K'$  being “near” or “far,” something special happens in two dimensions: there can be a natural notion of whether  $K$  is to the left or right of  $K'$ . The product  $KK'$  with  $K$  to the left of  $K'$  may be different from the product  $K'K$  with  $K$  to the right. Thus, in two dimensions, the OPE of line operators is not necessarily commutative.<sup>16</sup>

Although not necessarily commutative in two dimensions, the OPE of line operators is always associative. That is because (Fig. 16(b)) given three parallel line operators  $K, K',$  and  $K''$ , there is no natural notion of  $K$  and  $K'$  being closer or farther than  $K'$  and  $K''$ . There is just one product of line operators that depends only on how they are arranged from left to right.

<sup>16</sup> In three dimensions, there is a more subtle analog of this:  $KK'$  and  $K'K$  are always isomorphic, but there can be differ-

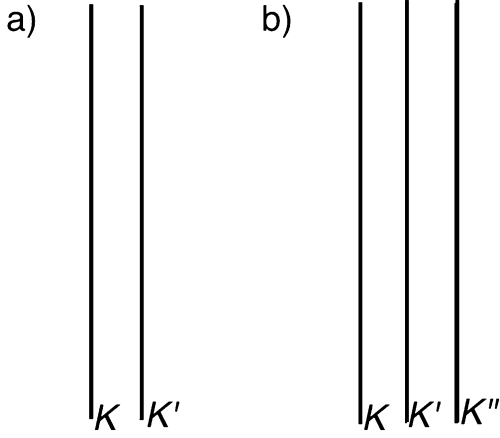


Figure 16. (a) In a diffeomorphism-invariant theory, two parallel line operators can be considered to be arbitrarily close, so they behave as a single line operator. This is the operator product expansion for line operators. (b) In two dimensions, this operator product expansion is not necessarily commutative, but it is always associative, because given three parallel line operators, there is no natural sense of one pair being “closer” than the other. The triple product  $KK'K''$  can be identified as either  $(KK')K''$  or  $K(K'K'')$  by moving together one pair of adjacent line operators or the other.

Our problem has a few special features. The line operators that we will be studying are indeed supported on a line  $K$  in the smooth two-manifold  $\Sigma$  which has diffeomorphism symmetry,<sup>17</sup> but they are also supported at a point in the complex Riemann surface  $C$ . The OPE for line operators supported at distinct points in  $C$  is trivial – and in particular commutative – as they can pass through each other in  $\Sigma$  without any singularity. The interesting case is the OPE for line operators that are supported at the same point in  $C$ . We may as well take this point to be  $z = 0$ .

Although abstractly the product  $KK'$  of line operators  $K$  and  $K'$  will always be another line operator, if we try to consider too small a class of line operators, we might find that the product  $KK'$  is not in the class that we started with. That is actually what happens in the theory described in the present paper if we consider only the most obvious class of line operators: Wilson line operators associated to representations of the finite-dimensional Lie algebra  $\mathfrak{g}$ . We will see in section 5.2 that this class of line operators is not closed under operator products. To get an

ent isomorphisms between them, depending on the direction in which  $K$  and  $K'$  are moved around each other, leading to a notion of braiding of line operators.

<sup>17</sup> This diffeomorphism symmetry is mildly broken by a framing anomaly, but not in a way that affects the present discussion.

OPE for Wilson line operators, we have to consider the more general class of Wilson operators associated to representations of  $\mathfrak{g}[[z]]$ .

Given two parallel Wilson line operators, on general grounds their product in a diffeomorphism-invariant theory is another line operator. But it is non-trivial to exhibit this new line operator as another Wilson operator for some representation of  $\mathfrak{g}[[z]]$ . To exhibit this, we have to do a calculation, starting with two parallel line operators, taking the limit, in a concrete quantization scheme, as they approach each other, and searching for a single Wilson operator that will reproduce the effects of the two Wilson operators that we started with.

At the classical level, the OPE for Wilson operators is trivial. Given Wilson operators associated to representations  $\rho$  and  $\rho'$  of  $\mathfrak{g}[[z]]$ , their product is the Wilson line associated to the tensor product representation  $\rho \otimes \rho'$ . There is a quantum correction to this and we will compute it to lowest order in  $\hbar$ .

We will perform the computation assuming that  $\rho$  and  $\rho'$  are representations of the finite-dimensional algebra  $\mathfrak{g}$ , or equivalently that they are representations of  $\mathfrak{g}[[z]]$  in which the generators  $t_{a,n}$  vanish for  $n > 0$ . What we will show is that in order  $\hbar$ , the tensor product representation acquires a nonzero  $t_{a,1}$  (there is no correction to  $t_{a,n}$  for any  $n \neq 1$ ). It is fundamentally because of this fact that it is important to consider Wilson operators associated to representations of  $\mathfrak{g}[[z]]$  that do not come from representations of  $\mathfrak{g}$ .

## 5.2 Lowest Order Computation

We start with a Wilson line in the representation  $\rho'$  of  $\mathfrak{g}$  at  $y = 0$ , and one in the representation  $\rho''$  at  $y = \epsilon$ . As just explained, we assume that these are “ordinary” Wilson lines, associated to representations of  $\mathfrak{g}$ .

The leading order Feynman diagram is given in Fig. 17. This diagram represents the coupling of an external gauge field to the two Wilson lines. We will calculate what happens, to leading order in  $\hbar$ , when we put these lines beside each other and send  $\epsilon \rightarrow 0$  from above. Modulo  $\hbar$ , the result will simply be the Wilson line associated to the tensor product representation  $\rho = \rho' \otimes \rho''$  of  $\mathfrak{g}$ . There is an order  $\hbar$  correction, in which the  $z$ -derivative of the gauge field  $A$  is coupled to the Wilson line in a non-trivial way.

For the evaluation of this diagram we need an extra Feynman rule not needed so far, namely the bulk interaction vertex, away from any Wilson lines. This can be read off from the action (3.3), and is

$$(5.1) \quad \frac{i}{2\pi} \int^{abc} dz.$$

Note that this vertex contains a 1-form  $dz$ , since this is present in the action. To evaluate the diagram of

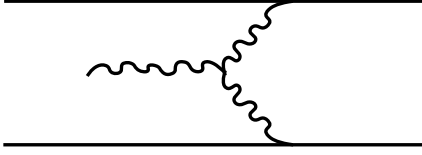


Figure 17. The Feynman diagram representing the  $\mathcal{O}(\hbar)$  correction to the OPE of Wilson lines. The two lines are at  $y = 0$  and  $y = \epsilon$ , respectively. We consider the limit  $\epsilon \rightarrow 0$ .

Fig. 17, we have to integrate one interaction vertex over the line  $y = 0$ , one over the line  $y = \epsilon$ , and one over  $\mathbb{R}^4 = \mathbb{R}^2 \times \mathbb{C}$ . Connecting the vertices by propagators, we have to integrate

$$(5.2) \quad I_2 = \frac{i}{2\pi} \hbar (t^a \otimes t^b f_{abc}) \int_{x_1, x_2 = -\infty}^{\infty} \int_{x, y, z, \bar{z}} P(x_1 - x, y, z) \wedge dz \wedge A^c(x, y, z) \wedge P(x_2 - x, y - \epsilon, z),$$

where  $P$  is the four-dimensional propagator two-form introduced previously in (4.5). Since we have two propagators and one bulk vertex, we have a factor of  $\hbar^{2-1} = \hbar$  in front, as in the discussion of the  $R$ -matrix. (The meaning of this is that the diagram is of order  $\hbar$  compared to a contribution in which the gauge boson couples directly to one of the Wilson operators, with the bulk interaction playing no role.)

Integrating first over  $x_1$  and  $x_2$ , we obtain

$$(5.3) \quad I_2 = \frac{i}{2\pi} \hbar (t^a \otimes t^b f_{abc}) \times \int_{x, y, z, \bar{z}} dz \wedge A^c(x, y, z, \bar{z}) \wedge P'(y, z, \bar{z}) \wedge P'(y - \epsilon, z, \bar{z}),$$

where we defined a three-dimensional one-form propagator with color indices stripped off on the plane  $\mathbb{R}^3$  parametrized by  $(y, z, \bar{z})$ :

$$(5.4) \quad P'(y, z, \bar{z}) := \int dx P(x, y, z, \bar{z}) = -\frac{1}{4} \left( -d\bar{z} \frac{\partial}{\partial y} + 4dy \frac{\partial}{\partial z} \right) \frac{1}{(y^2 + z\bar{z})^{\frac{1}{2}}}.$$

Note that we obtained a numerical factor of  $\pi$  from the  $x$  integral.

The most fundamental thing to explain about eqn. (5.3) is why, for  $\epsilon \rightarrow 0$ , it produces a local coupling of  $A$  to a line operator at  $y = z = 0$ . The reason for this is simply that for  $\epsilon \rightarrow 0$ , the integrand in eqn. (5.3) vanishes as long as  $y$  and  $z$  are not both 0. That is true simply because the integrand for  $\epsilon = 0$  is proportional to  $P'(y, z, \bar{z}) \wedge P'(y, z, \bar{z})$ , and this trivially vanishes because  $P'$  is a 1-form.

Accordingly, the small  $\epsilon$  limit of  $dz \wedge P'(y, z, \bar{z}) \wedge P'(y - \epsilon, z, \bar{z})$  is a distribution supported at  $y = z = 0$ .

By dimensional analysis and rotation symmetry in the  $z$  plane, this distribution must be a multiple of  $\partial_z \delta^3(y, z, \bar{z})$ , where  $\delta^3(y, z, \bar{z})$  is a three-form delta function satisfying  $\int_{\mathbb{R} \times \mathbb{C}} \delta^3(y, z, \bar{z}) = 1$ . We will now show that

$$(5.5) \quad \lim_{\epsilon \rightarrow 0} dz \wedge P'(y, z, \bar{z}) \wedge P'(y - \epsilon, z, \bar{z}) = \frac{\pi}{i} \frac{\partial}{\partial z} \delta^3(y, z, \bar{z}).$$

This will imply that

$$(5.6) \quad I_2 = \frac{i}{2\pi} \hbar (t^a \otimes t^b f_{abc}) \int_{x, y, \bar{z}} A^c(x, y, z, \bar{z}) \left( \frac{\pi}{i} \frac{\partial}{\partial z} \delta^3(y, z, \bar{z}) \right) = -\hbar \frac{1}{2} (t^a \otimes t^b f_{abc}) \int dx \frac{\partial}{\partial z} A^c(x, y, z, \bar{z}).$$

The coupling to  $\partial_z A$  shows that the composite Wilson line operator obtained by bringing two such operators together has  $t_{a,1} \neq 0$ , that is it is associated to a representation of  $\mathfrak{g}[[z]]$  but not to a representation of  $\mathfrak{g}$ . Moreover, it is easy to see from this formula that although we will take  $\epsilon > 0$  in our calculation, the result is actually proportional to the sign of  $\epsilon$ . Changing the sign of  $\epsilon$  would lead, after replacing  $y$  with  $y + \epsilon$ , to the same calculation but with the roles of the two line operators exchanged, replacing  $t^a \otimes t^b$  in eqn. (5.6) with  $t^b \otimes t^a$ . Since this changes the sign of  $f_{abc} t^a \otimes t^b$ , eqn. (5.6) implies that the quantum OPE for Wilson operators is noncommutative: it gives a result that depends on which of the two Wilson operators is on the “left” of the other.

To show (5.5), let us express the three-dimensional one-form propagator (5.4) as

$$(5.7) \quad P'(y, z, \bar{z}) = -d\bar{z} \partial_y Q(y, z, \bar{z}) + 4dy \partial_z Q(y, z, \bar{z})$$

with

$$(5.8) \quad Q(y, z, \bar{z}) := -\frac{1}{4} \frac{1}{(y^2 + z\bar{z})^{\frac{1}{2}}}.$$

Then the left hand side of (5.5) reads (we here drop  $z, \bar{z}$  from the arguments of  $Q$ , to simplify the expressions)

$$(5.9) \quad -4dy \wedge dz \wedge d\bar{z} (\partial_y Q(y) \partial_z Q(y - \epsilon) - \partial_z Q(y) \partial_y Q(y - \epsilon)) = dy \wedge dz \wedge d\bar{z} \left[ \frac{\partial}{\partial z} (a) + \frac{\partial}{\partial y} (b) \right],$$

with

$$(5.10) \quad \begin{aligned} (a) &:= -2\partial_y Q(y) Q(y - \epsilon) + 2Q(y) \partial_y Q(y - \epsilon) \\ &= \frac{2}{4^2} \frac{\epsilon(y(\epsilon - y) + |z|^2)}{(y^2 + |z|^2)^{\frac{3}{2}} ((y - \epsilon)^2 + |z|^2)^{\frac{3}{2}}}, \\ (b) &:= -2Q(y) \partial_z Q(y - \epsilon) + 2\partial_z Q(y) Q(y - \epsilon) \\ &= \frac{1}{4^2} \frac{-\epsilon \bar{z} (\epsilon - 2y)}{(y^2 + |z|^2)^{\frac{3}{2}} ((y - \epsilon)^2 + |z|^2)^{\frac{3}{2}}}. \end{aligned}$$

These are all explicitly proportional to  $\epsilon$ , and thus, as claimed earlier, everything vanishes for  $\epsilon \rightarrow 0$  as



long as  $y, z$  are not both zero. We expect to extract from (5.9) in the limit  $\epsilon \rightarrow 0$  a multiple of  $\partial_z \delta^3(y, z)$ . Since the  $(a)$  term appears in eqn. (5.9) in the form  $\partial_z(a)$ , we expect that  $(a)$  by itself without this derivative would simply produce a multiple of  $\delta^3(y, z)$ .

This means that the contribution of the  $(a)$  term to the coefficient of  $\partial_z \delta^3(y, z)$  can be obtained by evaluating the integral of  $(a)$ :

$$(5.11) \quad C := \int dy dz d\bar{z} \bar{z}(a).$$

To demonstrate eqn. (5.5), it suffices to show that  $C = \frac{\pi}{i}$ . This integral is absolutely convergent. A simple scaling argument shows that it is independent of  $\epsilon$ , so we may as well set  $\epsilon = 1$ .

In polar coordinates ( $dzd\bar{z} = -2irdrd\theta$ ), after integration over  $\theta$ , we get

$$(5.12) \quad C = -\frac{8i\pi}{4^2} \int_{-\infty}^{\infty} dy \int_0^{\infty} dr \frac{r(y(1-y)+r^2)}{(y^2+r^2)^{\frac{3}{2}}((y-1)^2+r^2)^{\frac{3}{2}}}.$$

We can evaluate this integral with a formula familiar from the evaluation of Feynman diagrams:

$$(5.13) \quad \frac{1}{A^\alpha B^\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 dt \frac{t^{\alpha-1}(1-t)^{\beta-1}}{(tA+(1-t)B)^{\alpha+\beta}}.$$

Choosing  $\alpha = \beta = 3/2$  and remembering  $\Gamma(3)/\Gamma(3/2)^2 = 8/\pi$ , we have

$$(5.14) \quad C = \frac{-64i}{4^2} \int_0^1 dt \sqrt{t(1-t)} \times \int_{-\infty}^{\infty} dy \int_0^{\infty} dr \frac{r(y(1-y)+r^2)}{((1-t)y^2+t(y-1)^2+r^2)^3}.$$

Using

$$(5.15) \quad \int_0^{\infty} dr \frac{r(a+r^2)}{(b+r^2)^3} = \frac{a+b}{4b^2},$$

we obtain

$$(5.16) \quad \begin{aligned} C &= \frac{-16i}{4^2} \int_0^1 dt \sqrt{t(1-t)} \int dy \frac{y(1-y) + (1-t)y^2 + t(y-1)^2}{((1-t)y^2 + t(y-1)^2)^2} \\ &= \frac{-16i}{4^2} \int_0^1 dt \sqrt{t(1-t)} \frac{\pi}{\sqrt{t(1-t)}} = \frac{\pi}{i}. \end{aligned}$$

We can do a similar analysis for  $(b)$ , but  $(b)$  takes the form

$$(5.17) \quad \bar{z}F(y, |z|)$$

and hence vanishes when integrated over the angle  $\theta$ . This means that there is no singular contribution from  $(b)$ . This concludes the verification of (5.5).

### 5.3 Relation to the $R$ -Matrix

Though we have performed an independent computation, the result can actually be deduced from a knowledge of the  $\mathcal{O}(\hbar)$  contribution to the  $R$ -matrix, computed in section 4. Consider (Fig. 18) two parallel Wilson lines  $K$  and  $K'$  both supported at  $z = 0$  and a third Wilson line  $K''$  at some other point  $z = z_0$  that is crossing them. Assuming that  $K$  and  $K'$  are far apart (compared to  $|z_0|$ ) there is a unique lowest order Feynman diagram in which  $K''$  interacts nontrivially with both  $K$  and  $K'$ . This is the diagram sketched in the figure with a gluon exchanged from  $K''$  to  $K$  and another from  $K''$  to  $K'$ . The group theory factors associated to the two gluon exchanges are respectively  $t_a \otimes 1 \otimes t_a$  and  $1 \otimes t_b \otimes t_b$  (where the three factors refer respectively to  $K, K',$  and  $K''$ ). The product of these is  $t_a \otimes t_b \otimes t_b t_a$ , where the factors acting on  $K''$  are ordered as  $t_b t_a$  because  $K'$  is to the right of  $K$  (so that path ordering along  $K''$  puts  $t_b$  to the left of  $t_a$ ).

Now suppose that  $K$  and  $K'$  are brought closer together. Two-dimensional diffeomorphism invariance means that we have to get the same result after summing over all diagrams, but the contributions of individual diagrams can depend on the distance from  $K$  to  $K'$ . In particular, the diagram that we considered before still contributes when  $K$  and  $K'$  are close compared to  $|z_0|$ , but their contribution is not the same as before. That is because (Fig. 19(a)) even though  $K'$  is to the right of  $K$ , the gluon emitted from  $K'$  might be absorbed on  $K''$  to the left of the gluon emitted from  $K$ . So the contribution of the diagram that we considered previously is now modified by a term proportional to  $t_a \otimes t_b \otimes [t_a, t_b] = f_{abc} t_a \otimes t_b \otimes t_c$ . So there must be another diagram that is significant when  $K$  and  $K'$  are nearby and that gives a group theory factor of this form. That diagram is shown in Fig. 19(b).

But this last diagram is just the one (Fig. 17) that we studied to find the quantum correction to the OPE, except that now instead of considering an arbitrary external  $A$  field, as in the previous discussion, we have provided a third Wilson operator that is the source of this  $A$  field. The analysis of Fig. 19(b) for  $K \rightarrow K'$  is essentially the same as the analysis that we have already performed of Fig. 17.

The upshot is that the quantum correction to the OPE is an inevitable consequence of the quantum correction to the  $R$ -matrix, and vice-versa.

### 5.4 Interpretation of the Result

In the analysis of Fig. 17, we started with two representations  $\rho'$  and  $\rho''$  of  $\mathfrak{g}$ , that is representations of  $\mathfrak{g}[[z]]$  with generators  $t'_{a,n} = t''_{a,n} = 0$  for  $n > 0$ . Let us write just  $t'_a$  and  $t''_a$  for  $t'_{a,0}$  and  $t''_{a,0}$ . The above computation showed that the Wilson operator obtained by fusing

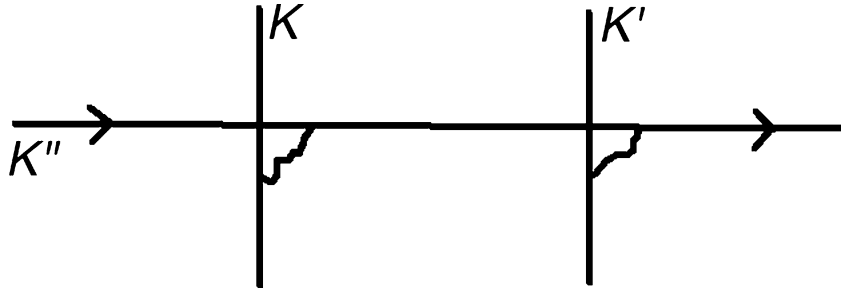


Figure 18. A Wilson operator  $K''$  crossing two parallel Wilson operators  $K$  and  $K'$ .  $K$  and  $K'$  are supported at  $z = 0$  and  $K''$  at some other point  $z = z_0$ . Drawn is a Feynman diagram with a gluon exchanged between  $K$  and  $K''$  as well as one between  $K'$  and  $K''$ .

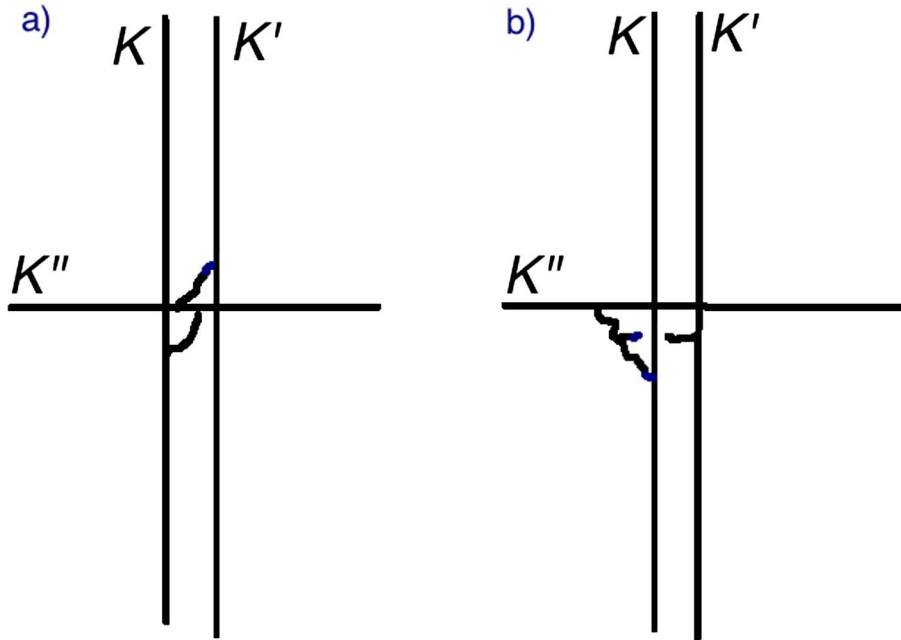


Figure 19. When  $K$  and  $K'$  are brought nearby, the contribution of the diagram in (a) is modified relative to Fig. 18. To maintain two-dimensional topological invariance, this is balanced by a new contribution in (b), which is related to the existence of a quantum correction to the OPE.

the two we started with couples to  $\partial_z A$  and therefore has a nonzero  $t_{a,1}$ . In fact  $t_{a,1}$  of the fused Wilson line can be read off from eqn. (5.6):

$$(5.18) \quad t_{a,1} = -\frac{\hbar}{2} \sum_{b,c} f_{abc} t'_b \otimes t''_c.$$

On the other hand, the computation gave no contributions to generators  $t_{a,n}$  of the composite Wilson operator for  $n \neq 1$ . So to this order,  $t_{a,0}$  is given by the classical formula

$$(5.19) \quad t_{a,0} = t'_a \otimes 1 + 1 \otimes t''_a,$$

and the higher generators vanish,  $t_{a,n} = 0$ ,  $n > 1$ . Here  $t_{a,0}$  are the generators of  $\mathfrak{g}$  in the classical tensor product  $\rho = \rho' \otimes \rho''$ .

This result certainly shows that to get a closed OPE, we must consider Wilson operators derived from

representations of  $\mathfrak{g}[[z]]$ , not just  $\mathfrak{g}$ . But its consequences go far beyond that. The formula (5.18) actually implies that quantum corrections to the theory actually deform  $\mathfrak{g}[[z]]$  itself - or to be more precise that at the quantum level, Wilson operators of the theory correspond to representations not of  $\mathfrak{g}[[z]]$  itself but of a quantum deformation of this algebra (or more accurately, of its universal enveloping algebra, as we will see).

The basic reason for this is that the formula (5.18) is not consistent with the commutation relations of  $\mathfrak{g}[[z]]$ , so it implies further deformations. In  $\mathfrak{g}[[z]]$ , one has the commutation relation

$$(5.20) \quad [t_{a,1}, t_{b,1}] = f_{abc} t_{c,2}.$$

Recalling the Jacobi identity

$$(5.21) \quad f_{uva}f_{abc} + f_{vba}f_{auc} + f_{bua}f_{avc} = 0,$$

we deduce from this that

$$(5.22) \quad f_{uva}[t_{a,1}, t_{b,1}] + (\dots) = 0,$$

where the omitted terms are obtained by cyclic permutations of the indices  $uvb$ .

This is an identity in  $\mathfrak{g}[[z]]$ , but a short calculation will show that  $t_{a,1}$  as defined in eqn. (5.18) does not obey this identity. Instead it satisfies a deformed version of this identity that we will describe presently.

Because  $\mathfrak{g}[[z]]$  is being deformed, what did we mean in claiming that to get a closed OPE for line operators, we should start with representations of  $\mathfrak{g}[[z]]$ ? The precise statement is not that line operators in the quantum theory correspond to representations of  $\mathfrak{g}[[z]]$ , but that had we started at the classical level with arbitrary representations of the  $\mathfrak{g}[[z]]$ , then consideration of products of line operators would not have forced us to consider new objects. By contrast, if we start at the classical level with representations of  $\mathfrak{g}$  only, then getting a closed OPE does require introducing many more line operators.

Now let us go back to the question of how to interpret the failure of the quantum-induced  $t_{a,1}$  to obey the commutation relations of  $\mathfrak{g}[[z]]$ . Since  $t_{a,1}$  in eqn. (5.18) is bilinear in the generators  $t'_a$  and  $t''_a$ , the left hand side of eqn. (5.21), after evaluating the commutator, is cubic in these generators; more specifically it is a sum of terms of bidegree (2, 1) and (1, 2) in  $t'_a$  and  $t''_a$ .

The upshot is that to account for the failure of eqn. (5.22), we have to deform this commutation relation by adding, in order  $\hbar^2$ , a certain cubic polynomial in  $t_{a,0}$ . The generators  $t_{a,0}$  and  $t_{a,1}$  of the deformed algebra satisfy

$$(5.23) \quad f_{uva}[t_{a,1}, t_{b,1}] + (\dots) = \hbar^2 Q_{uvb}(t_{a,0}),$$

where  $Q_{uvb}(t_{a,0})$ , which is completely antisymmetric in its indices  $uvb$ , is for each  $uvb$  a homogeneous symmetric cubic polynomial in the  $t_{a,0}$ .

What polynomial is needed can be deduced by trying to make sure that eqn. (5.23) is consistent with eqns. (5.18) and (5.19). However, when one tries to do this, one runs into a snag. No matter what  $Q_{uvb}$  may be, if  $t_{a,0} = t'_a \otimes 1 + 1 \otimes t''_a$ , then  $Q_{uvb}(t_{a,0})$  is a sum of terms of bidegree (3, 0), (2, 1), (1, 2), and (0, 3) in  $t'_a$  and  $t''_a$ . We can pick  $Q_{uvb}$  so that the terms in eqn. (5.23) of bidegree (2, 1) and (1, 2) work out correctly. But there is nothing we can do about the terms of bidegree (3, 0) and (0, 3), unless they vanish by themselves.

The interpretation of this is as follows. The quantum deformed algebra satisfies (5.23) (with analogous

deformations of the commutation relations involving other generators). In contrast to  $\mathfrak{g}[[z]]$ , a representation of  $\mathfrak{g}$  does not automatically lift or extend to a representation of the deformed algebra. If we start with matrices  $t_a$  that represent  $\mathfrak{g}$ , we can always get a representation of  $\mathfrak{g}[[z]]$  by setting  $t_{a,0} = t_a$  and  $t_{a,n} = 0$  for  $n > 0$ . But this only works in the deformed algebra if  $Q_{uvb}(t_a) = 0$ . Otherwise, the deformed commutation relations force  $t_{a,1}$  to be nonzero. So quantum mechanically, the Wilson operators whose product we are trying to determine are anomalous, and need to be modified<sup>18</sup> with a contribution to  $t_{a,1}$  of  $\mathcal{O}(\hbar)$ , unless  $Q_{uvb}(t'_a) = Q_{uvb}(t''_a) = 0$ . Once we impose this restriction, we do not need to worry about the terms in eqn. (5.23) of bidegree (3, 0) or (0, 3). We need consider only the (2, 1) and (1, 2) terms in that equation.

The explicit polynomial  $Q_{uvb}(t_{a,0})$ , though not very illuminating, can be worked out by analyzing those terms.<sup>19</sup> A notable fact is that because this polynomial is of degree greater than 1, the deformation from  $\mathfrak{g}[[z]]$  by including  $\hbar^2 Q_{uvb}$  on the right hand side of eqn. (5.23) cannot be understood as a deformation of the Lie algebra  $\mathfrak{g}[[z]]$ . It can instead be understood as part of an associative algebra deformation of the universal enveloping algebra of  $\mathfrak{g}[[z]]$ , denoted  $U(\mathfrak{g}[[z]])$ . This makes sense because a polynomial in elements of  $U(\mathfrak{g}[[z]])$  is itself an element of  $U(\mathfrak{g}[[z]])$ , so  $Q_{uvb}(t_{a,0})$  is an element of  $U(\mathfrak{g}[[z]])$ . Eqn. (5.23) (and its analogs for other components that we have not calculated) can be understood as giving an associative algebra deformation (which in particular entails a Lie algebra deformation) of  $U(\mathfrak{g}[[z]])$ .

## 5.5 The Yangian

The deformation of  $U(\mathfrak{g}[[z]])$  that was uncovered in section 5.4 is known as the Yangian. In fact, eqn. (5.18) is a standard formula describing the difference in lowest order between the tensor product of representations of the Yangian and the tensor product of representations of  $\mathfrak{g}[[z]]$ .

How can we know that the relevant deformation of  $U(\mathfrak{g}[[z]])$  is the Yangian, given that we have only performed some simple computations in lowest non-trivial order? The answer to this question is that according to general theorems, the Yangian is the only deformation of the OPE that agrees with the lowest order deformation that we have found in eqn. (5.18) and possesses certain general properties. Thus (as already stated in [11, 12]), the associativity of the OPE, together with the fact that the OPE in  $\mathcal{O}(\hbar)$  receives

<sup>18</sup> If they cannot be so modified, they are simply anomalous and do not have counterparts in the quantum theory.

<sup>19</sup> We will also derive it by a direct Feynman diagram analysis in section 8. Section 8.6.1 contains a fairly thorough analysis of this polynomial. See also [8], p. 376.

a non-trivial correction, determines the OPE uniquely to all orders in perturbation theory, up to changes of variables. This is guaranteed by a theorem of Drinfeld, which also says that the resulting 1-parameter deformation of the  $U(\mathfrak{g}[[z]])$  is the Yangian  $Y_{\hbar}(\mathfrak{g})$ , the algebra underlying the rational solution of the Yang-Baxter equation associated to  $\mathfrak{g}$ .

To make these statements more precise and closer in spirit to those in [11, 12], we can use the language of category theory to say that the “category of Wilson lines”  $\mathcal{C}_{\hbar}$  is given by the category of representations of  $Y_{\hbar}(\mathfrak{g})$ .

This category  $\mathcal{C}_{\hbar}$  has the following four properties.

First,  $\mathcal{C}_{\hbar}$  is a non-trivial one-parameter deformation of the category of representations<sup>20</sup> of  $U(\mathfrak{g}[[z]])$ .

Second, the OPE of the Wilson lines defines a functor, by fusing two objects to produce a third object

$$(5.24) \quad \mathcal{C}_{\hbar} \times \mathcal{C}_{\hbar} \rightarrow \mathcal{C}_{\hbar}.$$

Moreover, the OPE is associative, though not commutative. These two conditions define a *monoidal category*, and hence  $\mathcal{C}_{\hbar}$  is monoidal.<sup>21</sup>

Third, the category  $\mathcal{C}_{\hbar}$  allows for a kind of braiding,<sup>22</sup> because it has certain additional structure. Here we again consider two Wilson lines parallel in the topological plane, but now at different points  $z$  and  $z + \lambda$  of the complex plane. Since  $z \neq z + \lambda$ , the Wilson lines never coincide with each other, even when they coincide in the topological plane. Moreover, since we have two directions at our disposal in the holomorphic plane  $C$ , we can move the two Wilson lines around in  $C$ , to replace  $z$  with  $z + \lambda$ , without encountering any singularities. In the language of [11, 12], the translation  $z \mapsto z + \lambda$  of the Wilson line by a parameter  $\lambda$  defines a functor  $T_{\lambda} : \mathcal{C}_{\hbar} \rightarrow \mathcal{C}_{\hbar}$ . If  $W, V$  are two Wilson lines supported at 0, then the explanations above means that we have a natural isomorphism

$$(5.25) \quad R_{V,W} : T_{\lambda}W \otimes V \cong V \otimes T_{\lambda}W,$$

where  $\otimes$  indicates the monoidal structure on  $\mathcal{C}_{\hbar}$  coming from the OPE of parallel Wilson lines.

<sup>20</sup> At the classical level, representations of  $\mathfrak{g}[[z]]$  are the same as representations of  $U(\mathfrak{g}[[z]])$ , so Wilson operators can be identified with representations of either of these algebras. But the statement in the text must be formulated for  $U(\mathfrak{g}[[z]])$ , because the Yangian is a deformation of  $U(\mathfrak{g}[[z]])$ , not of  $\mathfrak{g}[[z]]$ .

<sup>21</sup> By the same argument, the category of line operators in a two-dimensional topological quantum field theory (TQFT) is always a monoidal category, as is the category of boundary line operators in a 3-dimensional TQFT with a topological boundary condition.

<sup>22</sup> This is different from the braiding familiar from line operators in three-dimensional topological field theories.

A theorem of Drinfeld [7] now implies that  $\mathcal{C}_{\hbar}$  is uniquely fixed by these properties to be the category of representations of the Yangian  $Y_{\hbar}(\mathfrak{g})$ . The monoidal structure on  $\mathcal{C}_{\hbar}$  comes from the coproduct on the Yangian, and the map  $R_{V,W}$  comes from the  $R$ -matrix of the Yangian  $Y_{\hbar}(\mathfrak{g})$ .

These arguments are admittedly somewhat abstract. In a companion paper [25], we will explain, in a concrete and down to earth way, how to extract a representation of the Yangian – exactly, not just to order  $\hbar$  – from any Wilson line operator of this theory. The ability to do this is somewhat analogous to the ability in section 3.5 to determine in an elementary way some solutions of the Yang-Baxter equation for  $GL_N$ . Hopefully this will help convince the reader that the conclusions from the abstract arguments really apply to the quantum field theory.

## 6. The Framing Anomaly and Curved Wilson Lines

In section 3.6, we deduced from some explicit elementary examples that there must be a framing anomaly for Wilson lines. The framing anomaly says that what is constant along a Wilson line is not  $z$  but  $z - \hbar h^{\vee} \varphi / \pi$ , where  $h^{\vee}$  is the dual Coxeter number of the gauge group, and  $\varphi$  is the angle between the tangent vector to the Wilson line and some chosen direction in the topological plane. The purpose of the present section is to perform a direct Feynman diagram computation exhibiting this effect.

In contrast to the calculations that we performed for the  $R$ -matrix and for the OPE of Wilson operators, for the framing anomaly, a lowest order computation gives the exact result. For the case of the Yangian – that is for  $C = \mathbb{C}$  – this is clear from the fact that this theory is invariant under adding a constant to  $z$  and under a common rescaling of  $z$  and  $\hbar$ . The assertion that  $z - \hbar h^{\vee} \varphi / \pi$  is constant along a Wilson operator is consistent with those symmetries, but a modified statement with additional terms of higher order in  $\hbar$  would not be. The fact that the lowest order framing anomaly gives the full answer is also evident in the examples considered in section 3.6.

For other choices of  $C$  related to trigonometric or elliptic solutions of the Yang-Baxter equation, there is no scaling symmetry. However, because the analysis of the framing anomaly is local along  $C$ , the form of the anomaly is the same for the other cases as long as one picks variables so that locally along  $C$  the action takes the same form as for  $C = \mathbb{C}$ . (In the coordinates of eqn. (3.14), this is automatically true if  $C$  is an elliptic curve, and it is true for  $C = \mathbb{C}^{\times}$  after replacing  $z$  with  $\log z$ .)

The relevant diagram that will give the anomaly is shown in Fig. 20. This diagram makes a convergent



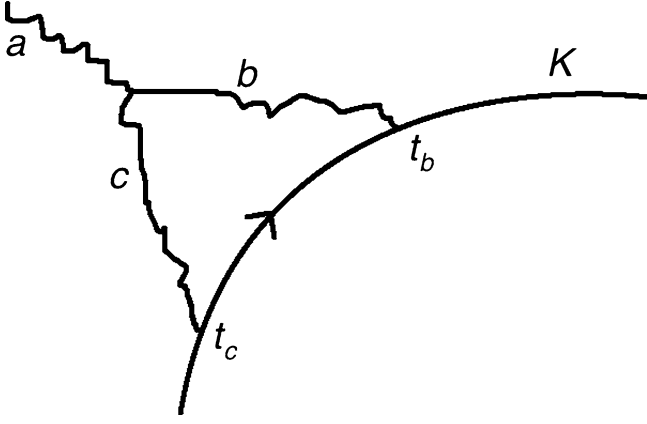


Figure 20. The one-loop diagram that leads to a framing anomaly for Wilson operators.

and well-defined contribution to the effective action of the theory. However, this contribution is not gauge-invariant. A failure of gauge-invariance will arise from the behavior when all three interaction vertices coincide at a common point, which will inevitably be along the support  $K$  of the Wilson operator.

Before describing what form the anomaly will take, we first do a small exercise in group theory. We write the gauge field  $A$  explicitly as  $A = \sum_a A_a t_a$ , and recall that the trilinear vertex involving a coupling of three gauge fields is proportional to  $f_{abc}$ . A gauge boson of type  $b$  or  $c$  couples to the Wilson line via a factor of  $t_b$  or  $t_c$  (understood here as a matrix acting in the appropriate representation). So as far as group theory is concerned, the indicated diagram generates a coupling of an external gauge field  $A_a$  to the Wilson operator via a matrix

$$(6.1) \quad \sum_{a,b,c} A_a f_{abc} t_b t_c.$$

Here because of antisymmetry of  $f_{abc}$ , we can replace  $t_b t_c$  by  $\frac{1}{2}[t_b, t_c] = \frac{1}{2} f_{bcd} t_d$ . Then as  $\sum_{b,c} f_{abc} f_{bcd} = 2h^\vee \delta_{ad}$ , where  $h^\vee$  is known as the dual Coxeter number<sup>23</sup> of  $\mathfrak{g}$ . Thus just from the point of view of group theory, we get a coupling  $h^\vee \sum_a A_a t_a$  of the external gauge field to the Wilson line. Thus the dependence on  $\mathfrak{g}$  is only an overall factor of  $h^\vee$ .

To calculate the anomaly, it suffices to consider the case of a Wilson line supported on a curve  $K \subset \mathbb{R}^2$  that is nearly a straight line, and at a point  $z = z_0$  in  $\mathbb{C}$ . In fact, since a simple description in terms of Yangians and integrable systems only arises in a limit in which the metric of  $\mathbb{R}^2$  is scaled up, we are really supposed to consider only the case that  $K$  has a very large radius of curvature and therefore can everywhere be

<sup>23</sup> Recall our normalization of the Killing form in eqn. (3.2). The factor of 2 here is inserted to match with the standard definition of  $h^\vee$  in the literature.

locally well-approximated as a straight line. We can pick coordinates so that  $K$ , in the region of interest, is very close to the  $x$ -axis in  $\mathbb{R}^2$ , and is described by a curve  $y = y(x)$ , with  $y$  everywhere small. We will find that the amplitude  $I_1$  that comes from the diagram of Fig. 20 is not gauge-invariant. Under a gauge transformation  $A \rightarrow A + D\varepsilon$ , with  $\varepsilon = \sum_a t_a \varepsilon_a(x, y, z)$  a gauge parameter, the variation of  $I_1$  will be

$$(6.2) \quad \delta I_1 = -\frac{\hbar h^\vee}{2\pi} \int_K dx \left( \frac{d^2 y}{dx^2} \partial_z \varepsilon(x, y(x), z_0) \right).$$

Here as above  $h^\vee$  is the dual Coxeter number. (A factor of  $\hbar$  is present because this is the anomaly in a diagram with two propagators, each giving a factor of  $\hbar$ , and one bulk vertex, proportional to  $1/\hbar$ .) This formula for  $\delta I_1$  as an integral over  $K$  is written in a way that is only valid in a portion of  $K$  that is close to the  $x$ -axis. Since we will assume  $y(x)$  small, we can replace  $\varepsilon(x, y(x), z_0)$  by  $\varepsilon(x, 0, z_0)$ , which we will write simply as  $\varepsilon(x, z_0)$ . Now let us discuss how to cancel this anomaly by a correction to the classical definition of the Wilson operator. We recall that, in linear order, an external gauge field  $A$  couples to the Wilson operator via

$$(6.3) \quad I_{cl} = \int_K dx A_x(x, z_0).$$

This is gauge-invariant, as long as  $z_0$  is constant. To cancel the anomaly  $\delta I_1$ , we must modify the classical coupling  $I_{cl}$  so that it is no longer gauge-invariant. We do this by replacing  $z_0$  by  $z_0 - \hbar h^\vee \frac{1}{2\pi} dy/dx$ . Thus we replace  $I_{cl}$  with

$$(6.4) \quad I'_{cl} = \int_K dx A_x \left( x, z_0 - \frac{\hbar h^\vee}{2\pi} \frac{dy}{dx} \right).$$

Under a gauge transformation  $\delta A_x(x, y, z) = \partial_x \varepsilon(x, y, z)$ , the variation of  $I'_{cl}$  is<sup>24</sup>

$$(6.5) \quad \delta I'_{cl} = -\frac{\hbar h^\vee}{2\pi} \int_K dx \left( \frac{d^2 y}{dx^2} \partial_z \varepsilon(x, z_0) \right).$$

This cancels the anomaly  $\delta I_1$ .

We learn, then, that including the anomaly, we must set not  $z = z_0$  but  $z = z_0 - \frac{1}{2\pi} \hbar h^\vee dy/dx$ . The formulas as written, however, are only valid under the assumption that  $dy/dx$  is small. To get a more general formula, we should re-express  $dy/dx$  in terms of the angle  $\varphi$  between the tangent vector to  $K$  and some chosen direction in the  $xy$  plane, for instance the direction of increasing  $x$ . If we define the sign of  $\varphi$  to increase when  $K$  rotates in a clockwise direction, then  $dy/dx$  can be identified (when it is small) with  $-\varphi$ . So a more general form of the relation between  $z$  and

<sup>24</sup> We use  $d\varepsilon/dx = \partial_x \varepsilon + (\partial_x z) \partial_z \varepsilon$ . We work to first order in  $y$  and replace  $\partial_z \varepsilon(x, z_0 - \hbar h^\vee dy/dx)$  with  $\partial_z \varepsilon(x, z_0)$ .

