Proceedings of the Royal Society of Edinburgh, **140A**, 857–895, *2010*

Geometric Langlands duality and the equations of Nahm and Bogomolny

Edward Witten*
Theory Group, CERN CH1211, Geneva 23, Switzerland

(MS received 2 June 2009; accepted 6 November 2009)

Geometric Langlands duality relates a representation of a simple Lie group $G^\vee$ to the cohomology of a certain moduli space associated with the dual group G. In this correspondence, a principal SL_2 subgroup of $G^\vee$ makes an unexpected appearance. This can be explained using gauge theory, as this paper will show, with the help of the equations of Nahm and Bogomolny.

1. Introduction

This paper is intended as an introduction to the gauge theory approach [15] to the geometric Langlands correspondence. But, rather than a conventional overview, which I have attempted elsewhere [25, 26], the focus here is on understanding a very particular result, which I learned of from [13]. (Another standard reference on closely related matters is [17].)

This introduction is devoted to describing the facts that we wish to explain. In § 2, gauge theory, in the form of new results about how duality acts on boundary conditions [11,12], will be brought to bear to explain them. Finally, some technical details are reserved for § 3. In § 3.3, we also briefly discuss the compactification of the relevant gauge theory to three dimensions, showing some novel features that appear to be relevant to recent work [6]. In § 3.5 we discuss the universal kernel of geometric Langlands from a gauge theory point of view.

1.1. The dual group

Let us start with a compact, simple Lie group G and its Langlands or Goddard–Nuyts–Olive dual group $G^\vee$. (In gauge theory, we start with a compact gauge group G, but by the time we make contact with the usual statements of geometric Langlands, G is replaced by its complexification $G_{\mathbb{C}}$.) If we write T and $T^\vee$ for the respective maximal tori, then the basic relation between them is

$$\mathrm{Hom}(T^\vee, U(1)) = \mathrm{Hom}(U(1), T), \tag{1.1}$$

and vice versa. Modulo some standard facts about simple Lie groups, this relation defines the correspondence between G and $G^\vee$.

*Present address: School of Natural Sciences, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA (`witten@ias.edu`).

Now let $R^\vee$ be an irreducible representation of $G^\vee$. Its highest weight is a homomorphism $\rho^\vee : T^\vee \to U(1)$. Using (1.1), this corresponds to a homomorphism in the opposite direction $\rho : U(1) \to T$.

We can think of $\mathrm{U}(1) \cong S^1$ as the equator in $S^2 \cong \mathbb{CP}^1$. With this understood, we can view $\rho : S^1 \to G$ as a 'clutching function' that defines a holomorphic $G_\mathbb{C}$ bundle $E_\rho \to \mathbb{CP}^1$. Every holomorphic $G_\mathbb{C}$ bundle over $\mathbb{CP}^1$ arises this way, up to isomorphism, for a unique choice of $R^\vee$. Thus, isomorphism classes of such bundles correspond to isomorphism classes of irreducible representations of $G^\vee$. In the language of Goddard, Nuyts and Olive this is the correspondence between the electric charge of $G^\vee$ and the magnetic charge of G.

Actually, the homomorphism $\rho : U(1) \to G$ can be complexified to a homomorphism $\rho : \mathbb{C}^* \to G_\mathbb{C}$. Here we can view $\mathbb{C}^*$ as the complement in $\mathbb{CP}^1$ of two points p and q (the north and south poles). So the bundle E_ρ is naturally made by gluing a trivial bundle over $\mathbb{CP}^1 \setminus p$ to a trivial bundle over $\mathbb{CP}^1 \setminus q$. In particular, E_ρ is naturally trivial over the complement of the point $p \in \mathbb{CP}^1$. So E_ρ is 'a Hecke modification at p of the trivial $G_\mathbb{C}$ bundle over $\mathbb{CP}^1$'. By definition, such a Hecke modification is simply a holomorphic $G_\mathbb{C}$ bundle $E \to \mathbb{CP}^1$ with a trivialization over the complement of p. E is said to be of type ρ if, forgetting the trivialization, it is equivalent holomorphically to E_ρ.