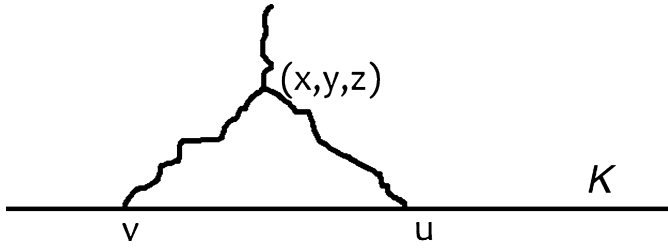


Figure 21. The one-loop diagram with  $K$  taken to be a straight line - the  $x$  axis.

the slope of  $K$  is  $z = z_0 + \hbar h^\vee \varphi / (2\pi)$ . A different way to express this result is to say that along  $K$ , what is constant is not  $z$  but  $z - \hbar h^\vee \varphi / (2\pi)$ . This is the way the anomaly was stated in section 3.6.

Before trying to explain how this anomaly comes from the diagram of Fig. 20, let us first explain how the amplitude coming from that diagram is defined. The only possible problem in defining the integral associated to this diagram comes from the region of integration in which all three interaction vertices - one in bulk and two along the curve  $K$  - coincide in  $\mathbb{R}^4$ . In this region,  $K$  can be approximated as a straight line. Let us take the line to be the  $x$ -axis, that is the curve  $y = 0$ , and discuss the integral in this case. For simplicity, we choose  $z_0 = 0$ . As in the figure, we take the vertices along  $K$  to be at  $x = u, v$  with  $v \leq u$ , and we place the bulk vertex at a generic point  $(x, y, z)$ . Making use of the explicit form of the propagators, the necessary integral turns out to be (up to a constant factor, about which we will be more precise later)

$$(6.6) \quad \int_{-\infty < v \leq u < \infty} du dv \int_{\mathbb{R}^4} dx dy dz d\bar{z} A_x(x, y, z) \times \frac{2y\bar{z}}{((x-v)^2 + y^2 + |z|^2)^2 ((x-u)^2 + y^2 + |z|^2)^2}.$$

Here  $A = A_x dx + A_y dy + A_z d\bar{z}$  is a background field. However, for the case of a straight line in a flat  $\mathbb{R}^2$  (and only for this case), only the  $A_x$  component contributes, as is indicated in eqn. (6.6). This depends on the detailed form of the propagator.

Any possible divergence in the integral will be a local effective action along  $K$  that will be represented as an integral along  $K$ . Thus in assessing the convergence of the integral, we should leave one variable un-integrated, say  $x$ . (Alternatively, we could ensure that there is no divergence in the integral over  $x$  by taking  $A$  to have compact support in the  $x$  direction.) So we set, for example,  $x = 0$ . If  $A_x(0, y, z)$  is independent of  $y$  and  $z$ , then the integral (6.6) is linearly divergent by power counting. (We have to integrate over five variables  $u, v, y, z, \bar{z}$ , and the integrand scales as  $(\text{length})^{-4}$ .) However, the integrand is odd under  $y \rightarrow -y$  and separately under  $z \rightarrow -z$ . Accordingly, the integral is well-defined by a sort of principal value

prescription. We take  $\mathbb{R}_\eta^4$  to be the subspace of  $\mathbb{R}^4$  defined by  $(y^2 + |z|^2)^{1/2} \geq \eta$ , and we replace the integral by

$$(6.7) \quad \lim_{\eta \rightarrow 0} \int_{-\infty < v \leq u < \infty} du dv \int_{\mathbb{R}_\eta^4} dx dy dz d\bar{z} A_x(x, y, z) \times \frac{2y\bar{z}}{((x-v)^2 + y^2 + |z|^2)^2 ((x-u)^2 + y^2 + |z|^2)^2}.$$

This integral is well-defined. Concretely, if we expand  $A_x(0, y, z)$  in powers of  $y, z$  and  $\bar{z}$  near  $y = z = 0$ , the first term that can contribute is proportional to  $yz$  and leads to a convergent integral. (Other nonvanishing contributions come from terms with additional factors of  $y^2$  or  $z\bar{z}$ .)

We have described this for the case that  $K$  is a straight line, but since any curve can be approximated locally as a straight line, the general case is similar. The one-loop integral can always be defined by a principal value procedure in which one constrains the bulk vertex to be a distance at least  $\eta$  from  $K$  and then takes the limit  $\eta \rightarrow 0$ . The only change from the discussion in the last paragraph is that instead of breaking the symmetry under  $y \rightarrow -y$  by picking out a term in  $A(0, y, z)$  that is linear in  $y$ , we could make use of the curvature of  $K$  to break the symmetry. This again leads to a convergent integral.

Now that we have defined the 1-loop amplitude that we want to study, we can assess its gauge-invariance. For this, it is convenient to perform the  $u$  and  $v$  integrals in (6.7), leading to an expression of the general form

$$(6.8) \quad \int_{\mathbb{R}^4} A \wedge \Theta_0,$$

where  $\Theta_0$  is 3-form on  $\mathbb{R}^4$ . To be more precise,  $\Theta_0$  is a distributional 3-form on  $\mathbb{R}^4$ , that is, it is a distribution on smooth 1-forms  $A$ . We note that although we have described a specific procedure to define  $\Theta_0$ , there was nothing really distinguished about this procedure and we could have used a different one. The effect of using a different procedure would be to add to  $\Theta_0$  a distribution  $\Theta'_0$  supported on  $K$ . But

$$(6.9) \quad \int_{\mathbb{R}^4} A \wedge \Theta'_0 = \int_K A \beta,$$

where  $\beta$  is some 0-form on  $K$ , which depends on  $\Theta'_0$ . A shift of this type in the effective action can be interpreted as a change in the classical line operator whose quantum properties we are trying to study. We are only interested in anomalies modulo those that can be removed by such a redefinition of the underlying classical line operator.

We want to assess gauge-invariance of eqn. (6.7) under  $A \rightarrow A + D\varepsilon$ , where  $D = dx\partial_x + dy\partial_y + d\bar{z}\partial_{\bar{z}}$ . It suffices to consider the case  $\varepsilon = z\varepsilon'$  where  $\varepsilon'$  is regular

at  $z = 0$ . In fact, any  $\varepsilon$  can be written  $\varepsilon = z\varepsilon'(x, y, z, \bar{z}) + \varepsilon''(x, y, \bar{z})$ , where  $\varepsilon''$  is holomorphic in  $\bar{z}$ . In the following computation, any contribution from  $\varepsilon''$  will disappear after integration over the phase or argument of  $z$ . (It will also become clear that we can assume  $\varepsilon'$  to be independent of  $z$  and  $\bar{z}$  since terms proportional to  $z$  or  $\bar{z}$  are not singular enough to be relevant, and that likewise only the restriction of  $\varepsilon'$  to  $K$  matters.) Since  $D$  commutes with  $z$ , it is equivalent to replace  $\Theta_0$  with  $\Theta = z\Theta_0$  and to assess the invariance of

$$(6.10) \quad \int_{\mathbb{R}^4} A \wedge \Theta$$

under  $A \rightarrow A + D\varepsilon'$ . After integration by parts, this means that we need to study the distributional four-form  $D\Theta$ . This form vanishes away from  $K$  because of the classical gauge-invariance of the theory. So we expect  $D\Theta$  to be a distribution with support on  $K$ . Finally, as  $\Theta$  is proportional to an explicit factor of  $dz$  which comes from the bulk interaction vertex, we can replace  $D\Theta$  by  $d\Theta$ , where  $d$  is the ordinary exterior derivative  $d = D + dz\partial_z$ .

Thus, we are reduced to studying  $d\Theta$ , which should equal a distribution supported along  $K$ . In fact, we claim

$$(6.11) \quad d\Theta = \alpha\delta_K,$$

where  $\delta_K$  is a three-form delta function that is Poincaré dual to  $K$ , and  $\alpha$  is a one-form supported on  $K$ . The anomaly is then given by  $\alpha$ . (A priori, instead of  $\delta_K$ , we might have gotten an expression involving normal derivatives of  $\delta_K$ . Indeed, this would have happened had we not included an explicit factor of  $z$  in the definition of  $\Theta$ . With that term extracted, we will see that the most singular contribution is a delta function rather than a derivative of one.) We will use the following simple procedure to study  $d\Theta$ . We assume that  $K$  is close to the  $x$ -axis, so that it can be parametrized by  $x$ . Then we expand the three-form  $\Theta$  as the sum of two terms, one proportional to  $dx$  and one not:

$$(6.12) \quad \Theta = dx\Lambda + \Lambda'.$$

Here  $\Lambda'$  cannot contribute to the anomaly. The reason is that as  $\Lambda'$ , by definition, is a 3-form that vanishes if contracted with  $\partial_x$ , and is smooth away from  $K$ ,  $d\Lambda'$  cannot generate a delta function supported on  $K$  unless  $\Lambda'$  already has a delta function supported on  $K$ . But such a term is irrelevant; it could be eliminated by redefining  $\Theta$  along the lines of eqn. (6.9).

So in studying  $d\Theta$ , we can replace  $\Theta$  with  $dx\Lambda$ , in other words with the part of  $\Theta$  that is proportional to  $dx$ . We restrict  $\Lambda$  to, say,  $x = 0$ , and look for a delta function contribution in  $d\Lambda$  at  $y = z = 0$ . We can assume that  $K$  is tangent to the  $x$ -axis at  $x = 0$  and is described near

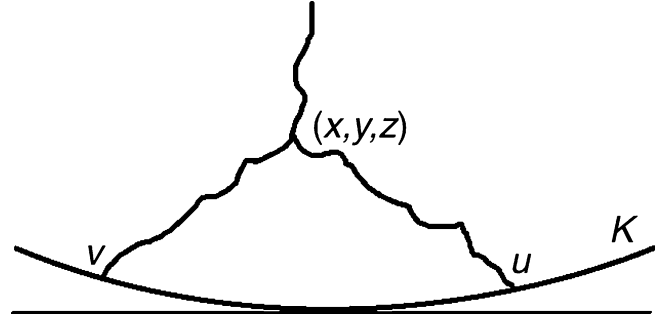


Figure 22. Here we take  $K$  to be tangent to the  $x$ -axis (the horizontal line in the figure) at, say,  $x = 0$ .

$x = 0$  by  $y = \frac{1}{2}fx^2$ , where  $f = d^2y/dx^2$ . Note that  $\Lambda = 0$  for the case that  $K$  is the  $x$ -axis. (This is equivalent to the statement that the only component of  $A$  that appears in eqn. (6.6) above is  $A_x$ .) So we expand  $\Lambda$  in powers of  $f$  near  $f = 0$ . We will see momentarily that the term in  $\Lambda$  linear in  $f$  does indeed lead to a delta function in  $d\Lambda$ . Higher order terms in  $f$  are not singular enough to make such a contribution.

Explicit calculation (see Appendix D) shows that the contribution to  $\Lambda$  that is linear in  $f$  is at  $x = 0$ ,

$$(6.13) \quad \Lambda = -\frac{3fi}{32\pi^2} \hbar h^\vee \frac{-yz\bar{z}(dz \wedge d\bar{z}) + 2z\bar{z}^2(dy \wedge dz)}{(y^2 + |z|^2)^{\frac{5}{2}}}.$$

By Stokes's theorem, the coefficient of the delta function can be extracted as

$$(6.14) \quad \int_{S_\eta} \Lambda,$$

where  $S_\eta$  is the two-sphere  $(y^2 + |z|^2)^{1/2} = \eta$  (the integral is independent of  $\eta$ , so we can choose  $\eta = 1$ ). We then need to evaluate

$$(6.15) \quad \int_{S_{\eta=1}} (-yz\bar{z}(dz \wedge d\bar{z}) + 2z\bar{z}^2(dy \wedge dz)).$$

By choosing  $y = \cos\theta$ ,  $z = \sin\theta e^{i\phi}$ , this becomes<sup>25</sup>

$$(6.16) \quad 2\pi \int_0^\pi d\theta (-2i \sin^3\theta \cos^2\theta - 2i \sin^5\theta) \\ = (2\pi)(-2i) \left( \frac{4}{15} + \frac{16}{15} \right) = -\frac{16\pi i}{3}.$$

Therefore the coefficient in front of the delta function is  $-f\hbar h^\vee/(2\pi)$ :

$$(6.17) \quad d\Lambda' = -\frac{1}{2\pi} f\hbar h^\vee \delta^3(y, z),$$

where  $\delta^3(y, z)$  is the Poincaré dual to the point  $x = y = z = 0$  in the hypersurface  $x = 0$ .

<sup>25</sup> We use  $dz \wedge d\bar{z} = -2i \sin\theta \cos\theta d\theta \wedge d\phi$ ,  $dy \wedge dz = -i \sin^2\theta e^{i\phi} d\theta \wedge d\phi$ .

Since  $f = d^2y/dx^2$ , this result when inserted back in (6.10) corresponds to an anomaly

$$(6.18) \quad -\frac{\hbar h^\vee}{2\pi} \int_K dx \frac{d^2y}{dx^2} \varepsilon'.$$

But  $\varepsilon'$  was defined by removing a factor of  $z$  from  $\varepsilon$ . So an equivalent and more illuminating way to describe the result is that the anomaly in the 1-loop amplitude under a gauge transformation generated by  $\varepsilon$  is

$$(6.19) \quad -\frac{1}{2\pi} \hbar h^\vee \int_K dx \frac{d^2y}{dx^2} \partial_z \varepsilon.$$

This is the form of the anomaly that was promised in eqn. (6.2).

## 7. Networks of Wilson Lines

### 7.1 Overview

By a “network” of Wilson lines, we mean simply a graph (Fig. 23) made of Wilson lines. Each line segment in the graph is labeled by a representation of the gauge group  $G$  (in general a different representation for each segment), and a vertex in the graph represents, in physics language, a gauge-invariant coupling among the representations that meet at that vertex. At the classical level – modulo the framing anomaly and its generalization for networks – the whole network will be at a fixed value<sup>26</sup> of the spectral parameter  $z$ .

Since we are already familiar with Wilson lines, the new ingredient in building such a network is the vertex. So let us discuss this in more detail. In general, if we are given a collection of representations  $V_1, \dots, V_n$ , with a  $G$ -invariant element

$$(7.1) \quad v \in V_1 \otimes \dots \otimes V_n,$$

then we can at the classical level form a Wilson line vertex in which  $n$  Wilson lines labeled by these  $n$  representations meet, as depicted in Fig. 24.

This construction is possible, at least classically, in any gauge theory. Classically, the vertex is defined as follows. For each Wilson line, we define the holonomy operator as usual, starting at the vertex where all the Wilson lines meet. The  $G$ -invariant tensor  $v \in V_1 \otimes \dots \otimes V_n$  provides an initial state to which we apply the holonomy operator on each Wilson line. The resulting operator is invariant under gauge transformations which vanish at the other end of the Wilson line. If each Wilson line has its second end either at

<sup>26</sup> The analog of this in purely three-dimensional Chern-Simons theory is to consider a not necessarily planar graph made from Wilson lines and embedded in spacetime in an arbitrary fashion. For example, the quantum 6j symbol is the expectation value of a tetrahedral graph [32].

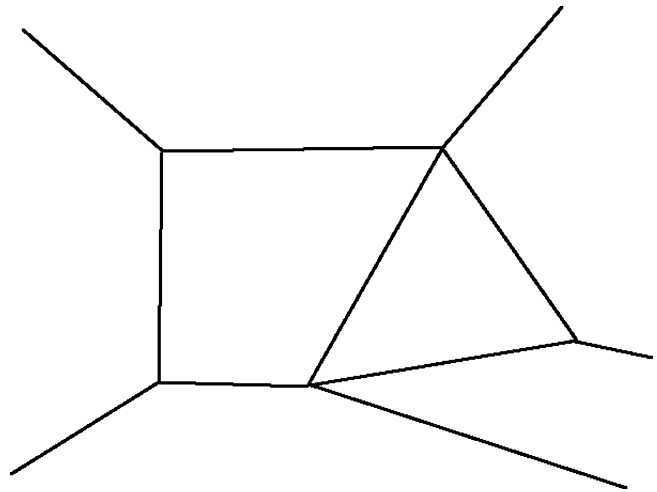


Figure 23. A “network” of Wilson lines is a graph in which the line segments are Wilson lines, labeled by representations of  $G$ , and a vertex represents a gauge-invariant coupling among the representations that label the lines that meet at that vertex.

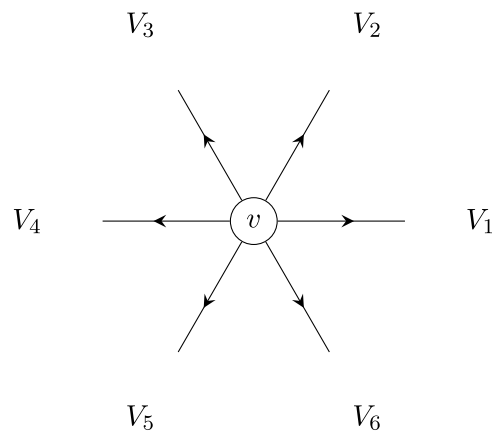


Figure 24. A vertex labeled by an invariant tensor  $v \in V_1 \otimes \dots \otimes V_6$ . Here and later, lines are labeled by the corresponding representations.

another gauge-invariant vertex or at infinity (where we assume fields and gauge parameters to be trivial), then we get a gauge-invariant network of Wilson lines, as in Fig. 23.

In the theory under discussion in the present paper, not every classical Wilson line corresponds to a line operator of the quantum theory. There is an obstruction that we first encountered in section 5.4. Likewise, it turns out that not every classical vertex corresponds to a vertex in the quantum theory.

As we will see, the condition that a classical vertex should correspond to a vertex in the quantum theory is that a certain  $\mathcal{O}(\hbar)$  anomaly should vanish. This anomaly comes from a lowest order Feynman diagram with one gluon exchange.

In this section, we will establish the result just stated; moreover, we will compute the  $\mathcal{O}(\hbar)$  obstruction to “quantizing” a vertex and find useful conditions under which it vanishes.

Finally, we will give concrete, interesting, and (as we will see in a companion paper) useful examples of quantum vertices.

Of course, to give examples of vertices, we first need examples of Wilson lines that we are allowed to work with. We will use the following sufficient condition that a Wilson line associated to a representation of  $G$  (as opposed to a more general Wilson line associated to a representation of  $\mathfrak{g}[[z]]$ ) can be quantized. Suppose the representation  $V$  satisfies the following algebraic condition:<sup>27</sup>

- (†) Every  $G$ -invariant map from  $\wedge^2 \mathfrak{g}$  to  $\text{End}(V)$  factors through a copy of the adjoint representation.

Then a Wilson line associated to the representation  $V$  can be “quantized,” that is, it exists in the quantum theory. Unfortunately, the proof of this statement is rather technical, and we have relegated it to Appendix C. We should point out that the condition (†) is only a sufficient criterion for a representation to be quantizable in this sense, but is far from being necessary. For example, for  $G = SL_n$ , it is known from other arguments that all representations of  $G$  are quantizable, but in general the condition (†) is not satisfied.

We should perhaps remark that what we call vertices correspond, in the theory of integrable relativistic scattering, to couplings of external particles to bound states or poles of the  $S$ -matrix. However, we will not try to make contact with the insight that comes from that point of view.

## 7.2 Vanishing of Higher Order Anomalies

Our first task is to show that an anomaly obstructing quantization of a vertex can arise only in lowest nontrivial order, from one-gluon exchange.

The anomaly that describes the failure of the configuration in Fig. 24 to be gauge-invariant at the quantum level will be a local operator of ghost number 1, made from the gauge field and the ghost field  $c$ , supported at the vertex. It will be valued in the vector space  $\widehat{V} = \otimes_i V_i$  which lives at the vertex. As in our discussion of the framing anomaly, the anomaly is determined by local considerations, so the choice of the complex Riemann surface  $C$  will not matter. We may as well take  $C = \mathbb{C}$  and take the vertex to be supported at  $z = 0$ . Likewise we can take the topological two-manifold  $\Sigma$  to be the  $xy$  plane with the vertex at the origin.

<sup>27</sup> An equivalent statement is that the only irreducible representation of  $G$  that appears both in  $\wedge^2 \mathfrak{g}$  and in  $\text{End}(V) = V \otimes V^*$  is the adjoint representation.

A key constraint is that the anomaly will be invariant under any symmetry of the theory that is also a symmetry of the classical vertex. A special case of diffeomorphism invariance in the  $\Sigma$  directions is invariance under scaling of the  $xy$  plane. The configuration in Fig. 24 with  $n$  Wilson lines emerging radially from a common vertex is invariant under this scaling, which therefore must be a symmetry of the anomaly. This tells us that the anomaly cannot depend on the  $x$  and  $y$  components of the gauge field  $A$ , and it cannot involve any  $x$  and  $y$  derivatives.

The classical theory is also invariant under the symmetry that simultaneously scales  $z, \bar{z}$  and  $\hbar$  by a real number, and under the symmetry in which  $z$  is rotated through  $\theta$ ,  $\bar{z}$  through  $-\theta$ , and  $\hbar$  through  $\theta$ . It follows that the anomaly is also preserved by these symmetries. From this it follows that the anomaly cannot have any  $\bar{z}$  derivatives, nor can it depend on  $A_{\bar{z}}$ . Thus the anomaly must be constructed from the ghost field  $c$  only, and it is linear in  $c$  since it has ghost number 1. Moreover, in order  $\hbar^k$ , the anomaly must have  $k$   $z$ -derivatives, that is, it must be proportional to  $\partial_z^k c(0)$ .

Finally, the anomaly must be invariant under constant gauge transformations. Since  $\partial_z^k c$  transforms in the adjoint representation, it must be combined with a copy of the adjoint representation in  $\widehat{V} = \otimes_i V_i$ . Thus, an order  $\hbar^k$  contribution to the anomaly must be an operator of the form

$$(7.2) \quad \sum_a \partial_z^k c^a(0) \alpha_a$$

for some collection of elements  $\alpha_a \in \widehat{V}$  that transform in the adjoint representation of  $\mathfrak{g}$ . Here  $c^a$  are the components of the ghost field  $c$  relative to a basis  $t_a$ ,  $a = 1, \dots, \dim \mathfrak{g}$ .

It remains to impose the condition that the anomaly must be BRST closed. We will see that this condition is satisfied if and only if  $k = 1$ . In implementing the condition of BRST invariance, we have to remember that the local operator that represents the anomaly lives at the endpoint of  $n$  Wilson lines. In general, the BRST transformation of a Wilson operator  $W(p, q)$  with ends at  $p$  and  $q$  is

$$(7.3) \quad \{Q, W(p, q)\} = c(p)W - Wc(q).$$

Here  $c(p)$  and  $c(q)$  are operators acting in the representation carried by the Wilson line. We are interested in the case that  $q$  is the location  $x = y = z = 0$  of the vertex under study, and we are only interested in the  $c(q)$  term in (7.3) (the other term will participate in a similar cancellation at the other end of the Wilson line in question). For the  $i^{\text{th}}$  Wilson line that ends at the vertex, we can write  $c(q)$  in more detail as

$$(7.4) \quad c(q) = \sum_a c^a(0) t_{a,i},$$



where  $t_{a;i}$  is the operator by which the Lie algebra generator  $t_a$  acts in the representation  $V_i$ .

The quantity that must vanish for BRST invariance of the anomaly (7.2) is therefore

$$(7.5) \quad -\sum_{i=1}^n c^a(0)t_{a;i}\partial_z^k c^b(0)\alpha_b + \{Q, \partial_z^k c^b(0)\}\alpha_b.$$

The statement that the  $\alpha_b$  transform in the adjoint representation means that

$$(7.6) \quad \sum_{i=1}^n t_{a;i}\alpha_b = f_{ab}^c \alpha_c.$$

Using the standard commutator relation

$$(7.7) \quad \{Q, c^a(0)\} = \frac{1}{2} f_{bc}^a c^b(0)c^c(0),$$

from which it follows that

$$(7.8) \quad \{Q, \partial_z^k c^a(0)\} = \frac{1}{2} \sum_{r=0}^k f_{bc}^a \partial_z^r c^b(0)\partial_z^{k-r} c^c(0),$$

we see that eqn. (7.5) is satisfied if and only if  $k = 1$ .

We conclude that anomalies can occur only in order  $\hbar$ , that is in the lowest nontrivial order, due to one-gluon exchange.

### 7.3 Calculating the Anomaly

Let us now turn to calculating the anomaly. Any  $\mathcal{O}(\hbar)$  anomaly must come from the Feynman diagrams depicted in Fig. 25.

This diagram involves only two of the Wilson lines in the network, and the anomaly will receive a contribution from each pair of Wilson lines.

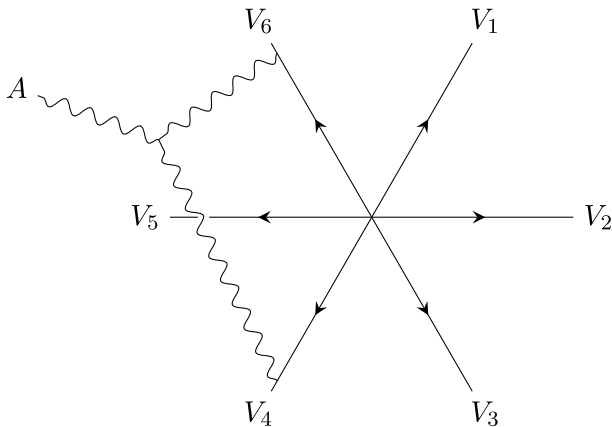


Figure 25. One-loop anomaly to the Wilson network in Fig. 24. Gluons are attached to two of the outgoing Wilson lines, in this case the ones labeled by  $V_4$  and  $V_6$ .

The full anomaly comes from a sum of such diagrams, with gluons attached to any two distinct Wilson lines.

For a fixed pair of Wilson lines, we have already done essentially this computation in section 6: it is the same as the calculation leading to the framing anomaly. There are a few small differences. First, in the calculation we are now considering, both Wilson lines are labelled with arrows pointing away from the vertex. At the classical level, this configuration is the same as one where one arrow is incoming, and the other is outgoing, but we use the representations  $V_i$  and  $V_j^*$  instead of  $V_i$  and  $V_j$ .

Also, in our calculation of the framing anomaly, we labelled both segments of the line by a single representation of the group, but now the two representations may be different. This affects only the group theory factor, and even that factor can still be written similarly to what we had before.

Finally, in our previous analysis, the Wilson line was allowed to bend in an arbitrary fashion, but now the tangent direction is changing only by a delta function jump at the origin.

Taking these details into account, we can readily write down the anomaly associated to Fig. 25, where the two gluons attach to Wilson lines  $i$  and  $j$ . To get the full anomaly, we will have to sum over pairs  $i, j$ . Let

$$(7.9) \quad v \in V_1 \otimes \cdots \otimes V_n$$

denote the invariant tensor that we place at the vertex classically. The group theory factor associated to the diagram in which the gluons are attached to lines  $i$  and  $j$  is

$$(7.10) \quad \sum f_a^{bc} t_{b;i} t_{c;j} v.$$

(This reduces to eqn. (6.1) if  $v$  is an invariant bilinear form that establishes an isomorphism between  $V_i$  and  $V_j^*$ .)

Let  $\theta_i$  denote the angle between the  $i^{\text{th}}$  Wilson line and the vertical, measured in the clockwise direction. Let us assume that  $\theta_i < \theta_j$  for  $i < j$ . This means that  $V_1$  is the first line segment we encounter if we start at the vertical and move clockwise. Then, the anomaly associated to the  $i$  and  $j$  line segments with  $i < j$  is proportional to  $\theta_j - \theta_i - \pi$ . This follows from our calculation of the framing anomaly, which was the special case of this calculation when the representations  $V_j$  was  $V_i^*$ . In that case, a straight line does not have an anomaly, and the anomaly is proportional to the deviation of the line from being straight. This deviation for the geometry we are considering now is  $\theta_j - \theta_i - \pi$ .

Thus, the total anomaly to the existence of the vertex at the quantum level, summing all pairs of Wilson lines and otherwise borrowing our previous result, is

$$(7.11) \quad \frac{\hbar}{2\pi} \partial_z c^a \left( \sum_{1 \leq i < j \leq n} (\theta_j - \theta_i - \pi) f_a^{bc} t_{b;i} t_{c;j} \right) v.$$

This can be viewed as a  $G$ -invariant linear map from  $\mathfrak{g}$  to  $V_1 \otimes \cdots \otimes V_n$ .

A possibly counter-intuitive feature of the anomaly is that it is discontinuous in the angles  $\theta_i$ . If  $\theta_i$  increases so that it crosses  $\theta_{i+1}$ , there is a discontinuity in the anomaly of the form  $-\hbar f_a^{bc} t_{b;i} t_{c;i+1} v$  (an exchange of  $i$  and  $i+1$ . This can be explained as follows. When  $\theta_i = \theta_{i+1}$ , what we are computing is the anomaly for a system with  $n-1$  Wilson lines, where the lines  $V_i$  and  $V_{i+1}$  have been fused to a single Wilson line. In section 5, we have calculated the  $\mathcal{O}(\hbar)$  correction to the action  $\mathfrak{g}[[z]]$  on the fusion of two Wilson lines. The level 1 generator acting on the fused Wilson line turned out to be

$$(7.12) \quad \mp \frac{1}{2} \hbar f_a^{bc} t_{b;i} t_{c;i+1},$$

where the sign depends on whether  $V_i$  is brought to  $V_{i+1}$  from the left or the right. Thus the configurations with  $V_i$  approaching  $V_{i+1}$  from left or right are different, and therefore they can have different anomalies.

The above formula for the fusion of  $V_i$  with  $V_{i+1}$  shows that the anomaly for  $n$  Wilson lines with  $\theta_i$  tending to  $\theta_{i+1}$  from below coincides with the anomaly for  $n-1$  Wilson lines where  $V_i$  and  $V_{i+1}$  have been fused, with  $V_i$  merging with  $V_{i+1}$  from the right. (The fused Wilson line has a level one Yangian generator. A vertex in which it couples to ordinary Wilson lines that lack such a generator is not gauge-invariant at the classical level, and this contributes to the anomaly for such a vertex.)

Similarly, the anomaly for  $n$  Wilson lines where  $\theta_i$  tends to  $\theta_{i+1}$  from above coincides with the anomaly for  $n-1$  Wilson lines where  $V_i$  has been fused with  $V_{i+1}$  from the left.

Thus the fact that the anomaly has a discontinuity when  $\theta_i$  crosses  $\theta_{i+1}$  reflects the fact that the algebra of Wilson line operators is non-commutative: the result of fusing two parallel Wilson lines depends on whether they are brought together from the left or the right.

## 7.4 Cancelling the Anomaly

There are two ways to try to cancel this anomaly. We can shift the relative positions of the Wilson lines in the  $z$ -plane by an amount of order  $\hbar$ . Suppose we shift the  $i^{\text{th}}$  Wilson line from  $z_i = 0$  to  $z_i = b_i \hbar$ , for some constants  $b_i$ . This introduces a new anomaly  $(\sum b_i \hbar t_{a;i}) v$ , and one can hope that with a judicious choice of  $b_i$  this will cancel the anomaly.

The other thing we can try to do is to change the angles of the Wilson lines. However, it turns out that moving the angle of the  $i^{\text{th}}$  Wilson line has the same effect as shifting it in the  $z$  plane. This is clear from the framing anomaly, but we can derive it from

eqn. (7.11). Suppose that we change  $\theta_i$  to  $\theta_i + w_i$  for some fixed value of  $i$ , with  $w_i$  a constant, leaving the other Wilson lines unchanged. Then, the anomaly shifts by

$$(7.13) \quad \frac{\hbar}{2\pi} \left( + \sum_{j<i} w_i f_a^{bc} t_{b;j} t_{c;i} - \sum_{i<j} w_i f_a^{bc} t_{b;i} t_{c;j} \right) v.$$

Using the fact that  $\sum_j t_{j;a} v = 0$ , we can rewrite this as

$$(7.14) \quad \frac{\hbar}{2\pi} \left( w_i f_a^{bc} t_{b;i} t_{c;i} \right) v = \frac{\hbar h^\vee}{\pi} w_i t_{a;i} v.$$

This is the same shift we find if we move the  $i^{\text{th}}$  Wilson line by  $w_i$  in the  $z$  plane, up to a factor of the dual Coxeter number  $h^\vee$  which arises from the framing anomaly.<sup>28</sup>

Comparing the two formulas makes it clear that if it is possible to cancel the anomaly by adjusting the angles, then it is also possible to do so by shifting the relative positions in the  $z$ -plane. The converse is not quite true. In our computation of the anomaly, we have made an assumption about the cyclic order of the Wilson lines in the  $xy$  plane. This assumption only allows us to vary the angles while preserving certain inequalities. In some cases, the anomaly can be canceled by shifting the  $z_i$  but not by varying the angles, because the relevant inequalities would be violated.

## 7.5 Indecomposable Vertices

Before discussing specific examples, let us discuss a useful condition that can be satisfied by a classical vertex.

For each  $i$ , the states  $t_{a;i} v$  transform in the adjoint representation of  $G$ , since  $v$  is  $G$ -invariant. This gives  $n$  copies of the adjoint representation, but  $G$ -invariance of  $v$  is equivalent to one relation between them:

$$(7.15) \quad \sum_{i=1}^n t_{a;i} v = 0.$$

Thus, these states form at most  $n-1$  linearly independent copies of the adjoint representation. We will call a vertex *indecomposable* if there are no further relations between the states  $t_{a;i} v$ , so that the number of linearly independent copies of the adjoint representation made in this way is precisely  $n-1$ . The motivation for the terminology is that if there is an additional relation, one can show by general considerations of group theory that we can decompose the set of representations into disjoint subsets  $R$  and  $S$  with the feature that the operators  $\sum_{r \in R} t_{a;r}$  and  $\sum_{s \in S} t_{a;s}$  each

<sup>28</sup> If  $h^\vee = 0$ , then shifting the angles has no effect and we have to try to cancel the anomaly by shifting the values of  $z$ . For simple Lie groups,  $h^\vee$  is always positive, but there actually are interesting solutions of the Yang-Baxter equation for supergroups with  $h^\vee = 0$  [33].

annihilate the element  $v$ . This means that  $v$  is the tensor product (or a sum of tensor products) of invariant elements in  $\otimes_{r \in R} V_r$  and in  $\otimes_{s \in S} V_s$ . Thus, our vertex decomposes as a tensor product of two vertices, which can be shifted to different values of  $z$  and analyzed independently.

Accordingly, it is reasonable to restrict our attention to indecomposable vertices. An indecomposable vertex has at least  $n - 1$  possible anomalies (since the  $t_{a,i}v$  provide  $n - 1$  copies of the adjoint representation in  $\widehat{V} = V_1 \otimes V_2 \otimes \dots \otimes V_n$ ). If there are only these  $n - 1$  linearly independent copies of the adjoint in  $\widehat{V}$ , then by shifting the  $n - 1$  relative positions of the  $n$  Wilson lines in the  $z$ -plane, there is a unique way to cancel the anomaly. Indeed, anomalies of the form  $t_{a,i}V$  are precisely the ones that can be eliminated by shifting the values of  $z$ . If certain inequalities are obeyed, the anomalies can also be cancelled by shifting the relative angles rather than the values of  $z$ .

If the number of copies of the adjoint representation in  $\widehat{V}$  is actually greater than  $n - 1$ , one would expect that generically anomaly cancellation is not possible.

## 7.6 Vertices Constrained by Symmetries

We will describe various concrete examples of anomaly-free quantum vertices.<sup>29</sup> The simplest examples, which do not require any computations, arise when there are enough symmetry constraints to determine the angles between the Wilson lines and ensure that the anomaly vanishes.

For a simple and also useful case, suppose that  $V_1, \dots, V_n$  are all the same representation  $V$ , and consider a vertex associated to a  $G$ -invariant vector  $v \in V^{\otimes n}$  that is either cyclically invariant or cyclically anti-invariant (in other words, assume that it either is invariant or changes sign under a cyclic permutation of the  $n$  copies of  $V$ ). Such a vertex is anomaly-free, assuming that the angles between successive Wilson lines are equal, so as to respect the symmetry. (Fig. 24 has been drawn with equal angles, so it is invariant under a  $2\pi/n$  rotation and potentially represents a vertex with cyclic symmetry or antisymmetry.) The proof is simple. Since the anomaly depends linearly on  $v$ , it is cyclically invariant or anti-invariant if  $v$  is. By assumption, there are no cyclically invariant or anti-invariant copies of the adjoint representation in  $\widehat{V}$ , so the anomaly is zero.

In most cases, a vertex with cyclic symmetry or antisymmetry has a further symmetry. This happens because a cyclically symmetric configuration of  $n$  lines meeting at equal angles in the plane, as in

<sup>29</sup> These detailed examples will not be needed in the rest of the present paper, though some of them play an important role in the companion paper [25].

Fig. 24, is actually invariant under suitable reflections of the plane. In the gauge theory under study in the present paper, a reflection of the  $xy$  plane is a symmetry if accompanied by  $z \rightarrow -z$ . The latter is also a symmetry of ordinary Wilson lines (associated to representations of  $\mathfrak{g}$ , not  $\mathfrak{g}[[z]]$ ) that are supported at  $z = 0$ . So as long as the vector  $v \in V^{\otimes n}$  used in constructing the vertex is either even or odd under the reflection symmetry (all examples we will consider will have this property), the corresponding cyclically symmetric classical vertex actually has dihedral symmetry, generated by  $2\pi/n$  rotations and also certain reflections. If the vertex is anomaly-free because of the conditions stated in the last paragraph, it will automatically possess the dihedral symmetry.

We will now describe explicit examples of anomaly-free vertices that are cyclically invariant or anti-invariant (and thus also dihedrally invariant). In each case, we have to first make sure that the representation  $V$  that we want to use is itself anomaly-free, in other words that there is a quantum Wilson line in this representation. For this, we will use the criterion ( $\dagger$ ) that was stated at the end of section 7.1: any  $G$ -invariant map from  $\wedge^2 \mathfrak{g}$  to  $\text{End}(V) = V \otimes V^*$  factors through the adjoint representation, or equivalently the only irreducible representation of  $\mathfrak{g}$  that appears both in  $\wedge^2 \mathfrak{g}$  and in  $V \otimes V^*$  is the adjoint representation.

Here are some examples:

1.  $V$  is the fundamental representation of  $SL_n$ , which satisfies condition ( $\dagger$ ), and  $v \in V^{\otimes n}$  is the essentially unique<sup>30</sup> invariant tensor. There are  $n - 1$  copies of the adjoint in  $V^{\otimes n}$ , so this vertex can be quantized. Since  $v$  is cyclically invariant or anti-invariant (depending on whether  $n$  is odd or even), the vertex can be quantized so the Wilson lines all have the same value of  $z$  and the angle between them is  $2\pi/n$ .
2. Take  $V$  to be the  $\mathbf{7}$  of  $G_2$ , which satisfies condition ( $\dagger$ ), and take

$$v \in \wedge^3 \mathbf{7} \subset \mathbf{7}^{\otimes 3}$$

to be the essentially unique invariant tensor. Using the tables on p. 298 of [34], one finds that every map from  $\wedge^2 \mathbf{14}$  to  $\mathbf{7} \otimes \mathbf{7}$  factors through  $\mathbf{14}$ . This implies that the Wilson line associated to  $\mathbf{7}$  can be quantized. Further, the adjoint appears precisely twice in  $\mathbf{7}^{\otimes 3}$ . Therefore, the vertex associated to the tensor  $v$  can be quantized with angles  $2\pi/3$  between the Wilson lines.

3. Consider the three 8 dimensional representations  $\mathbf{8}_v, \mathbf{8}_c, \mathbf{8}_s$  of  $\text{Spin}(8)$ , which are permuted by triality. For each representation, criterion ( $\dagger$ )

<sup>30</sup> In such a statement, we always mean unique up to a constant multiple.

holds, so that there are no anomalies to quantizing the corresponding Wilson lines. Let

$$v \in \mathbf{8}_v \otimes \mathbf{8}_c \otimes \mathbf{8}_s$$

be the essentially unique invariant tensor. Note that  $v$  is also invariant under triality, together with a cyclic permutation of the representations. There are only two copies of the adjoint in  $\mathbf{8}_v \otimes \mathbf{8}_c \otimes \mathbf{8}_s$ , so this vertex can be quantized. If the Wilson lines have relative angles  $2\pi/3$ , then a rotation through  $2\pi/3$  together with an application of the triality symmetry of Spin(8) is a symmetry of the configuration. This tells us that the only consistent quantization is the one where the Wilson lines have relative angles  $2\pi/3$ .

4. Take  $V$  to be the **26** of  $F_4$ . Using table 45 of [35], we find that this representation satisfies condition  $(\dagger)$  and so the corresponding Wilson line can be quantized. We consider the vertex associated to the essentially unique invariant tensor  $v \in \text{Sym}^3 \mathbf{26}$ . Again using table 45 of [35], one can check that there are only two copies of the adjoint in  $\mathbf{26}^{\otimes 3}$ . Therefore this vertex quantizes, with angles  $2\pi/3$  between adjacent lines.
5. Take  $V$  to be the **27** of  $E_6$ . Table 48 of [35] implies that this representation satisfies condition  $(\dagger)$ . (See section 7.10.) We take  $v \in \text{Sym}^3 \mathbf{27}$  to be the unique invariant element. Again using table 48 of [35], we find that there are only two copies of the adjoint in  $\mathbf{27}^{\otimes 3}$ . Therefore this vertex quantizes, with angles  $2\pi/3$  between adjacent lines.

## 7.7 A General Formula for the Angles at a Trivalent Vertex

Now we turn our attention to general trivalent vertices. Assuming that only two copies of the adjoint occur in  $\widehat{V} = V_1 \otimes V_2 \otimes V_3$ , a classical trivalent vertex can be quantized. Moreover, it is possible to find a simple general formula for the relative angles that are required. (We will express our results in terms of angles rather than in terms of shifts in  $z$  because that corresponds to a simpler classical picture, but when some relative angles come out to be negative and thus inconsistent with an assumed cyclic ordering of the vertices, one can use the alternative approach in terms of shifting the relative values of the  $z_i$ .)

Suppose that  $V_{\rho_1}, V_{\rho_2}, V_{\rho_3}$  are three irreducible highest weight representations of a group  $G$  of highest weights  $\rho_1, \rho_2, \rho_3$ , and suppose that the corresponding Wilson lines can be quantized. Consider a classical vertex associated to an invariant tensor  $v \in V_{\rho_1} \otimes V_{\rho_2} \otimes V_{\rho_3}$ . As we have seen, if there are exactly two copies of the adjoint representation in  $V_{\rho_1} \otimes V_{\rho_2} \otimes V_{\rho_3}$ , then this vertex can be quantized.

Suppose the Wilson lines are arranged in the plane with cyclic order  $V_{\rho_1}, V_{\rho_2}, V_{\rho_3}$ . Let  $\theta_{12}, \theta_{23}, \theta_{31}$  be the angles between the Wilson lines. We will derive a formula for the angles  $\theta_{ij}$ .

Let  $c(\rho_i)$  denote the action of the quadratic Casimir of  $\mathfrak{g}$  on  $V_{\rho_i}$ . We define

$$(7.16) \quad \begin{aligned} \beta_1 &= c(\rho_1) - c(\rho_2) - c(\rho_3), \\ \beta_2 &= c(\rho_2) - c(\rho_3) - c(\rho_1), \\ \beta_3 &= c(\rho_3) - c(\rho_1) - c(\rho_2). \end{aligned}$$

We will show that the angles between the three Wilson lines are given by the formula

$$(7.17) \quad \begin{aligned} \theta_{12} &= \pi - \pi \frac{\beta_1 \beta_2}{\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3}, \\ \theta_{23} &= \pi - \pi \frac{\beta_2 \beta_3}{\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3}, \\ \theta_{31} &= \pi - \pi \frac{\beta_1 \beta_3}{\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3}. \end{aligned}$$

The derivation of this formula is as follows. The anomaly to quantizing the vertex vanishes if

$$(7.18) \quad \begin{aligned} &((\theta_{12} - \pi) f_a^{bc} t_{b;1} t_{c;2} + (\pi - \theta_{31}) f_a^{bc} t_{b;1} t_{c;3} \\ &+ (\theta_{23} - \pi) f_a^{bc} t_{b;2} t_{c;3}) v = 0. \end{aligned}$$

This equation should hold for every value of  $a$ .

Let us choose our basis  $t_a$  to be orthonormal with respect to the chosen invariant pairing on the Lie algebra  $\mathfrak{g}$ . Applying the operator  $t_{a;1}$  to eqn. (7.18) and summing over  $a$  we find

$$(7.19) \quad \begin{aligned} &((\theta_{12} - \pi) f_{bc}^a t_{a;1} t_{b;1} t_{c;2} + (\pi - \theta_{31}) f_{bc}^a t_{a;1} t_{b;1} t_{c;3} \\ &+ (\theta_{23} - \pi) f_{bc}^a t_{a;1} t_{b;2} t_{c;3}) v = 0. \end{aligned}$$

Since

$$(7.20) \quad \sum_a f_{bc}^a t_{a;1} t_{b;1} = 2h^\vee t_{c;1},$$

we can rewrite this equation as

$$(7.21) \quad \begin{aligned} &((\theta_{12} - \pi) h^\vee t_{c;1} t_{c;2} + (\pi - \theta_{31}) h^\vee t_{c;1} t_{c;3} \\ &+ (\theta_{23} - \pi) f_{bc}^a t_{a;1} t_{b;2} t_{c;3}) v = 0. \end{aligned}$$

Next, since

$$(7.22) \quad t_{c;3} v = -t_{c;1} v - t_{c;2} v,$$

we have

$$(7.23) \quad \begin{aligned} f_{abc} t_{a;1} t_{b;2} t_{c;3} v &= -f_{abc} t_{a;1} t_{b;2} (t_{c;1} + t_{c;2}) \\ &= f_{acb} t_{a;1} t_{c;1} t_{b;2} - f_{bca} t_{a;1} t_{b;2} t_{c;2} \\ &= t_{b;1} t_{b;2} h^\vee - t_{a;1} t_{a;2} h^\vee = 0. \end{aligned}$$

Thus, our equation becomes

$$(7.24) \quad ((\theta_{12} - \pi) h^\vee t_{c;1} t_{c;2} + (\pi - \theta_{31}) h^\vee t_{c;1} t_{c;3}) v = 0.$$

Now,

$$(7.25) \quad \sum_c t_{c;1} t_{c;2} = \sum_c \frac{1}{2} (t_{c;1} + t_{c;2})^2 - \frac{1}{2} t_{c;1}^2 - \frac{1}{2} t_{c;2}^2.$$

Let  $c(\rho_i)$  denote the eigenvalue of the quadratic Casimir  $\sum_a t_a^2$  on the representation  $V_{\rho_i}$ . The operator  $\sum_a t_{a;i}^2$  acts on  $V_{\rho_1} \otimes V_{\rho_2} \otimes V_{\rho_3}$  by  $c(\rho_i)$ .

Acting on an element  $v$ , we have

$$(7.26) \quad \begin{aligned} \sum_c t_{c;1} t_{c;2} v &= \sum_c \frac{1}{2} (t_{c;1} + t_{c;2})^2 v - \frac{1}{2} t_{c;1}^2 v - \frac{1}{2} t_{c;2}^2 v \\ &= \frac{1}{2} t_{c;3}^2 v - \frac{1}{2} t_{c;1}^2 v - \frac{1}{2} t_{c;2}^2 v \\ &= \frac{1}{2} c(\rho_3) v - \frac{1}{2} c(\rho_1) v - \frac{1}{2} c(\rho_2) v. \end{aligned}$$

Thus (after dividing by  $\hbar^\vee/2$ ), our equation becomes

$$(7.27) \quad \begin{aligned} (\theta_{12} - \pi)(c(\rho_3) - c(\rho_1) - c(\rho_2)) \\ = (\theta_{31} - \pi)(c(\rho_2) - c(\rho_1) - c(\rho_3)). \end{aligned}$$

Similar arguments give us two additional equations, which are the cyclic permutations of the equation we have just derived:

$$(7.28) \quad \begin{aligned} (\theta_{12} - \pi)(c(\rho_3) - c(\rho_1) - c(\rho_2)) \\ = (\theta_{23} - \pi)(c(\rho_1) - c(\rho_2) - c(\rho_3)), \\ (\theta_{12} - \pi)(c(\rho_3) - c(\rho_1) - c(\rho_2)) \\ = (\theta_{31} - \pi)(c(\rho_2) - c(\rho_1) - c(\rho_3)). \end{aligned}$$

Evidently, one of these equations is redundant.

In terms of  $\beta_i$  defined in eqn. (7.16), our equations become

$$(7.29) \quad (\theta_{12} - \pi)\beta_3 = (\theta_{23} - \pi)\beta_1$$

plus its cyclic permutations. Their unique solution is given by (7.17).

As a consistency check, note that

$$\theta_{12} + \theta_{23} + \theta_{31} = 2\pi.$$

Note also that if the three representations are the same, then the angle between any two Wilson lines is  $2\pi/3$ , as expected for symmetry reasons.

Let us specialize the formula to the case that the two representations  $V_{\rho_1}, V_{\rho_2}$  are the same (or at least have the same eigenvalue of the quadratic Casimir). In that case, letting  $c(\rho) = c(\rho_1) = c(\rho_2)$ , we have

$$(7.30) \quad \begin{aligned} \beta_1 = \beta_2 = -c(\rho_3), \\ \beta_3 = c(\rho_3) - 2c(\rho). \end{aligned}$$

The angles between the Wilson lines become

$$\theta_{12} = \pi - \pi \frac{c(\rho_3)}{4c(\rho) - c(\rho_3)},$$

$$(7.31) \quad \begin{aligned} \theta_{31} &= \pi \frac{2c(\rho)}{4c(\rho) - c(\rho_3)}, \\ \theta_{23} &= \pi \frac{2c(\rho)}{4c(\rho) - c(\rho_3)}. \end{aligned}$$

As an example, let analyze the vertex connecting two copies of the fundamental representation of  $\mathfrak{sl}_n$  with the dual of the exterior square of the fundamental representation. Let us normalize the quadratic Casimir so that its value on the fundamental representation  $V$  is 1 (the normalization plays no role in our formula). Then, its value on  $\wedge^2 V^*$  is  $\frac{2(n-2)}{n-1}$ . If we take  $V_{\rho_1}, V_{\rho_2}$  to be the two copies of the fundamental representation in the above calculation, then we have

$$(7.32) \quad \begin{aligned} \theta_{12} &= \pi - \pi \frac{\frac{2(n-2)}{n-1}}{4 - \frac{2(n-2)}{n-1}} = \pi \frac{2}{n}, \\ \theta_{31} &= \pi \frac{2}{4 - \frac{2(n-2)}{n-1}} = \pi \frac{n-1}{n}, \\ \theta_{23} &= \pi \frac{n-1}{n}. \end{aligned}$$

This is a special case of a more general formula that we compute next.

## 7.8 Trivalent Vertices Linking Fundamental Representations of $\mathfrak{sl}_n$

We will describe trivalent vertices involving three of the fundamental representations of  $\mathfrak{sl}_n$ , which are the  $k^{\text{th}}$  rank antisymmetric tensors  $\wedge^k V$ , where  $V$  is the fundamental representation and  $1 \leq k \leq n-1$ . These representations are all quantizable (since in fact all representations of  $\mathfrak{sl}_n$  are quantizable), though criterion (†) generally does not hold.

In what follows, it is useful to recall that, in an appropriate normalization, the eigenvalue  $c(V)$  of the quadratic Casimir on a representation  $V = V_\rho$  satisfies

$$(7.33) \quad c(V) = \frac{l(v)}{\dim(V)},$$

where  $l(v)$  is the Dynkin index of the representation. Dynkin indices of various representations can be found in tables such as those in [35].

Let  $V$  be the vector representation of  $\mathfrak{sl}_n$  so that  $\wedge^k V$ ,  $k = 1, \dots, n-1$  are the fundamental representations. The Dynkin index of  $\wedge^k V$  is  $\binom{n-2}{k-1}$ , so that

$$(7.34) \quad c(\wedge^k V) = \frac{\binom{n-2}{k-1}}{\binom{n}{k}} = \frac{k(n-k)}{n(n-1)}.$$

Consider three fundamental representations  $\wedge^{k_i} V$  where  $k_1 + k_2 + k_3 = n$ . There is an invariant element of the tensor product of these representations coming from the map

$$(7.35) \quad \wedge^{k_1} V \otimes \wedge^{k_2} V \otimes \wedge^{k_3} V \rightarrow \wedge^n V = \mathbb{C}.$$



The adjoint representation only appears twice in the tensor product of the three representations  $\wedge^k V$ . It follows that the vertex corresponding to the invariant tensor can be quantized. From formulae (7.17), we can determine the angles.

We set

$$(7.36) \quad \beta_1 = c(\wedge^{k_1} V) - c(\wedge^{k_2} V) - c(\wedge^{k_3} V)$$

plus cyclic permutation. Then

$$\beta_1 = \frac{1}{n(n-1)}(k_1(n-k_1) - k_2(n-k_2) - k_3(n-k_3)) = \frac{-2k_2k_3}{n(n-1)}$$

plus cyclic permutations. Bearing in mind that the a scaling of all the  $\beta_i$  will not affect the angles, we find

$$(7.37) \quad \begin{aligned} \theta_{12} &= \pi \frac{k_1 + k_2}{n}, \\ \theta_{23} &= \pi \frac{k_2 + k_3}{n}, \\ \theta_{31} &= \pi \frac{k_1 + k_3}{n}. \end{aligned}$$

## 7.9 Vertices Related to $\Gamma$ -Matrices

Consider the group  $\text{Spin}(n)$  where  $n$  is even. Let  $V$  denote the vector representation, and  $S_+$ ,  $S_-$  the two irreducible (complex) spin representations of opposite chirality. These are both of dimension  $2^{\frac{n}{2}-1}$ . We will use criterion  $(\dagger)$  to show that these representations can be quantized.

If  $n = 2 \pmod 4$ , then the vector representation appears in  $S_+ \otimes S_+$  and in  $S_- \otimes S_-$ . If  $n = 0 \pmod 4$ , then the vector representation appears in  $S_+ \otimes S_-$ . In each case, we can try to quantize the vertex linking the vector representation to two spin representations.

In the calculation that follows, we will assume  $n \geq 8$  to avoid any low-dimensional coincidences.

Note that the endomorphisms of the direct sum  $S_+ \oplus S_-$  make up the Clifford algebra  $\text{Cl}_n$  built from the vector representation  $V$ . As a representation of  $\mathfrak{so}(n)$ , the Clifford algebra is isomorphic to direct sum of the exterior powers of  $V$ .

To use condition  $(\dagger)$  to show that the spinor representations can be quantized, we need to classify maps

$$(7.38) \quad \wedge^2 \mathfrak{so}(n) \rightarrow \wedge^k V$$

for all values of  $k$ . An exercise in classical invariant theory tells us that the only such maps that can possibly exist are when  $k = 2, 4, n-4, n-2$ . If  $k = 2, n-2$  then  $\wedge^k V$  is the adjoint representation, so such maps will not contribute anomalies. If  $k = 4, n-4$  then the only map  $\mathfrak{so}(n)^{\otimes 2} \rightarrow \wedge^k V$  comes from the wedge product map

$$(7.39) \quad \mathfrak{so}(n) \otimes \mathfrak{so}(n) = \wedge^2 V \otimes \wedge^2 V \rightarrow \wedge^4 V$$

(noting that  $\wedge^4 V = \wedge^{n-4} V$ ). This map is symmetric, not antisymmetric, and so is not relevant to condition  $(\dagger)$ . We conclude that in all cases, the Wilson lines attached to the spin representations exist in the quantum theory.

Next, let us analyze whether the vertex linking the vector representation  $V$  with two copies of a spin representation can be quantized. We will start with the case when  $n = 2 \pmod 4$ , in which case the vector representation appears in  $S_+ \otimes S_+$ . We need to show that there are exactly two copies of the adjoint representation in  $V \otimes S_+ \otimes S_+$ .

To do this, we need to recall how to describe  $S_+ \otimes S_+$  in terms of exterior powers of  $V$  (this computation is familiar from the study of central extensions of supersymmetry algebras). Since  $n = 2 \pmod 4$ ,  $S_+$  and  $S_-$  are dual representations. The endomorphisms of  $S_+ \oplus S_-$  which reverse the parity of a spinor are the odd elements of the Clifford algebra built from  $V$ . This space of endomorphisms is  $S_+ \otimes S_+ \oplus S_- \otimes S_-$ . Thus,

$$(7.40) \quad S_+ \otimes S_+ \oplus S_- \otimes S_- = \bigoplus_{k \text{ odd}} \wedge^k V.$$

We can further decompose the right hand side of this equation to find

$$(7.41) \quad S_+ \otimes S_+ = V \oplus \wedge^3 V \oplus \cdots \oplus \wedge_+^{n/2} V,$$

where  $\wedge_+^{n/2} V$  indicates those elements which are self-dual under the Hodge star operator.

In  $V \otimes S_+ \otimes S_+ = V \otimes (V \oplus \wedge^3 V \oplus \cdots \oplus \wedge_+^{n/2} V)$ , the only copies of the adjoint representation are those in  $V \otimes V$  and in  $V \otimes \wedge^3 V$ , each of which contains one copy of the adjoint. So overall there are precisely two copies, and the vertex connecting  $V$ ,  $S_+$  and  $S_+$  can be quantized.

Next let us check whether the vertex can be quantized in the case that  $n = 0 \pmod 4$ . In this case  $V$  appears in  $S_+ \otimes S_-$  and the representations  $S_+$ ,  $S_-$  are self-dual. We again find that  $S_+ \otimes S_- \oplus S_- \otimes S_+$  is the space of odd elements in the Clifford algebra, so that

$$(7.42) \quad S_+ \otimes S_- = V \oplus \wedge^3 V \oplus \cdots \oplus \wedge^{n/2-1} V.$$

The only copies of the adjoint that appear in  $V \otimes S_+ \otimes S_-$  are those in  $V \otimes V$  and  $V \otimes \wedge^3 V$ , so again the vertex can be quantized.

Now we know that the spinor representations and the vertices connecting spinor and vector representations can be quantized. The final step is to calculate the angles between the lines at the vertex. To do this, we need to know the Dynkin indices of the representations  $V, S_{\pm}$ . The Dynkin index of  $V$  is 2. We can calculate the Dynkin indices of  $S_{\pm}$  as follows.

Recall that the Dynkin index is additive under direct sums of representations, and under tensor product there is the following formula:

$$(7.43) \quad l(R_1 \otimes R_2) = l(R_1) \dim(R_2) + \dim(R_1) l(R_2),$$

where  $R_1, R_2$  are representations,  $l(R_i)$  is the Dynkin index and  $\dim(R_i)$  is the dimension. Further, if  $R$  is any representation, then

$$(7.44) \quad l(\wedge^k R) = \binom{\dim(R) - 2}{k - 1} l(R)$$

The representations  $S_+$  and  $S_-$  are related by a diagram automorphism of the Dynkin diagram  $D_{n/2}$  of the Lie algebra  $\mathfrak{so}(n)$ . Therefore they have the same Dynkin index. Using the fact that  $(S_+ \oplus S_-)^{\otimes 2}$  is the sum of all the exterior powers of  $V$ , we find

$$(7.45) \quad 2^{n/2+2} l(S_{\pm}) = 2 \sum_{k=1}^{n-1} \binom{n-2}{k-1} = 2^{n-1}.$$

Thus,

$$(7.46) \quad l(S_{\pm}) = 2^{\frac{n}{2}-3}.$$

Now let us compute the angles between the three Wilson lines. We label the lines where  $V_1 = S_+$ ,  $V_2 = S_-$  and  $V_3 = V$  is the vector representation. The quadratic Casimirs in each representation are the ratios of the Dynkin index to the dimension. They are

$$(7.47) \quad \begin{aligned} c(V) &= \frac{2}{n}, \\ c(S_{\pm}) &= \frac{2^{\frac{n}{2}-3}}{2^{\frac{n}{2}-1}} = 2^{-2}. \end{aligned}$$

According to formula (7.31), we have

$$(7.48) \quad \begin{aligned} \theta_{12} &= \pi \frac{n-4}{n-2}, \\ \theta_{31} &= \pi \frac{n}{2n-4}, \\ \theta_{23} &= \pi \frac{n}{2n-4}. \end{aligned}$$

Note that when  $n = 8$ , the angle between any two lines is  $2\pi/3$ , which is consistent with what we determined earlier using the triality symmetry of the vertex in this case. When  $n = 6$ , the vertex we are considering is that relating two copies of the vector representation of  $\mathfrak{sl}(4)$  with the exterior square of the fundamental. The formula for the angle for a vertex connecting three fundamental representations of  $\mathfrak{sl}_n$  agrees, in this case, with the formula given here.

### 7.10 A Vertex Connecting Representations of $E_6$

Just for fun, let us use our formula to calculate the angles in a vertex associated to representations of the exceptional group  $E_6$ . The fundamental representation of  $E_6$  will be denoted by  $\mathbf{27}$ , and its dual by  $\overline{\mathbf{27}}$ . There are four representations of dimension 351, which come in dual pairs. We will use the conventions of [35] and denote them by  $\mathbf{351}$ ,  $\overline{\mathbf{351}}$ ,  $\mathbf{351}'$ ,  $\overline{\mathbf{351}'}$ . The  $\mathbf{27}^3$  vertex was already considered in section 7.6, so here

we primarily consider a more elaborate example. (We also will complete the discussion of the  $\mathbf{27}^3$  by showing that the  $\mathbf{27}$  can be quantized.)

According to table 48 of [35],  $\overline{\mathbf{351}'}$  appears once in  $\mathbf{27} \otimes \mathbf{27}$ . Thus there is an invariant tensor in  $\mathbf{27} \otimes \mathbf{27} \otimes \mathbf{351}'$ .

We would like to quantize this to a vertex connecting three Wilson lines. To do this, we first need to show that the Wilson lines themselves quantize. It is sufficient, according to condition ( $\dagger$ ), to show that any map from the exterior square of the adjoint representation to the endomorphisms of the  $\mathbf{27}$  or  $\mathbf{351}'$  factors through the adjoint representation.

According to table 48 of [35], the exterior square of the adjoint representation decomposes as

$$(7.49) \quad \wedge^2 \mathbf{78} = \mathbf{78} \oplus \mathbf{2925}.$$

To show that the  $\mathbf{27}$  and  $\mathbf{351}'$  quantize, we need to show that  $\mathbf{2925}$  does not appear in  $\overline{\mathbf{27}} \otimes \mathbf{27}$  or in  $\overline{\mathbf{351}'}$   $\otimes$   $\mathbf{351}'$ . Table 48 of [35] shows that it does not, so these representations quantize.

Next, to show that the vertex in  $\mathbf{27} \otimes \mathbf{27} \otimes \mathbf{351}'$  quantizes, we need to show that the adjoint representation appears precisely twice in this tensor product. Table 48 of [35] tells us that

$$(7.50) \quad \mathbf{27} \otimes \mathbf{27} = \overline{\mathbf{27}} \oplus \mathbf{351} \oplus \mathbf{351}'.$$

If we tensor this with  $\mathbf{351}'$ , table 48 of [35] tells us that the adjoint appears once in  $\overline{\mathbf{351}'}$   $\otimes$   $\mathbf{351}'$ , once in  $\overline{\mathbf{351}}$   $\otimes$   $\mathbf{351}'$ , and not at all in  $\overline{\mathbf{27}}$   $\otimes$   $\mathbf{351}$ . Therefore the vertex quantizes.

Next, let us compute the angles. Let us label the representations as  $V_1 = \mathbf{27}$ ,  $V_2 = \overline{\mathbf{27}}$ ,  $V_3 = \mathbf{351}$ . Table 47 of [35] tells us that the Dynkin index of  $\mathbf{27}$  is 6 and that of  $\mathbf{351}'$  is  $6 \times 28$ . The values of the quadratic Casimirs are  $6/27$  and  $6 \times 28/351$ . We can change the normalization so that the values of the quadratic Casimirs are 1 and  $28/13$ . The angles are

$$(7.51) \quad \begin{aligned} \theta_{12} &= \pi - \pi \frac{28}{13} \frac{1}{4 - \frac{28}{13}} = -\pi \frac{1}{6}, \\ \theta_{23} = \theta_{31} &= \pi \frac{2}{4 - \frac{28}{13}} = \pi \frac{13}{12}. \end{aligned}$$

Since, in this example,  $\theta_{12} < 0$ , we have to shift the Wilson lines in the  $z$ -plane instead of just placing them at angles in the topological plane.

Many more examples can be analyzed in a similar way.

### 7.11 A 4-Valent Vertex for the $\mathbf{56}$ of $E_7$

The smallest representation of  $E_7$  is the  $\mathbf{56}$ . It is a pseudoreal or symplectic representation, so there

is an invariant antisymmetric form  $\omega \in \wedge^2 \mathbf{56}$ . In addition, there is a completely symmetric quartic invariant  $\psi \in \text{Sym}^4 \mathbf{56}$ . It is natural to ask whether this vertex can be quantized, like the  $\mathbf{27}^3$  of  $E_6$ , which was one of our examples in section 7.6. The answer is that it can, though the proof is not as simple as for the  $\mathbf{27}^3$  vertex. Both of these examples will be useful in [25].

A configuration of four Wilson lines cannot have full  $S_4$  permutation symmetry. The maximum possible symmetry is a dihedral subgroup  $D_4$ . This is simply the symmetry group of four Wilson lines with equal relative angles  $\pi/2$ , say running along the  $\pm x$  and  $\pm y$  axes. Dihedral symmetry was discussed in section 7.6. In this case, the group  $D_4$  is generated by rotations of the  $xy$  plane by an angle  $\pi/2$ , along with a reflection that preserves the given configuration of Wilson lines. Thus in all  $D_4$  has eight elements.

We would like to understand the possible vertices that are  $D_4$  invariant, and the possible anomalies that are compatible with the  $D_4$  symmetry. Let us first enumerate the  $E_7$ -invariant elements in  $\mathbf{56}^{\otimes 4}$  that are also  $D_4$ -invariant. Note that we can identify  $\mathbf{56}$  with its dual, using the  $E_7$ -invariant symplectic form  $\omega$ . We can therefore identify  $E_7$ -invariant elements of  $\mathbf{56}^{\otimes 4}$  with maps of  $E_7$  representations

$$(7.52) \quad \mathbf{56}^{\otimes 2} \rightarrow \mathbf{56}^{\otimes 2}.$$

According to the tables of [35] or [36],  $\mathbf{56}^{\otimes 2}$  decomposes as a sum of 4 distinct irreducible representations. Therefore, there are 4  $E_7$  invariant linear operators on  $\mathbf{56}^{\otimes 2}$ , given by the projectors onto these 4 irreducible subrepresentations. Correspondingly, there are 4 invariant tensors in  $\mathbf{56}^{\otimes 4}$ .

We can enumerate these tensors as follows. One of them is the completely symmetric invariant  $\psi$  with which we began. The other three are more elementary. Let  $e_i$  be a basis of  $\mathbf{56}$ ; in this basis, the antisymmetric form  $\omega$  corresponds to a matrix  $\omega^{ij}$ . The other three  $E_7$  invariants in  $\mathbf{56}^{\otimes 4}$  are given by the formulas

$$(7.53) \quad \begin{aligned} &\omega^{ij} \omega^{kl} e_i \otimes e_j \otimes e_k \otimes e_l, \\ &\omega^{ik} \omega^{lj} e_i \otimes e_j \otimes e_k \otimes e_l, \\ &\omega^{il} \omega^{jk} e_i \otimes e_j \otimes e_k \otimes e_l. \end{aligned}$$

Among these three tensors, there is a single linear combination which is invariant under  $D_4$ , namely

$$(7.54) \quad (\omega^{ij} \omega^{kl} + \omega^{jk} \omega^{li}) e_i \otimes e_j \otimes e_k \otimes e_l.$$

We conclude that there are a total of two dihedrally invariant tensors in  $\mathbf{56}^{\otimes 4}$ .

Given the explicit form of the invariant (7.54), an elementary computation using the general formula (7.11) for the anomaly shows that a vertex constructed using this invariant has an anomaly.

This means that it will be possible to use this invariant as a counterterm to help in canceling an anomaly.

We will prove that there is a unique linear combination of these two dihedrally invariant tensors which quantizes to a vertex linking the Wilson lines. To show this, we have to show that there is precisely one possible anomaly.

#### 7.11.1 Anomalies

Anomalies are proportional to  $\partial_z c$ . Since an element of  $D_4$  that acts as a reflection of the  $xy$  plane also acts as  $z \rightarrow -z$ , changing the sign of  $\partial_z c$ , the group theory invariant that multiplies  $\partial_z c$  in an anomaly is not  $D_4$ -invariant. Rather, it is  $D_4$  anti-invariant, that is, invariant under rotations in  $D_4$  but odd under reflections. We will show that the adjoint representation of  $E_7$  (which is the  $\mathbf{133}$ ) occurs precisely once in the  $D_4$  anti-invariant part of  $\mathbf{56}^{\otimes 4}$ . Since the invariant (7.54) does have a nonzero anomaly, this implies that by adding a multiple of (7.54) to the invariant  $\psi$ , one can construct an anomaly-free  $\mathbf{56}^{\otimes 4}$  vertex.

As a first step, let us compute the part of  $\mathbf{56}^{\otimes 4}$  that is anti-invariant under a dihedral subgroup  $D_2 \subset D_4$ . There actually are two possible embeddings of  $D_2$  in  $D_4$ . We make the following choice. If the four Wilson lines run along the  $\pm x$  and  $\pm y$  axes, we consider a subgroup  $D_2 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$  that is generated by a reflection that acts by  $(x, y) \rightarrow (x, -y)$  and one that acts as  $(x, y) \rightarrow (-x, y)$ . Thus if we label the four Wilson lines in cyclic order as 1, 2, 3, and 4, one reflection acts by exchanging 1 and 3, keeping fixed 2 and 4, and the other exchanges 2 and 4, keeping fixed 1 and 3. Thus, the part of  $\mathbf{56}^{\otimes 4}$  that is odd under each reflection is  $\wedge^2 \mathbf{56} \otimes \wedge^2 \mathbf{56}$ . The two factors are associated to the pairs 13 and 24.

From the tables in [35] or [36], one has  $\wedge^2 \mathbf{56} \cong \mathbf{1} \oplus \mathbf{1539}$ . So the  $D_2$  anti-invariant part of  $\mathbf{56}^{\otimes 4}$  is  $(\mathbf{1} \oplus \mathbf{1539}) \otimes (\mathbf{1} \oplus \mathbf{1539})$ .

Now we want to identify the part of this that is anti-invariant under  $D_4$ , not just under  $D_2$ . So we have to consider the action of a  $\pi/2$  rotation. The  $D_4$  anti-invariants are simply the  $D_2$  anti-invariants that are invariant under a  $\pi/2$  rotation. However, there is a small surprise when we try to impose invariance under a  $\pi/2$  rotation on the above description of the  $D_2$  anti-invariants.

A  $\pi/2$  rotation exchanges the two factors of  $\wedge^2 \mathbf{56}$  that we used in the above analysis, but with an important minus sign. This happens as follows. We recall that the two factors of  $\wedge^2 \mathbf{56}$  are associated respectively to the pair of Wilson lines 13 and 24. A  $\pi/2$  rotation maps 13 to 24, but it maps 24 to 31; replacing 31 with 13 acts as  $-1$  on one of the two copies of  $\wedge^2 \mathbf{56}$ . Thus the  $D_4$  anti-invariant part of  $\mathbf{56}^{\otimes 4}$  is the antisym-

metric part of  $(\mathbf{1} \oplus \mathbf{1539}) \otimes (\mathbf{1} \oplus \mathbf{1539})$ , or more explicitly it is<sup>31</sup>  $\mathbf{1539} \otimes \wedge^2 \mathbf{1539}$ .

From the tables of [36], one learns that the adjoint or  $\mathbf{133}$  of  $E_7$  occurs precisely once in  $\mathbf{1539}^{\otimes 2}$ . This one occurrence actually is in  $\wedge^2 \mathbf{1539}$ , not in  $\text{Sym}^2 \mathbf{1539}$ , because, as  $\mathbf{1539}$  is a real representation of  $E_7$ , the adjoint must occur at least once in  $\wedge^2 \mathbf{1539}$ . So as claimed above, there is precisely one possible anomaly.

## 8. Two-Loop Correction to Gauge Invariance

### 8.1 Preliminaries

Consider a Wilson line in our four-dimensional theory in a general representation  $V$ . If we choose a basis  $t_a$  of the Lie algebra  $\mathfrak{g}$ , the Wilson line is characterized classically by matrices

$$(8.1) \quad t_{a,k} : V \rightarrow V,$$

where, as before, the matrices  $t_{a,k}$  tell us how  $\partial_z^k A$  is coupled to the Wilson line. If  $V$  is a representation of  $G$  and not of  $\mathfrak{g}[[z]]$ , then  $t_{a,k} = 0$  for  $k > 0$ .

At the classical level, gauge-invariance requires that these matrices must satisfy the commutation relations

$$(8.2) \quad [t_{a,k}, t_{b,l}] = f_{ab}^c t_{c,k+l}.$$

These relations receive quantum corrections; the condition for a Wilson operator to be anomaly-free at the quantum level is different from eqn. (8.2). We have seen in section 5.4 that a correction must occur at order  $\hbar^2$ . We gave one derivation of this statement, which relied on an analysis of the fusion of parallel Wilson lines. In a companion paper [25], we will give another derivation based on the RTT presentation of the Yangian algebra.

Both derivations are a little indirect, and one might wish for a more direct one. In this section we will provide a direct derivation via Feynman diagrams.

Some elementary remarks may help one anticipate what sort of diagrams will be relevant. In an ordinary gauge theory with a finite dimensional Lie algebra  $\mathfrak{g}$ , how does one usually see in perturbation theory that the matrices  $t_a$  by which gauge fields  $A^a$  couple to a charged particle or a Wilson line must satisfy the expected commutation relations? At tree level, there are three diagrams (Fig. 26). They are proportional respectively to  $t_a t_b$ ,  $t_b t_a$ , and  $f_{ab}^c t_c$ , where  $f_{abc}$

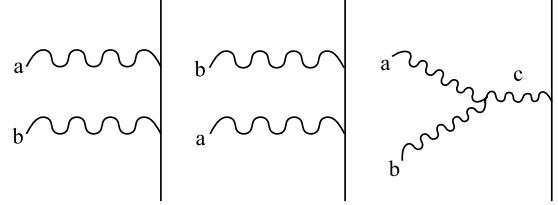


Figure 26. The three tree-level diagrams describing coupling of a pair of gauge bosons to a charged particle or a Wilson line operator.

are the structure constants that appear in the bulk  $A^3$  vertex. One tests these diagrams for gauge-invariance or BRST invariance by making a gauge transformation  $A^a \rightarrow A^a + dc^a$  for one of the external gauge bosons, with  $c^a$  the ghost field. A standard calculation shows that the sum of the three diagrams is gauge-invariant and BRST-invariant if and only if  $[t_a, t_b] = f_{ab}^c t_c$ . The violation of BRST-invariance if this condition is not satisfied is bilinear in  $A$  and  $c$ , because the diagrams have two external bosons, one of which is replaced by  $dc^a$  when one tests for BRST-invariance. Likewise, in the analysis that follows, the anomalies come from diagrams with two external gauge bosons, and are bilinear in  $A$  and  $c$ .

After verifying gauge invariance at tree level, one should go on to determine whether quantum corrections to the coupling of two gauge bosons to a charged particle or to a Wilson line preserve gauge invariance. In conventional four-dimensional quantum field theory, the answer in general is that there is a problem: in certain theories, one-loop triangle diagrams have an anomaly that irreparably spoils gauge invariance. In the model that we will study, there is no problem in order  $\hbar$ , but we will find an anomaly in order  $\hbar^2$  – roughly speaking, in two-loop order. This anomaly, however, will not represent a complete breakdown of gauge invariance. Rather, it will represent a deformation of the gauge symmetry algebra – a quantum correction to the classical commutation relations (8.2).

Such a quantum deformation as the outcome of an analysis of anomalies is perhaps unfamiliar, for the following reason. It does not occur in the case of a semi-simple gauge group, because the classification of semi-simple groups is discrete. The infinite-dimensional gauge algebra  $\mathfrak{g}[[z]]$  is, however, susceptible to continuous deformation. (Actually, because the anomalous diagrams have more than one gauge boson attached to the Wilson line, what we will get is a deformation of the universal enveloping algebra of  $\mathfrak{g}[[z]]$ , not a deformation of  $\mathfrak{g}[[z]]$  as a Lie algebra.)

In our theory, because  $\mathfrak{g}[[z]]$  will be deformed, the gauge invariance of a Wilson line depends on modifying the matrices  $t_{a,n}$  to provide a representation of the

<sup>31</sup> As a check on this, the dimension of  $\mathbf{1539} \oplus \wedge^2 \mathbf{1539}$  is 1185030. This is the right dimension for the  $D_4$  anti-invariant part of  $\mathbf{56}^{\otimes 4}$ . For example, an exercise using the character table of  $D_4$  (or based on elementary considerations) tells us that the  $D_4$  anti-invariants in  $(\mathbb{C}^d)^{\otimes 4}$  are of dimension  $\frac{1}{8}(d^4 - 2d^3 - d^2 + 2d)$ .



deformed algebra rather than of  $\mathfrak{g}[[z]]$ . If this is possible for a given Wilson line, we say that this Wilson line (or the associated representation of  $\mathfrak{g}[[z]]$ ) “quantizes.” A general Wilson line does not quantize in this sense. In section 8.7, we describe an explicit counterexample.

## 8.2 Anomalies on Wilson Lines

We will first provide a general analysis of what anomalies can occur in the coupling of gauge fields to a given Wilson line operator. We will derive a cohomological interpretation of the anomaly. Combined with a cohomology calculation given in appendix C, this allows us to prove the following results.

**Theorem 8.1.** *Let  $V$  be a representation of a simple Lie algebra  $\mathfrak{g}$  which has no Abelian factors and which is not  $\mathfrak{sl}_2$ .*

1. *The first possible anomaly to quantizing the Wilson line arises at 2 loops, that is in order  $\hbar^2$ .*
2. *Let  $\wedge^2_0 \mathfrak{g}$  be the kernel of the bracket map  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$ . If there are no non-trivial  $G$ -invariant maps  $\wedge^2_0 \mathfrak{g} \rightarrow \text{End}(V)$ , then there are no anomalies.*
3. *At two loops, the number of possible anomalies is exactly the dimension of the space of  $G$ -invariant maps  $\wedge^2_0 \mathfrak{g} \rightarrow \text{End}(V)$ .*

These results imply, for instance, that the vector representations of the classical groups  $SL_n$ ,  $SO_n$  and  $Sp_{2n}$  all quantize.

Let us analyze the possible forms of the anomalies to quantizing a Wilson line. By the anomaly of a Wilson line, we mean its BRST variation. This will always be the integral over the Wilson line of a local operator of ghost number 1. Thus, this operator will be linear in the ghost field  $c$  (and its derivatives) with *a priori* an arbitrary polynomial dependence on the gauge field  $A$  (and its derivatives).

As usual, we work on  $\mathbb{R}^2 \times C$ , with the usual real coordinates  $x, y$  and complex coordinate  $z$ . We take a Wilson line to be parametrized by  $x$ , at  $y = z = 0$ . Such a Wilson line is invariant under the classical symmetries that independently rescale  $x$  and  $y$ . The anomaly will have the same invariances. This implies that the anomaly must be linear in  $A_x$  and independent of  $A_y$ , and cannot have any  $x$  or  $y$  derivatives.

Because the classical symmetry which scales the  $x$  and  $y$  directions is unbroken by the coupling of a Wilson line, any anomaly<sup>32</sup> must respect this symmetry.

<sup>32</sup> In general, given a quantum system defined up to order  $k$  in  $\hbar$ , the anomaly to quantizing the system to order  $k + 1$  in  $\hbar$  must respect any symmetries present in the system at order  $k$ . To see this, we note that if we change the order  $k + 1$  counter-terms, the anomaly changes by a BRST-exact term. Therefore the cohomology class of the anomaly only depends on the theory up to order  $k$ , and is preserved by any symmetries present in the system up to order  $k$ .

Since the anomaly must be an integral over the Wilson line, it must involve only the  $x$ -component  $A_x$  of the connection, and can not involve the  $y$ -component or any  $x$  and  $y$  derivatives. Otherwise, it would not respect the scaling symmetries in the  $xy$  plane. For the same reason, the anomaly must be at most a linear function of  $A_x$ .

This argument does not exclude the possibility that the anomaly can include some polynomial in  $A_{\bar{z}}$ , or its  $\bar{z}$  derivatives. Let us now see why these can not occur (in section 7.2 we have performed a similar analysis for the anomaly to the existence of a vertex). The classical action functional of the theory is invariant under the symmetry which scales  $z, \bar{z}$ , and  $\hbar$  by a real parameter  $c$ . It is also invariant under the symmetry by which  $z$  and  $\hbar$  are rotated through an angle  $e^{2\pi i \theta}$ , and where  $\bar{z}$  is therefore rotated by  $e^{-2\pi i \theta}$ . Any anomaly must be invariant under these symmetries (where we include the factor of  $\hbar^k$  naturally present in a  $k$ -loop anomaly).

The most general anomaly at  $k$  loops will involve  $n_1$  copies of  $A_{\bar{z}}$ ,  $n_2$   $\bar{z}$  derivatives, and  $n_3$   $z$ -derivatives. A copy of  $A_{\bar{z}}$  in the anomaly behaves the same with respect to these symmetries as a  $\bar{z}$ -derivative, as it involves contracting the connection  $A$  with the vector field  $\partial_{\bar{z}}$ . The symmetries under scaling and rotation in the  $z$ -plane tell us that

$$(8.3) \quad \begin{aligned} n_1 + n_2 + n_3 - k &= 0, \\ -n_1 - n_2 + n_3 - k &= 0. \end{aligned}$$

From this we see that  $n_1 = n_2 = 0$  and  $n_3 = k$ . Therefore, the anomaly can not involve the  $\bar{z}$ -component of the connection or any  $\bar{z}$  derivatives, and the number of  $z$  derivatives in a  $k$ -loop anomaly must be precisely  $k$ .

As a further constraint, note that if the anomaly does not depend at all on  $A_x$ , and so depends only on the ghost field  $c$ , then it must be given by an integral of  $\partial_x c$  (in view of the scaling symmetry in the  $x$  direction). We can write  $\partial_x c = D_x c - [A_x, c]$ , where  $D_x c$  is the covariant derivative of  $c$ . Since insertion of  $D_x c$  in a Wilson operator gives a total derivative that would not contribute, we can replace  $\partial_x c$  with  $-[A_x, c]$ , and thus it is not necessary to consider terms that are independent of  $A_x$ .

The constraints we have considered so far tell us that the most general  $k$ -loop anomaly is of the form

$$(8.4) \quad \sum_{k_1+k_2=k} \Theta_{a,k_1,b,k_2} \int_{y=z=0} \partial_z^{k_1} A_x^a \partial_z^{k_2} c^b dx,$$

where for each value of  $k_1, k_2$ ,  $\Theta_{a,k_1,b,k_2}$  is a  $G$ -invariant linear map

$$\mathfrak{g} \otimes \mathfrak{g} \rightarrow \text{End}(V),$$

where  $V$  is the representation of  $G$  from which we build the Wilson line.



There is one more constraint. Since the anomaly is the BRST variation of the Wilson line, it is itself BRST-invariant. Applying this constraint leads us to the equation

$$(8.5) \quad \begin{aligned} & \sum_{k_1+k_2=k} \Theta_{a,k_1,b,k_2} \int_{y=z=0} \partial_z^{k_1} (d\mathbf{c}^a + f^a_{cd} \mathbf{c}^c A_x^d) (\partial_z^{k_2} \mathbf{c}^b) dx \\ & + \sum_{k_1+k_2=k} \Theta_{d,k_1,a,k_2} \int_{y=z=0} \partial_z^{k_1} A_x^d \partial_z^{k_2} \left( \frac{1}{2} f^a_{cb} \mathbf{c}^c \mathbf{c}^b \right) dx \\ & + \sum_{k_1+k_2=k} [\rho_c, \Theta_{d,k_1,b,k_2}] \int_{y=z=0} \mathbf{c}^c (\partial_z^{k_1} A_x^d) (\partial_z^{k_2} \mathbf{c}^b) dx = 0 \end{aligned}$$

In the third term in this equation, the operator  $\rho_a : V \rightarrow V$  indicates the classical action of the operator  $t_a$  on  $V$ . The appearance of this term was explained in detail in a similar context in section 7.2.

This equation can be separated into a term which is linear in  $A$  and one which has no  $A$ -dependence. Both of these terms must vanish. Let us first analyze the term which is independent of  $A$ . If we integrate by parts, and use the fact that the ghost field is an anti-commuting variable (i.e. a fermionic field of spin 0), we find that this equation tells us

$$(8.6) \quad \Theta_{a,k_1,b,k_2} + \Theta_{b,k_2,a,k_1} = 0.$$

Next, let us analyze the term in eqn. (8.5) which is linear in  $A$ . We obtain equations by setting the coefficients of  $f(\partial_z^{k_0} \mathbf{c}^c)(\partial_z^{k_1} A^d)(\partial_z^{k_2} \mathbf{c}^b)$  to zero. These equations are

$$(8.7) \quad \begin{aligned} & \Theta_{a,k_0+k_1,b,k_2} f^a_{cd} \binom{k_0+k_1}{k_0} - \Theta_{a,k_2+k_1,c,k_0} f^a_{bd} \binom{k_1+k_2}{k_1} \\ & + \Theta_{d,k_1,a,k_0+k_2} f^a_{cb} \binom{k_0+k_2}{k_0} + \delta_{k_0=0} [\rho_c, \Theta_{d,k_1,b,k_2}] \\ & - \delta_{k_2=0} [\rho_b, \Theta_{d,k_1,c,k_0}] = 0. \end{aligned}$$

Note that this quantity is totally antisymmetric under permutation of  $(k_0, c)$ ,  $(k_1, d)$  and  $(k_2, b)$ , as follows from eqn. (8.6).

We can summarize this result as follows. Let us view the collection  $\Theta_{a,k_1,b,k_2}$  as a linear map

$$(8.8) \quad \begin{array}{ccc} \Theta : \wedge^2 \mathfrak{g}[[z]] & \rightarrow & \text{End}(V), \\ \psi & & \psi \\ (t_a z^{k_1}) \wedge (t_b z^{k_2}) & \mapsto & k_1! k_2! \Theta_{a,k_1,b,k_2}. \end{array}$$

Let  $C^*(\mathfrak{g}[[z]], \text{End}(V))$  denote the Chevalley-Eilenberg cochain complex, which in degree  $n$  consists of linear maps  $\wedge^n \mathfrak{g}[[z]] \rightarrow \text{End}(V)$ .

Then, eqn. (8.7) is precisely the condition that  $\Theta$  is a closed element of this Chevalley-Eilenberg cochain complex.

### 8.2.1 Cancellation by Counter-Terms

Let us now turn to analyzing when such an anomaly can be cancelled by a counterterm. Suppose that the first anomaly arises at  $k$  loops and is given by the expression in eqn. (8.4) involving some  $\Theta_{a,k_1,b,k_2}$  satisfying 8.6.

To try to cancel the anomaly, we change the coupling of the gauge field  $A$  so that  $\partial_z^k A^a$  is coupled by some operator  $\rho_{a,k} \in \text{End}(V)$ . We will assume that  $\rho_{a,k}$  transforms under the adjoint representation of  $G$  on  $\text{End}(V)$ . Simple Lie algebras and their representations have a discrete classification and cannot be deformed, so we will assume that  $\rho_{a,k}$  vanishes for  $k=0$ .

The term  $\rho_{a,k} \int \partial_z^k A^a$  that we add to the action may not be gauge invariant at the classical level. The failure to be gauge invariant is given by the expression

$$(8.9) \quad \rho_{a,k} \int \partial_z^k d\mathbf{c}^a + \rho_{a,k} f_{abc} \int \partial_z^k (\mathbf{c}^b A^c) + [\rho_b, \rho_{a,k}] \int \mathbf{c}^b \partial_z^k A^a.$$

In order to cancel the anomaly at  $k$  loops, we need

$$(8.10) \quad \begin{aligned} & \sum_{k_1+k_2=k} \Theta_{a,k_1,b,k_2} \int_{y=z=0} \partial_z^{k_1} A_x^a \partial_z^{k_2} \mathbf{c}^b dx \\ & = \rho_{a,k} f_{abc} \int \partial_z^k (\mathbf{c}^b A^c) + [\rho_b, \rho_{a,k}] \int \mathbf{c}^b \partial_z^k A^a. \end{aligned}$$

This can happen if and only if

$$(8.11) \quad \Theta_{a,k_1,b,k_2} = \rho_{c,k} f_{ab} \binom{k}{k_1} + [\rho_a, \rho_{b,k}].$$

This is precisely the condition that the cocycle in  $C^*(\mathfrak{g}[[z]], \text{End}(V))$  associated to the anomaly (as in eqn. (8.8)) is exact.

In sum, we have found that possible anomalies to coupling to a Wilson line, modulo counter-terms which can cancel these anomalies, are given by the Chevalley-Eilenberg cohomology group

$$(8.12) \quad H^2(C^*(\mathfrak{g}[[z]], \text{End}(V))).$$

We will repeatedly use the fact that for  $G$  a simple Lie group, the Chevalley-Eilenberg cohomology (of any degree) with values in any representation can always be represented by a  $G$ -invariant cocycle, roughly by averaging over a maximal compact subgroup of  $G$ . (Of course, the expressions that we generate from Feynman diagrams in expanding around the trivial flat connection will always be  $G$ -invariant.) Note that Chevalley-Eilenberg cohomology is also called Lie algebra cohomology, and under that name often appears in analysis of BRST cohomology in quantum field theory.

One nice consequence of the description of anomalies by Chevalley-Eilenberg cohomology is that

we can understand what anomalies can exist by applying standard arguments from homological algebra. When translated into field-theory language, the arguments would be quite complicated.

For instance, the following is a relatively easy result:

**Lemma 8.2.** *For a Wilson line in a representation  $V$  of a simple group  $G$  (as opposed to a representation of  $\mathfrak{g}[[z]]$ ), there are no one-loop anomalies which cannot be cancelled by the introduction of a counter-term.*

*Further, all two-loop anomalies are equivalent to ones given by expressions of the form*

$$(8.13) \quad \sum \Theta_{a,1,b,1} \int_{y=z=0} \partial_z A_x^a \partial_z c^b dx,$$

where, as before,  $\Theta_{a,1,b,1}$  is a linear operator on the vector space  $V$  and is antisymmetric under the exchange of the  $a, b$  indices. Viewed as a linear map  $\wedge^2 \mathfrak{g} \rightarrow \text{End}(V)$ ,  $\Theta$  can be assumed to be  $G$ -invariant.

*Two-loop anomalies which factor through a map  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$  can be cancelled by a counter-term.*

*Proof.* The proof of the lemma uses the notion of relative Lie algebra cochains of  $\mathfrak{g}[[z]]$  with respect to the subalgebra  $\mathfrak{g}$ . Relative Lie algebra cochains are  $G$ -invariant elements of  $C^*(z\mathfrak{g}[[z]], \text{End}(V))$ , that is,  $G$ -invariant cochains which do not involve  $\mathfrak{g}$ . The relative cochains are equipped with the same Chevalley-Eilenberg differential as before. We let  $H^*((\mathfrak{g}[[z]]|\mathfrak{g}), \text{End}(V))$  denote the relative cohomology. General results on the cohomology of simple Lie algebras imply that there is an isomorphism

$$(8.14) \quad \oplus_{i+j=k} H^i(\mathfrak{g}) \otimes H^j((\mathfrak{g}[[z]]|\mathfrak{g}), \text{End}(V)) \cong H^k(\mathfrak{g}[[z]], \text{End}(V)).$$

Since  $\mathfrak{g}$  is simple,  $H^*(\mathfrak{g})$  is zero in degrees 1 and 2, and  $H^0(\mathfrak{g}) = \mathbb{C}$ . Therefore relative and absolute Lie algebra cohomology coincide in degrees 1 and 2.

At weight 1 under the  $\mathbb{C}^\times$  action that scales  $z$ , the relative cohomology is zero in degrees 2 and higher. This is simply because we cannot build a bilinear of weight 1 from  $z\mathfrak{g}[[z]]$ . Therefore there are no anomalies of weight 1.

The cohomology of weight 2 under this  $\mathbb{C}^\times$  action is the part that contributes to the two-loop anomaly. Relative 2-cocycles of weight 2 must be given by a  $G$ -invariant bilinear and antisymmetric map  $z\mathfrak{g} \otimes z\mathfrak{g} \rightarrow \text{End}(V)$ . Because each tensor factor involves one  $z$ , the corresponding two-loop anomaly is of the form given in eqn. (8.13) (with one  $z$  derivative of  $A$  and one of  $c$ ). So every two-loop anomaly is of this form.

Concerning the last statement in the lemma, we observe first that for simple  $G$ , any  $G$ -invariant map  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$  is actually a multiple of the commutator map. We do not know a unified proof of this statement,

but for classical groups it follows from invariant theory,<sup>33</sup> and for exceptional groups, it can be verified by consulting tables (such as those in [36]). Given this, we need to show that if a two-cocycle

$$(8.15) \quad \wedge^2(z\mathfrak{g}) \rightarrow \text{End}(V)$$

factors through the commutator map  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$ , then it is the coboundary of a  $G$ -invariant map  $z^2 \mathfrak{g} \rightarrow \text{End}(V)$ . But the coboundary of a  $G$ -invariant map  $z^2 \mathfrak{g} \rightarrow \text{End}(V)$  is given by the composition

$$(8.16) \quad \wedge^2(z\mathfrak{g}) \rightarrow z^2 \mathfrak{g} \rightarrow \text{End}(V),$$

where the first map is the commutator. Therefore any two-cocycle that factors through the commutator is a coboundary.  $\square$

This lemma tells us that find that the first possible anomalies can arise at two loops, from diagrams with two external gauge fields.

In appendix C, we prove the following result (relying heavily on the results of [31]). This result generalizes the previous lemma to all loops. The proof is much more difficult.

**Proposition 8.3.** *Let  $\mathfrak{g}$  be a simple Lie algebra which is not<sup>34</sup>  $\mathfrak{sl}_2$ . Let  $\wedge_0^2 \mathfrak{g}$  be the kernel of the Lie bracket map from  $\wedge^2 \mathfrak{g}$  to  $\mathfrak{g}$ . (Note that there are no copies of the adjoint in the representation  $\wedge_0^2 \mathfrak{g}$ ). Then the inclusion map*

$$(8.17) \quad H^2(\mathfrak{g}[[z]], \wedge_0^2 \mathfrak{g}) \rightarrow H^2(\mathfrak{g}[[z], \mathfrak{g} \otimes \mathfrak{g}])$$

*is an isomorphism.*

In other words, if we decompose  $\mathfrak{g} \otimes \mathfrak{g}$  as  $\text{Sym}^2 \mathfrak{g} \oplus \mathfrak{g} \oplus \wedge_0^2 \mathfrak{g}$ , then all of the cohomology we are interested in comes from  $\wedge_0^2 \mathfrak{g}$ .

We are interested in  $H^2(\mathfrak{g}[[z]], \text{End}(V))$  for a representation  $V$ . In general,  $\text{End}(V)$  is a complicated, highly reducible representation of  $G$ . Chevalley-Eilenberg cohomology is defined for every representation, and if  $\text{End}(V)$  is a direct sum of irreducibles, then the desired cohomology is the direct sum of cohomology with values in these irreducibles.

However, the desired cohomology can be represented by  $G$ -invariant cocycles and such a cocycle, ignoring the  $z$ -dependence, is a  $G$ -invariant map from  $\mathfrak{g} \otimes \mathfrak{g}$  to  $\text{End}(V)$ . Hence, any two-cocycle valued in  $\text{End}(V)$  must come from a two-cocycle valued in

<sup>33</sup> That is, one explicitly considers an element of  $\mathfrak{g}$  as a two-index tensor, and one considers the possible contractions of indices to make a  $G$ -invariant map from  $\wedge^2 \mathfrak{g}$  to  $\mathfrak{g}$ .

<sup>34</sup> The case of  $\mathfrak{sl}_2$  is slightly different. There we find that  $H^2(\mathfrak{g}[[z]], \mathfrak{g} \otimes \mathfrak{g})$  is isomorphic to  $H^2(\mathfrak{g}[[z]], \text{Sym}^2 \mathfrak{g})$ , which is one-dimensional and in weight 3 under the  $\mathbb{C}^\times$  action scaling  $z$ . Since it is known by other methods that every representation of  $\mathfrak{sl}_2$  quantizes, we will avoid the special case of  $\mathfrak{sl}_2$  in what follows.

$\mathfrak{g} \otimes \mathfrak{g}$  under some map of  $G$ -representations from  $\mathfrak{g} \otimes \mathfrak{g}$  to  $\text{End}(V)$ . The same is true for the corresponding second cohomology classes. The proposition gives an important refinement of this statement: any non-trivial second cohomology class must come from a map  $\wedge^2 \mathfrak{g} \rightarrow \text{End}(V)$ . This means that when we study Feynman diagrams, we can ignore terms in which two gauge bosons  $A^a$  and  $A^b$  are coupled symmetrically in  $a$  and  $b$ , and we can also ignore terms in which they are coupled via  $f_{ab}^c$ .

As a corollary, we find that if there are no maps of representations  $\wedge^2 \mathfrak{g} \rightarrow \text{End}(V)$ , then the Wilson line living in  $V$  can not have an anomaly. This is the criterion (†) that was introduced in section 7.1 and used in various applications.

### 8.3 Enumerating Feynman Diagrams

So far, we have performed a cohomological analysis of possible anomalies that can appear when one tries to quantize a Wilson line. We have seen that the first such anomalies can appear at two loops. Next, we will enumerate all possible two-loop diagrams that may contribute an anomaly. We will find that, for various reasons, all but one of the diagrams we enumerate can not produce an anomaly.

Throughout our analysis, we will assume our Wilson line is placed at  $z = 0$  and is invariant under the symmetry which rescales  $z$  and  $\hbar$ . This means that  $t_{a,k}$  is accompanied by  $\hbar^k$ , and in particular at the classical level  $t_{a,k} = 0$  for  $k > 0$ .

Let us now analyze possible two-loop diagrams which can contribute anomalies of this form. The only diagrams that are relevant are two-loop diagrams which may be attached to the Wilson line at an arbitrary number of points, but which have precisely two external lines to which the gauge field is coupled.

We must also bear in mind an additional subtlety. In principle one can introduce one-loop counter-terms whereby  $\partial_z A$  is coupled to the Wilson line in some way. It could happen that these are forced on us to cancel any one-loop anomalies, but even if this is not the case, such one-loop counter-terms do not violate any symmetries and so one is always free to introduce them. Therefore there are two possible vertices at which a gluon can meet a Wilson line. There is the usual classical interaction, but also the interaction coming from a one-loop counter-term. We will depict these interactions as an ordinary vertex, and a circle labelled 1, respectively. Counting of the loop parameter tells us that one-loop diagrams, one of whose vertices on the Wilson line is the one-loop vertex, are counted as two-loop diagrams.

One can cut down the number of diagrams we need to consider by observing that every diagram must have at least two vertices on the Wilson line.

Diagrams with one vertex on the Wilson line can not contribute, because the corresponding anomaly will be given by a two-cocycle valued in some copy of  $\mathfrak{g} \subset \text{End}(V)$ . Our cohomological analysis in proposition 8.3 tells us that any anomaly of this form can be cancelled by a counter-term.

In evaluating whether a diagram can contribute to the anomaly, note that the anomaly must have one  $z$ -derivative at each external line, and no other derivatives. We can thus calculate whether or not a diagram contributes to the anomaly by taking the two external fields to be the gauge field  $A = z\delta_{x=0}t_a$ , and the ghost field  $c = zt_b$ . This tells us, in particular, that it is not possible to have an anomaly from a diagram in which the external lines are connected directly to the Wilson line by a tree-level vertex. This is because the Wilson line is at  $z = 0$ , and there are no  $z$ -derivatives in how the Wilson line is coupled classically.

A further constraint on the diagrams that can appear is provided by the observation that if we evaluate the propagator with both ends on a straight Wilson line we find zero. Therefore in any diagram which can contribute to the anomaly this configuration can not occur.

These observations cut down substantially the diagrams that can appear. We will further exclude a few simple diagrams which are topologically trees, but whose coupling to the Wilson line involves one and two-loop counter-terms. These diagrams are similar to those in Fig. 26. In section 8.6 we will analyze the effect of these diagrams, and see that they contribute to the two-loop correction to the algebra  $\mathfrak{g}[[z]]$ .

All remaining diagrams are depicted in Fig. 27. Each figure admits a number of variants where the order of the vertices on the Wilson line is permuted. These variants are not depicted.

We will find that the only diagram that can contribute is diagram (A1). All other diagrams will be excluded by a variety of arguments.

#### 8.3.1 Excluding Diagrams (A2), (C1), (C2) and (D2)

Consider any two-loop diagram with the feature that the two vertices attached to the external lines are connected by a propagator. The only non-trivial anomalies are antisymmetric in the external lines. After antisymmetrizing and using the Jacobi identity, the anomaly from such a diagram is of the form

$$(8.18) \quad \int f^{cab} \partial_z c_a \partial_z A_b \mu_c$$

for some matrices  $\mu_c : V \rightarrow V$ . Any anomaly of this form can be cancelled, because it comes from a map  $\wedge^2 \mathfrak{g} \rightarrow \text{End}(V)$  which factors through the adjoint representation.

This tells us that diagrams (A2), (C1), (C2) and (D2) can not contribute a nontrivial anomaly.

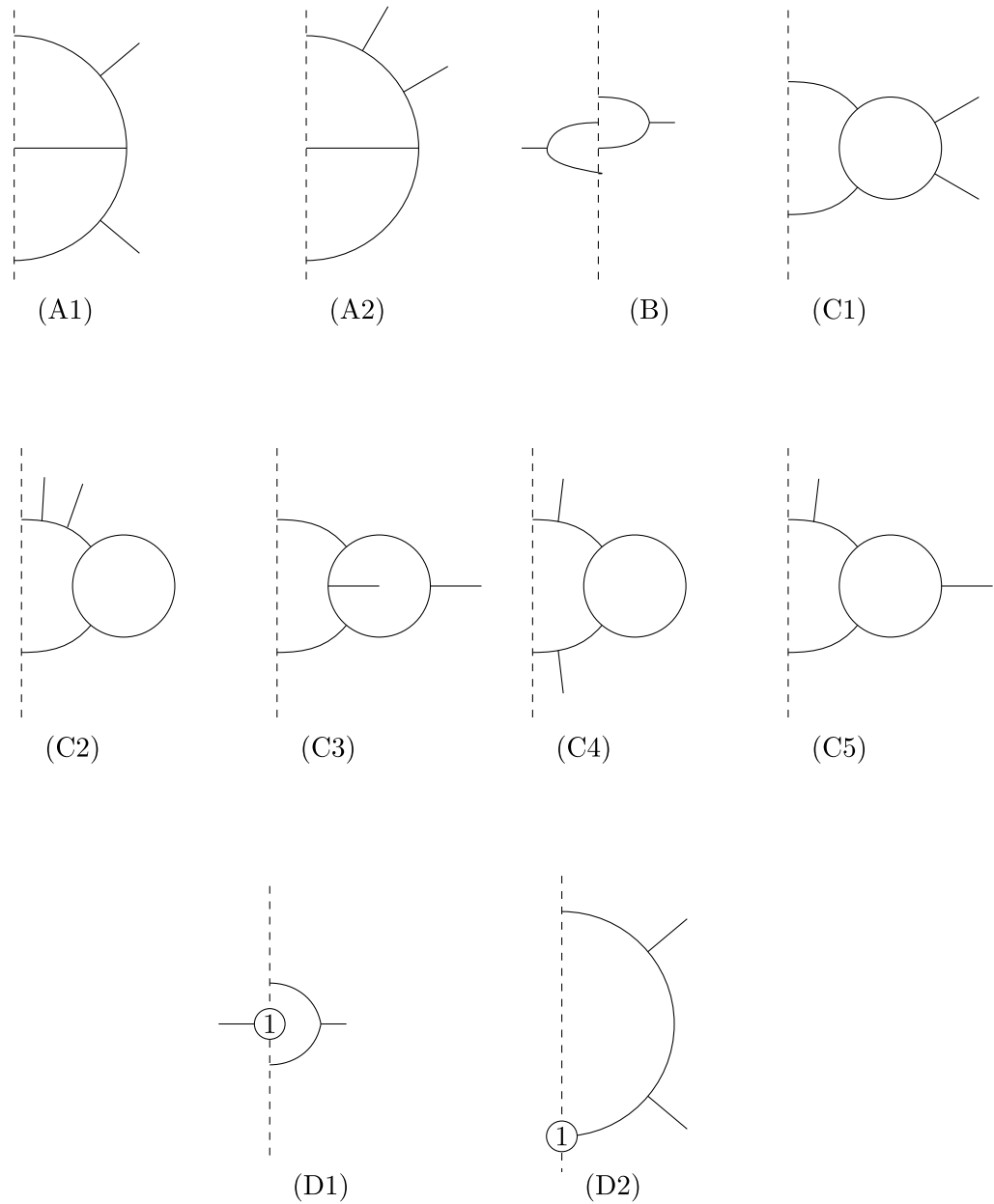


Figure 27. Two-loop Feynman diagrams with two external lines which can potentially contribute to the anomaly. The dashed vertical line represents a Wilson line, and the vertices on the Wilson line labelled 1 indicate a coupling of the gauge field and Wilson line by a one-loop counter-term. Each diagram admits a number of variants, not shown, in which the positions of the vertices on the Wilson line are permuted.

### 8.3.2 Excluding Diagrams of Type (C3)

Diagram of type (C3) can be excluded for Lie algebraic reasons. Any anomaly that can appear from a diagram of type (C3) is of the form:

$$(8.19) \quad \text{Tr}_{\mathfrak{g}}(t_a t_c t_b t_d) \int \partial_z c_a \partial_z A_b \rho(t_c) \rho(t_d).$$

(Here the trace is taken in the adjoint representation, and the external lines have indices  $a, b$ .)

This anomaly can be cancelled by a two-loop counter-term if the tensor  $\text{Tr}_{\mathfrak{g}}(t_a t_c t_b t_d)$  is symmetric in

$a$  and  $b$ , or if it can be written in terms of the commutator  $f_{ab}^c t_c$  of  $t_a$  and  $t_b$ . Since the adjoint representation is equipped with a symmetric non-degenerate pairing, we have the identity

$$(8.20) \quad \text{Tr}_{\mathfrak{g}}(t_a t_c t_b t_d) = \text{Tr}_{\mathfrak{g}}(t_d t_b t_c t_a).$$

Together with cyclic symmetry of the trace, this tells us that  $\text{Tr}_{\mathfrak{g}}(t_a t_c t_b t_d)$  is symmetric in  $a$  and  $b$ . This implies that any anomaly associated to the tensor  $\text{Tr}_{\mathfrak{g}}(t_a t_c t_b t_d)$  can be cancelled by a counter-term.

### 8.3.3 Excluding Diagrams (C4) and (C5)

Diagrams (C4) and (C5) contain an internal gauge particle loop. In (C4), this loop has two external lines, both of which are internal lines in the full diagram, while (C5) has an internal loop with three external lines, of which in the full diagram two are internal and one is external.

The group theory factor that comes from the internal loop in (C4) is an invariant in  $\mathfrak{g} \otimes \mathfrak{g}$ . Any such invariant is a multiple of  $\delta_{cd}$ , that is, the Killing form, and accordingly, the overall group theory factor of the diagram in (C4) is a multiple of what it would be if the internal loop were collapsed to a point. Then in (C4) the two external gauge fields would be connected by a propagator. So this diagram can be excluded by the same argument as in section 8.3.1.

In (C5), since the internal loop has three external lines, it produces an invariant in  $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$ . For any simple Lie algebra other than  $SL_N$ ,  $N > 2$ , such an invariant actually lies in  $^3\wedge^3\mathfrak{g}$  and is a multiple of the invariant associated to the structure constants  $f_{abc}$ . When this is the case, the group theory factor of (C5) is the same as if the internal loop were collapsed to a point. When this is done, the two external gauge fields are connected by a propagator and this diagram can be excluded as before.

For  $SL_N$ ,  $N > 2$ , there actually is another invariant in  $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$ ; in fact, it lies in  $\text{Sym}^3\mathfrak{g}$ . However, this invariant is odd under the outer automorphism of  $SL_N$ , which is a symmetry of the theory under discussion, and therefore cannot appear.

### 8.3.4 Excluding Diagrams (B) and (D1)

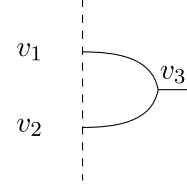
Diagrams (B) and (D1) cannot be excluded by Lie algebraic arguments. Instead we will show by an explicit calculation that the amplitudes for such diagrams are zero, so they can not contribute an anomaly. The computation involves only the part of the diagram on the right in the figures; this is the part that (B) and (D1) have in common.

To evaluate the amplitudes for these diagrams, we will employ a point-splitting regularization in which any pair of vertices on the Wilson lines must be separated by a distance of at least  $\epsilon$ . Once we employ such a point-splitting regularization, we can compute the regularized amplitude for each connected component of the diagrams (B) and (D1) separately. The two

<sup>35</sup> Invariants in  $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$  correspond to  $G$ -invariant maps  $\mathfrak{g} \otimes \mathfrak{g} \rightarrow \mathfrak{g}$ . That there is only one such invariant for most  $G$  and two for  $SL_N$ ,  $N > 2$ , can be shown by invariant theory for classical  $G$  and by consulting tables such as those in [36] for exceptional  $G$ . In  $SL_N$ ,  $N > 2$ , the two invariants are  $\text{Tr}A\{B,C\}$  and  $\text{Tr}A\{B,C\}$  and are respectively even and odd under the outer automorphism. The fact that an invariant in  $\text{Sym}^3\mathfrak{g}$  exists only for  $SL_N$ ,  $N > 2$  is important in particle physics in classifying possible anomalies in gauge theory.

connected components only interact with each other by how they affect the domain of integration of the vertices on the Wilson lines. We will get a vanishing that does not depend on this domain.

We therefore need to compute the amplitude of the diagram



The coordinates for the vertex  $v_i$  will be  $(x_i, y_i, z_i)$ . For vertices  $v_1, v_2$  we have  $y_i = z_i = 0$ .

Recall (see eqn. (4.5)) that the propagator two-form is

$$(8.21) \quad P = \frac{1}{2\pi} \frac{xdy \wedge d\bar{z} - ydx \wedge d\bar{z} + 2\bar{z}dx \wedge dy}{(x^2 + y^2 + |z|^2)^2}.$$

The amplitude for the diagram is then

$$(8.22) \quad c \int_{x_1, x_2, x_3, y_3, z_3} Adz_3 \wedge \frac{(x_3 - x_1)dy_3 d\bar{z}_3 - y_3 d(x_3 - x_1)d\bar{z}_3 + 2\bar{z}_3 d(x_3 - x_1)dy_3}{d(v_3, v_2)^4} \wedge \frac{(x_3 - x_2)dy_3 d\bar{z}_3 - y_3 d(x_3 - x_2)d\bar{z}_3 + 2\bar{z}_3 d(x_3 - x_2)dy_3}{d(v_3, v_2)^4},$$

where  $d(v_i, v_j)$  indicates the distance between the vertices  $v_i$  and  $v_j$ , and  $A$  is the gauge field on the external line.

Since we are integrating over  $x_1, x_2$ , we need only keep terms which involve  $dx_1$  and  $dx_2$ . If we retain only those terms, we find that the integrand is  $dx_1 \wedge dx_2$  times the square of the one-form

$$(8.23) \quad y_3 d\bar{z}_3 - 2\bar{z}_3 dy_3$$

and is therefore zero.

## 8.4 The Anomaly Associated to Diagram (A1)

Let us now come to the only remaining diagram, namely (A1) in Fig. 27. We need to note that we first need to include two other diagrams, as depicted in Fig. 28, where we permute the three vertices along the Wilson lines.

### 8.4.1 Color Factor

It turns out that all these possibilities have the same color factor, up to those contributions which can be canceled by counterterms.

To see this, let us assign color indices  $a, b$  to external lines (ghost field and the gauge field) and  $c, d, e$  to



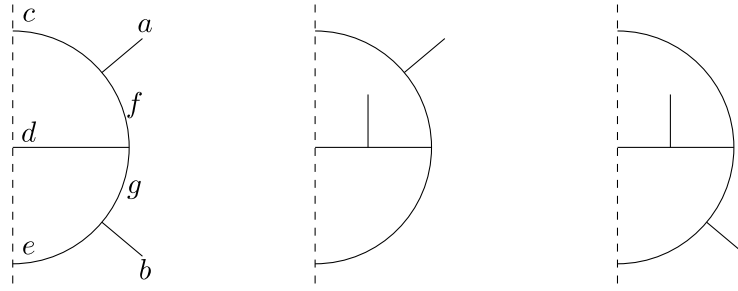


Figure 28. Permutations of diagram (A1).

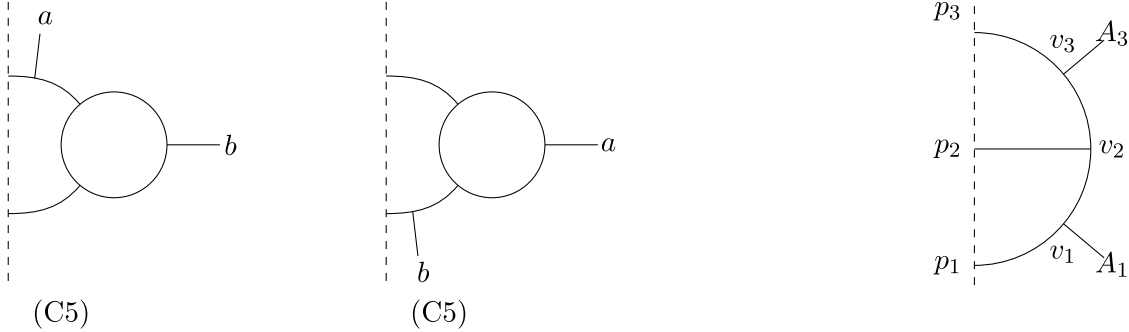


Figure 29. The differences of the color factors between first and second, and first and third diagrams in Fig. 28 have the color structures as shown in this figure.

Figure 30. The Feynman diagram with vertices labelled. The  $A_i$  on the external lines indicate gauge fields.

points on Wilson lines. In the first diagram in Fig. 28 we have  $c, d, e$  along the Wilson line from top to bottom, but in the second we have  $c, e, d$  and third  $d, c, e$ .

Let us consider the difference of color factors between first and the second diagram. Since this involves the change in the relative position of  $d$  and  $e$ , the difference gives the commutator  $[\rho(t^d), \rho(t^e)] = f^{de}{}_f \rho(t^f)$ . Hence the difference in color factors can be represented graphically as the diagram on the left of Fig. 29. Similarly, the difference of the color factor between the second and third diagram gives the second diagram of Fig. 29.

We verified above (section 8.3.3) that the color factor for diagrams of this type is such that any anomaly can always be cancelled by a counter-term.

This tells us that for each of the three diagrams in Fig. 28, the anomaly is symmetric under permutation of the order of the vertices connected to the Wilson line, up to anomalies which can be cancelled by a counter-term. Since the three diagrams in Fig. 28 are exchanged by such a permutation, we conclude that they all have the same color factor (again, modulo terms which can be cancelled by a counter-term).

The non-trivial part of the color factor is (recall the normalization of the Killing form in eqn. (3.2))

$$(8.24) \quad f^{acf} f^{fgd} f^{geb} \{\rho(t^c), \rho(t^d), \rho(t^e)\}$$

$$= ([t^a, t^c], t^d], [t^b, t^e]) \{\rho(t^c), \rho(t^d), \rho(t^e)\},$$

where  $a, b$  are indices for the external lines and

$$(8.25) \quad \{\rho(t^{a_1}), \rho(t^{a_2}), \rho(t^{a_3})\} = \frac{1}{3!} \sum_{\sigma \in \mathcal{S}_3} \rho(t^{a_{\sigma(1)}}) \rho(t^{a_{\sigma(2)}}) \rho(t^{a_{\sigma(3)}}),$$

where the sum is over the permutations of the indices 1, 2, 3.

### 8.5 Numerical Factor

Let us next initiate the evaluation of the numerical factor. We choose the Wilson line to be placed along a straight line at  $y = z = 0$ .

We let  $p_1, p_2, p_3$  denote the position of the vertices on the Wilson line, and  $v_1, v_2, v_3 \in \mathbb{R} \times \mathbb{C}$  be the positions of the internal vertices. The coordinates for vertex  $v_i$  are denoted  $x_i, y_i, z_i$  for  $i = 1, 2, 3$ . The labeling on the vertices for the first diagram in Fig. 28 is given in Fig. 30.

Let us write down the numerical factor of the amplitude, where we include all possible orderings of the vertices on the Wilson lines. We let  $P$  be the propagator viewed as a two-form on  $\mathbb{R}^2 \times \mathbb{C}$  (with singularities at the origin). The space over which we integrate consists of a copy of  $\mathbb{R}^2 \times \mathbb{C}$  for each internal vertex, and  $\mathbb{R}$  for each vertex on the Wilson line. Every vertex gives

a map from this space to  $\mathbb{R}^2 \times \mathbb{C}$ , where we project on the relevant factor for an internal vertex and project to  $\mathbb{R}$  and then include on the  $x$ -axis for the vertices on the Wilson lines. For a pair of vertices  $(v_i, v_j)$  or  $(v_i, p_j)$  we define two-forms by

$$(8.26) \quad P(v_i, v_j) = (v_i - v_j)^* P, \quad P(v_i, p_j) = (v_i - p_j)^* P,$$

where we pull-back the two-form  $P$  on  $\mathbb{R}^2 \times \mathbb{C}$  via the map  $v_i - v_j$  or  $v_i - p_j$ .

The complete amplitude, including all the relevant diagrams and the color factor, is then

$$(8.27) \quad \frac{1}{2} \left( \frac{i}{2\pi} \right)^3 \int_{p_1, p_2, p_3 \in \mathbb{R}} \int_{v_1, v_2, v_3} (P(0, v_1)) dz_1 A_1^q P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3) \times \int f^{afc} ffgd fgbec \{p(t^c), p(t^d), p(t^e)\},$$

where  $A_2, A_3$  indicate the external gauge fields. The prefactor of  $\frac{1}{2}$  comes from the diagram automorphism.

### 8.5.1 A Comment on Signs

In the numerical factor, we have implicitly oriented each edge of the graph. If we reverse the orientation of the graph, the propagator changes sign, because  $P(v_i, v_j) = -P(v_j, v_i)$ . Similarly, in defining the integral, we have implicitly ordered the set of 6 vertices (three on the Wilson line and three integrated over the four-dimensional space-time). If we permute the order of these vertices we change the integral by a sign. For the vertices on the Wilson line, the sign arises because a permutation can reverse the orientation. For the bulk vertices, since each is accompanied by the one-form  $dz$ , permuting them will change the order in which we take the wedge product of these forms, and therefore introduce a sign.

These signs are also implicit in the definition of the color factor, so there is no overall sign ambiguity. One way to see this is to treat the elements of the Lie algebra as being anti-commuting variables. This is a natural thing to do, because the form degree of the forms we are integrating differs by one from the ghost number of the corresponding fields of the physical theory. For example, a one-form is of odd degree but is a field of ghost number zero. We can force Lie algebra valued forms to have the correct parity if we declare that the Lie algebra factor is fermionic.

If we treat the Lie algebra elements as fermionic, then the color factor  $\delta_{ab} t_{v_i}^a \otimes t_{v_j}^b$  associated to the propagator  $P(v_i, v_j)$  changes sign if we reverse the order of an edge, just like the  $P(v_i, v_j)$  does. The color factor associated to a vertex is a graded symmetric function of the three Lie algebra elements. Because it is a fermionic function – since it depends on three

fermionic Lie algebra elements – the overall color factor depends via a sign on the ordering of the set of vertices. Similarly, the color factor associated to a vertex on the Wilson line depends on a single Lie algebra element, together with an endomorphism of the bosonic vector space of states on the Wilson line. Therefore the color factor depends on the ordering of the set of vertices on the Wilson line, just as the numerical factor does.

To get the correct signs in the total amplitude, the rule is to choose an orientation of each edge on the graph and an ordering on the set of vertices. Then, compute both the color factor and the numerical factor using this choice. Their product is the amplitude. Because the color factor and the numerical factor change signs in the same way if we change the ordering of the vertices and the orientation of the edges, there is no ambiguity. The color factor in eqn. (8.27) was computed in this way.

### 8.5.2 Regularizing the Integral

We will use a point-splitting regulator in which we restrict the domain of integration to the region where  $p_3 - p_1 \geq \epsilon$ , if we assume that the three points are ordered so that  $p_1 < p_2 < p_3$ . To verify that this is a good regulator, we need to show that the integral converges absolutely in this domain. We are only concerned with UV divergences, so we will further restrict the domain of integration to a region where all vertices are in some ball around the origin.

To verify convergence, we will use the following bound in absolute value of the propagator:

$$(8.28) \quad |P| \leq (x^2 + y^2 + z\bar{z})^{-3/2}.$$

This bound arises because the denominator in  $P$  is  $(x^2 + y^2 + |z|^2)^{-2}$ , while the numerator is a linear function of the variables  $x, y, \bar{z}$ . Any linear function on  $\mathbb{R}^4$  is bounded in absolute value by some multiple of  $(x^2 + y^2 + |z|^2)^{1/2}$ . It suffices to verify that the integral converges when each propagator  $P(v_i, v_j)$  or  $P(p_i, v_j)$  is replaced by  $d(v_i, v_j)^{-3}$  or  $d(p_i, v_j)^{-3}$ , where  $d$  is the Euclidean distance.

We can bound the external gauge fields by a constant, so we can drop them from the integral. We are thus left with the integral

$$(8.29) \quad \int_{p_2, p_3 > \epsilon, v_1, v_2, v_3} \frac{1}{d(p_3, v_3)^3 d(v_2, v_3)^3 d(v_2, p_2)^3 d(v_2, v_1)^3 d(v_1, p_1 = 0)^3},$$

where by overall translation invariance we set  $p_1 = 0$ . In our domain of integration, the most divergent region is when  $v_1, v_2, v_3, p_2$  are all near 0 (or all near  $p_3$ ). Focusing on this region is equivalent to taking  $p_3$

to be far away from the other points, which allows us to drop the  $d(p_3, v_3)^{-3}$  term since it is non-singular. We are then considering the integral

$$(8.30) \quad \int_{p_2, v_1, v_2, v_3} \frac{1}{d(0, v_1)^3 d(v_1, v_2)^3 d(v_2, p_1)^3 d(v_2, v_3)^3} d^4 v_1 d^4 v_2 d^4 v_3 d p_2.$$

There are 13 integration variables, and the integrand has weight  $-12$  under scaling all the variables. Therefore the integral converges absolutely on a domain where the integration variables  $v_i, p_2$  are bounded from above.

### 8.5.3 The Failure of the Amplitude to be Gauge Invariant

If we include all three diagrams that contribute to the anomaly, we find that the regularized amplitude is

$$(8.31) \quad \frac{1}{2} \left( \frac{i}{2\pi} \right)^3 \int_{p_{\min} < p_{\max} - \epsilon}^{p_i, v_j} \times P(0, v_1) dz_1 A_1^a P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3) \times f^{afc} f^{fgd} f^{gbe} \{ \rho(t^c), \rho(t^d), \rho(t^e) \},$$

where  $p_{\min}, p_{\max}$  are the coordinates on the Wilson line with the minimum and maximum value. The factor of  $\frac{1}{2}$  appears because there are 6 ways of ordering three points on the line, but only 3 diagrams we need to consider.

Let us now investigate the failure of the integral to be gauge invariant. Let us perform a linearized gauge transformation  $A \mapsto A + dc$  to the external gauge field  $A$ . By integration by parts, the result can be written as a sum of terms where we apply the exterior derivative  $d$  to one of the propagators, or else we integrate over one of the boundary components of the domain of integration. We will show that all such terms vanish except the term where we integrate over the boundary component in which  $p_{\max} = p_{\min} + \epsilon$ .

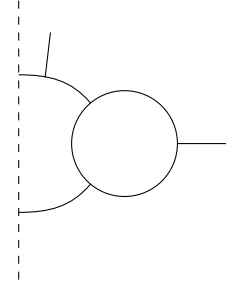
We use the identity (4.7). Applying the exterior derivative to a propagator has the effect of yielding a contribution where we integrate over the region where the vertices  $v_i, v_j$  (or  $p_i, v_j$ ) at either end of the propagator are identified.

If we contract the propagator connecting  $p_3$  and  $v_3$ , the result must vanish. This is because we can assume that the external gauge field  $A$  and the ghost  $c$  are both divisible by  $z$ . This assumption is justified because the anomaly always involves a  $z$ -derivative on each external field. If we integrate over the region where  $v_3 = p_3$ , then since  $z_3 = 0$  on this domain, the integrand vanishes.

If we contract the propagator connecting  $p_2$  and  $v_2$ , the result vanishes by the argument presented in section 8.3.4.

If we contract the propagator connecting  $v_1$  and  $v_2$ , or  $v_3$  or  $v_1$ , the results cancel by the Jacobi identity.

Next, let us consider the boundary components where two adjacent points on the Wilson line can meet. They can meet from above or below, with different signs. After using the Jacobi identity, the color factor associated to a boundary component like this is that given by the diagram



We have seen in section 8.3.3 that anomalies with a color factor of this form can be cancelled by a counter-term, and so are not relevant.

The remaining boundary component is the one with  $p_{\max} = p_{\min} + \epsilon$ . This is the one that will contribute to the anomaly.

The integral describing the anomaly is then

$$(8.32) \quad \left( \frac{i}{2\pi} \right)^3 \int_{p_{\min} = p_{\max} - \epsilon}^{p_i, v_j} \times P(p_1, v_1) dz_1 c_1^a P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3) \times f^{afc} f^{fgd} f^{gbe} \{ \rho(t^c), \rho(t^d), \rho(t^e) \},$$

where on vertex  $v_1$  the external gauge field has been replaced by the ghost  $c$ . The prefactor of  $\frac{1}{2}$  has been cancelled by the factor of 2 coming from the two possible external lines to the ghost field. The diagram automorphism interchanges these two, so we are left with the single integral above but without the prefactor of  $\frac{1}{2}$ .

This integral can be written as a sum of six terms, according to the six possible orderings of the points on the Wilson line. A reflection in a plane orthogonal to the Wilson line shows that there are only three independent integrals, which we can take to be the three where the points  $p_1, p_2, p_3$  are cyclically ordered. We find that we need to compute

$$(8.33) \quad 2 \left( \frac{i}{2\pi} \right)^3 \int_{p_1 < p_2 < p_3 = p_1 + \epsilon} \int_{v_1, v_2} \times P(p_1, v_1) dz_1 c_1^a P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3)$$

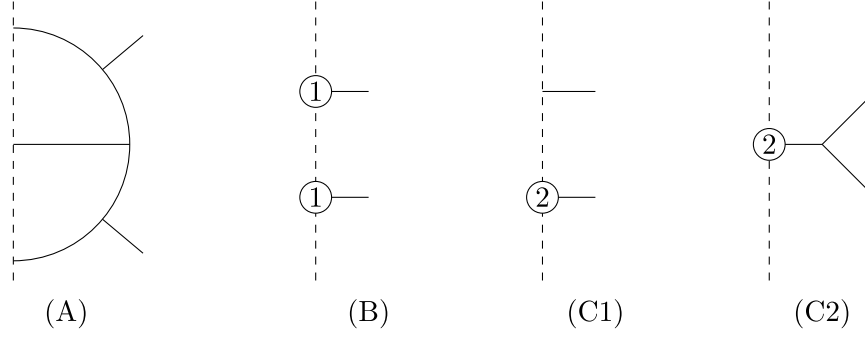


Figure 31. Four Feynman diagrams contributing to the anomaly.

$$\begin{aligned}
& + 2 \left( \frac{i}{2\pi} \right)^3 \int_{p_2 < p_3 < p_1 = p_2 + \epsilon} \int_{v_1, v_2} \\
& \times P(p_1, v_1) dz_1 c_1^a P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3) \\
& + 2 \left( \frac{i}{2\pi} \right)^3 \int_{p_3 < p_1 < p_2 = p_3 + \epsilon} \int_{v_1, v_2} \\
& \times P(p_1, v_1) dz_1 c_1^a P(v_1, v_2) dz_2 P(v_2, p_2) P(v_2, v_3) dz_3 A_3^b P(v_3, p_3).
\end{aligned}$$

The factor of 2 here is because each integral represents the contribution from one of two possible orderings on the points  $p_i$ .

#### 8.5.4 Calculating the Anomaly Integral

Because we know that the anomaly must involve a  $z$ -derivative for each external line and no other derivatives, we can detect the anomaly by assuming that the external ghost field  $c$  is  $z$ , and the external gauge field  $A$  is  $z\delta_{\bar{x}=0}$ .

The choice of external gauge field fixes the location of the vertex  $v_3$ ; by translation invariance we choose instead to fix the location of the vertex  $p_1$  to be  $p_1 = 0$ .

We will first evaluate the integral of the diagram in Fig. 30, where  $p_1 < p_2 < p_3$ , and we integrate over the region where  $p_3 = p_1 + \epsilon$ , and we set  $p_1 = 0$ . We denote  $p_2$  by  $p$ .

The integral for the first diagram of Fig. 28 is

$$(8.34) \quad 2 \left( \frac{i}{2\pi} \right)^3 \int_{p=0}^{\epsilon} \int_{v_1, v_2, v_3} P(0, v_1) \wedge z_1 dz_1 \wedge P(v_1, v_2) \wedge dz_2 \\
\wedge P(v_2, p) \wedge P(v_2, v_3) \wedge z_3 dz_3 \wedge P(v_3, \epsilon).$$

Now that the integrand has 5 propagator 2-forms, and 1 vertex 1-forms, making a 13-form, which is consistent since we wish to integrate over  $1 + 4 \times 3 = 13$  variables.

We evaluate this integral in Appendix E. Reintroducing the color factor and the factor of two present in eqn. (8.33), we find that the anomaly is given by

$$(8.35) \quad \frac{\hbar^2}{12} f^{afc} f^{fgd} f^{gbe} \{ \rho(t^c), \rho(t^d), \rho(t^e) \}.$$

## 8.6 The Anomaly as a Correction to the Algebra

In this section we will show that for a Wilson line to be defined at the quantum level, there must be quantum corrections to the coupling of the gauge field which satisfy a certain algebraic relation. This relation forces the Wilson line to be built from a representation of the Yangian, not of the Lie algebra  $\mathfrak{g}[[z]]$ .

Suppose we have a Wilson line in a representation  $V$  of our gauge group. Suppose that the level-one generators  $t_{a,1}$  of the Yangian act by some operators  $\rho(t_{a,1})$ . Consider the four two-loop diagrams in Fig. 31.

As before, a vertex on the Wilson line labelled with the number 1 or 2 indicates a coupling of the first or second  $z$ -derivative of the gauge field. (We tacitly include the variations on these diagrams where the ordering of the vertices on the Wilson lines has been permuted.)

The amplitude of each of these diagrams can fail to be gauge invariant. For the first diagram (A), this involves the two-loop computation that we have been describing. For the other diagrams, the failure to be gauge invariant is much more straightforward, as we will now see.

Suppose that at the quantum level, the first and second derivative of the gauge field are coupled to the Wilson line by operators  $\rho_{a,1} : V \rightarrow V$  and  $\frac{1}{2}\rho_{a,2} : V \rightarrow V$ , where  $a$  is an adjoint index.

Then, the amplitude for diagram (B) is given by

$$(8.36) \quad \int_{p_1 < p_2 \in \mathbb{R}} \partial_z A^a(p_1) \partial_z A^b(p_2) \rho_{a,1} \circ \rho_{b,1}.$$

If we change the gauge field by a gauge transformation  $A^a \mapsto A^a + dc^a$  we find (by integration by parts and Stokes' theorem) that the integral becomes

$$(8.37) \quad \int_{p \in \mathbb{R}} \partial_z c^a(p) \partial_z A^b(p) [\rho_{a,1}, \rho_{b,1}].$$

Similarly, for diagram (C1), the amplitude is

$$(8.38) \quad \frac{1}{2} \int_{p_1 < p_2 \in \mathbb{R}} (\partial_z^2 A^a(p_1)) A^b(p_2) \rho_{a,2} \circ \rho_{b,0}$$

$$+ \frac{1}{2} \int_{p_1 < p_2 \in \mathbb{R}} (A^a(p_1)) \partial_z^2 A^b(p_2) \rho_{a,0} \circ \rho_{b,2}.$$

The failure of this to be gauge invariant is given by

$$(8.39) \quad \frac{1}{2} \int_{p \in \mathbb{R}} (\partial_z^2 c^a(p)) A^b(p) [\rho_{a,2}, \rho_{b,0}] + \frac{1}{2} \int_{p \in \mathbb{R}} c^a(p) \partial_z^2 A^b(p) [\rho_{a,0}, \rho_{b,2}].$$

Next, diagram (C2) has amplitude

$$(8.40) \quad \frac{1}{4} \int_{p \in \mathbb{R}, v=(x,y,z)} dz \partial_z^2 P(p, v) A^a(v) A^b(v) f_{ab}^c \rho_{c,2}.$$

Here  $v$  indicates the position of the interior vertex, and  $\partial_{z_p}$  indicates we apply a  $z$ -derivative to the  $p$ -coordinate of the propagator. The failure of this to be gauge invariant is, using the fact that  $dz_v, dP(p, v) = -\delta_{p=v}$  and imposing the equations of motion for  $A$ ,

$$(8.41) \quad -\frac{1}{2} \int_{p \in \mathbb{R}} \partial_z^2 (c^a(p) A^b(p)) f_{ab}^c \rho_{c,2}.$$

Finally, the anomaly for diagram (A) is, as we have seen,

$$(8.42) \quad \frac{\hbar^2}{12} \int_{p \in \mathbb{R}} \partial_z c^a(p) \partial_z A^b(p) f^{afc} f^{fgd} f^{gbe} \{\rho(t^c), \rho(t^d), \rho(t^e)\}.$$

The sum of the anomalies for the diagrams (A), (B) and (C) must vanish. This gives rise to the equations

$$(8.43) \quad [\rho_{a,2}, \rho_{b,0}] = f_{ab}^c \rho_{c,2},$$

$$(8.44) \quad [\rho_{a,1}, \rho_{b,1}] = f_{ab}^c \rho_{c,2} - \frac{\hbar^2}{12} Q_{ab}(\rho_{\bullet,0}),$$

where the first equation is from the cancellation of the coefficients of  $(\partial_z^2 c^a(p)) A^b(p)$  and  $c^a(p) \partial_z^2 A^b(p)$ , and the second from those of  $(\partial_z c^a(p)) \partial_z A^b(p)$ . In the equations above we used the short-hand notation  $\rho_{a,n} = \rho(t_{a,n})$  and we defined

$$(8.45) \quad Q_{ab}(\rho_{\bullet,0}) = f^{afc} f^{fgd} f^{gbe} \{\rho(t^c), \rho(t^d), \rho(t^e)\}.$$

The first equation is the commutation relation we found classically. The second equation tells us that the operators  $\rho_{a,1}$  do not commute to give  $\rho_{c,2}$ , as we would find classically, but a linear combination of  $\rho_{c,2}$  and a certain cubic polynomial in the level 0 generators  $\rho_{d,0}$ .

We can decompose the exterior square  $\wedge^2 \mathfrak{g}$  into  $\mathfrak{g} \oplus \wedge_0^2 \mathfrak{g}$ , where  $\wedge_0^2 \mathfrak{g}$  is the kernel of the Lie bracket map from  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$ . By using this decomposition, we can write the second relation as a sum of two independent relations:

$$(8.46) \quad f^{ab}_c [\rho_{a,1}, \rho_{b,1}] = \hbar^\vee \rho_{c,2} - \frac{\hbar^2}{12} f^{ab}_c Q_{ab}(\rho_{\bullet,0}),$$

$$(8.47) \quad \Lambda^{ab}[\rho_{a,1}, \rho_{b,1}] = -\frac{\hbar^2}{12} \Lambda^{ab} Q_{ab}(\rho_{\bullet,0}),$$

where  $\Lambda^{ab} \in \wedge^2 \mathfrak{g}$ . The first relation can always be satisfied by redefining the operator  $\rho_{c,2}$  to

$$(8.48) \quad \rho'_{c,2} = \rho_{c,2} - \frac{\hbar^2}{12 \hbar^\vee} f^{ab}_c Q_{ab}(\rho_{\bullet,0}).$$

The second relation, however, can not be satisfied in such a trivial way.

What we have found from this analysis is that, in order for a classical Wilson line to quantize modulo  $\hbar^3$ , we must be able to define the operators  $\rho_{a,1}$  so that they satisfy the relation (8.47). This relation is one of the relations in the Yangian algebra which makes it into a non-trivial deformation of the universal enveloping algebra of  $\mathfrak{g}[[z]]$ .

The reader will note that the diagrams in Fig. 31 are simply the classical diagrams of Fig. 26, with quantum corrections to the vertices, plus the anomaly diagram (A). Note that we did not add to diagrams (C2) in the figure additional contributions involving a quantum correction to the bulk vertex. Such contributions can be omitted, using Proposition 8.3, as they involve a cohomology class that factors through a map to  $\mathfrak{g}$ .

#### 8.6.1 Matching with Standard Relations in the Yangian

Let us explain how to match the relation we have found with one of the known descriptions of the Yangian.

For any simple Lie algebra  $\mathfrak{g}$ , the sequence

$$(8.49) \quad \wedge^3 \mathfrak{g} \rightarrow \wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$$

is exact in the middle, where each map in the sequence is obtained by applying the Lie bracket to two entries in the exterior power. This sequence is part of the Chevalley-Eilenberg homology complex of  $\mathfrak{g}$ . For any simple Lie algebra,  $H_2(\mathfrak{g}, \mathbb{C}) = 0$ , which is why this sequence is exact in the middle.

Therefore, the kernel of the map from  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$  is the image of the map from  $\wedge^3 \mathfrak{g} \rightarrow \wedge^2 \mathfrak{g}$ . The antisymmetric tensor  $\Lambda^{ab} \in \wedge^2 \mathfrak{g}$  appearing in eqn. (8.47) is in the kernel of this map. Therefore, we lose no information by considering the relation in eqn. (8.47) when  $\Lambda^{ab}$  is taken to be in the image of the map from  $\wedge^3 \mathfrak{g}$ , that is, when  $\Lambda^{ab}$  is of the form

$$(8.50) \quad \Lambda^{ab} = \Gamma^{ade} f_{de}^b$$

for some antisymmetric tensor  $\Gamma^{ade} \in \wedge^3 \mathfrak{g}$ .

Relation (8.47) takes the form, in this notation,

$$(8.51) \quad \Gamma^{ade} f_{de}^b [\rho_{a,1}, \rho_{b,1}] = -\frac{\hbar^2}{12} \Gamma^{ade} f_{de}^b Q_{ab}(\rho_{\bullet,0}).$$



For an element  $t \in \mathfrak{g}$ , we let  $\rho(t)$  denote the action of  $t$  on our representation, and  $\rho(J(t))$  denote the action of the corresponding level one generator on our representation. Then our relation is

$$(8.52) \quad \begin{aligned} & \Gamma^{abc}[\rho(J(t_a)), \rho(J([t_b, t_c]))] \\ &= -\frac{\hbar^2}{12} \Gamma^{abc}([t_e, [t_a, t_d]], [t_f, [t_b, t_c]]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}, \end{aligned}$$

where as before  $\{\rho(t_d), \rho(t_e), \rho(t_f)\}$  is 1/6 times the sum of the products of these operators in each of the six possible orders (see eqn. (8.25)). The Jacobi identity and the  $\mathfrak{g}$ -invariance of the Killing form allows us to rewrite

$$\begin{aligned} & ([t_e, [t_a, t_d]], [t_f, [t_b, t_c]]) \\ &= ([t_e, [t_a, t_d]], [[t_f, t_b], t_c]) + ([t_e, [t_a, t_d]], [t_b, [t_f, t_c]]) \\ &= ([t_c, [t_e, [t_a, t_d]]], [t_f, t_b]) - ([t_b, [t_e, [t_a, t_d]]], [t_f, t_c]). \end{aligned}$$

Thus anti-symmetrizing in  $a, b, c$  and again using the Jacobi identity, we obtain

$$\begin{aligned} & \Gamma^{abc}([t_e, [t_a, t_d]], [t_f, [t_b, t_c]]) \\ &= 2\Gamma^{abc}([t_c, [t_e, [t_a, t_d]]], [t_f, t_b]) \\ &= 2\Gamma^{abc}([[t_c, t_e], [t_a, t_d]], [t_f, t_b]) + 2\Gamma^{abc}([t_e, [t_c, [t_a, t_d]]], [t_f, t_b]) \\ &= -2\Gamma^{abc}([[t_c, t_e], [t_a, t_d]], [t_b, t_f]) \\ &+ \Gamma^{abc}([t_e, [[t_c, t_a], t_d]], [t_f, t_b]), \end{aligned}$$

where on the last line we have again used anti-symmetry of  $\Gamma^{abc}$  to conclude that we can replace  $[t_c, [t_a, t_d]]$  by  $\frac{1}{2}[[t_c, t_a], t_d]$ . The second factor in the final line, after combining with the factor  $\{\rho(t_d), \rho(t_e), \rho(t_f)\}$ , gives back the expression we started with, with an overall minus sign:

$$\begin{aligned} & \Gamma^{abc}([t_e, [[t_c, t_a], t_d]], [t_f, t_b]) \{\rho(t_d), \rho(t_e), \rho(t_f)\} \\ &= -\Gamma^{abc}([t_e, [t_f, t_b]], [[t_c, t_a], t_d]) \{\rho(t_d), \rho(t_e), \rho(t_f)\} \\ &= -\Gamma^{abc}([t_e, [t_a, t_d]], [t_f, [t_b, t_c]]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}, \end{aligned}$$

where we used  $\mathfrak{g}$ -invariance of the Killing form, anti-symmetry in  $a, b, c$  and symmetry in  $d, e, f$ . This allows us to conclude that

$$(8.53) \quad \begin{aligned} & \Gamma^{abc}([t_e, [t_a, t_d]], [t_f, [t_b, t_c]]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}, \\ &= -\Gamma^{abc}([[t_c, t_e], [t_a, t_d]], [t_b, t_f]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}. \end{aligned}$$

Thus the quantum-corrected commutation relation (8.52) becomes

$$(8.54) \quad \begin{aligned} & \Gamma^{abc}[\rho(J(t_a)), \rho(J([t_b, t_c]))] \\ &= \frac{\hbar^2}{12} \Gamma^{abc}([[t_c, t_e], [t_a, t_d]], [t_b, t_f]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}. \end{aligned}$$

Note that by anti-symmetry of  $\Gamma^{abc}$  and symmetry of  $\{\rho(t_d), \rho(t_e), \rho(t_f)\}$  we have

$$(8.55) \quad \begin{aligned} & \Gamma^{abc}[\rho(J(t_a)), \rho(J([t_b, t_c]))] \\ &= \frac{1}{3} \Gamma^{abc}([\rho(J(t_a)), \rho(J([t_b, t_c]))] + [\rho(J(t_b)), \rho(J([t_c, t_a]))] \\ &+ [\rho(J(t_c)), \rho(J([t_a, t_b]))]). \end{aligned}$$

Also,  $([[t_c, t_e], [t_a, t_d]], [t_f, t_b]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}$  is antisymmetric in  $a, b, c$ . Therefore our relation is equivalent to

$$(8.56) \quad \begin{aligned} & [\rho(J(t_a)), \rho(J([t_b, t_c]))] \\ &+ [\rho(J(t_b)), \rho(J([t_c, t_a]))] + [\rho(J(t_c)), \rho(J([t_a, t_b]))] \\ &= \frac{\hbar^2}{4} ([[t_c, t_e], [t_a, t_d]], [t_b, t_f]) \{\rho(t_d), \rho(t_e), \rho(t_f)\} \\ &= \frac{\hbar^2}{4} ([t_a, t_d], [[t_b, t_e], [t_c, t_f]]) \{\rho(t_d), \rho(t_e), \rho(t_f)\}. \end{aligned}$$

This is the same as the known relation for the Yangian, see e.g. eqn. (4) in [8, Theorem 12.1.1].

## 8.7 An Example of an Anomalous Wilson Line

It is known that the adjoint representation of any simple Lie algebra not of type  $A$  does not lift to a representation of the Yangian. This indicates that a Wilson line in the adjoint representation should be anomalous. In this section we will see explicitly that our two-loop anomaly cannot be cancelled for the adjoint representation of  $SO_N$ .

Let us start with some group-theory background. If  $M_1, \dots, M_k \in \mathfrak{so}_N$  (for  $N$  sufficiently large), the ring of invariant function of these  $k$  elements of  $\mathfrak{so}_N$  is generated by the invariant functions given by traces, that is, the functions  $\text{Tr}(M_{i_1} \dots M_{i_r})$  where  $i_1, \dots, i_r$  are in the set  $\{1, \dots, k\}$ . Repeated indices are allowed and the trace is taken in the fundamental representation.

For finite  $N$ , there are trace relations among these functions. For  $N$  large compared to the degree of the polynomial in the variables  $M_i$ , the only relation that survives is the relation that says

$$(8.57) \quad \text{Tr}(M_{i_1} \dots M_{i_r}) = (-1)^r \text{Tr}(M_{i_r} \dots M_{i_1}).$$

We will need to understand the decomposition of  $\mathfrak{so}_N \otimes \mathfrak{so}_N$  into irreducible representations. The number of irreducible representations is the same as the number of invariant tensors in  $\mathfrak{so}_N^{\otimes 4}$ . Let  $M_1, M_2, N_1, N_2$  denote elements of  $\mathfrak{so}_N$ . The analysis above allows us to enumerate the invariant functions of these matrices which are linear in each variable. We find that there are 6 invariant tensors, three with a single trace and three with two traces. Of these, 4 of them are

symmetric when we exchange  $M_1$  and  $M_2$ , or when we exchange  $N_1$  and  $N_2$ . These correspond to projectors onto irreducible representations inside  $\text{Sym}^2 \mathfrak{so}_N$ . Two of them are antisymmetric under the same permutations, and correspond to projectors onto irreducible representations in  $\wedge^2 \mathfrak{so}_N$ . The two antisymmetric tensors are

$$(8.58) \quad \begin{aligned} & \text{Tr}([M_1, M_2][N_1, N_2]), \\ & \text{Tr}(M_1 N_1) \text{Tr}(M_2 N_2) - \text{Tr}(M_1 N_2) \text{Tr}(M_2 N_1). \end{aligned}$$

The first antisymmetric tensor gives rise to the projection from  $\wedge^2 \mathfrak{so}_N$  onto the adjoint representation, and the second to the projection onto the irreducible representation  $\wedge_0^2 \mathfrak{so}_N$ .

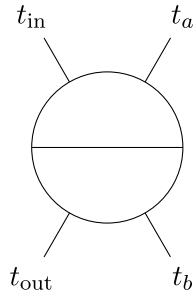
Note also that there is only one copy of the adjoint representation in  $\mathfrak{so}_N \otimes \mathfrak{so}_N$ . Any copy of the adjoint representation is associated to an invariant element in  $\mathfrak{so}_N^{\otimes 3}$ , and so an invariant function of  $M_1, M_2, M_3$  which is linear in each  $M_i$ . The discussion above tells us that the only such invariant function is  $\text{Tr}(M_1 M_2 M_3)$ .

If we have a Wilson line in the adjoint representation of  $SO_N$ , it follows from this that the level one generators must be coupled by a multiple of the level 0 generators. Therefore, the commutator of the level 1 generators must be proportional to the level 0 generators. Relation (8.44) can never hold unless

$$(8.59) \quad Q_{ab}(\rho_{\bullet,0}) = f^{afc} f^{fgd} f^{gbe} \{\rho(t^c), \rho(t^d), \rho(t^e)\}$$

is a multiple of some  $\rho(t^c)$ . We will show the projection of the tensor  $Q_{ab}$  onto  $\wedge_0^2(\mathfrak{so}_N)$  is non-zero, so the anomaly is non-zero.

Let us introduce incoming and outgoing states  $t_{\text{in}}, t_{\text{out}}$  on the adjoint Wilson line. The three diagrams which give us an anomaly all, as we have seen, have the same color factor. Since  $\rho$  is in the adjoint representation, we can compute the color factor using a diagram in which the Wilson line is placed on the same footing as the other lines in the Feynman diagram. If we do this, the diagram looks like



where the segment connecting  $t_{\text{in}}, t_{\text{out}}$  is what was the Wilson line. Here we are implicitly anti-symmetrizing in  $a$  and  $b$ , since the anomaly is always antisymmetric in the external lines.

This makes it clear that the color factor of the anomaly in the adjoint representation is

$$(8.60) \quad (t_{\text{out}}, Q_{ab}(\rho_{\bullet,0}) t_{\text{in}}) = \text{Tr}(t_{\text{in}} t_a t_f t_b t_{\text{out}} t_f) - \text{Tr}(t_{\text{in}} t_b t_f t_a t_{\text{out}} t_f),$$

where the trace is taken in the adjoint representation, and we sum over  $f$ .

Let us view the anomaly (8.60) as a linear operator from  $\mathfrak{so}_N^{\otimes 2} \rightarrow \mathfrak{so}_N^{\otimes 2}$ , where the first two copies of  $\mathfrak{so}_N$  are given by  $t_a, t_b$  and the second two by  $t_{\text{in}}, t_{\text{out}}$ . Since this is an  $SO(N)$  invariant operator, it is a linear combination of the projection onto the 6 irreducible representations in  $\mathfrak{so}_N^{\otimes 2}$ . The anomaly is non-zero as long as the coefficient of the projection onto  $\wedge_0^2 \mathfrak{so}_N$  is non-zero.

Equivalently, we can expand the expression (8.60) as a linear combination of the 6 invariant tensors in  $\mathfrak{so}_N^{\otimes 4}$ . To show that the anomaly is non-zero it suffices to show that, in this expansion, the coefficient of

$$(8.61) \quad \text{Tr}(t_{\text{in}} t_a) \text{Tr}(t_{\text{out}} t_b) - \text{Tr}(t_{\text{in}} t_b) \text{Tr}(t_{\text{out}} t_a)$$

is non-zero. This is because this invariant tensor corresponds to projection onto  $\wedge_0^2 \mathfrak{so}_N$ . Since the anomaly is antisymmetric in  $a$  and  $b$  we need to verify that the coefficient of  $\text{Tr}(t_{\text{in}} t_a) \text{Tr}(t_{\text{out}} t_b)$  is non-zero.

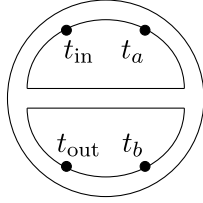
We can evaluate the expression in equation (8.60) for  $\mathfrak{so}_N$  using the double-line notation familiar from evaluation of the color factors for  $\mathfrak{gl}_N$  gauge theories. Since we are using  $\mathfrak{so}_N$  instead of  $\mathfrak{gl}_N$ , the double-line technique works a little differently. There are two types of double-line edges, which we can think of as a flat ribbon and a ribbon with a half-twist. This is because the Casimir for  $\mathfrak{so}_N$  is represented as

$$(8.62) \quad \sum E_{ij} \otimes E_{ji} - \sum E_{ij} \otimes E_{ij},$$

where  $E_{ij}$  is the elementary matrix. Only the first term appears in the Casimir of  $\mathfrak{gl}_N$ , which is why only the first type of double-line edge appears when we study  $\mathfrak{gl}_N$  color factors.

Using the double line notation, we find that the color factor we are computing is given by a sum over connected unoriented surfaces of Euler characteristic  $-1$ , with 4 marked points on the boundary. Since we are interested in the double-trace terms in the anomaly, we need to consider surfaces where two of the marked points (labelled by  $t_{\text{in}}, t_a$ ) are on one boundary component, and the other two on a different component.

The order  $N$  two-trace term in the anomaly is given by the planar diagram



which contributes

$$(8.63) \quad N \text{Tr}(t_{in} t_a) \text{Tr}(t_b t_{out}).$$

(We have not drawn the similar diagram which contributes the same expression with  $a$  and  $b$  exchanged and a different sign.)

There is also an order 1 term, given by a double-line diagram which has the topology of a Möbius band with a circle removed. For  $N$  sufficiently large ( $N = 10$  certainly suffices) the coefficient of this term is much smaller than that of the order  $N$  term, so the anomaly is non-zero.

## 9. Trigonometric Solutions of the Yang-Baxter Equations

### 9.1 Preliminaries

Our goal in this section will be to understand from the present perspective the trigonometric solutions of the Yang-Baxter equation. For this we have to consider the case that  $C = \mathbb{C}^\times$ , or equivalently  $C = \mathbb{CP}^1$  with a differential  $\omega = dz/z$  that has two simple poles at 0 and  $\infty$ .

The basic consequence of a pole in  $\omega$  is the following. One has

$$(9.1) \quad \bar{\partial} \frac{dz}{z} = 2\pi i \delta^2(z),$$

where  $\delta^2(z)$  is a delta function normalized by  $\int |d^2z| \delta^2(z) = 1$ . Accordingly, when we vary the action (3.3) to derive the equations of motion, we pick up a “boundary term” supported at  $z = 0$ :

$$(9.2) \quad \delta S = \dots = -\frac{i}{\hbar} \int_{\Sigma \times \{0\}} \text{Tr} A \wedge \delta A,$$

where bulk terms have been omitted. The equations of motion of the classical theory tell us to set to zero the bulk terms and also the boundary term  $\text{Tr} A \wedge \delta A|_{\Sigma \times \{0\}}$ . We will abbreviate the latter as  $\text{Tr} A \wedge \delta A|$  (and we use a similar notation in general for restriction to a singularity or boundary).

Actually, the boundary terms in the equations of motion play a distinguished role. Quantum field theory is constructed by integrating over fields that do not necessarily satisfy the equations of motion in bulk. The bulk equations of motion are only satisfied in the classical limit. But boundary terms in the

equations of motion have to vanish exactly. This is needed in order to get in bulk a symmetric propagator, as expected for bosons (or an antisymmetric one in an analogous problem with fermions). In general, in quantum field theory in the presence of boundaries or defects, one typically runs into “boundary terms” analogous to  $\text{Tr} A \wedge \delta A$ . As a starting point in quantization, one always needs to impose a minimal condition that sets the boundary terms to 0. For example, if one quantizes on a manifold  $X$  an ordinary scalar field with the usual Lagrangian  $\frac{1}{2}|d\phi|^2$ , the boundary term is  $\int_{\partial X} \delta\phi \partial_n \phi$  (where  $\partial_n$  is the normal derivative). The simplest way to dispose of this boundary term is to set either  $\phi| = 0$  (Dirichlet boundary conditions) or  $\partial_n \phi| = 0$  (Neumann boundary conditions). With either of these conditions, the theory can be quantized. By contrast, a stronger condition such as  $\phi = \partial_n \phi = 0$  is too strong and does not lead to a quantum theory of a scalar field.

What condition on  $A$  will we use to set  $\text{Tr} A \wedge \delta A| = 0$ ? The most obvious condition might seem to be  $A| = 0$ . However, this condition is too strong, analogous to  $\phi| = \partial_n \phi| = 0$  for the scalar field. If we impose it, a suitable propagator will not exist and we will not be able to do perturbation theory.<sup>36</sup>

We can gain some intuition by considering the case of a double pole. Of course, we have already studied double poles in the context of rational solutions of the Yang-Baxter equation. In that analysis, we considered the case that  $\omega = dz$  with a double pole at infinity. For our present purposes, it is more convenient to place the pole at a finite point, so we take  $\omega = dz/z^2$ , with a double pole at the origin. Since

$$(9.3) \quad \bar{\partial} \frac{dz}{z^2} = -2\pi i \partial_z \delta^2(z),$$

the vanishing of the surface term now requires

$$(9.4) \quad 0 = (\partial_z \text{Tr} A \wedge \delta A)|.$$

A completely natural way to satisfy this condition is to set  $A| = 0$ . Varying  $A$  with the constraint  $A| = 0$ , we have also  $\delta A| = 0$ , so that  $\text{Tr} A \wedge \delta A$  has a double zero at  $z = 0$  and eqn. (9.4) is obeyed. The condition  $A| = 0$  is gauge-invariant if we likewise constrain the generator  $\varepsilon$  of a gauge transformation to obey  $\varepsilon| = 0$ .

Not only is  $A| = 0$  a natural way to satisfy eqn. (9.4), it is what we have actually done in studying rational solutions of the Yang-Baxter equation. Starting in section 4, we used a propagator which vanishes at infinity, which amounts to taking  $A$  to vanish at infinity (where we placed the double pole). As

<sup>36</sup> Later, we will impose a weaker condition and find a unique solution for the  $r$ -matrix. A similar analysis assuming that  $A| = 0$  will show that no possible  $r$ -matrix exists. This is actually clear from the fact that the unique  $r$ -matrix we get with a weaker condition does not obey  $A| = 0$ .

explained in section 3.4, this choice has the advantage of ensuring that up to gauge transformation, the only classical solution is the trivial one  $A = 0$ , a condition that is needed if we want to obtain something as simple as the usual Yang-Baxter equation. As we will explain below, there are other reasonable ways to satisfy  $(\partial_z \text{Tr} A \wedge \delta A)| = 0$ , but they lead to something more complicated.

If  $A| = 0$  is a satisfactory condition in the presence of a double pole, it can hardly be the right condition for a simple pole. If we start with a differential  $\omega = dz/z^2$  with a double pole, and perturb it to, say,  $\omega' = dz/(z - z_1)(z - z_2)$ , with small  $z_1, z_2$ , the condition  $A| = 0$  at a double pole must somehow be split between the two simple poles at  $z_1$  and  $z_2$ .

As remarked above, when one studies field theory in the presence of a boundary or defect, the right procedure is always to impose a minimally restrictive condition that sets this term to zero. But the condition  $A| = 0$  is not a minimal condition to ensure that  $\text{Tr} A \wedge \delta A| = 0$ . For this, it suffices to pick a middle-dimensional complex subspace  $\mathfrak{l}_0 \subset \mathfrak{g}$  that is “isotropic” (or “Lagrangian”) for the quadratic form  $\text{Tr}$ , in the sense that for  $a, b \in \mathfrak{l}_0$ ,  $\text{Tr} ab = 0$ . (We take  $\mathfrak{l}_0$  to be middle-dimensional because this is the maximum possible dimension for an isotropic subspace, leading to the weakest possible condition on  $A$ .) Then we require that  $A|$  is  $\mathfrak{l}_0$ -valued, that is, that it is an  $\mathfrak{l}_0$ -valued 1-form along  $\Sigma$ . Having imposed this condition on  $A$ , we impose it also on  $\delta A$ , and then we see that with  $A|$  and  $\delta A|$  being  $\mathfrak{l}_0$ -valued 1-forms,  $\text{Tr} A \wedge \delta A| = 0$ .

One may object that the condition for  $A|$  to be  $\mathfrak{l}_0$ -valued is not gauge-invariant. However, what we actually need is not to maintain the full gauge symmetry along the locus of the pole, but only to maintain enough gauge symmetry so that the usual “longitudinal” part of  $A|$  can be gauged away. For this, we simply ask that  $\mathfrak{l}_0$  should be a subalgebra of  $\mathfrak{g}$  (and not just a subspace) and we ask that the generator  $\varepsilon$  of a gauge transformation should satisfy the condition that  $\varepsilon|$  is  $\mathfrak{l}_0$ -valued. These conditions on  $A$  and  $\varepsilon$  are compatible in the sense that, with both  $A|$  and  $\varepsilon|$  constrained to be  $\mathfrak{l}_0$ -valued, the usual gauge transformation law  $\delta A = D\varepsilon$  makes sense, and moreover the action (3.3) is gauge-invariant. In effect, what has happened is the following. Along the locus of the pole, we have set to zero some components of  $A$ , and the other components have their usual gauge-invariance.

We are really interested in a situation in which  $\omega$  has two simple poles, say  $\omega = dz/z$  with poles at 0 and at  $\infty$ . Treating each simple pole as above, we pick two middle-dimensional isotropic subalgebras of  $\mathfrak{g}$ , say  $\mathfrak{l}_0$  and  $\mathfrak{l}_\infty$ , in general with no relation between them. We require  $A$  and  $\varepsilon$  to be  $\mathfrak{l}_0$ -valued when restricted to  $\Sigma \times \{0\}$ , and  $\mathfrak{l}_\infty$ -valued when restricted to  $\Sigma \times \{\infty\}$ .

To get from this construction a solution of the usual Yang-Baxter equation (as opposed to the “dynamical Yang-Baxter equation,” which we study in section 11), we need a further condition that is familiar from section 3.4: the trivial solution  $A = 0$  should have no deformations and no continuous unbroken gauge symmetries. The two conditions are equivalent for the following reason. Let  $\mathfrak{g}_{0,\infty}$  be the sheaf of holomorphic  $\mathfrak{g}$ -valued functions on  $\mathbb{CP}^1$  that are  $\mathfrak{l}_0$ -valued at 0 and  $\mathfrak{l}_\infty$ -valued at  $\infty$ . The Lie algebra of the group of gauge symmetries of the trivial solution  $A = 0$  is  $H^0(\mathbb{CP}^1, \mathfrak{g}_{0,\infty})$ , and the tangent space to  $A = 0$  in the moduli space of classical solutions of the theory is  $H^1(\mathbb{CP}^1, \mathfrak{g}_{0,\infty})$ . Thus the condition that  $A = 0$  has no deformations is  $H^1(\mathbb{CP}^1, \mathfrak{g}_{0,\infty}) = 0$ , and the condition that it has no continuous gauge symmetries is  $H^0(\mathbb{CP}^1, \mathfrak{g}_{0,\infty}) = 0$ . These two conditions are equivalent because of the Riemann-Roch theorem, which in the present situation implies that

$$(9.5) \quad \dim H^0(\mathbb{CP}^1, \mathfrak{g}_{0,\infty}) - \dim H^1(\mathbb{CP}^1, \mathfrak{g}_{0,\infty}) = 0.$$

Concretely,  $H^0(\mathbb{CP}^1, \mathfrak{g}_{0,\infty})$  is simply<sup>37</sup>  $\mathfrak{l}_0 \cap \mathfrak{l}_\infty$ . Thus, our two conditions are equivalent to

$$(9.6) \quad \mathfrak{l}_0 \cap \mathfrak{l}_\infty = 0.$$

On dimensional grounds, this is equivalent to

$$(9.7) \quad \mathfrak{l}_0 + \mathfrak{l}_\infty = \mathfrak{g}.$$

We can understand directly the role of this last condition. If  $\mathfrak{g}' = \mathfrak{l}_0 + \mathfrak{l}_\infty$  and  $\mathfrak{g} = \mathfrak{g}' \oplus \mathfrak{g}''$ , then the  $\mathfrak{g}''$ -valued part of  $A$  cannot be gauged away<sup>38</sup> and therefore the trivial solution  $A = 0$  has deformations. So absence of deformations means that  $\mathfrak{g}'' = 0$  and  $\mathfrak{g} = \mathfrak{l}_0 + \mathfrak{l}_\infty$  (since  $\mathfrak{l}_0 \cap \mathfrak{l}_\infty = 0$ , this is equivalent to  $\mathfrak{g} = \mathfrak{l}_0 \oplus \mathfrak{l}_\infty$ ).

Accordingly, to get a trigonometric solution of the Yang-Baxter equation, rather than its “dynamical” generalization, we need  $\mathfrak{l}_0 + \mathfrak{l}_\infty = \mathfrak{g}$ . We have arrived precisely at the notion of a “Manin triple” (see for example [8, p. 26]). A Manin triple is a complex Lie algebra  $\mathfrak{g}$  with an invariant, nondegenerate quadratic form  $\text{Tr}$ , and a decomposition  $\mathfrak{g} = \mathfrak{l}_0 + \mathfrak{l}_\infty$ , where  $\mathfrak{l}_0$  and  $\mathfrak{l}_\infty$  are middle-dimensional isotropic subalgebras. So trigonometric solutions of the Yang-Baxter equation, or at least the ones that we will study, are associated to Manin triples.<sup>39</sup>

<sup>37</sup> A global holomorphic section of  $\mathfrak{g}_{0,\infty}$  is a  $\mathfrak{g}$ -valued constant (or it would have singularities somewhere) that must be valued in  $\mathfrak{l}_0 \cap \mathfrak{l}_\infty$  because of the conditions at 0 and at  $\infty$ .

<sup>38</sup> In trying to do so, one runs into the fact that  $H^1(\mathbb{CP}^1, \mathcal{O}(-p-q)) \neq 0$ , where  $\mathcal{O}(-p-q)$  is the sheaf of holomorphic functions that vanish at two points  $p$  and  $q$  (here those two points are  $z = 0$  and  $\infty$ ).

<sup>39</sup> The construction as we have described it involves a Manin triple of  $\mathfrak{g}$ , but one could actually in a somewhat similar way use a Manin triple of  $\mathfrak{g}[z, z^{-1}]$  (endowed with the nondegener-



At this stage, it is perhaps also clear that in contrast to the rational solutions of Yang-Baxter, the trigonometric ones do not have  $G$  as a group of global symmetries. In the rational case, given the differential  $\omega = dz/z^2$ , the condition on the generator of a gauge transformation at  $z = 0$  was  $\varepsilon| = 0$ . This condition leaves constant gauge transformations at  $z = 0$  as global symmetries. In the trigonometric case, a constant unbroken gauge symmetry, to be compatible with the conditions at 0 and at  $\infty$ , has to be an element of  $G$  that conjugates  $l_0$  to itself, and also conjugates  $l_\infty$  to itself. In practice, in the main example that will be introduced in section 9.2, this means that the group of global symmetries is the maximal torus of  $G$ . (This symmetry accounts for the block diagonal form of eqn. (9.14).)

Before leaving this subject, let us note that we stated the condition  $l_0 + l_\infty = \mathfrak{g}$  in a somewhat naive way. In gauge theory, any comparison between  $l_0$  and  $l_\infty$  involves parallel transport from 0 to  $\infty$ . What the condition  $l_0 + l_\infty = \mathfrak{g}$  really means is that, after conjugating  $l_0$  and  $l_\infty$  into general position,  $l_0 + l_\infty = \mathfrak{g}$ . The case that would lead to a solution of the dynamical Yang-Baxter equation rather than the ordinary one is that  $l_0$  and  $l_\infty$  are such that even after conjugating them into general position,  $l_0 + l_\infty \neq \mathfrak{g}$ .

## 9.2 Example

Given  $\mathfrak{g}$ , can we pick  $l_0$  and  $l_\infty$  to make a Manin triple? In general, the answer is certainly “no,” since if  $\mathfrak{g}$  is of odd dimension – for example  $\mathfrak{sl}_2$  – it has no middle-dimensional isotropic subspaces. However, the following is a useful construction of examples that are related to the usual trigonometric solutions of the Yang-Baxter equation.

For any simple Lie algebra  $\mathfrak{g}$ , add to  $\mathfrak{g}$  another copy  $\tilde{\mathfrak{h}}$  of the Cartan subalgebra  $\mathfrak{h}$  of  $\mathfrak{g}$  to make a Lie algebra  $\tilde{\mathfrak{g}} = \mathfrak{g} \oplus \tilde{\mathfrak{h}}$ . Equip  $\tilde{\mathfrak{h}}$  with a quadratic form given by the restriction of the Killing form on  $\mathfrak{g}$ . In this way  $\tilde{\mathfrak{g}}$  acquires an invariant nondegenerate pairing. We will construct a Manin triple for  $\tilde{\mathfrak{g}}$ .

Choose a decomposition  $\mathfrak{g} = \mathfrak{n}_- \oplus \mathfrak{h} \oplus \mathfrak{n}_+$  into nilpotent and Cartan subalgebras. Set

$$(9.8) \quad \begin{aligned} \mathfrak{h}_+ &= \{(X, iX) | X \in \mathfrak{h}\} \subset \mathfrak{h} \oplus \tilde{\mathfrak{h}}, \\ \mathfrak{h}_- &= \{(X, -iX) | X \in \mathfrak{h}\} \subset \mathfrak{h} \oplus \tilde{\mathfrak{h}}. \end{aligned}$$

Then we can choose a pair of complementary Lagrangian subalgebras of  $\tilde{\mathfrak{g}}$  by

ate quadratic form  $(a, b) = \oint \frac{dz}{z} \text{Tr} a(z)b(z)$ ). This would involve a construction somewhat like that above, with a more complicated set of conditions on  $A$  and  $\varepsilon$ . The relation between Manin triples of  $\mathfrak{g}$  and of  $\mathfrak{g}[z, z^{-1}]$  is that a Manin triple of  $\mathfrak{g}$  determines a Manin triple of  $\mathfrak{g}[z, z^{-1}]$ , namely  $\mathfrak{g}[z, z^{-1}] = \widehat{l_0} \oplus \widehat{l_\infty}$ , with  $\widehat{l_0} = z\mathfrak{g}[z] \oplus l_0$ ,  $\widehat{l_\infty} = z^{-1}\mathfrak{g}[z^{-1}] \oplus l_\infty$ .

$$(9.9) \quad \begin{aligned} l_0 &= \mathfrak{n}_+ \oplus \mathfrak{h}_+, \\ l_\infty &= \mathfrak{n}_- \oplus \mathfrak{h}_-. \end{aligned}$$

This gives a Manin triple for  $\tilde{\mathfrak{g}}$ .

## 9.3 The $r$ -Matrix

Now we consider our four-dimensional theory on  $\mathbb{R}^2 \times \mathbb{P}^1$ , where we use the 1-form  $dz/z$  on  $\mathbb{P}^1$  and the Manin triple just described. Let us place Wilson lines at  $z_1, z_2 \in \mathbb{C}^\times$ , and have them cross in the topological plane as in Fig. 15. Suppose the Wilson lines are in representations  $V, W$  of  $\mathfrak{g}$ . We will show that the result of this crossing will be the trigonometric  $R$ -matrix.

Let us first calculate explicitly what happens to leading order in  $\hbar$ . To leading order in  $\hbar$ , the result of the crossing Wilson lines is described by  $R = 1 + \hbar r$ , where  $r$  is a  $\tilde{\mathfrak{g}} \otimes \tilde{\mathfrak{g}}$ -valued meromorphic function. To obtain  $r(z_1, z_2)$ , we can pick a gauge and then repeat the Feynman diagram computations of the previous section. Instead let us here note that  $r(z_1, z_2)$  must satisfy the following properties:

1.  $r(z_1, z_2)$  has a first-order pole at  $z_1 = z_2$ , and is regular elsewhere. The residue of this pole is  $\hbar c$  where  $c \in \tilde{\mathfrak{g}} \otimes \tilde{\mathfrak{g}}$  is the Casimir element, dual to the chosen invariant pairing. This is the same singular behavior as in the rational case, since OPE singularities are local and are not affected by global topology.
2. At  $z_1 = 0$ ,  $r(z_1, z_2)$  is in  $l_0 \otimes l_\infty$ , and at  $z_1 = \infty$  it is in  $l_\infty \otimes l_0$ . This just reflects the corresponding conditions on  $A$ .
3.  $r(z_1, z_2)$  is sent to  $-r(z_1, z_2)$  if we simultaneously swap  $z_1$  and  $z_2$  and exchange the tensor factors in  $\tilde{\mathfrak{g}} \otimes \tilde{\mathfrak{g}}$ . (This reflects the fact that the action of the theory – and likewise the propagator and the Feynman diagram used to compute  $r(z_1, z_2)$  – are all odd under an orientation-reversing symmetry of the topological plane that exchanges the two oriented Wilson lines that are crossing. Note that the rational  $r$ -matrix  $r = t_a \otimes t_a / (z_1 - z_2)$  has the same property.)
4. Finally,  $r(z_1, z_2)$  is a function only of the ratio  $z_1/z_2$ , since the differential  $\omega = dz/z$  the conditions we placed at 0 and  $\infty$  are all invariant under the  $\mathbb{C}^\times$  action on  $C = \mathbb{C}^\times$ .

There is a unique function satisfying these properties. To write it down, we observe that the Casimir  $c = \sum_a t_a \otimes t_a \in \tilde{\mathfrak{g}} \otimes \tilde{\mathfrak{g}}$  can be written in a unique way as a sum

$$(9.10) \quad c = c(l_0, l_\infty) + c(l_\infty, l_0),$$

where  $c(l_0, l_\infty) \in l_0 \otimes l_\infty$ , and similarly  $c(l_\infty, l_0) \in l_\infty \otimes l_0$ .



Then the unique function  $r(z_1, z_2)$  satisfying the desired conditions is

$$(9.11) \quad 2\pi i r(z_1, z_2) = \frac{c(l_0, l_\infty)}{1 - \frac{z_1}{z_2}} - \frac{c(l_\infty, l_0)}{1 - \frac{z_2}{z_1}}.$$

Let us write this out more explicitly. Let us choose a basis  $X_\alpha^\pm, H_i$  of  $\mathfrak{g}$ , so that  $X_\alpha^\pm \in \mathfrak{n}_\pm$  and  $H_i$  form a basis of  $\mathfrak{h}$ . We assume that in this basis the chosen pairing on  $\mathfrak{g}$  is such that  $\langle X_{\alpha^+}, X_{\beta^-} \rangle = \delta_{\alpha\beta}$ , and  $\langle H_i, H_j \rangle = \delta_{ij}$ . Let  $\tilde{H}_i$  be the corresponding basis of the other copy  $\tilde{\mathfrak{h}}$  of the Cartan. Then, in this basis, we have

$$\begin{aligned} & 2\pi i r(z_1, z_2) \\ &= \frac{1}{1 - \frac{z_1}{z_2}} \sum_{\alpha} X_{\alpha}^+ \otimes X_{\alpha}^- + \frac{1}{1 - \frac{z_1}{z_2}} \frac{1}{2} \sum (H_j + i\tilde{H}_j) \otimes (H_j - i\tilde{H}_j) \\ &\quad - \frac{1}{1 - \frac{z_2}{z_1}} \sum_{\alpha} X_{\alpha}^- \otimes X_{\alpha}^+ - \frac{1}{1 - \frac{z_2}{z_1}} \frac{1}{2} \sum (H_j + i\tilde{H}_j) \otimes (H_j - i\tilde{H}_j) \\ &= \frac{1}{1 - \frac{z_1}{z_2}} \sum_{\alpha} X_{\alpha}^+ \otimes X_{\alpha}^- - \frac{1}{1 - \frac{z_2}{z_1}} \sum_{\alpha} X_{\alpha}^- \otimes X_{\alpha}^+ \\ &\quad + \frac{1}{2} \frac{z_2 + z_1}{z_2 - z_1} (\sum H_j \otimes H_j + \sum \tilde{H}_j \otimes \tilde{H}_j) \\ &\quad + \frac{i}{2} \sum (\tilde{H}_j \otimes H_j - H_j \otimes \tilde{H}_j). \end{aligned}$$

To evaluate what happens when Wilson lines in particular representations  $V, W$  of  $\tilde{\mathfrak{g}}$  cross, one applies the homomorphism

$$(9.12) \quad \rho_V \otimes \rho_W : \mathfrak{g} \otimes \mathfrak{g} \rightarrow \text{End}(V) \otimes \text{End}(W)$$

to the function  $r(z_1, z_2)$  (where  $\rho_V, \rho_W$  indicate the maps coming from the  $\mathfrak{g}$ -action on  $V$  and  $W$ ).

## 9.4 Specializing to $\mathfrak{sl}_2$

As an example, let us consider the case that  $\mathfrak{g} = \mathfrak{sl}_2$ . Let us use the standard basis  $e, f, h$  of  $\mathfrak{sl}_2$  where  $\langle e, f \rangle = 1$  and  $\langle h, h \rangle = 2$ . In this basis,

$$(9.13) \quad \begin{aligned} 2\pi i r(z_1, z_2) &= \frac{1}{1 - \frac{z_1}{z_2}} e \otimes f - \frac{1}{1 - \frac{z_2}{z_1}} f \otimes e \\ &\quad + \frac{1}{4} \frac{z_2 + z_1}{z_2 - z_1} (\sum h \otimes h + \sum \tilde{h} \otimes \tilde{h}) \\ &\quad + \frac{i}{4} \sum (\tilde{h} \otimes h - h \otimes \tilde{h}). \end{aligned}$$

Let us see what this looks like if our representations  $V, W$  are both the spin 1/2 representation of  $\mathfrak{sl}_2$ , and the basis element  $\tilde{H}$  of the second copy of the Cartan of  $\mathfrak{sl}_2$  acts with constants  $s_1, s_2$  in the two representations. Let us choose a basis  $e_+, e_-$  of the spin 1/2 representation of  $\mathfrak{sl}_2$ , and a corresponding basis  $e_+ \otimes e_+, \text{etc.}$ , of the tensor product  $V \otimes W$ . In this basis, the matrix  $r(z_1, z_2)$  looks like

$$(9.14) \quad \frac{1}{2\pi i} \begin{pmatrix} r_{++}^{++} & & & \\ & r_{+-}^{++} & r_{+-}^{-++} & \\ & r_{-+}^{++} & r_{-+}^{-++} & \\ & & & r_{--}^{++} \end{pmatrix},$$

$$(9.15) \quad \begin{aligned} r_{++}^{++} &= (1 + s_1 s_2) \frac{z_2 + z_1}{z_2 - z_1} + i(s_1 - s_2), \\ r_{+-}^{++} &= (-1 + s_1 s_2) \frac{z_2 + z_1}{z_2 - z_1} + i(-s_1 - s_2), \\ r_{+-}^{-++} &= \frac{4}{1 - \frac{z_1}{z_2}}, \\ r_{-+}^{-++} &= \frac{-4}{1 - \frac{z_2}{z_1}}, \\ r_{-+}^{++} &= (-1 + s_1 s_2) \frac{z_2 + z_1}{z_2 - z_1} + i(s_1 + s_2), \\ r_{--}^{++} &= (1 + s_1 s_2) \frac{z_2 + z_1}{z_2 - z_1} + i(s_2 - s_1). \end{aligned}$$

If we set  $s_1 = s_2 = 0$ , we get the usual trigonometric  $r$ -matrix for  $\mathfrak{sl}_2$ , associated to the six-vertex model of statistical mechanics. For comparison with the literature, one might want to symmetrize our  $r$ -matrix by conjugation, replacing the off-diagonal components by

$$(9.16) \quad r_{+-}^{-++} = \frac{4(\frac{z_1}{z_2})^{\frac{1}{2}}}{1 - \frac{z_1}{z_2}} = r_{-+}^{-++}.$$

without changing the diagonal entries.

In the above basis, the  $s$ -dependent terms only contribute to the diagonal matrix elements of  $r$ . They come in two groups. First, we have  $s_1 s_2 \frac{z_2 + z_1}{z_2 - z_1}$  times the identity operator. This we can absorb into the definition of the overall constant normalization factor of the  $R$ -matrix. The remaining terms are nontrivial. They reproduce the known generalization of the six-vertex model to include horizontal and vertical fields given by  $is_1$  and  $-is_2$  (see eqn. (B.5) in Appendix).

## 10. Elliptic Solutions of the Yang-Baxter Equation

### 10.1 Preliminaries

In this section we discuss the elliptic case, where the holomorphic curve  $C$  is an elliptic curve  $E$ . Since the differential  $dz$  does not have any poles we do not need to consider boundary terms like those encountered in section 9. However, there is an important topological subtlety to consider.

In formulating the theory with gauge group  $G$ , we begin by considering a topological  $G$ -bundle over  $\Sigma \times E$ . Our considerations leading to a solution of the Yang-Baxter equation are local along  $\Sigma$ , so we can take  $\Sigma = \mathbb{R}^2$ , in which case the choice of a topological  $G$ -bundle  $\mathcal{V} \rightarrow \Sigma \times E$  amounts to the choice of topological  $G$ -bundle  $\mathcal{V} \rightarrow E$ . However, in general there is a choice to be made, because, for a connected group  $G$ , a  $G$ -bundle over  $E$  is classified topologically by an invariant<sup>40</sup>  $\zeta \in \pi_1(G)$ .

<sup>40</sup> For example, for  $G = SO(3)$ , this invariant is the second

Therefore, in setting up the theory, we have the freedom to make an arbitrary choice of the element  $\zeta$ . Once we make this choice, a classical solution will be defined by a  $\bar{\partial}$  operator  $\partial_{\bar{z}} + [A_{\bar{z}}, \dots]$  on  $\mathcal{V}$ , modulo gauge transformations. Such an operator gives a holomorphic structure to the bundle  $\mathcal{V}$  and the gauge-invariant data is precisely the holomorphic equivalence class of this bundle.

The moduli space of classical solutions is therefore the moduli space of holomorphic  $G$ -bundles over  $E$  with topological class  $\zeta$ . The tangent space to the moduli space is  $H^1(E, \text{ad}(\mathcal{V}))$  (where  $\text{ad}(\mathcal{V})$  is the adjoint bundle associated to  $\mathcal{V}$ ). Thus the condition of the theory having a unique classical solution – and therefore leading to straightforward perturbation theory and a solution of the classical Yang-Baxter equation – is equivalent to  $H^1(E, \text{ad}(\mathcal{V})) = 0$ .

By Serre duality,  $H^1(E, \text{ad}(\mathcal{V}))$  is dual to  $H^0(E, \text{ad}(\mathcal{V}))$  for an elliptic curve  $E$ . Here  $H^0(E, \text{ad}(\mathcal{V}))$  is the Lie algebra of the automorphism group of a holomorphic  $G_{\mathbb{C}}$  bundle  $\mathcal{V}$ . Thus if and only if  $\mathcal{V}$  has no infinitesimal deformations, the Lie algebra of its automorphism group will be trivial and  $\mathcal{V}$  has only a finite group of automorphisms. (Instead of Serre duality, we could have invoked here the Riemann-Roch theorem, as we did at a similar point in section 9.1.)

To find, therefore, an elliptic solution of the Yang-Baxter equation (as opposed to its “dynamical” generalization), we need to find a complex Lie group  $G$  and a holomorphic  $G$ -bundle  $\mathcal{V} \rightarrow E$  such that  $H^0(E, \text{ad}(\mathcal{V})) = H^1(E, \text{ad}(\mathcal{V})) = 0$ . However, the options for such a  $G$  and  $\mathcal{V}$  are very limited. The only cases are  $G = PGL_N = GL_N/GL_1$  with  $N \geq 2$ , with  $\mathcal{V}$  chosen so that  $\zeta$  is a generator of the finite group  $\pi_1(G) \cong \mathbb{Z}_N$ . For such  $G$  and  $\zeta$ , there is a unique stable holomorphic vector bundle  $\mathcal{V}$ , and it does obey  $H^0(E, \text{ad}(\mathcal{V})) = H^1(E, \text{ad}(\mathcal{V})) = 0$ .

Thus elliptic solutions of the Yang-Baxter equation (or at least those that we can construct) are classified by a choice of  $N \geq 2$  – determining the group  $PGL_N$  – and a generator  $\zeta$  of  $\mathbb{Z}_N$ , or equivalently a primitive  $N^{\text{th}}$  root of 1. The automorphism group of such a solution is the automorphism group of  $\mathcal{V}$ , which in all cases is  $\mathbb{Z}_N \times \mathbb{Z}_N$ .

## 10.2 Rigid Holomorphic Bundles

We will now describe concretely the rigid holomorphic bundles  $\mathcal{V} \rightarrow E$  associated a choice of  $N$  and  $\zeta$ .

Stiefel-Whitney class  $w_2(\mathcal{V})$ . To define it in general, we observe that for  $G$  connected and  $p$  a point in  $E$ , the restriction of  $\mathcal{V}$  to  $E \setminus p$  is trivial, so  $\mathcal{V}$  can be constructed by gluing a trivial bundle over  $E \setminus p$  to a trivial bundle over a small disc  $D$  containing  $p$ . Since  $D \setminus p$  is homotopic to a circle, the gluing function that is used here is classified up to homotopy by a class  $\zeta \in \pi_1(G)$ .

It is well-known that there exist pairs of  $N \times N$  matrices  $A, B$  obeying

$$(10.1) \quad AB = BAe^{2\pi i/N}.$$

Moreover,  $A$  and  $B$  are unique up to conjugation and multiplication by scalars. For example, we can pick  $A = \text{diag}(1, e^{2\pi i/N}, e^{4\pi i/N}, \dots, e^{2\pi i(N-1)/N})$ , and  $B$  a matrix that cyclically permutes the eigenspaces of  $A$ . In the opposite direction, the equation shows that (after possibly multiplying  $A$  by a scalar) the eigenvalues of  $A$  are the  $N^{\text{th}}$  roots of 1, each with multiplicity 1. So  $A$  can be put in the claimed form, and then it is not hard to see that (up to conjugation and multiplication by a scalar)  $B$  must be as claimed.

Now given  $\zeta \in \mathbb{Z}_N$ , the matrices  $A^\zeta$  and  $B$  do not commute, but obey  $A^\zeta B = BA^\zeta \exp(2\pi i \zeta/N)$ . But this means that  $A^\zeta$  and  $B$  commute if projected to  $PGL_N$ . So we can define a flat  $PGL_N$  bundle  $\mathcal{V} \rightarrow E$  whose monodromies around a pair of generators of  $\pi_1(E) = \mathbb{Z} \oplus \mathbb{Z}$  are  $A^\zeta$  and  $B$ . Being flat, this bundle is automatically stable and holomorphic.

The automorphism group of this bundle always includes a subgroup  $\mathbb{Z}_N \times \mathbb{Z}_N$  generated by  $A$  and  $B$ . This is the full automorphism group if  $\zeta$  is a generator of  $\mathbb{Z}_N$ . (One may prove this starting with the fact that if  $\zeta$  is a generator of  $\mathbb{Z}_N$ , then the eigenvalues of the monodromy  $A^\zeta$  are nondegenerate.) Accordingly, if  $\zeta$  is a generator, then  $H^0(E, \mathcal{V}) = H^1(E, \mathcal{V}) = 0$ , and we are in the favorable situation that will lead to a solution of the Yang-Baxter equation.

If  $\zeta$  is not a generator, then this automorphism group has a strictly positive dimension and likewise the bundle  $\mathcal{V}$  can be deformed (as a holomorphic bundle and even as a flat one).

To conclude this section, we will describe  $\text{ad}(\mathcal{V})$  as a holomorphic bundle over  $E$ . We consider  $A^\zeta$  and  $B$  as matrices acting by conjugation on  $\mathfrak{g}$ , the Lie algebra of  $PGL_N$ . The eigenvalues are pairs  $\mu, \mu'$  of  $N^{\text{th}}$  roots of 1, with each pair occurring exactly once except the trivial pair  $\mu = \mu' = 1$ . Each joint eigenspace of  $A^\zeta$  and  $B$  corresponds to a holomorphic line bundle over  $E$ , and these line bundles are all of order  $N$  because the eigenvalues of  $A^\zeta$  and  $B$  are of order  $N$ . There are  $N^2$  equivalence classes of holomorphic line bundle over  $E$  of order  $N$ . The trivial line bundle is of order  $N$ , but there are  $N^2 - 1$  nontrivial ones, and  $\text{ad}(\mathcal{V}) \rightarrow E$  is the direct sum of all  $N^2 - 1$  nontrivial line bundles of order  $N$ . (Thus as a holomorphic bundle,  $\text{ad}(\mathcal{V})$  does not depend on the generator  $\zeta$  of  $\mathbb{Z}_N$ , though its Lie algebra structure does depend on  $\zeta$ .)

## 10.3 Specializing to $N = 2$

To achieve some minor simplifications, we will specialize to  $N = 2$ . There is now only one choice of

$\zeta$ . Accordingly,  $\mathcal{V}$  is unique up to isomorphism. As a holomorphic vector bundle over  $E$ ,  $\text{ad}(\mathcal{V})$  is the direct sum of the three non-trivial line bundles over  $E$ . However, we would like to describe the Lie algebra structure on  $\text{ad}(\mathcal{V})$ .

We identify  $E$  with the complex  $z$ -plane modulo  $z \cong z + 1$  and  $z \cong z + \tau$ , where the complex number  $\tau$  is constrained to have  $\text{Im } \tau > 0$  and is the modulus of  $E$ . A flat line bundle can be described by its monodromies  $\alpha$  and  $\beta$  under  $z \rightarrow z + 1$  and  $z \rightarrow z + \tau$ . A flat line bundle is of order 2 if  $\alpha^2 = \beta^2 = 1$ . To describe the 3 non-trivial line bundles of order 2, we take

$$(10.2) \quad (\alpha_i, \beta_i) = \begin{cases} (1, -1) & \text{for } \mathcal{L}_1, \\ (-1, 1) & \text{for } \mathcal{L}_2, \\ (-1, -1) & \text{for } \mathcal{L}_3, \end{cases}$$

Since  $(\alpha_1 \alpha_2, \beta_1 \beta_2) = (\alpha_3, \beta_3)$ , there is a natural isomorphism  $\phi_{12}: \mathcal{L}_1 \otimes \mathcal{L}_2 \cong \mathcal{L}_3$ . The same holds with any cyclic permutations of the labels 123. The Lie algebra structure on  $\mathcal{V}$  can be defined as follows. If  $s_1$  and  $s_2$  are local sections of  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , then  $[s_1, s_2]$  is the local section of  $\mathcal{L}_3$  defined by

$$(10.3) \quad [s_1, s_2] = \phi_{12}(s_1 \otimes s_2).$$

This statement and its cyclic permutations define the Lie algebra structure.

However, it may be useful to describe this structure in a slightly more explicit way. The matrices  $A$  and  $B$  of section 10.2 commute as matrices acting on the Lie algebra  $\mathfrak{sl}_2$ . Their joint eigenfunctions, up to a choice of basis, are the standard generators  $t_1, t_2, t_3$  of  $\mathfrak{sl}_2$ , obeying  $[t_1, t_2] = t_3$  and cyclic permutations. Thus one might prefer to think of a section of  $\text{ad}(\mathcal{V})$  more explicitly as  $\sum_{i=1}^3 s_i t_i$ , where  $s_i$  is a section of  $\mathcal{L}_i$  and  $t_i$  is an element of a standard basis of  $\mathfrak{sl}(2)$ . In this description, the commutator of  $s = \sum_{i=1}^3 s_i t_i$  with  $s' = \sum_{i=1}^3 s'_i t_i$  is  $s'' = \sum_{i=1}^3 s''_i t_i$  where

$$(10.4) \quad s''_3 = \phi_{12}(s_1 \otimes s'_2 - s_2 \otimes s'_1),$$

and cyclic permutations. Thus, one can think of eqn. (10.3) as a formula for the commutator of  $s_1 t_1$  with  $s_2 t_2$ .

Now let us determine the lowest order nontrivial contribution to the  $r$ -matrix. It will be a (meromorphic) section of<sup>41</sup>  $\text{ad}(\mathcal{V}) \boxtimes \text{ad}(\mathcal{V})$  over  $\Sigma \times \Sigma$ . This section can be determined by reasoning similar to what we used in section 9.3, though the details are simpler.

*A priori*, a general form of  $r$  would be  $r(z_1, z_2) = \sum_{i,j=1}^3 w_{i,j}(z_1, z_2) t_i \otimes t_j$ , where  $w_{i,j}(z_1, z_2)$  is a section of

<sup>41</sup> Here  $\text{ad}(\mathcal{V}) \boxtimes \text{ad}(\mathcal{V})$  is simply the tensor product of the  $\text{ad}(\mathcal{V})$  bundle over the first copy of  $\Sigma$  with the  $\text{ad}(\mathcal{V})$  bundle over the second copy; if  $\pi_i: \Sigma \times \Sigma \rightarrow \Sigma$ ,  $i = 1, 2$  are the two projections, then  $\text{ad}(\mathcal{V}) \boxtimes \text{ad}(\mathcal{V}) = \pi_1^*(\text{ad}(\mathcal{V})) \otimes \pi_2^*(\text{ad}(\mathcal{V}))$ .

$\mathcal{L}_i \boxtimes \mathcal{L}_j$ . However, the  $\mathbb{Z}_2 \times \mathbb{Z}_2$  automorphism group ensures that  $w_{i,j} = 0$  for  $i \neq j$ . Thus we reduce to  $r(z_1, z_2) = \sum_{i=1}^3 w_i(z_1, z_2) t_i \otimes t_i$ . Here the functions  $w_i(z_1, z_2)$  are determined by the following (somewhat redundant) set of conditions:

1.  $r(z_1, z_2)$  has a first-order pole at  $z_1 = z_2$ , with residue  $\hbar c = \hbar \sum_k t_k \otimes t_k$  as in the rational case, since OPE singularities are local.
2.  $r(z_1, z_2) = -r(z_2, z_1)$  (we need not combine this with an exchange of the two tensor factors of  $\mathfrak{sl}_2 \otimes \mathfrak{sl}_2$ , since  $r = \sum_i w_i t_i \otimes t_i$  is invariant under this switch).
3. Finally,  $r(z_1, z_2)$  is a function only of  $z_1 - z_2$ , because of the translation symmetry of the elliptic curve  $E$  and the differential  $\omega = dz$ .

The functions  $w_k(z_1, z_2)$  that satisfy these conditions actually have a simple interpretation. Since the canonical bundle  $K$  of  $E$  is trivial, its possible square roots  $K^{1/2}$  are line bundles of order 2, and thus we can think of  $\mathcal{L}_k$ , for  $k = 1, 2$  or 3, as one of the three even spin bundles of  $\Sigma$ , that is, as representing a possible  $K^{1/2}$ . Let  $\psi_k$  be a (holomorphic) fermion field on  $\Sigma$  valued in this spin bundle, with action

$$(10.5) \quad I = \frac{1}{2\pi} \int_E \psi_k \bar{\partial} \psi_k.$$

Then the two-point function  $\langle \psi_k(z_1) \psi_k(z_2) \rangle$  satisfies precisely the conditions of the functions  $w_k(z_1, z_2)$ .

With or without this interpretation, it is straightforward to write a formula for  $w_k$ :

$$(10.6) \quad w_k(z_1, z_2) = \sum_{n,m \in \mathbb{Z} \setminus \{(0,0)\}} \frac{\alpha_k^n \beta_k^m}{z_1 - z_2 - n - m\tau}.$$

The factors of  $\alpha_k^n \beta_k^m$  ensure that  $w_k(z_1, z_2)$  is a section of  $\mathcal{L}_k \boxtimes \mathcal{L}_k$ , and the desired properties are all obvious. We have defined  $w_k$  so that for  $z_1 \rightarrow z_2$  it behaves as  $1/(z_1 - z_2)$ . Thus, taking the normalization from the rational case, the  $r$ -matrix is

$$(10.7) \quad r = \hbar \sum_{k=1}^3 w_k(z_1 - z_2) t_k.$$

This coincides with the known expression in the literature, see e.g. [37, p. 539].

## 11. The Dynamical Yang-Baxter Equation

### 11.1 Reduction to an Abelian Subgroup

Our goal in this section is to take a first look at the case that  $G$  and  $C$  and the other relevant choices are such as to lead to a moduli space of classical solutions, not just a single isolated classical solution. This makes matters fundamentally more difficult and our analysis here will be preliminary.

A basic case to bear in mind is that  $C$  is an elliptic curve  $E$ ,  $G$  is simple, and the topological invariant  $\zeta$  introduced in section 10.1 vanishes, so that the  $G$ -bundle  $\mathcal{V} \rightarrow E$  is topologically trivial. Under this condition, the structure group of a stable holomorphic  $G$ -bundle  $\mathcal{V} \rightarrow E$  reduces always to the (complex) maximal torus  $T \subset G$ . The moduli space  $\mathcal{M}$  of stable holomorphic  $G$ -bundles over  $E$  is then simply the corresponding moduli space  $\mathcal{M}'$  for  $T$ , divided by the Weyl group  $\mathcal{W} : \mathcal{M} = \mathcal{M}'/\mathcal{W}$ . The automorphism group of a generic stable holomorphic  $G$ -bundle is simply  $T$ , though the symmetry is enhanced at points in  $\mathcal{M}'$  at which the Weyl action is not free. For our purposes, we can just identify the moduli space of classical solutions with  $\mathcal{M}'$  and ignore the Weyl group action and possible symmetry enhancement. (That is because we will eventually impose a constraint that keeps us away from special points.)

For  $C$  an elliptic curve, the story is similar<sup>42</sup> for any  $\zeta$ , with  $T$  replaced by another torus  $T_\zeta \subset T$ ,  $\mathcal{M}'$  replaced with the moduli space  $\mathcal{M}'_\zeta$  of  $T_\zeta$ -bundles over  $E$ , and the Weyl group replaced by a subgroup. Rather than explain these details, we will simply continue with the case that  $\zeta = 0$ .

A complex line bundle over  $E$  can be represented by a very simple gauge field  $A_{\bar{z}} = b$ , with  $b$  a complex constant. Here  $b$  is subject to some equivalences, but they will not be important in what follows. Similarly, for gauge group  $G$ , a classical solution whose structure group reduces to  $T$  can be represented by  $A_{\bar{z}} = b$ , where now  $b$  is a  $\mathfrak{t}$ -valued constant. For example, if  $G = SL_N$ , this means that  $A_{\bar{z}} = \text{diag}(b_1, b_2, \dots, b_N)$ , with  $\sum_{i=1}^N b_i = 0$ . In particular, this classical solution is invariant under the translation symmetries of  $E$ .

In general in nonabelian gauge theory with gauge group  $G$ , when we expand around a classical solution whose structure group reduces to a proper subgroup  $H$  of  $G$ , we should look for a low energy description in the form of an  $H$  gauge theory. In the present case, we have the further fact that to study the Yang-Baxter equation or its analog, we can take the topological plane  $\Sigma$  to be simply a plane  $\mathbb{R}^2$ , and thus we work on  $\mathbb{R}^2 \times E$  where  $E$  is compact and  $\mathbb{R}^2$  is not. In this situation, in quantum field theory in general, we can look for a low energy effective field theory on  $\mathbb{R}^2$ . Our present problem is actually diffeomorphism-invariant in the  $\mathbb{R}^2$  direction, so the low energy description is really a topological field theory.

Combining these facts, we can aim to find an effective description in the form of a two-dimensional topological gauge theory with structure group  $T$ . To find this description at least at a formal level, we make use of the translation symmetry of the classical solutions and throw away “massive modes”

<sup>42</sup> More exotic examples with moduli, such as the case  $\mathfrak{l}_0 + \mathfrak{l}_\infty \neq \mathfrak{g}$  mentioned in section 10.1, might be more complicated.

that lack this symmetry. Likewise, we keep only the  $\mathfrak{t}$ -valued part of the four-dimensional gauge field  $A$ . This means that  $A_{\bar{z}}$  is reduced to the  $\mathfrak{t}$ -valued field  $b$ , which now depends on coordinates  $x$  and  $y$  of  $\mathbb{R}^2$  but not on  $z$  and  $\bar{z}$ , and likewise the rest of the gauge field, namely  $A_x dx + A_y dy$ , becomes a purely two-dimensional gauge field that we will denote simply as  $A$ .

Specializing the underlying action (3.3) to this situation, we get an effective twodimensional abelian action

$$(11.1) \quad S = -\kappa \int_{\Sigma} \text{Tr} b F,$$

where  $\text{Tr}$  is the restriction to  $\mathfrak{t}$  of the quadratic form of the same name on  $\mathfrak{g}$ ,  $F = dA$  is the two-dimensional gauge curvature, and  $\kappa = i \text{Im} \tau / \pi \hbar$ . This is the action that we will use somewhat formally in analyzing the present problem.

The reader may object that in quantum field theory in general, it is not correct to simply set the massive modes to zero, as we have done. Instead, one has to integrate them out, producing in general corrections to the effective action of the fields that are retained in the effective description. Moreover, if we wish to include Wilson loops and their crossings, we have to integrate out the massive fields in the presence of those operators. All this is true, but the implications for our problem are limited because two-dimensional diffeomorphism symmetry severely constrains what couplings can be generated by integrating out massive modes. If  $\Sigma$  is not  $\mathbb{R}^2$  but is a curved two-manifold with scalar curvature  $R$ , then it is possible to generate a coupling  $\int_{\Sigma} d^2x \sqrt{g} R f(b)$ , for some function  $f(b)$ . One should expect such a term, but it will not be important for our purposes because in analyzing the crossing of line operators, one can assume that  $\Sigma$  is flat. In the presence of line operators, integrating out the massive modes will generate the framing anomaly of  $\mathfrak{g}$ . That is an important effect, but we are already familiar with it and will not discuss it further in the present section. Finally, integrating out the massive modes certainly affects the  $R$ -matrix that governs crossing of two line operators. But here we will just discuss the formal properties of this  $R$ -matrix, and will not try to calculate it. So we will not have to explicitly discuss the contribution of the massive modes to the  $R$ -matrix.

If  $A$  is regarded as a two-dimensional gauge field with structure group a compact torus  $T$ , and similarly  $b$  is real-valued, taking values in the corresponding Lie algebra  $\mathfrak{t}_c$ , then eqn. (11.1) becomes the action of a simple but much-studied two-dimensional topological field theory, often called  $BF$  theory.<sup>43</sup> We are not

<sup>43</sup> In that context,  $\kappa$  is real, assuming we want a unitary theory. In the context of reduction from the four-dimensional theory,  $\kappa$  is a complex number, since  $\hbar$  is complex.



quite in that situation, since in eqn. (11.1),  $b$ ,  $A_x$ , and  $A_y$  are all complex-valued fields and the action is a holomorphic function of those fields. Moreover, we are in a situation in which there is a nontrivial moduli space of classical solutions (parameterized locally by  $b \in \mathfrak{t}$ ), so if we want to define correlation functions, we cannot just do perturbation theory in a formal way; we need some recipe for what to do with the moduli. One point of view that may seem somewhat formal but that seems to provide a satisfactory framework for our considerations below (for instance, the derivation of the dynamical Yang-Baxter equation makes sense in this framework) is to view the theory as a machine that generates a differential form on the moduli space of classical solutions, without worrying about how to integrate it. Alternatively, it might be possible to define an operator that fixes the values of the moduli. This idea is explained at the end of section 11.3. Finally, in principle one may use the D4-NS5 system to define a nonperturbative integration cycle for the underlying four-dimensional theory (3.3). Specializing this to the present situation, one would then learn in principle what to do with the moduli in the effective theory (11.1).

In practice, we will simply draw inspiration from conventional two-dimensional  $BF$  theory.<sup>44</sup> Based on this, we will suggest a simple procedure to make contact with the dynamical Yang-Baxter equation [21, 22, 23, 24]. Given that we assume  $\zeta = 0$ , the solution of the dynamical Baxter equation that arises in this way will have  $T$  as a group of automorphisms, since a generic classical solution with  $\zeta = 0$  has automorphism group  $T$ , and we will be avoiding the exceptional cases.

## 11.2 $BF$ Theory and the Dynamical Yang-Baxter Equation

The classical equation of motion of the gauge field  $A$  in  $BF$  theory, in the absence of line operators, simply says that  $db = 0$ . Thus  $b$  is simply constant.

What happens in the presence of a Wilson line operator? To start with, we take the gauge group to be  $G = U(1)$ . An irreducible representation of  $U(1)$  is determined by the choice of an integer  $n$ , the “charge.” The Wilson operator supported on a curve  $K$  for the representation of charge  $n$  is  $\exp(in \int_K A)$ . Including this factor, the argument of the path integral becomes  $\exp(iS) \exp(in \int_K A) = \exp(iS')$ , with

$$(11.2) \quad S' = -\kappa \int_{\Sigma} \text{Tr} bF + n \int_K A.$$

In other words, the Wilson operator for  $U(1)$  effectively contributes an additional term to the classical action. The classical equation of motion now becomes

<sup>44</sup> The purely two-dimensional analysis that follows is similar to section 2.5 of [38].

$$(11.3) \quad \kappa db + n \delta_K = 0,$$

where  $\delta_K$  is a one-form delta function Poincaré dual to  $K$ . The import of this is that  $b' = \kappa b$  jumps by  $n$  in crossing  $K$  from right to left (Fig. 32(a)).

Now we can consider the case of two Wilson operators crossing. But in doing so, we may as well generalize<sup>45</sup> beyond a simple crossing of two line operators of charges  $n_1$  and  $n_2$ . Since we are discussing a purely two-dimensional theory, any crossing really does involve a physical intersection of the two line operators, and charge exchange is possible. In general, we can consider a case with charges  $n_1$  and  $n_2$  coming in and charges  $n_3$  and  $n_4$  going out, the only constraint being that  $n_1 + n_2 = n_3 + n_4$ . The behavior of  $b' = \kappa b$  is then as sketched in Fig. 32(b).

Now we consider the case of two-dimensional  $BF$  theory with a nonabelian gauge group. First let us take the gauge group to be  $SU(2)$ . As long as the field  $b$  is nonzero, the gauge symmetry is spontaneously broken from  $SU(2)$  to  $U(1)$ . As a result, apart from possible subtleties when  $b$  vanishes, we can aim for a description of  $BF$  theory of  $SU(2)$  in terms of  $BF$  theory of  $U(1)$ . It is actually possible (see section 2.3 of [39]) to compute rather precisely the effective  $U(1)$  action that arises from integrating out the charged modes.<sup>46</sup> We omit this, as the results are not very essential for us, as explained in section 11.1. (However, we describe some aspects of a precise quantum treatment in section 11.3.) What is important for us is that a line operator of the  $SU(2)$  theory can be written as a sum of line operators of the effective  $U(1)$  theory. For Wilson line operators, this decomposition is fairly evident. A Wilson operator associated to an irreducible representation  $\rho$  of  $SU(2)$  decomposes in the  $U(1)$  description as a sum of line operators corresponding to the weight spaces of  $\rho$ . Let us denote a Wilson operator for the representation  $\rho$  of  $SU(2)$  as  $W_\rho$  and one for the charge  $n$  representation of  $U(1)$  as  $W'_n$ . Then, for example, if  $\rho$  is the two-dimensional representation of  $SU(2)$ , it decomposes under  $U(1)$  as the sum of one-dimensional weight spaces of charges 1 and  $-1$ , so the corresponding formula for Wilson operators is  $W_\rho = W'_1 + W'_{-1}$ .

Now we can reconsider Fig. 32(b), with a slightly different interpretation. We consider the crossing of

<sup>45</sup> The purpose of discussing this generalization here is to give the simplest possible motivation for the picture of Fig. 32(b). A similar generalization is not possible in the context of the four-dimensional theory, assuming that the lines that are crossing have different values of the spectral parameter. Such a generalization is possible in two-dimensional  $BF$  theory with a nonabelian gauge group, though we will not need that case.

<sup>46</sup> The relevant procedure also has an analog in three-dimensional Chern-Simons theory on a Seifert manifold [40].



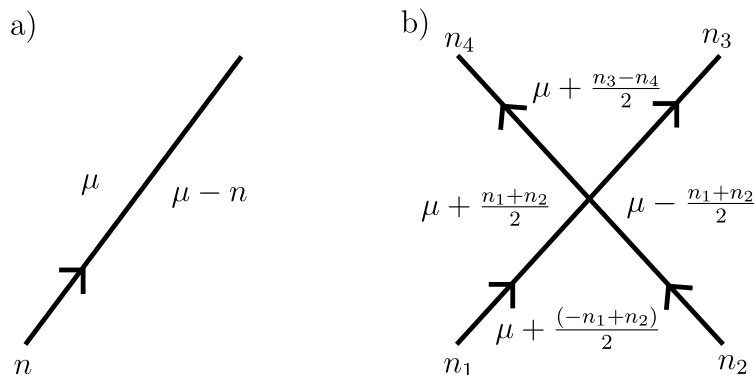


Figure 32. (a) The field  $b' = \kappa b$  is constant away from Wilson operators and jumps by  $n$  in crossing a charge  $n$  Wilson operator from right to left. In the example shown,  $b'$  jumps from  $\mu - n$  to  $\mu$ . (b) Two Wilson operators of charges  $n_1$  and  $n_2$  come in to this intersection and two of charges  $n_3$  and  $n_4$  go out (here  $n_1 + n_2 = n_3 + n_4$ ). The field  $b'$  jumps as shown.

two Wilson operators of  $SU(2)$  in some representations  $\rho$  and  $\rho'$ . On either of the two Wilson operators, on either side of the crossing, we can replace  $W_\rho$  or  $W_{\rho'}$  by a sum of  $U(1)$  Wilson operators, labeled by the weight spaces in  $\rho$  or  $\rho'$ . Thus the integers  $n_1, \dots, n_4$  in Fig. 32(b) are now not arbitrary but denote weights of  $\rho$  or  $\rho'$ . In passing through the point at which the two Wilson operators cross, the charges obey  $n_1 + n_2 = n_3 + n_4$  as before, because the effective description has  $U(1)$  symmetry.

The case of a compact nonabelian gauge group  $G$  of any rank  $r$  is similar. Away from special values of  $b$ , the  $G$  gauge symmetry is broken to the maximal torus  $T$ . The charges of an irreducible representation of  $T$  are now an  $r$ -plet of integers  $\vec{n}$ . Likewise  $b' = \kappa b$  becomes an  $r$ -plet  $\vec{b}'$ . Its jumping in crossing a Wilson operator is  $\vec{b}' \rightarrow \vec{b}' + \vec{n}$ , just as before. One detail is slightly different. For  $r > 1$ , the weight spaces of an irreducible representation may have multiplicity greater than 1. (For example, the adjoint representation of  $SU(3)$  has six weight spaces each of multiplicity 1 and one of multiplicity 2.) Accordingly, when we write  $W_\rho$  as a sum of Wilson operators of an effective  $T$  gauge theory, some charges may appear more than once. So in general away from crossings, an effective Wilson operator of the low energy theory carries some additional labels in addition to its “charges”  $\vec{n}$ .

Finally we come to the main point. In actually calculating a path integral appropriate to the situation of Fig. 32(b), we would run into a factor associated to the crossing. This factor would depend on the charges that label the incoming and outgoing lines in the effective abelian theory and on the parameter  $\mu$  that determines the labels of the bulk regions in the figure. We can denote this factor as a generalized  $R$ -matrix element  $R_{n_1 n_2}^{n_3 n_4}(\mu)$ . (For simplicity, in writing the  $R$ -matrix in detail as a matrix, we take  $G = SU(2)$  so that each line is simply labeled by an integer.)

Having defined the generalized  $R$ -matrix of two-dimensional  $BF$  theory, one can ask if it obeys a generalized Yang-Baxter equation, now with bulk labels as well as labels for line segments between crossings (Fig. 33). However, there is no reason to expect this. In a purely two-dimensional theory, one has to pass through a singularity to interpolate between the left and right hand sides of the figure. As far as we know, the generalized Yang-Baxter equation is not satisfied in this situation.

Hopefully the reader can anticipate what comes next. We consider not a purely two-dimensional  $BF$  theory, but an effective abelian  $BF$  theory in two dimensions that arises as in the discussion of eqn. (11.2) from a four-dimensional theory on  $\mathbb{R}^2 \times E$  with complex gauge group  $G$  and with  $\zeta = 0$ . A Wilson operator is now labeled by a representation  $\rho$  of  $G$  (or more generally by a representation of a quantum deformation of  $\mathfrak{g}[[z]]$ ) and also by a spectral parameter  $z \in E$ . Between crossings, a Wilson operator is further labeled by a basis vector of  $\rho$ . The bulk parameter  $b' = \kappa b$  jumps in crossing a Wilson operator of the low energy theory. Now, we associate to a crossing a generalized  $R$ -matrix element  $R_{n_1 n_2}^{n_3 n_4}(z_1 - z_2; \mu)$ . The difference from before is that as long as  $z_1 \neq z_2$ , there is no singularity associated to a crossing. Therefore the two sides of Fig. 33 (with spectral parameters now included) are equivalent. The equivalence is known as the dynamical Yang-Baxter equation.

### 11.3 Quantum Treatment

We have treated the parameter  $b'$  classically, starting with purely two-dimensional  $BF$  theory. Actually, the locally constant value of  $b'$  has a natural meaning in the quantum theory.

Let us ask what are the quantum states when  $BF$  theory - to begin with for  $G = U(1)$  - is quantized on a

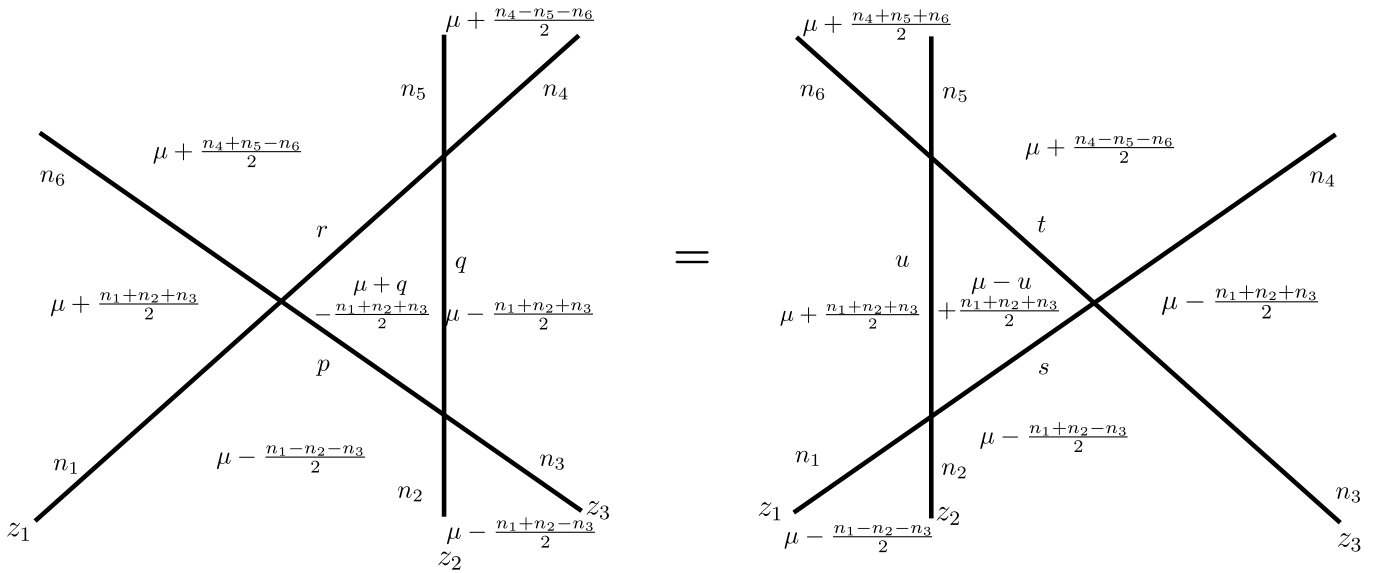


Figure 33. The dynamical Yang-Baxter equation. Both line segments and bulk regions are labeled. A line segment is labeled by a basis vector in a representation of  $G$  attached to that line, and a bulk region is labeled by a weight of  $G$ . The picture is drawn here for  $G = SU(2)$ , so that a basis vector is determined by its integer weight. Two-dimensional BF theory has all the ingredients to draw such a picture (without spectral parameters) but there is no reason to expect this equivalence to hold. When we go to four dimensions, the lines are labeled by spectral parameters in the familiar way (as shown here) and the usual arguments show an equivalence between the two pictures.

circle  $S$ . A physical state is a gauge-invariant function of the connection  $A$  restricted to  $S$ . Such a function is  $\Psi_n = \exp(im \oint_S A)$  for some integer  $n$ . On the other hand,  $b'$  is the momentum conjugate to  $A$ , so (for  $p$  a point in  $S$ ),  $b'(p)$  can be identified with  $-i\delta/\delta A(p)$ . Explicitly acting with this on  $\Psi_n$ , we find that  $b'(p)\Psi_n = n\Psi_n$ , for any  $p$ , independent of  $p$ .

Quantum mechanically, the constant value of  $b'(p)$  should be interpreted as an eigenvalue of this operator, and (for  $U(1)$ ) the eigenvalues are integers, as we have just seen. We can also understand in this language why crossing a Wilson line operator has the effect of shifting  $b'$  by an integer. For example, we can regard the Wilson line operator  $W_m = \exp(im \oint_S A)$  as an operator that acts on physical states on  $S$ . Since  $W_m \Psi_n = \Psi_{n+m}$ , acting with  $W_m$  shifts the value of  $b'$  by  $m$ .

Now let us repeat this analysis for a compact but possibly nonabelian gauge group  $G$ . Of course, the analog of  $\Psi_n$  is what we might call  $\Psi_\rho$ , the trace of the holonomy around  $S$  in the representation  $\rho$ :

$$(11.4) \quad \Psi_\rho = \text{Tr}_\rho U, \quad U = P \exp \left( \oint_S A \right).$$

However, we would prefer to express this in a language that is better suited for the reduction to an effective abelian description. For this, we observe that a gauge-invariant function of  $A$  is precisely a function of the holonomy  $U$  that is invariant under conjugation. In other words, gauge-invariant functions of  $A$

are functions on  $G/G$ , where  $G$  acts on itself by conjugation.

The quotient  $G/G$  is the same as  $T/\mathcal{W}$ , the quotient of the maximal torus  $T$  by the Weyl group  $\mathcal{W}$ . This might lead one to expect that gauge-invariant functions of  $A$  would correspond to Weyl-invariant functions on  $T$ , but actually they correspond in a natural way to Weyl *anti*-invariant functions. We say that a function on  $T$  is Weyl anti-invariant if it is odd under each of the elementary reflections that generate  $\mathcal{W}$ . The association of a representation  $\rho$  of  $G$  with a Weyl anti-invariant function on  $T$  is given by the Weyl character formula; the Weyl anti-invariant function corresponding to  $\rho$  is the numerator of the usual Weyl character formula for  $\rho$ . Let us just explain what this means for  $SU(2)$ . A maximal torus of  $SU(2)$  is the  $U(1)$  subgroup  $\text{diag}(e^{i\theta}, e^{-i\theta})$ . The character of the  $n$ -dimensional representation  $\rho_n$  of  $SU(2)$  is

$$(11.5) \quad e^{i(n-1)\theta} + e^{i(n-3)\theta} + \dots + e^{-i(n-1)\theta} = \frac{\sin(n\theta)}{\sin\theta}.$$

The Weyl anti-invariant function corresponding to  $\rho_n$  is the numerator, or

$$(11.6) \quad \sin(n\theta) = \frac{1}{2i}(e^{in\theta} - e^{-in\theta}).$$

The functions  $\sin(n\theta)$ ,  $n = 1, 2, 3, \dots$  are a basis for the Hilbert space of Weyl anti-invariant functions on  $T$ . This is the Hilbert space of BF theory of  $SU(2)$ ,

quantized on a circle. In this description, we can conveniently see the effective  $U(1)$  description of  $SU(2)$   $BF$  theory. In this effective  $U(1)$  description, the holonomy is  $e^{i\theta}$  and  $b'$  is the canonical momentum  $b' = -i\partial/\partial\theta$ . We see that  $\sin(n\theta)$  is not an eigenstate of  $b'$  but rather is a linear combination of eigenstates  $e^{in\theta}$  and  $e^{-in\theta}$  with eigenvalues  $\pm n$ . These values are Weyl conjugate, since the Weyl group of  $SU(2)$  is  $\mathbb{Z}_2$ , acting as  $-1$  on the Lie algebra  $\mathfrak{t}$  of  $T$ . The implication is that  $b'$  should not really be regarded as an integer (or a real number, as in the classical description) but as a Weyl orbit of nonzero integers. It is fairly natural to pick from each Weyl orbit  $n, -n$  the positive representative, and if we do this then the values of  $b'$  in the effective abelian description are positive integers. For any compact simple  $G$ , the analog is that the values of  $b'$  are dominant weights in the interior of a positive Weyl chamber.

Now let us see what happens to the value of  $b'$  in crossing a Wilson operator, say the operator  $W_m$  associated to the  $m$ -dimensional representation  $\rho_m$ . The character of  $\rho_m$  is  $F_m(\theta) = e^{i(m-1)\theta} + \dots + e^{-i(m-1)\theta}$ . Crossing  $W_m$  has the effect of multiplying the quantum state by  $F_m(\theta)$ . We have

$$(11.7) \quad F_m(\theta) \sin(n\theta) = \sum_{j=-m+1, -m+3, \dots, m-1} \sin((n+j)\theta),$$

and if  $m \leq n$ , then  $n+j$  is always positive. This means that in crossing  $W_m$ ,  $b'$  can jump by  $j$  for any  $j = -m+1, -m+3, \dots, m-1$ , that is, any weight of the representation  $\rho_m$ . This is the result that was claimed in section 11.2. For  $m > n$ , it is possible for  $n+j$  to be non-positive, and some terms on the right hand side of eqn. (11.7) vanish or cancel. This leads to some modification of the formalism when  $b'$  is not large.

These subtleties do not really affect the discussion in section 11.2 very much. In that discussion,  $b'$  was treated as a generic real number, but it would not have been much different to regard  $b'$  as a generic positive integer, where here “generic” is equivalent to “sufficiently large.” Thus, in studying any concrete collection of Wilson operators, associated with representations of dimensions  $m_1, \dots, m_s$ , the reasoning in section 11.2 is valid if  $b'$  is sufficiently large compared to those dimensions.

This quantum treatment of purely two-dimensional  $BF$  theory, however, highlights what is missing in our understanding of the four-dimensional theory. We certainly do not have available a quantum treatment that would identify definite allowed values of  $b'$  in the theory obtained by compactification from four dimensions. In fact, since the four-dimensional theory is a somewhat formal construction with a holomorphic action, it is not clear to what extent one should expect to have such a quantum treatment.

Neither – at least at first sight – do we wish to integrate over the possible values of  $b'$ , as a natural cycle for such an integral does not present itself. In section 11.1, we already mentioned several possible ways to deal with this issue. Here we just elaborate on one possibility. In purely two-dimensional  $BF$  theory on a two-manifold  $\Sigma$ , it is fairly natural to pick a point  $p \in \Sigma$ , not in the support of any Wilson operator, and specify the value of  $b'$  there. The values of  $b'$  elsewhere would then be determined by the crossing rules of section 11.2. It is tempting to believe that a similar constraint should be imposed in the four-dimensional theory, to spare ourselves from having to sum or integrate over  $b'$ . Moreover, a constraint setting  $b'$  to a generic value at some particular point  $p$ , when supplemented with the crossing rules, will ensure that  $b'$  never takes a value at which the automorphism group is enhanced and the effective abelian description breaks down. However, we have not seriously tried to study a quantum field theory operator that would impose this constraint.

## Acknowledgments

The authors would like to thank Davide Gaiotto, Robbert Dijkgraaf, Michio Jimbo, Cumrun Vafa, and especially Jacques H. H. Perk for discussions. K. C. is supported by the NSERC Discovery Grant program and by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. E. W. is partially supported by National Science Foundation grant NSF Grant PHY-1606531. M. Y. is partially supported by WPI program (MEXT, Japan), by JSPS Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers, by JSPS KAKENHI Grant No. 15K17634, and by JSPS-NRF Joint Research Project.

## Appendix A. Rational $R$ -Matrix for $SO_N$

In this Appendix we discuss the rational  $R$ -matrix for the fundamental representation of  $G = SO_N$ , extending the similar discussion for fundamental and anti-fundamental representations for  $G = GL_N$  (or  $SL_N$ ) in section 3.5. While the result in itself is known since the old work of [6], in the framework of this paper it is an illuminating exercise to check the consistency between unitarity, crossing and the framing anomaly.

Let us consider the  $R$ -matrix  $R(z_1 - z_2) : V \otimes V \rightarrow V \otimes V$  for the fundamental representation  $V$  of  $G = SO_N$ . Since this representation is equivalent to its own conjugate, a Wilson line in this representation carries no orientation.

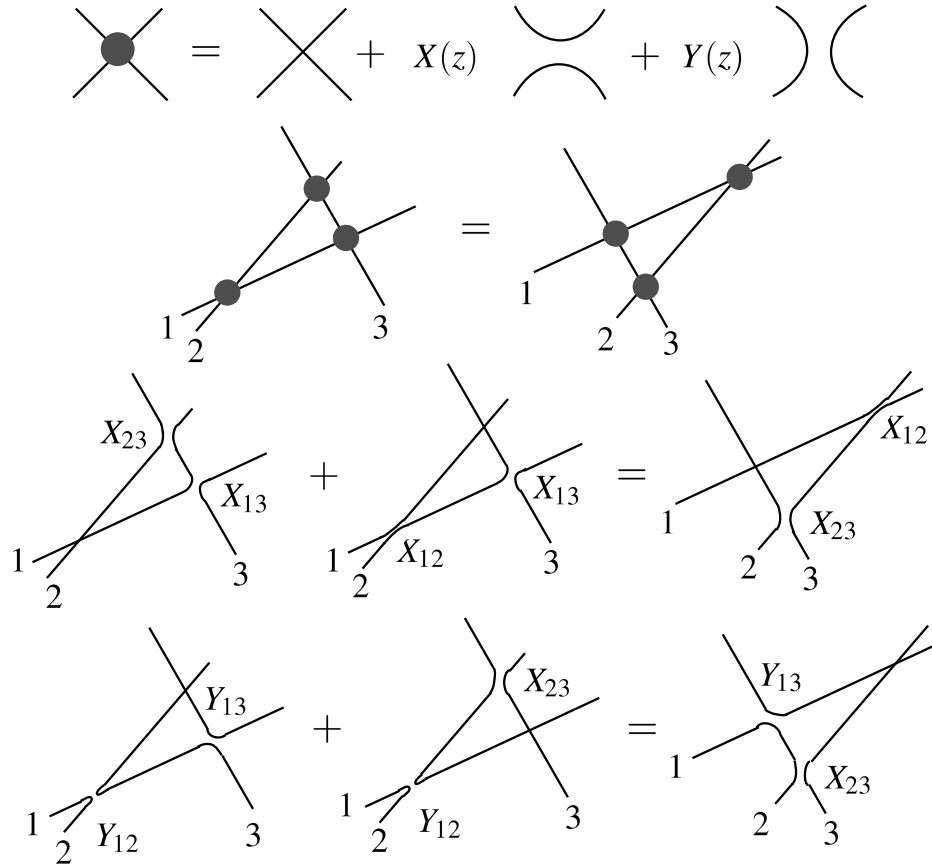


Figure 34. Graphical representation for the Yang-Baxter equation for the  $SO_N$  R-matrix given in eqn. (A.1). Here we used the shorthand notation  $X_{ij} = X(z_i - z_j)$  and  $Y_{ij} = Y(z_i - z_j)$ , with  $z_i$  associated with the  $i$ -th line. The two equations here give eqns. (A.3).

As we discussed in the main text, the rational R-matrix has  $G$  as a symmetry. Consequently, the R-matrix is constrained to be of the form

$$(A.1) \quad R_{ij}^{i'j'}(z) = \delta_i^{i'} \delta_j^{j'} E(z) + \delta_i^{j'} \delta_j^{i'} F(z) + \delta_{ij} \delta^{i'j'} G(z),$$

with three unknown functions  $E(z)$ ,  $F(z)$  and  $G(z)$ . In this literature this is also written as

$$(A.2) \quad R_{ij}^{i'j'}(z) = E(z)I + F(z)P + G(z)Q,$$

where  $P$  is a permutation, and  $Q = {}^t P$  its transposition. Compared to the case of  $GL_N$  in eqn. (3.22), we here have one extra structure  $\delta_{ij} \delta^{i'j'}$  consistent with  $G$ -symmetry.

The Yang-Baxter equation constrains only the ratios of the three functions:  $X(z) = F(z)/E(z)$ ,  $Y(z) = G(z)/E(z)$ . As we can see from Fig. 34, the resulting constraints takes precisely the same form as in (3.23) and (3.26), where the role of  $U(z)$  and  $W(z)$  are played

by  $X(z)$  and  $Y(z)$  respectively:

$$(A.3) \quad \begin{aligned} X(z_1 - z_3)X(z_2 - z_3) + X(z_1 - z_2)X(z_1 - z_3) \\ = X(z_1 - z_2)X(z_2 - z_3), \\ Y(z_1 - z_3)X(z_2 - z_3) \\ = Y(z_1 - z_2)Y(z_1 - z_3) + Y(z_1 - z_2)X(z_2 - z_3). \end{aligned}$$

By the similar reasoning as before we obtain

$$(A.4) \quad X(z) = \frac{\hbar}{z}, \quad Y(z) = -\frac{\hbar}{z - c\hbar},$$

where  $c$  is some undetermined constant.

The undetermined constant  $c$  is constrained by unitarity (recall eqn. (2.3)), which gives the relation (see Fig. 34; note the difference from eqn. (3.28)):

$$(A.5) \quad NY(z)Y(-z) + Y(z) + Y(-z) + X(z)Y(-z) + X(-z)Y(z) = 0.$$

This leads to

$$(A.6) \quad X(z) = \frac{\hbar}{z}, \quad Y(z) = -\frac{\hbar}{z + \frac{N-2}{2}\hbar}.$$

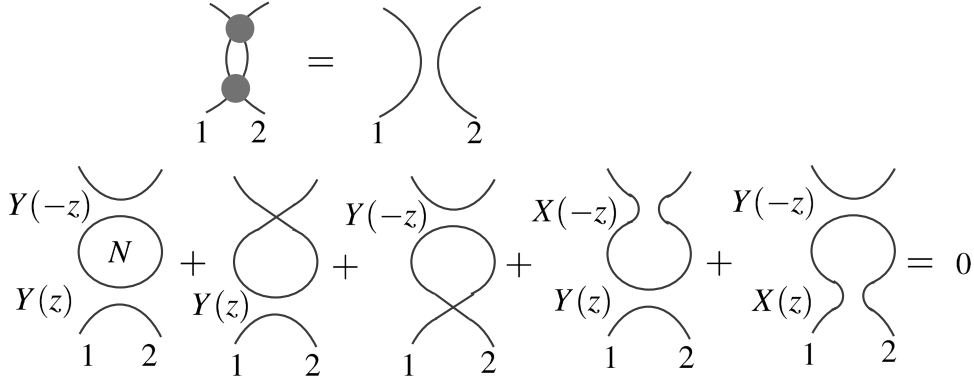


Figure 35. Graphical representation for the unitarity relation for the  $SO_N$  R-matrix given in eqn. (A.1), with  $z = z_1 - z_2$ . The factor of  $N$  comes from the color factor for the bubble in the first term of the equation.

These two functions satisfy

$$(A.7) \quad X\left(z - \hbar \frac{N-2}{2}\right) = Y(-z),$$

which is consistent with the crossing and framing anomaly, since  $N-2$  is the dual Coxeter number of  $SO_N$ .

We can also try to determine the overall factor  $E(z)$ . The diagonal component of the unitarity relation gives

$$(A.8) \quad E(z)E(-z) + F(z)F(-z) = E(z)E(-z)(1 + X(z)X(-z)) = 1,$$

or equivalently

$$(A.9) \quad E(z)E(-z) = \frac{z^2}{z^2 - \hbar^2}.$$

This solution can easily be solved, for example, by

$$(A.10) \quad E(z) = \frac{z}{z \pm \hbar}.$$

However, this is actually not the solution we want, since we also want to impose the condition that this factor does not spoil the unitarity relation:

$$(A.11) \quad E\left(z - \hbar \frac{N-2}{2}\right) = E(-z), \quad \text{or} \quad E\left(-\hbar \frac{N-2}{2} - z\right) = E(z).$$

The minimal solutions for (A.9) and (A.11) are

$$(A.12) \quad E(z) = \frac{Q_{\pm\hbar, -\hbar(N-2)/2}(z)}{Q_{0, -\hbar(N-2)/2}(z)},$$

where the function  $Q_{\alpha,\beta}(z)$ , satisfying the functional identities

$$(A.13) \quad Q_{\alpha,\beta}(z)Q_{\alpha,\beta}(-z) = \frac{1}{\frac{(z+\alpha)}{2\beta} \frac{(-z+\alpha)}{2\beta}}, \quad Q_{\alpha,\beta}(z) = Q_{\alpha,\beta}(\beta - z),$$

is written as

$$(A.14) \quad Q_{\alpha,\beta}(z) = \frac{\Gamma\left(\frac{z+\alpha}{2\beta}\right)\Gamma\left(\frac{-z+\alpha+\beta}{2\beta}\right)}{\Gamma\left(\frac{z+\alpha+\beta}{2\beta}\right)\Gamma\left(\frac{-z+\alpha+2\beta}{2\beta}\right)}.$$

Note that both expressions in (A.12), with either plus or minus sign, are minimal. In the literature, one is associated with the S-matrix for the  $O(N)$  sigma model, while another to that for the Gross-Neveu model [6].

To study the behavior at  $\hbar \sim 0$ , let us first note that

$$(A.15) \quad Q_{\gamma\delta\hbar, \delta\hbar}(z) = \frac{\Gamma\left(\frac{z}{2\hbar\delta} + \frac{1}{2}\right)\Gamma\left(-\frac{z}{2\hbar\delta} + \frac{1+\gamma}{2}\right)}{\Gamma\left(\frac{z}{2\hbar\delta} + \frac{1+\gamma}{2}\right)\Gamma\left(-\frac{z}{2\hbar\delta} + \frac{2+\gamma}{2}\right)}.$$

From the Stirling's formula

$$(A.16) \quad \log\Gamma(z+c) \sim z \log z - z + c \log z - \frac{1}{2} \log z + \frac{1}{2} \log 2\pi + \frac{c^2 - c}{z} + \mathcal{O}\left(\frac{1}{z^2}\right)$$

( $c$  being constant), we obtain

$$(A.17) \quad Q_{\gamma\delta\hbar, \delta\hbar}(z) \sim e^{-\log \frac{z}{2\hbar\delta} - \hbar \frac{(\gamma^2 - 2\gamma - 1)\delta}{2z} + \mathcal{O}\left(\left(\frac{\hbar}{z}\right)^2\right)},$$

and hence we learn that the overall factor has a perturbative expansion in powers of  $\hbar/z$ , starting with identity:

$$(A.18) \quad E(z) = \frac{Q_{\pm\hbar, -\hbar(N-2)/2}(z)}{Q_{0, -\hbar(N-2)/2}(z)} = \begin{cases} 1 - \frac{N^2 - 8}{4(N-2)} \frac{\hbar}{z} + \mathcal{O}\left(\left(\frac{\hbar}{z}\right)^2\right) & \text{(plus sign),} \\ 1 - \frac{N^2 - 8N + 8}{4(N-2)} \frac{\hbar}{z} + \mathcal{O}\left(\left(\frac{\hbar}{z}\right)^2\right) & \text{(minus sign).} \end{cases}$$

In perturbation theory, these formulas appear satisfactory, but nonperturbatively, one would like to understand the meaning of the poles of the function  $E(z)$ . As remarked in the introduction, the D4-NS5 system is likely to provide a nonperturbative framework



in which such questions could be addressed, but we will not pursue that in the present paper.

We can repeat a similar exercise for the gauge group  $G = Sp_{2N}$ . We again find that the constraint from the Yang-Baxter equation and unitarity are consistent with crossing symmetry and the framing anomaly, with the dual Coxeter number given by  $N + 1$ .

## Appendix B. Examples of Trigonometric and Elliptic $R$ -Matrix

In section (3.5) and appendix A, we discussed several examples of rational  $R$ -matrices. In this appendix, we present examples of trigonometric and elliptic  $R$ -matrices known in the literature, for the simplest case of the fundamental representation of  $G = SL_2$ . We verify that their classical limit reproduces the classical  $r$ -matrices discussed in the sections 9.4 and 10.3. The  $R$ -matrices below are quasi-classical, and are normalized to be  $R_{\hbar=0} = I$ , which is the canonical normalization in gauge theory. If expanded in perturbation theory, they lead to  $r$ -matrices that we computed in sections 9 and 10.

As discussed in appendix A the complete understanding of the overall factor of the  $R$ -matrix goes beyond the perturbative analysis of this paper. Expressions below should be understood modulo an overall scalar factor ambiguities for the  $R$ -matrix, and hence modulo the shift by identity matrices for the classical  $r$ -matrix.

### B.1 Rational

As discussed in the main text, there are three types of quasi-classical  $R$ -matrix known in the literature: rational, trigonometric and elliptic. Let us begin with the rational  $R$ -matrix, which we already discussed in section 3.5:

$$(B.1) \quad R_{\hbar}^{\text{rational}}(z) = \left( z + \frac{\hbar}{2} \right) I + \hbar c = zI + \hbar P,$$

where  $z \in \mathbb{C}$ ,  $c = (\vec{\sigma} \cdot \vec{\sigma})/2$  is the quadratic Casimir (4.12) for  $\mathfrak{g} = \mathfrak{sl}_2$ , and  $P = c + I/2$  is the permutation operator of the two spins.

This  $R$ -matrix has manifest  $SL_2$  symmetry. Choosing a basis, the  $R$ -matrix reads

$$(B.2) \quad R_{\hbar}^{\text{rational}}(z) = \begin{matrix} & \begin{matrix} |\uparrow\uparrow\rangle & |\uparrow\downarrow\rangle & |\downarrow\uparrow\rangle & |\downarrow\downarrow\rangle \end{matrix} \\ \begin{matrix} |\uparrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{matrix} & \begin{pmatrix} z + \hbar & & & \\ & z & \hbar & \\ & \hbar & z & \\ & & & z + \hbar \end{pmatrix} \end{matrix}.$$

Note that up to normalization only the ratio of  $z$  and  $\hbar$  enters into this solution, as expected on dimensional grounds.

### B.2 Trigonometric Solution

The trigonometric solution is

$$(B.3) \quad R_{\hbar}^{\text{trig}}(z) = \begin{pmatrix} a(z, \hbar) & & & \\ & b(z, \hbar) & c(z, \hbar) & \\ & c(z, \hbar) & b(z, \hbar) & \\ & & & a(z, \hbar) \end{pmatrix},$$

where

$$(B.4) \quad \begin{aligned} a(z, \hbar) &= \sinh(w + \hbar), \\ b(z, \hbar) &= \sinh(w), \\ c(z, \hbar) &= \sinh(\hbar), \end{aligned}$$

with  $z = e^w \in \mathbb{C}^\times$  and  $q := e^{\hbar}$ .

In terms of the quasi-classical parameter  $\hbar$  and the cylindrical coordinate  $w$ , this  $R$ -matrix is written in terms of sine functions, as the name “trigonometric” suggests. When we impose unitarity, this  $R$ -matrix should be multiplied by a suitable overall normalization factor.

The six-vertex model allows for an integrable deformation, corresponding to an inclusion of the “electric field”  $h, v$ , in vertical and horizontal directions:

$$(B.5)$$

$$R_{\hbar, \theta}^{\text{trig}}(z) = \begin{pmatrix} a(z, \hbar)e^{h+v} & & & \\ & b(z, \hbar)e^{-h+v} & c(z, \hbar) & \\ & c(z, \hbar) & b(z, \hbar)e^{h-v} & \\ & & & a(z, \hbar)e^{-h-v} \end{pmatrix}.$$

The parameters  $h, v$  correspond to  $s_1, s_2$  in section 9.4:  $h_1 = is_1, v = -is_2$ .

When this  $R$ -matrix is used to construct a model in lattice statistical mechanics in the usual way, its six non-zero matrix elements become the Boltzmann weights for six allowed configurations at a vertex. The model accordingly is generally called the six-vertex model. Among the six matrix elements of the  $R$ -matrix, two combinations are inessential; one parameter is the overall normalization of the  $R$ -matrix and one can be removed by conjugation. This leaves four parameters, which we have here denoted  $z, \hbar$  and  $h, v$ . The one-dimensional quantum spin chain associated with this model is the XXZ spin chain.

### B.3 Elliptic Solution

The elliptic  $R$ -matrix has a new parameter, namely the modulus of the torus  $\tau$ :

(B.6)

$$R_{\hbar,\tau}^{\text{elliptic}}(z) = e^{i\pi\hbar F(z,\tau)} \times \begin{pmatrix} a(z,\tau,\hbar) & & & d(z,\tau,\hbar) \\ & b(z,\tau,\hbar) & c(z,\tau,\hbar) & \\ & c(z,\tau,\hbar) & b(z,\tau,\hbar) & \\ d(z,\tau,\hbar) & & & a(z,\tau,\hbar) \end{pmatrix},$$

where we defined

$$(B.7) \quad \begin{aligned} a(z,\tau,\hbar) &= \frac{\vartheta_0(\pi\hbar|2\tau)}{\vartheta_0(0|2\pi)} \frac{\vartheta_1(\pi(z+\hbar)|2\tau)}{\vartheta_1(\pi z|2\tau)}, \\ b(z,\tau,\hbar) &= \frac{\vartheta_0(\pi\hbar|2\tau)}{\vartheta_0(0|2\pi)} \frac{\vartheta_0(\pi(z+\hbar)|2\tau)}{\vartheta_0(\pi z|2\tau)}, \\ c(z,\tau,\hbar) &= \frac{\vartheta_1(\pi\hbar|2\tau)}{\vartheta_0(0|2\pi)} \frac{\vartheta_0(\pi(z+\hbar)|2\tau)}{\vartheta_1(\pi z|2\tau)}, \\ d(z,\tau,\hbar) &= \frac{\vartheta_1(\pi\hbar|2\tau)}{\vartheta_0(0|2\pi)} \frac{\vartheta_1(\pi(z+\hbar)|2\tau)}{\vartheta_0(\pi z|2\tau)}. \end{aligned}$$

Here  $\vartheta_0(z|\tau)$  and  $\vartheta_1(z|\tau)$  are Jacobi theta functions

(B.8)

$$\begin{aligned} \vartheta_0(z|\tau) &= \prod_{n=1}^{\infty} (1 - 2e^{(2n-1)\pi i\tau} \cos 2z + e^{(4n-2)\pi i\tau})(1 - e^{2n\pi i\tau}), \\ \vartheta_1(z|\tau) &= 2e^{\pi i\tau/2} \sin u \prod_{n=1}^{\infty} (1 - 2e^{2n\pi i\tau} \cos 2z + e^{4n\pi i\tau}) \\ &\quad \times (1 - e^{2n\pi i\tau}), \end{aligned}$$

and  $F(z,\tau)$  is a normalization factor on which we will comment later. Note also that, for a better match with the main text, we have used the Jacobi theta function with elliptic modulus  $2\tau$  (not  $\tau$ ); care is need for comparison with the literature.

This  $R$ -matrix has eight non-zero matrix elements and represents the eight-vertex model solved by Baxter [5]. The associated one-dimensional quantum spin chain is the XYZ spin chain.

The periodicity of the Jacobi theta functions are given by

$$(B.9) \quad \begin{aligned} \vartheta_1(\pi(z+1)|2\tau) &= -\vartheta_1(\pi z|2\tau), \\ \vartheta_1(\pi(z+\tau)|2\tau) &= ie^{-\frac{i\pi\tau}{2}} e^{-i\pi z} \vartheta_0(\pi z|2\tau), \\ \vartheta_0(\pi(z+1)|2\tau) &= \vartheta_0(\pi z|2\tau), \\ \vartheta_0(\pi(z+\tau)|2\tau) &= ie^{-\frac{i\pi\tau}{2}} e^{-i\pi z} \vartheta_1(\pi z|2\tau). \end{aligned}$$

This means if we choose the normalization factor  $F(z,\hbar)$  to be a function on the two-torus satisfying

$$(B.10) \quad F(z+1,\tau) = F(z,\tau), \quad F(z+\tau,\tau) = F(z,\tau) + 1,$$

then all the entries of the  $R$ -matrix are periodic under the shift  $z \rightarrow z+1$ ,  $z \rightarrow z+\tau$ , up to a possible overall sign factor; for such  $F$  we can for example choose

$$(B.11) \quad F(z,\tau) = \frac{i}{\pi} \log \vartheta_0 \left( \pi \left( z - \frac{\tau-1}{2} \right) | 2\tau \right).$$

Note also that the normalization of the  $R$ -matrix here is chosen so that  $a(z,\tau) = b(z,\tau) = 1$  when  $\hbar = 0$ .

The classical  $r$ -matrix is then (note  $\vartheta_1(0) = 0$ )

$$(B.12) \quad r_{\tau}^{\text{elliptic}}(z,\tau) = \sum_{k=1}^3 w_k(z) \sigma_k \otimes \sigma_k$$

with

$$(B.13) \quad \begin{aligned} w_1(z,\tau) &= \pi \frac{\vartheta_1'(0|2\tau)}{\vartheta_0(0|2\tau)} \left[ \frac{\vartheta_0(\pi z|2\tau)}{\vartheta_1(\pi z|2\tau)} + \frac{\vartheta_1(\pi z|2\tau)}{\vartheta_0(\pi z|2\tau)} \right], \\ w_2(z,\tau) &= \pi \frac{\vartheta_1'(0|2\tau)}{\vartheta_0(0|2\tau)} \left[ \frac{\vartheta_0(\pi z|2\tau)}{\vartheta_1(\pi z|2\tau)} - \frac{\vartheta_1(\pi z|2\tau)}{\vartheta_0(\pi z|2\tau)} \right], \\ w_3(z,\tau) &= \partial_z \log \frac{\vartheta_1(\pi z|2\tau)}{\vartheta_0(\pi z|2\tau)}. \end{aligned}$$

From (B.9) we find that<sup>47</sup>

$$(B.15) \quad \begin{aligned} w_1(z+1) &= -w_1(z), & w_1(z+\tau) &= w_1(z), \\ w_2(z+1) &= -w_2(z), & w_2(z+\tau) &= -w_2(z), \\ w_3(z+1) &= w_3(z), & w_3(z+\tau) &= -w_3(z). \end{aligned}$$

This is consistent with eqn. (10.7).

### B.4 Relations

Among the three solutions for  $G = SL_2$ , the elliptic solution is the most general, and the other two can be obtained by limits (together with suitable overall rescaling):

$$(B.16) \quad R_{\hbar,\tau}^{\text{elliptic}}(z) \xrightarrow{\tau \rightarrow i\infty} R_{\hbar}^{\text{trig}}(z) \xrightarrow{\hbar \rightarrow 0} R^{\text{rational}}(z/\hbar).$$

However, the trigonometric  $R$ -matrix has a generalization with a non-zero ‘‘field’’ turned on; this generalization cannot be obtained from a reduction of the eight-vertex model. In section 10.3, we explained this fact from an analysis of the four-dimensional gauge theory on  $\mathbb{R}^2 \times \mathbb{C}^V$ .

## Appendix C. Anomalies to Quantizing Wilson Lines

In this section we will perform the cohomology calculation referenced in section 8.

The main result we will show is the following.

<sup>47</sup> In particular, we have from (B.9)

$$(B.14) \quad \begin{aligned} \frac{\vartheta_1(\pi(z+1)|2\tau)}{\vartheta_0(\pi(z+1)|2\tau)} &= -\frac{\vartheta_1(\pi z|2\tau)}{\vartheta_0(\pi z|2\tau)}, \\ \frac{\vartheta_1(\pi(z+\tau)|2\tau)}{\vartheta_0(\pi(z+\tau)|2\tau)} &= \frac{\vartheta_0(\pi z|2\tau)}{\vartheta_1(\pi z|2\tau)}. \end{aligned}$$

**Theorem C.1.** *Let  $\mathfrak{g}$  be a simple Lie algebra with no Abelian factors which is not  $\mathfrak{sl}_2$ . Let  $\wedge_0^2 \mathfrak{g}$  denote the kernel of the Lie bracket map from  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$ . Let  $V$  be a representation of  $\mathfrak{g}$ . Then*

1. *If there are no  $G$ -invariant maps from  $\wedge_0^2 \mathfrak{g} \rightarrow \text{End}(V)$  then*

$$(C.1) \quad H^2(\mathfrak{g}[[z]], \text{End}(V)) = 0.$$

2. *Let  $H_{(k)}^2(\mathfrak{g}[[z]], \text{End}(V))$  denote the part of the cohomology of weight  $k$  under the scaling of  $z$  (this group appears in the study of  $k$ -loop anomalies). Then, there is an isomorphism*

$$(C.2) \quad H_{(2)}^2(\mathfrak{g}[[z]], \text{End}(V)) \cong \text{Hom}_G(\wedge_0^2 \mathfrak{g}, \text{End}(V)),$$

where on the right hand side we have the space of  $G$ -invariant linear operators from  $\wedge_0^2 \mathfrak{g}$  to  $\text{End}(V)$ .

If  $\mathfrak{g} = \mathfrak{sl}_2$ , then  $H_{(k)}^2(\mathfrak{g}[[z]], \text{End}(V)) = 0$  unless  $k = 3$ , and the dimension of  $H_{(3)}^2(\mathfrak{g}[[z]], \text{End}(V))$  is the number of copies of the 5-dimensional irreducible representation of  $\mathfrak{sl}_2$  in  $\text{End}(V)$ .

The proof will take a number of steps, which we will write as separate propositions. The first step is the following.

**Proposition C.2.** *Let  $\mathfrak{g}$  be a simple Lie algebra with no Abelian factors. Let  $V$  be a representation of  $\mathfrak{g}$  with the feature that the image of any  $G$ -invariant linear map  $\mathfrak{g}^{\otimes 2} \rightarrow \text{End}(V)$  consists of copies of the trivial and adjoint representations of  $\mathfrak{g}$ . Then,*

$$(C.3) \quad H^2(\mathfrak{g}[[z]], \text{End}(V)) = 0.$$

The vanishing of this cohomology group implies that there are no anomalies to quantizing the Wilson line in the representation  $V$ .

*Proof.* In this calculation, we will use the Lie algebra  $\mathfrak{g}[z]$  of polynomials with coefficients in  $\mathfrak{g}$ , instead of the Lie algebra  $\mathfrak{g}[[z]]$  of series with coefficients in  $\mathfrak{g}$ . As long as we focus on cohomology classes of a fixed weight under the  $\mathbb{C}^\times$  action which scales  $z$ , the cohomology groups are the same in each case. This is because a cochain of weight  $k$  under this  $\mathbb{C}^\times$  action involves  $z^m$  for  $m \leq k$ .

Let  $R \subset \text{End}(V)$  be the largest sub-representation which contains all the representations occurring in the decomposition of  $\mathfrak{g} \otimes \mathfrak{g}$  into irreducible components. There is a natural injective map

$$(C.4) \quad C^*(\mathfrak{g}[z], R) \rightarrow C^*(\mathfrak{g}[z], \text{End}(V)).$$

It is a standard fact of Lie algebra cohomology that we find the same cohomology groups if we use  $G$ -invariant cocycles. It follows from the definition of

$R$  that, on  $G$ -invariant cocycles, this map is an isomorphism in degrees  $\leq 2$ . From this we see that this map must be an isomorphism on degree 2.

This reduces the problem to showing that  $H^2(\mathfrak{g}[z], R) = 0$ . The assumptions we state in the proposition are that  $R$  is a direct sum of copies of the adjoint and trivial representations. We therefore need to show that  $H^2(\mathfrak{g}[z], \mathbb{C}) = 0$  and that  $H^2(\mathfrak{g}[z], \mathfrak{g}) = 0$ .

Since we are working in low cohomological degrees, both of these statements can be proved by hand with a direct computation. However, it is more convenient to invoke a general result of Fishel, Grojnowski, and Teleman [31], who calculate the Lie algebra cohomology of  $\mathfrak{g}[z]$  with various coefficients.

Their results (Theorem B, p. 6 of [31]) tell us that  $H^2(\mathfrak{g}[z], \mathbb{C}) = 0$ . We need to show that  $H^2(\mathfrak{g}[z], \mathfrak{g}) = 0$  as well. This follows a little indirectly from their analysis, as we will now explain.

We write  $\mathfrak{g}[z^{-1}]$  for the space of  $\mathfrak{g}$ -valued functions of  $z$  of the form

$$\alpha_{a_0} t_{a_0} + z^{-1} \alpha_{a_1} t_{a_1} + \dots$$

These are meromorphic functions on  $\mathbb{C}$ , which have a finite order pole at 0 and no other poles, and which are regular at infinity. There is a natural residue pairing

$$(C.5) \quad \langle t_a f(z), t_b g(z^{-1}) \rangle = 2\delta_{ab} \frac{1}{2\pi i} \oint f(z) g(z^{-1}) z^{-1} dz$$

between  $\mathfrak{g}[z^{-1}]$  and  $\mathfrak{g}[z]$ .

The space  $\mathfrak{g}[z^{-1}]$  has the structure of module for  $\mathfrak{g}[z]$  whereby

$$(C.6) \quad (t_a z^k) \cdot (t_b z^{-l}) = f_{ab}^c t_c z^{k-l} \delta_{k-l \leq 0}.$$

This identifies  $\mathfrak{g}[z^{-1}]$  with the dual representation to the adjoint representation of  $\mathfrak{g}[z]$ , that is, with the coadjoint representation. (Because these are infinite dimensional modules, one has to take care with the meaning of “dual”: we mean the restricted dual, where we take the direct sum over the weight spaces of the  $\mathbb{C}^\times$  action which scales  $z$ . This use of restricted duals and the corresponding restricted cohomology elides the difference between  $\mathfrak{g}[z]$  and  $\mathfrak{g}[[z]]$ , so we will use the algebra  $\mathfrak{g}[z]$  in the course of the proof.)

Fishel, Grojnowski and Teleman show (Theorem B, p. 6 of [31]) that the cohomology of  $\mathfrak{g}[z]$  with coefficients in the coadjoint representation  $\mathfrak{g}[z^{-1}]$  is a graded tensor product of the cohomology of the compact group  $G$  with a graded vector space consists of  $\mathbb{C}[z^{-1}]$  in degree 0 and  $z^{-1}\mathbb{C}[z^{-1}]$  in degree 1. We have written these cohomology groups in a way compatible with the  $\mathbb{C}^\times$  which scales  $z$ . In particular, since the cohomology of the group  $G$  lives in degree 0 and degrees  $\geq 3$ , we find that there is no second cohomology of  $\mathfrak{g}[z]$  with coefficients in  $\mathfrak{g}[z^{-1}]$ .

To describe the cocycles for these cohomology classes, it is more convenient to use cohomology relative to the subalgebra  $\mathfrak{g} \subset \mathfrak{g}[z]$ . A relative  $k$ -cocycle is given by a  $G$ -invariant linear map

$$(C.7) \quad \wedge^k(z\mathfrak{g}[z]) \rightarrow \mathfrak{g}[z^{-1}].$$

The absolute cohomology groups are the tensor product of the relative cohomology groups with the cohomology of the group  $G$ . Therefore they coincide in degrees less than 3.

The relative 1-cocycle associated to an element  $z^{-k} \in z^{-1}\mathbb{C}[z^{-1}]$  is a linear map

$$(C.8) \quad \begin{array}{ccc} z\mathfrak{g}[z] & \rightarrow & \mathfrak{g}[z^{-1}] \\ \psi & & \psi \\ t_a f(z) & \mapsto & t_a \pi_-(z^{-k} f(z)), \end{array}$$

where  $\pi_-$  indicates the operation of projection of an element of  $\mathfrak{g}[z, z^{-1}]$  onto the nonpositive powers of  $z$ .

There is a short exact sequence of  $\mathfrak{g}[z]$ -representations

$$(C.9) \quad 0 \rightarrow \mathfrak{g} \rightarrow \mathfrak{g}[z^{-1}] \xrightarrow{z} \mathfrak{g}[z^{-1}] \rightarrow 0,$$

where the second map is given by multiplying by  $z$ , using the convention that we only retain negative powers of  $z$ .

This short exact sequence leads to a long exact sequence on cohomology

$$(C.10) \quad \begin{array}{c} \cdots \rightarrow H^1(\mathfrak{g}[z], \mathfrak{g}) \rightarrow H^1(\mathfrak{g}[z], \mathfrak{g}[z^{-1}]) \\ \rightarrow H^1(\mathfrak{g}[z], \mathfrak{g}[z^{-1}]) \rightarrow H^2(\mathfrak{g}[z], \mathfrak{g}) \rightarrow 0, \end{array}$$

where we use the fact that  $H^2(\mathfrak{g}[z], \mathfrak{g}[z^{-1}]) = 0$ .

This leads to an exact sequence

$$(C.11) \quad \begin{array}{c} \cdots \rightarrow H^1(\mathfrak{g}[z], \mathfrak{g}) \rightarrow z^{-1}\mathbb{C}[z^{-1}] \xrightarrow{z} z^{-1}\mathbb{C}[z^{-1}] \\ \rightarrow H^2(\mathfrak{g}[z], \mathfrak{g}) \rightarrow 0. \end{array}$$

From this we conclude that  $H^2(\mathfrak{g}[z], \mathfrak{g}) = 0$ , as desired.  $\square$

For any representation  $R$  of  $\mathfrak{g}$ , let  $H_{(k)}^i(\mathfrak{g}[z], R)$  denote the cohomology group of weight  $-k$  under the  $\mathbb{C}^\times$  action which scales  $z$ . This is the cohomology group that will play a role in quantization at  $k$  loops.

Next, we will prove a more difficult cohomology vanishing result, also based on the results of [31].

**Proposition C.3.** *For any simple Lie algebra with no Abelian factors which is not  $\mathfrak{sl}_n$  with  $n > 2$ , we have*

$$(C.12) \quad H_{(k)}^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) = 0 \text{ unless } k = 3.$$

Further, there is an exact sequence

$$(C.13) \quad \begin{array}{c} 0 \rightarrow H_{(3)}^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow \mathbb{C} \rightarrow H_{(2)}^2(\mathfrak{g}[z], \wedge^2 \mathfrak{g}) \\ \rightarrow H_{(3)}^3(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow 0. \end{array}$$

Note that this proposition applies to  $\mathfrak{sl}_2$ .

*Proof.* The proof again uses the results of Fishel, Grojnowski and Teleman [31]. Let us consider the  $\mathfrak{g}[z]$ -module  $\text{Sym}^2 \mathfrak{g}[z^{-1}]$ . The results of [31] show that

$$(C.14) \quad \begin{array}{c} H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}[z^{-1}]) = \wedge^2 z^{-1} \mathbb{C}[z^{-1}], \\ H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}[z^{-1}]) = (\text{Sym}^3 \mathfrak{g})^G \otimes z^{-1} \mathbb{C}[z^{-1}], \\ H^0(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}[z^{-1}]) = \mathbb{C}[z^{-1}]. \end{array}$$

We have written this isomorphism in a way compatible with the  $\mathbb{C}^\times$ -action which scales  $z$ . Note that, unless  $\mathfrak{g}$  is  $\mathfrak{sl}_n$  for  $n > 2$ ,  $(\text{Sym}^3 \mathfrak{g})^G = 0$ . Since we restrict to Lie algebras which are not of this form, we have  $H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}[z^{-1}]) = 0$ .

Define operators

$$(C.15) \quad \begin{array}{c} z^n : \text{Sym}^2 \mathfrak{g}[z^{-1}] \mapsto \text{Sym}^2 \mathfrak{g}[z^{-1}] \\ z^n((z^{-k} t_a)(z^{-l} t_b)) = (z^{n-k} t_a)(z^{-l} t_b) \delta_{k-n \geq 0} \\ \quad + (z^{-k} t_a)(z^{n-l} t_b) \delta_{l-n \geq 0}. \end{array}$$

for  $n \geq 1$ . Note that the operator  $z$  is surjective. We let  $M$  denote the kernel of  $z$ . The space  $M$  is spanned by the elements

$$(C.16) \quad \begin{array}{c} A_{ab}(t_a)(t_b z^{-2k}) - A_{ab}(t_a z^{-1})(t_b z^{-2k}) \pm \dots \\ + (-1)^{k+1} A_{ab}(t_a z^{1-k})(t_b z^{-k-1}) + \frac{1}{2} (-1)^k A_{ab}(t_a z^{-k})(t_b z^{-k}), \end{array}$$

where  $A_{ab}$  is a symmetric tensor, and the elements

$$(C.17) \quad \begin{array}{c} B_{ab}(t_a)(t_b z^{-2k-1}) - B_{ab}(t_a z^{-1})(t_b z^{-2k}) \pm \dots \\ + (-1)^{k+1} B_{ab}(t_a z^{1-k})(t_b z^{-k-2}) + (-1)^k B_{ab}(t_a z^{-k})(t_b z^{-k-1}), \end{array}$$

where  $B_{ab}$  is antisymmetric.

Thus, as a representation of  $\mathfrak{g}$ , the module  $M$  decomposes as

$$(C.18) \quad M = \text{Sym}^2 \mathfrak{g} \oplus z^{-1} \wedge^2 \mathfrak{g} \oplus z^{-2} \text{Sym}^2 \mathfrak{g} \dots,$$

where the powers of  $z$  indicate the  $\mathbb{C}^\times$ -weights.

As a  $\mathfrak{g}[z]$ -module, the structure is more complicated: elements in  $z\mathfrak{g}$  moves us between summands of different weight.

Now, let us identify the cohomology of  $\mathfrak{g}[z]$  with coefficients in  $M$ . Because we have a short exact sequence

$$(C.19) \quad 0 \rightarrow M \rightarrow \text{Sym}^2(\mathfrak{g}[z^{-1}]) \xrightarrow{z} \text{Sym}^2(\mathfrak{g}[z^{-1}]) \rightarrow 0,$$

we get a long exact sequence in cohomology

$$(C.20) \quad \begin{array}{c} 0 \rightarrow H^0(\mathfrak{g}[z], M) \rightarrow \mathbb{C}[z^{-1}] \xrightarrow{z} \mathbb{C}[z^{-1}] \\ \rightarrow H^1(\mathfrak{g}[z], M) \rightarrow 0 \rightarrow 0 \\ \rightarrow H^2(\mathfrak{g}[z], M) \rightarrow \wedge^2(z^{-1}\mathbb{C}[z^{-1}]) \xrightarrow{z} \wedge^2(z^{-1}\mathbb{C}[z^{-1}]) \\ \rightarrow H^3(\mathfrak{g}[z], M) \rightarrow 0. \end{array}$$

The endomorphism  $z$  of  $\wedge^2(z^{-1}\mathbb{C}[z^{-1}])$  is defined by

$$(C.21) \quad z(z^{-k} \wedge z^{-l}) = z^{1-k} \wedge z^{-l} \delta_{k-1 \geq 1} + z^{-k} \wedge z^{-l-1} \delta_{l-1 \geq 1}.$$

The kernel of the operator  $z$  is spanned by the elements

$$(C.22) \quad z^{-1} \wedge z^{-2k} - z^{-2} \wedge z^{1-2k} \pm \dots + (-1)^k z^{-k} \wedge z^{-1-k}$$

for  $k \geq 1$ . Thus,  $H^2(\mathfrak{g}[z], M)$  is spanned by these elements, in weights 3, 5, .... To sum up, we find

$$(C.23) \quad \begin{aligned} H^0(\mathfrak{g}[z], M) &= \mathbb{C}, \\ H^1(\mathfrak{g}[z], M) &= 0, \\ H^2(\mathfrak{g}[z], M) &= \mathbb{C} \cdot z^{-3} \oplus \mathbb{C} \cdot z^{-5} \oplus \dots, \\ H^3(\mathfrak{g}[z], M) &= 0, \end{aligned}$$

where the powers of  $z$  indicate the weights under the  $\mathbb{C}^\times$ -action.

Let  $M_0 \subset M$  be the kernel of the operator  $z^2$ . This operator is surjective on  $H^2(M)$ , with kernel  $\mathbb{C}$  in weight  $-3$ , and surjective on  $M$ , but acts by zero on  $H^0(M)$  and  $H^1(M)$ . We have a long exact sequence

$$(C.24) \quad \begin{aligned} 0 \rightarrow H^0(\mathfrak{g}[z], M_0) &\rightarrow \mathbb{C} \xrightarrow{0} \mathbb{C} \\ &\rightarrow H^1(\mathfrak{g}[z], M_0) \rightarrow 0 \rightarrow 0 \\ &\rightarrow H^2(\mathfrak{g}[z], M_0) \rightarrow z^{-3}\mathbb{C}[z^{-2}] \xrightarrow{z^2} z^{-3}\mathbb{C}[z^{-2}] \\ &\rightarrow H^3(\mathfrak{g}[z], M_0) \rightarrow 0. \end{aligned}$$

This sequence allows us to calculate the cohomology groups of  $H^*(\mathfrak{g}[z], M_0)$ . Before we state the answer, we should note that the boundary maps in this long exact sequence shift the weight under the  $\mathbb{C}^\times$ -action by  $-2$ . This is because of the appearance of the operator  $z^2$ , of weight 2, in the short exact sequence of modules leading to this long exact sequence of cohomology groups.

From this exact sequence we see that

$$(C.25) \quad \begin{aligned} H^0(\mathfrak{g}[z], M_0) &= \mathbb{C}, \\ H^1(\mathfrak{g}[z], M_0) &= \mathbb{C} \cdot z^{-2}, \\ H^2(\mathfrak{g}[z], M_0) &= \mathbb{C} \cdot z^{-3}. \end{aligned}$$

Since  $M_0$  is the intersection of the kernel of  $z$  and the kernel of  $z^2$  in  $\text{Sym}^2 \mathfrak{g}[z^{-1}]$ , one can show that there is an exact sequence

$$(C.26) \quad 0 \rightarrow \text{Sym}^2 \mathfrak{g} \rightarrow M_0 \rightarrow z^{-1} \wedge^2 \mathfrak{g} \rightarrow 0,$$

from which we derive a long exact sequence

$$(C.27) \quad \begin{aligned} 0 \rightarrow H^0(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) &\rightarrow H^0(\mathfrak{g}[z], M_0) \rightarrow H^0(\mathfrak{g}[z], z^{-1} \wedge^2 \mathfrak{g}) \\ &\rightarrow H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow H^1(\mathfrak{g}[z], M_0) \rightarrow H^1(\mathfrak{g}[z], z^{-1} \wedge^2 \mathfrak{g}) \\ &\rightarrow H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow H^2(\mathfrak{g}[z], M_0) \rightarrow H^2(\mathfrak{g}[z], z^{-1} \wedge^2 \mathfrak{g}) \\ &\rightarrow H^3(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow 0. \end{aligned}$$

Now,  $H^0(\mathfrak{g}[z], \wedge^2 \mathfrak{g}) = 0$  because there are no  $G$ -invariant elements in  $\wedge^2 \mathfrak{g}$  for any simple Lie algebra  $\mathfrak{g}$ . Further  $H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$  is zero because there are no copies of the adjoint representation in  $\text{Sym}^2 \mathfrak{g}$  (recall that we are assuming that  $\mathfrak{g}$  is not  $\mathfrak{sl}_n$  for  $n > 2$ ).

We also have  $H^1(\mathfrak{g}[z], z^{-1} \wedge^2 \mathfrak{g}) = z^{-2}\mathbb{C}$ , because there is one copy of the adjoint in  $\wedge^2 \mathfrak{g}$ .

From this, we find that we have an exact sequence

$$(C.28) \quad \begin{aligned} 0 \rightarrow \mathbb{C} \cdot z^{-2} &\rightarrow \mathbb{C} \cdot z^{-2} \xrightarrow{\delta} H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \\ &\rightarrow H^2(\mathfrak{g}[z], M_0) \rightarrow H^2(\mathfrak{g}[z], z^{-1} \wedge^2 \mathfrak{g}) \rightarrow H^3(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g}) \rightarrow 0. \end{aligned}$$

The boundary map

$$(C.29) \quad \delta : \mathbb{C} \cdot z^{-2} \rightarrow H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$$

must be zero.

We conclude that  $H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$  is a subspace of

$$(C.30) \quad H^2(\mathfrak{g}[z], M_0) = \mathbb{C} \cdot z^{-3}.$$

In particular,  $H^2(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$  is only non-zero in weight 3.  $\square$

Next, let us perform the same analysis in the case that  $\mathfrak{g}$  is  $\mathfrak{sl}_n$ .

**Proposition C.4.** *Suppose that  $\mathfrak{g} = \mathfrak{sl}_n$  for  $n > 2$ . Then,*

$$(C.31) \quad H_{(k)}^2(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n) = 0 \text{ unless } k = 3.$$

*As before, there is an exact sequence*

$$(C.32) \quad \begin{aligned} 0 \rightarrow H_{(3)}^2(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n) &\rightarrow \mathbb{C} \rightarrow H_{(2)}^2(\mathfrak{sl}_n[z], \wedge^2 \mathfrak{sl}_n) \\ &\rightarrow H_{(3)}^3(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n) \rightarrow 0. \end{aligned}$$

*Proof.* The cohomology groups  $H_{(k)}^*(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n)$  are acted on by the outer automorphism group  $\mathbb{Z}/2$  of  $\mathfrak{sl}_n$  (which is the automorphism group of the Dynkin diagram). We let  $H_{(k), \text{even}}^*$  and  $H_{(k), \text{odd}}^*$  refer to the cohomology groups which are even and odd under this action. We will first prove that the cohomology groups which are even under the outer automorphism satisfy the statement of the proposition, and then show that the groups which are odd under the outer automorphism are zero.

Note that in the proof of proposition C.3 we used the fact that  $\mathfrak{g}$  is not  $\mathfrak{sl}_n$  with  $n > 2$  in two places. Firstly, we used the fact that Lie algebras which are not of this type do not have an invariant element in  $\text{Sym}^3 \mathfrak{g}$ . An invariant element in  $\text{Sym}^3 \mathfrak{g}$  contributes, according to the results of [31], an element of  $H^1(\mathfrak{g}[z], \text{Sym}^2(\mathfrak{g}[z^{-1}]))$ . However, because the invariant element in  $\text{Sym}^3 \mathfrak{sl}_n$  is odd under the outer automorphism, the corresponding element of  $H^1(\mathfrak{sl}_n[z], \text{Sym}^2(\mathfrak{sl}_n[z^{-1}]))$  is also odd, and so does not contribute when we analyze the cohomology which is even under the outer automorphism.



The other place where we used the assumption that  $\mathfrak{g}$  is not  $\mathfrak{sl}_n$  with  $n > 2$  was when we asserted that  $H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$  must be zero. This is essentially the same point, as any element of  $H^1(\mathfrak{g}[z], \text{Sym}^2 \mathfrak{g})$  must come from an invariant element in  $\mathfrak{g} \otimes \text{Sym}^2 \mathfrak{g}$ , and the only invariant element (when  $\mathfrak{g} = \mathfrak{sl}_n$ ,  $n > 2$ ) is totally symmetric. Because this element is odd under the outer automorphism of  $\mathfrak{sl}_n$ , we find again that it can not contribute to the cohomology which is even under this automorphism.

Therefore, when we restrict to the cohomology which is even under the outer auto-morphisms, proposition C.3 holds with the same proof.

We need to calculate the cohomology which is odd under the outer automorphism. We are interested in  $H^2_{\text{odd}}(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n)$ . Any odd two-cocycle can be represented as a sum of  $\mathfrak{sl}_n$ -invariant linear operators

$$(C.33) \quad (z^k \mathfrak{sl}_n) \otimes (z^l \mathfrak{sl}_n) \rightarrow \text{Sym}^2 \mathfrak{sl}_n,$$

Any such linear operator can be thought of as an invariant tensor in  $\mathfrak{sl}_n^{\otimes 4}$ , which is invariant under the permutation of the first two factors. If  $A_1, A_2, B, C$  denote elements of  $\mathfrak{sl}_n$  then there are three such tensors, invariant under permutation of  $A_1$  and  $A_2$ :

$$\begin{aligned} & \text{Tr}(A_1 A_2 B C) + \text{Tr}(A_2 A_1 B C), \\ & \text{Tr}(A_1 A_2 C B) + \text{Tr}(A_2 A_1 C B), \\ & \text{Tr}(A_1 B A_2 C) + \text{Tr}(A_2 B A_1 C). \end{aligned}$$

The only linear combination of these three tensors which is odd under the outer automorphism is

$$(C.34) \quad \text{Tr}(A_1 A_2 [B, C]) + \text{Tr}(A_2 A_1 [B, C])$$

(recall that the outer automorphism sends  $A \in \mathfrak{sl}_n$  to  $-A^T$ ).

This invariant tensor can be viewed as a linear operator

$$(C.35) \quad (z^k \mathfrak{sl}_n) \otimes (z^l \mathfrak{sl}_n) \rightarrow \mathfrak{sl}_n \subset \text{Sym}^2 \mathfrak{sl}_n.$$

We have just shown that the natural map

$$(C.36) \quad H^2(\mathfrak{sl}_n[z], \mathfrak{sl}_n) \rightarrow H^2_{\text{odd}}(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n)$$

is surjective. However, we already know that  $H^2(\mathfrak{sl}_n[z], \mathfrak{sl}_n) = 0$ , so we conclude that  $H^2_{\text{odd}}(\mathfrak{sl}_n[z], \text{Sym}^2 \mathfrak{sl}_n) = 0$  also.  $\square$

The next result we need is the following.

**Proposition C.5.** *Let  $\wedge^2_0 \mathfrak{g} \subset \wedge^2 \mathfrak{g}$  denote the kernel of the Lie bracket map  $\wedge^2 \mathfrak{g} \rightarrow \mathfrak{g}$ . Then,*

$$(C.37) \quad H^2_{(2)}(\mathfrak{g}[z], \wedge^2 \mathfrak{g}) = \text{Hom}^G(\wedge^2_0 \mathfrak{g}, \wedge^2_0 \mathfrak{g}),$$

where on the right hand side we have the space of  $G$ -invariant maps from  $\wedge^2_0 \mathfrak{g}$  to itself.

Further, the map

$$(C.38) \quad \mathbb{C} = H^2_{(3)}(\mathfrak{g}[z], M_0) \rightarrow H^2_{(2)}(\mathfrak{g}[z], \wedge^2 \mathfrak{g})$$

is surjective for  $\mathfrak{g} \neq \mathfrak{sl}_2$  (where we are using the notation  $M_0$  from the proof of proposition C.3).

*Proof.* We can calculate  $H^2_{(2)}(\mathfrak{g}[z], \wedge^2 \mathfrak{g})$  using Lie algebra cohomology relative to  $\mathfrak{g}$ . A basis for the relative 2-cochains is the space of  $G$ -invariant maps  $\wedge^2(z\mathfrak{g}) \rightarrow \wedge^2 \mathfrak{g}$ . Every such relative two-cochain is closed, because there are no relative three-cochains of weight 2. Elements of the form

$$(C.39) \quad f^a_{bc}(z t_a \wedge z t_b) \otimes R^a_{ef}(t_e \wedge t_f)$$

are exact, where  $R^a_{ef}$  defines some  $G$ -invariant map from  $\mathfrak{g}$  to  $\wedge^2 \mathfrak{g}$ .

For all simple Lie algebras, there are no copies of the adjoint representation in  $\wedge^2_0(\mathfrak{g})$ . This argument tells us that the space of closed, but not exact, 2-cochains of weight  $-2$  is the space of  $G$ -invariant maps  $\wedge^2 \mathfrak{g} \rightarrow \wedge^2_0 \mathfrak{g}$ .

For  $G \neq SL_2$ , this space is non-trivial. The final thing we need to check is that for  $G \neq SL_2$ , the map

$$(C.40) \quad \mathbb{C} = H^2_{(3)}(\mathfrak{g}[z], M_0) \rightarrow H^2_{(2)}(\mathfrak{g}[z], \wedge^2 \mathfrak{g})$$

is non-zero. This follows from the explicit description of the cocycles of  $\mathfrak{g}[z]$  with values in  $\text{Sym}^* \mathfrak{g}[z^{-1}]$  given in [31].  $\square$

The following corollary sums up what we have learned about the cohomology groups controlling the obstructions to quantizing a Wilson line.

**Corollary C.6.** *For a representation  $V$  of a simple Lie algebra  $\mathfrak{g} \neq \mathfrak{sl}_2$ , consider the group  $H^2_{(k)}(\mathfrak{g}[z], \text{End}(V))$  which contains possible anomalies to quantizing the Wilson line associated to  $V$  at  $k$  loops. Then,  $H^2_{(k)}(\mathfrak{g}[z], \text{End}(V)) = 0$  unless there exists a non-trivial  $G$ -invariant map from  $\wedge^2_0 \mathfrak{g}$  to  $\text{End}(V)$ .*

*Proof.* Two-cocycles representing classes in  $H^2(\mathfrak{g}[z], \text{End}(V))$  are given by  $G$ -invariant linear maps  $\wedge^2 z\mathfrak{g}[z] \rightarrow \text{End}(V)$ . Any such  $G$ -invariant map must factor through some copies of  $\text{Sym}^2 \mathfrak{g}$ ,  $\mathfrak{g}$ , or  $\wedge^2_0 \mathfrak{g}$ . We have seen that the cohomology with coefficients in  $\text{Sym}^2 \mathfrak{g}$  and  $\mathfrak{g}$  is zero. So, assuming there are no non-trivial maps from  $\wedge^2_0 \mathfrak{g} \rightarrow \text{End}(V)$ , the cohomology groups are all zero.  $\square$

In fact, one can strengthen this result with some further work. For instance, one can show that  $H^2_{(k)}(\mathfrak{g}[z], \wedge^2_0 \mathfrak{g}) = 0$  unless  $k = 2, 4, 6$ . We have already described the rank of the cohomology group when  $k = 2$ . Computing explicitly the rank when  $k = 4, 6$  is a bit involved, and since we do not ultimately need to know the answer, we have not included these calculations.

## Appendix D. Derivation of Eqn. (6.13)

In the Feynman diagram of Fig. 22, we have three vertices, two vertices  $u, v$  on the Wilson line  $K$

$$(D.1) \quad \left(u, \frac{fu^2}{2}, 0, 0\right), \quad \left(v, \frac{fv^2}{2}, 0, 0\right),$$

and one vertex  $w$  in the bulk, which we parametrize as  $(x, y, z, \bar{z})$ . At the vertex  $u$  of the Wilson line, we have the line element

$$(D.2) \quad \int_K Ads = (A_x(u) + A_y(u)fu)du.$$

In the following, to save space we sometimes write  $A_x(u) := A_x(x(u), y(u), 0, 0)$ . The similar expression applies to the vertex  $v$ .

The diagram has a symmetry factor of 1/2, which we can incorporate by restricting the integration region to be  $u < v$ .

Now we connect the vertices by a propagator, to the vertex  $w$ . We are interested in linear order in  $f$ .

One possibility is to take  $A_x$  from both, and then obtain a factor of  $f$  from the expansion of the propagator. However, this contributes an expression of the form

$$(D.3) \quad \pm \langle A_x(u)A_y(w) \rangle \langle A_x(v)A_{\bar{z}}(w) \rangle A_x(w)$$

(or with  $u$  and  $v$  exchanged), and since we have  $A_x(w)$  this does not contribute to  $\Lambda$ .

Another possibility is to take one  $A_x$  and  $A_y$  one from  $u$  and  $v$ :

$$(D.4) \quad (fv)A_x(u)A_y(v) \quad \text{or} \quad (fu)A_y(u)A_x(v).$$

When we consider the propagator, there are three different possibilities for (D.4):

$$(D.5) \quad \begin{aligned} &(f(v-u)) \langle A_x(u)A_y(w) \rangle \langle A_y(v)A_{\bar{z}}(w) \rangle A_x(w), \\ &(f(v-u)) \langle A_x(u)A_{\bar{z}}(w) \rangle \langle A_y(v)A_x(w) \rangle A_y(w), \\ &(f(v-u)) \langle A_x(u)A_y(w) \rangle \langle A_y(v)A_x(w) \rangle A_{\bar{z}}(w). \end{aligned}$$

Using the expression for the propagator (4.5), the expressions (D.5) can be simplified, and when combined with differential forms we have

$$(D.6) \quad \begin{aligned} &\frac{1}{(2\pi)^2} (f(v-u)) \frac{(2\bar{z})(x-u)}{d(u,w)^4 d(v,w)^4} A_x(w) \cdot dz(dudy)(dvd\bar{z})dx, \\ &\frac{1}{(2\pi)^2} (f(v-u)) \frac{(y-\frac{fu^2}{2})(2\bar{z})}{d(u,w)^4 d(v,w)^4} A_y(w) \cdot dz(dud\bar{z})(dvdx)dy, \\ &\frac{1}{(2\pi)^2} (f(v-u)) \frac{(-1)(2\bar{z})(2\bar{z})}{d(u,w)^4 d(v,w)^4} A_{\bar{z}}(w) \cdot dz(dudy)(dvdx)d\bar{z}, \end{aligned}$$

with distance  $d(u, w)$  defined by

$$(D.7) \quad d(u, w)^2 = (x-u)^2 + \left(y - \frac{fu^2}{2}\right)^2 + |z|^2.$$

We also need to supplement these expressions by a factor  $\hbar h^\vee$ , where  $h^\vee$  is the color factor explained in the text and  $\hbar$  is the loop counting parameter. In the leading order in  $f$ , we have  $d(u, w)^2 \simeq (x-u)^2 + y^2 + |z|^2$ , and hence we obtain

$$(D.8) \quad -\frac{1}{(2\pi)^2} \frac{i}{2\pi} \hbar h^\vee \int_{u<v} dudv \int_{\mathbb{R}^4} dx dy dz d\bar{z} (f(v-u)) \times \frac{2\bar{z}(x-u)A_x(w) + 2y\bar{z}A_y(w) + 4\bar{z}^2 A_{\bar{z}}(w)}{((x-u)^2 + y^2 + |z|^2)^2 ((x-v)^2 + y^2 + |z|^2)^2}.$$

The equation above can be expressed in the form (6.8) if we define

$$(D.9) \quad \Theta_0 = -\frac{1}{(2\pi)^2} \frac{i}{2\pi} \hbar h^\vee \int_{u<v} dudv f(v-u) \times \frac{2(\bar{z})(x-u)(dy \wedge dz \wedge d\bar{z}) - 2y\bar{z}(dx \wedge dz \wedge d\bar{z}) + 4\bar{z}^2(dx \wedge dy \wedge dz)}{((x-u)^2 + y^2 + |z|^2)^2 ((x-v)^2 + y^2 + |z|^2)^2}.$$

In the notation of the main text, we have  $\Theta = z\Theta_0$  and  $\Theta = dx \wedge \Lambda + \Lambda'$  to obtain

$$(D.10) \quad \Lambda = -\frac{1}{(2\pi)^2} \frac{2fi}{2\pi} \hbar h^\vee \times \int_{u<v} dudv (v-u) \frac{-yz\bar{z}(dz \wedge d\bar{z}) + 2z\bar{z}^2(dy \wedge dz)}{((x-u)^2 + y^2 + |z|^2)^2 ((x-v)^2 + y^2 + |z|^2)^2}.$$

We obtain (6.13) after doing the  $u$  and  $v$  integrals:

$$(D.11) \quad \int_{-\infty \leq u < v \leq \infty} dudv (v-u) \frac{1}{(u^2 + a^2)^2 (v^2 + a^2)^2} = \frac{1}{a^5} \frac{3\pi}{8}.$$

## Appendix E. Evaluation of Eqn. (8.34)

In this appendix we present details on the evaluation of the integral (8.34).

Let us first evaluate the part relevant for the  $v_1$  integral inside the expression (8.34):

$$(E.1) \quad \int_{v_1} P(v_0, v_1) \wedge dz_1(z_1) \wedge P(v_1, v_2),$$

where we have introduced, temporarily, a vertex labelled  $v_0$  with coordinates  $v_0 = (x_0, y_0, z_0, \bar{z}_0)$ . Later we will set these coordinates to zero.

From the explicit expression for the propagator in eqn. (4.5) we have

$$(E.2) \quad P(v_i, v_j) = \frac{1}{2\pi} \frac{x_{ij} dy_{ij} \wedge d\bar{z}_{ij} - y_{ij} dx_{ij} \wedge d\bar{z}_{ij} + 2\bar{z}_{ij} dx_{ij} \wedge dy_{ij}}{d(v_i, v_j)^4},$$

where the distance  $d(v_i, v_j)$  between two points  $v_i, v_j$  is given by  $d(v_i, v_j)^2 = x_{ij}^2 + y_{ij}^2 + |z_{ij}|^2$ , and  $x_{ij} = x_i - x_j$ ,  $z_{ij} = z_i - z_j$ , etc.

We can calculate that

$$(E.3) \quad P(v_0, v_1) \wedge P(v_1, v_2) = \frac{1}{(2\pi)^2} dx_1 dy_1 dz_1 d\bar{z}_1 \\ \times \frac{1}{d(v_0, v_1)^4 d(v_1, v_2)^4} [(x_{01} y_{12} d\bar{z}_2 - 2x_{01} \bar{z}_{12} dy_2 + 2y_{01} \bar{z}_{12} dx_2 \\ - y_{01} x_{12} d\bar{z}_2 + 2\bar{z}_{01} x_{12} dy_2 - 2\bar{z}_{01} y_{12} dx_2)],$$

where we have dropped all terms involving  $dx_0$ ,  $dy_0$ ,  $d\bar{z}_0$  because we will not be integrating over the vertex  $v_0$ . From this we see that

$$(E.4) \quad P(v_0, v_1) \wedge z_1 dz_1 P(v_1, v_2) \\ = \frac{1}{(2\pi)^2} dx_1 dy_1 dz_1 d\bar{z}_1 \left[ \partial_{\bar{z}_0} \frac{1}{d(v_0, v_1)^2 d(v_1, v_2)^4} \right] \\ \times (x_{01} y_{12} d\bar{z}_2 - 2x_{01} \bar{z}_{12} dy_2 + 2y_{01} \bar{z}_{12} dx_2 - y_{01} x_{12} d\bar{z}_2 \\ + 2\bar{z}_{01} x_{12} dy_2 - 2\bar{z}_{01} y_{12} dx_2).$$

after we set the  $v_0$  coordinates to zero.

Moving the position of the derivative in  $\bar{z}_0$ , we obtain

$$(E.5) \quad P(v_0, v_1) \wedge z_1 dz_1 P(v_1, v_2) \\ = \frac{1}{(2\pi)^2} dx_1 dy_1 dz_1 d\bar{z}_1 \\ \times \partial_{\bar{z}_0} \left[ \frac{1}{d(v_0, v_1)^2 d(v_1, v_2)^4} [(x_{01} y_{12} d\bar{z}_2 - 2x_{01} \bar{z}_{12} dy_2 \\ + 2y_{01} \bar{z}_{12} dx_2 - y_{01} x_{12} d\bar{z}_2 + 2\bar{z}_{01} x_{12} dy_2 - 2\bar{z}_{01} y_{12} dx_2)] \right] \\ \frac{1}{(2\pi)^2} dx_1 dy_1 dz_1 d\bar{z}_1 \frac{(2x_{12} dy_2 - 2y_{12} dx_2)}{d(v_0, v_1)^2 d(v_1, v_2)^4}.$$

By integration by parts, the integral over  $x_1$ ,  $y_1$ ,  $z_1$ ,  $\bar{z}_1$  of all the terms on the first two lines vanishes. For example,

$$\int_{v_1} (x_{01} y_{12} - y_{01} x_{12}) \frac{1}{d(v_0, v_1)^2} \frac{1}{d(v_1, v_2)^4} \\ = \frac{1}{2} \int_{v_1} (x_{01}) \frac{1}{d(v_0, v_1)^2} \left( \partial_{y_1} \frac{1}{d(v_1, v_2)^2} \right) \\ + \frac{1}{2} \int_{v_1} y_{01} \frac{1}{d(v_0, v_1)^2} \left( \partial_{x_1} \frac{1}{d(v_1, v_2)^2} \right) \\ = \frac{1}{2} \int_{v_1} x_{01} \left( \partial_{y_1} \frac{1}{d(v_0, v_1)^2} \right) \frac{1}{d(v_1, v_2)^2} \\ - \frac{1}{2} \int_{v_1} y_{01} \left( \partial_{x_1} \frac{1}{d(v_0, v_1)^2} \right) \frac{1}{d(v_1, v_2)^2} \\ = \int_{v_1} (x_{01} y_{01} - x_{01} y_{01}) \frac{1}{d(v_0, v_1)^4} \frac{1}{d(v_1, v_2)^2} = 0.$$

We are left with the integral

$$(E.6) \quad -\frac{1}{(2\pi)^2} \int_{x_1, y_1, z_1} dx_1 dy_1 dz_1 d\bar{z}_1 \frac{(2x_{12} dy_2 - 2y_{12} dx_2)}{d(v_0, v_1)^2 d(v_1, v_2)^4}.$$

We can evaluate this integral with the help of eqn. (5.13). Choosing  $\alpha = 2$ ,  $\beta = 1$  and using  $\Gamma(3)/(\Gamma(2)\Gamma(1)) = 2$ , we obtain

$$(E.7) \quad -\frac{2}{(2\pi)^2} \int_0^1 dt \int dx_1 dy_1 dz_1 d\bar{z}_1 \frac{(2x_{12} dy_2 - 2y_{12} dx_2)}{(\star)^3},$$

where

$$(E.8) \quad (\star) = (1-t)[x_1^2 + y_1^2 + |z_1|^2] + t[x_{12}^2 + y_{12}^2 + |z_{12}|^2] \\ = (x_1 - tx_2)^2 + (y_1 - ty_2)^2 + |z_1 - tz_2|^2 \\ + t(1-t)[x_2^2 + y_2^2 + |z_2|^2].$$

After shifting integration variables we obtain

$$(E.9) \quad -\frac{2}{(2\pi)^2} \int_0^1 dt \int dx_1 dy_1 dz_1 d\bar{z}_1 \\ \times \frac{(2(x_1 - (1-t)x_2)dy_2 - 2(y_1 - (1-t)y_2)dx_2)}{(x_1^2 + y_1^2 + |z_1|^2 + t(1-t)[x_2^2 + y_2^2 + |z_2|^2])^3} \\ = \frac{4}{(2\pi)^2} \int_0^1 dt (1-t) \int dx_1 dy_1 dz_1 d\bar{z}_1 \\ \times \frac{(x_2 dy_2 - y_2 dx_2)}{(x_1^2 + y_1^2 + |z_1|^2 + t(1-t)[x_2^2 + y_2^2 + |z_2|^2])^3},$$

where we dropped the pieces odd in  $x_1$  or  $y_1$ , and hence do not contribute to the integral. After evaluating the integral over  $x_1$ ,  $y_1$  and then over  $z_1$ ,  $\bar{z}_1$ , the integral of  $t$  becomes trivial, and we obtain (note that  $d_z d\bar{z} = -2i r dr d\theta$  for the polar coordinate  $z = re^{i\theta}$ )

$$(E.10) \quad \frac{4}{(2\pi)^2} \frac{\pi}{2} (x_2 dy_2 - y_2 dx_2) \int_0^1 dt (1-t) \int dz_1 d\bar{z}_1 \\ \times \frac{1}{(|z_1|^2 + t(1-t)[x_2^2 + y_2^2 + |z_2|^2])^2} \\ = \frac{4}{(2\pi)^2} \frac{\pi}{2} (-2i) \pi (x_2 dy_2 - y_2 dx_2) \frac{1}{(x_2^2 + y_2^2 + |z_2|^2)}.$$

We therefore find

$$(E.11) \quad \int_{v_1} P(v_0, v_1) \wedge z_1 dz_1 \wedge P(v_1, v_2) = \frac{1}{i} \frac{(x_2 dy_2 - y_2 dx_2)}{d(v_0, v_2)^2}.$$

Similarly, the part of (8.34) relevant for the  $v_3$  integral is

$$(E.12) \quad \int_{v_3} P(v_2, v_3) \wedge z_3 dz_3 \wedge P(v_3, v_3) = -\frac{1}{i} \frac{((x_2 - \epsilon) dy_2 - y_2 dx_2)}{d(v_0, v_2)^2}.$$

We can now come back to the evaluation of (8.34). Using the results (E.11) and (E.12), we obtain

$$(E.13) \quad 2 \left( \frac{i}{2\pi} \right)^3 \frac{1}{2\pi(i)(-i)} \\ \times \int_{p=0}^{\epsilon} \int_{x, y, z, \bar{z}} \frac{(xdy - ydx)(dz)(y d\bar{z} dp)((x - \epsilon)dy - ydx)}{d(0, v)^2 d(p, v)^4 d(\epsilon, v)^2},$$

where we dropped the index 2 from  $v_2 = (x_2, y_2, z_2, \bar{z}_2)$  to simplify the expressions. This gives

$$(E.14) \quad 2 \frac{i}{(2\pi)^4} \int_{p=0}^{\epsilon} \int_{x,y,z,\bar{z}} \frac{\epsilon y^2 dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-\epsilon)^2 + y^2 + |z|^2)}.$$

We need to take into account two more diagrams in Fig. 28. It turns out that the evaluation is rather similar, with the only different exchange of the role of the points 0,  $p$  and  $\epsilon$  on the Wilson line. This means in addition to (E.14) we have two extra contributions

$$(E.15) \quad 2 \frac{i}{(2\pi)^4} \int_{p=0}^{\epsilon} \int_{x,y,z,\bar{z}} \frac{\epsilon y^2 dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-\epsilon)^2 + y^2 + |z|^2)},$$

$$2 \frac{i}{(2\pi)^4} \int_{p=0}^{\epsilon} \int_{x,y,z,\bar{z}} \frac{\epsilon y^2 dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-\epsilon)^2 + y^2 + |z|^2)}.$$

Summing all the three contributions, we obtain

$$(E.16) \quad - \frac{i}{(2\pi)^4} \int_{p=0}^{\epsilon} \int_{x,y,z,\bar{z}} (\epsilon y) \frac{\partial}{\partial y} \frac{dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-\epsilon)^2 + y^2 + |z|^2)},$$

which after integrating by parts gives

$$(E.17) \quad \frac{i}{(2\pi)^4} \int_{p=0}^{\epsilon} \int_{x,y,z,\bar{z}} \frac{\epsilon}{dx dy dz d\bar{z} dp} \frac{dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-\epsilon)^2 + y^2 + |z|^2)}.$$

After scaling the integration variables by  $\epsilon$  (assuming  $\epsilon > 0$ )  $\epsilon$  depends drops out, as expected:

$$(E.18) \quad \frac{i}{(2\pi)^4} \int_{p=0}^1 \int_{x,y,z,\bar{z}} \frac{dx dy dz d\bar{z} dp}{(x^2 + y^2 + |z|^2)((x-p)^2 + y^2 + |z|^2)((x-1)^2 + y^2 + |z|^2)}.$$

In angular coordinates in  $(y, z, \bar{z})$ -plane, the volume form  $dy dz d\bar{z}$  is  $-8\pi r^2 dr d\Omega_{S^2}$  where  $d\Omega_{S^2}$  is the volume form on the two-sphere of volume 1. Integrating over the two-sphere we get

$$(E.19) \quad \frac{8\pi}{(2\pi)^4} \int_{p=0}^1 dp \int_x \frac{dr r^2}{(x^2 + r^2)((x-p)^2 + r^2)((x-1)^2 + r^2)}.$$

After the  $r$  integral we obtain

$$(E.20) \quad \frac{8\pi}{(2\pi)^4} \frac{\pi}{2} \int_{p=0}^1 dp \times \int_x \frac{dx}{(|x| + |x-p|)(|x| + |x-1|)(|x-p| + |x-1|)}.$$

We can evaluate this integral by dividing into four cases  $x < 0$ ,  $0 < x < p$ ,  $p < x < 1$ ,  $1 < x$ . These integrals are interchanged by the change of coordinates  $x \mapsto 1-x$ ,  $p \mapsto 1-p$ , so that the only two independent integrals are the cases when  $p < x < 1$  and  $1 < x$ . We have

$$(E.21) \quad \int_{0 < p < x < 1} \frac{1}{(2x-p)(1-p)} = \frac{\pi^2}{8},$$

$$(E.22) \quad \int_{0 < p < 1 < x} \frac{1}{(2x-p)(2x-1)(2x-1-p)} = \frac{\pi^2}{24}.$$

Therefore we obtain

$$(E.23) \quad \frac{8\pi}{(2\pi)^4} \frac{\pi}{2} 2 \left( \frac{\pi^2}{8} + \frac{\pi^2}{24} \right) = \frac{1}{12}.$$

Including the factors of  $\hbar^2$  need for the two-loop diagram, we reproduce the numerical factor of (8.35).

## References

- [1] H. Bethe, "Zur Theorie der Metalle. I. Eigenwerte und Eigenfunktionen der linearen Atomkette," *Zeit. fur Physik* 71 (1931) 205-226.
- [2] L. Onsager, "Crystal Statistics I: A Two-Dimensional Model With An Order-Disorder Transition," *Phys. Rev.* 65 (1944) 117-49.
- [3] J. B. McGuire, "Study of Exactly Soluble One-Dimensional N-Body Problems," *J. Math. Phys.* 5, 622 (1964).
- [4] C. N. Yang, "Some Exact Results for the Many Body Problems in One Dimension With Repulsive Delta Function Interaction," *Phys. Rev. Lett.* 19, 1312 (1967).
- [5] R. J. Baxter, "Eight-Vertex Model in Lattice Statistics," *Phys. Rev. Lett.* 26, 832 (1971).
- [6] A. B. Zamolodchikov and A. B. Zamolodchikov, "Factorized s Matrices in Two-Dimensions as the Exact Solutions of Certain Relativistic Quantum Field Models," *Annals Phys.* 120, 253 (1979).
- [7] V. G. Drinfeld, "Quantum Groups," in *Proceedings of the International Congress of Mathematicians, Berkeley, 1986*, American Mathematical Society, 1987.
- [8] V. Chari and A. Pressley, "A Guide to Quantum Groups," Cambridge University Press, 1994.
- [9] M. Jimbo, ed., *Yang-Baxter Equation In Integrable Systems* (World-Scientific, 1989).
- [10] J. H. H. Perk and H. Au-Yang, "Yang-Baxter Equation," in *Encyclopedia of Mathematical Physics*, eds. J.-P. Francoise, G. L. Naber and S. T. Tsou (Oxford: Elsevier, 2006), Vol. 5, pp. 465-73, with an expanded version in arXiv:math-ph/0606053.
- [11] K. Costello, "Integrable Lattice Models From Four-Dimensional Field Theories," *Proc. Symp. Pure Math.* 88 (2014) 3-24, arXiv:1303.2632.
- [12] K. Costello, "Supersymmetric Gauge Theory and the Yangian," arXiv:1303.2632 [hep-th].

- [13] E. Witten, "Integrable Lattice Models From Gauge Theory," arXiv:1611.00592 [hep-th].
- [14] M. F. Atiyah, "New Invariants Of 3- and 4-Dimensional Manifolds," in R. O. Wells, ed., *The Mathematical Heritage Of Hermann Weyl* (American Mathematical Society, 1988).
- [15] H. Au-Yang, B. M. McCoy, J. H. H. Perk, S. Tang and M. L. Yan, "Commuting Transfer Matrices in the Chiral Potts Models: Solutions of Star Triangle Equations with Genus  $> 1$ ," Phys. Lett. A 123, 219 (1987).
- [16] R. J. Baxter, J. H. H. Perk and H. Au-Yang, "New Solutions of the Star Triangle Relations for the Chiral Potts Model," Phys. Lett. A 128, 138 (1988).
- [17] V. V. Bazhanov and S. M. Sergeev, "A Master solution of the quantum Yang-Baxter equation and classical discrete integrable equations," Adv. Theor. Math. Phys. 16, no. 1, 65 (2012), arXiv:1006.0651 [math-ph].
- [18] V. P. Spiridonov, "Elliptic beta integrals and solvable models of statistical mechanics," Contemp. Math. 563, 181 (2012) [arXiv:1011.3798 [hep-th]].
- [19] M. Yamazaki, "Quivers, YBE and 3-manifolds," JHEP 1205, 147 (2012), arXiv:1203.5784 [hep-th].
- [20] M. Yamazaki, "New Integrable Models from the Gauge/YBE Correspondence," J. Statist. Phys. 154, 895 (2014), arXiv:1307.1128 [hep-th].
- [21] J.-L. Gervais and A. Neveu, "Novel Triangle Relation And Absence of Tachyons in Liouville String Field Theory," Nucl. Phys. B238 (1984) 125-41.
- [22] G. Felder, "Elliptic Quantum Groups," arXiv:hep-th/9412207.
- [23] G. Felder, "Conformal Field Theory and Integrable Systems Associated to Elliptic Curves," in *Proceedings of the International Congress of Mathematicians, Zürich 1994*, Birkhäuser, 1994; "Elliptic quantum groups", hep-th/9412207.
- [24] P. Etinghof and O. Schiffman, "Lectures on the Dynamical Yang-Baxter Equations," arXiv:math/9908064.
- [25] K. Costello, E. Witten and M. Yamazaki, "Gauge Theory and Integrability, II," *to appear*.
- [26] E. Witten, "Fivebranes And Knots," Quantum Topol. 3 (2012) 1-137, arXiv:1101.3216 [hep-th].
- [27] N. A. Nekrasov and S. L. Shatashvili, "Supersymmetric Vacua and Bethe Ansatz," Nucl. Phys. Proc. Suppl. 192-3 (2009) 91-112, arXiv:0901.4744.
- [28] N. A. Nekrasov and S. L. Shatashvili, "Quantization Of Integrable Systems And Four Dimensional Gauge Theories," in P. Exner, ed. *Proceedings, 16th International Congress on Mathematical Physics*, arXiv:0901.4748.
- [29] A. A. Belavin and V. G. Drinfeld, "Solutions of the Classical Yang-Baxter equation for Simple Lie Algebras," Funktsional. Anal. i Prilozhen. 16, 1 (1982).
- [30] E. Witten, "Chern-Simons Gauge Theory as a String Theory," Prog. Math. 133, 637 (1995) [hep-th/9207094].
- [31] S. Fishel, I. Grojnowski and C. Teleman, "The strong Macdonald conjecture and Hodge theory on the loop Grassmannian", Ann. of Math. 168, 175 (2008).
- [32] E. Witten, "Gauge Theories, Vertex Models, and Quantum Groups," Nucl. Phys. B 330 (1990) 285-346.
- [33] N. Beisert, "The SU(2|2) Dynamic R-Matrix," Adv. Theor. Math. Phys. 12 (2008) 945-79.
- [34] P. Ramond, "Group Theory: A Physicist's Survey," Cambridge, UK: Univ. Pr. (2010) 310 p.
- [35] R. Slansky, "Group Theory for Unified Model Building," Phys. Rept. 79, 1 (1981).
- [36] W. G. McKay, J. Patera, and D. W. Rand, *Tables Of Representations Of Simple Lie Algebras: Exceptional Simple Lie Algebras* (Centre de Recherches Mathematiques, Univ. of Montréal, 1990).
- [37] L. D. Faddeev and L. A. Takhtajan, "Hamiltonian Methods In The Theory Of Solitons," Springer, 1987.
- [38] E. Witten, "On Quantum Gauge Theories In Two Dimensions," Commun. Math. Phys. 141 (1991) 153-209.
- [39] M. Blau and G. Thompson, "Lectures on  $2-d$  Gauge Theories: Topological Aspects and Path Integral Techniques," in *1993 Summer School in High-Energy Physics and Cosmology*, ed. E. Gava et al. (World Scientific, 1994), arXiv:hep-th/9402097.
- [40] M. Blau and G. Thompson, "Chern-Simons Theory on  $S^1$ -Bundles: Abelianization and  $q$ -Deformed Yang-Mills Theory," arXiv:hep-th/0601068.