More generally, for any Riemann surface C, point $p \in C$ and holomorphic $G_\mathbb{C}$ bundle $E_0 \to C$, a Hecke modification of E_0 at p is a holomorphic $G_\mathbb{C}$ bundle $E \to C$ with an isomorphism $\varphi : E \cong E_0$ away from p. As in [13, 17], loop groups and affine Grassmannians give a natural language for describing these notions and explaining in general what it means to say that a Hecke modification is of type ρ. We will not need this language here.

1.2. An example

Let us consider an example. Suppose that $G^\vee = \mathrm{SU}(N)$ for some N, and accordingly its complexification is $G^\vee_\mathbb{C} = \mathrm{SL}(N, \mathbb{C})$. Then $G = \mathrm{PSU}(N)$ and $G_\mathbb{C} = \mathrm{PSL}(N, \mathbb{C})$. We can think of a holomorphic $G_\mathbb{C}$ bundle as a rank N holomorphic vector bundle V, with an equivalence relation $V \cong V \otimes \mathcal{L}$, for any holomorphic line bundle $\mathcal{L}$. (The equivalence relation will not play an important role in what we are about to say.) Let us take the representation $R^\vee$ to be the obvious N-dimensional representation of $G^\vee_\mathbb{C} = \mathrm{SL}(N, \mathbb{C})$. With this data we should associate a rank-N holomorphic bundle $V \to \mathbb{CP}^1$ that is obtained by modifying the trivial bundle $U = \mathbb{C}^N \times \mathbb{CP}^1 \to \mathbb{CP}^1$ at a single point $p \in \mathbb{CP}^1$. More precisely, we will obtain a family of possible choices of V: the possible Hecke modifications of U of the appropriate type. To describe such a V, pick a one-dimensional complex subspace $S \subset \mathbb{C}^N$ and let z be a local coordinate near p. Declare that a holomorphic section v of V over an open set $\mathcal{U} \subset \mathbb{CP}^1$ is a holomorphic section of U over $\mathcal{U} \setminus p$, which, near p, looks like

$$v = a + \frac{s}{z}, \tag{1.2}$$

where a and s are holomorphic at $z = 0$ and $s(0) \in S$.

This gives a Hecke modification of U, since V is naturally equivalent to U away from $z = 0$. Clearly, the definition of V depends on S, so we have really constructed a family of possible choices of V, parametrized by $\mathbb{CP}^{N-1}$. This is the family of all possible Hecke modifications of the appropriate type.

There is an analogue of this for any choice of representation $R^\vee$ of the dual group. To such a representation, we associate as before the clutching function $\rho : U(1) \to T \subset G$ with this representation, leading to a holomorphic $G_\mathbb{C}$ bundle $E_\rho \to \mathbb{CP}^1$. Then we define $\mathcal{N}(\rho)$ as the space of all possible[1] Hecke modifications at p of the trivial bundle over $\mathbb{CP}^1$ that are of type ρ.

The moduli space $\mathcal{N}(\rho)$ of possible Hecke modifications has a natural compactification $\bar{\mathcal{N}}(\rho)$. In the description of $\bar{\mathcal{N}}(\rho)$ via the three-dimensional Bogomolny equations, which we come to in § 2, the compactification involves monopole bubbling,[2] which is analogous to instanton bubbling in four dimensions. $\mathcal{N}(\rho)$ is known as a Schubert cell in the affine Grassmannian, and $\bar{\mathcal{N}}(\rho)$ is known as a Schubert cycle in that Grassmannian. $\bar{\mathcal{N}}(\rho)$ parametrizes a family of Hecke modifications of the trivial bundle, but they are not all of type ρ; the compactification is achieved by allowing Hecke modifications dual to a representation of $G^\vee$ whose highest weight is 'smaller' than that of $R^\vee$.

1.3. The principal SL_2

Geometric Langlands duality associates the representation $R^\vee$ of the dual group to the cohomology of $\bar{\mathcal{N}}(\rho)$. Let us see how this works in our example.

In the example, $R^\vee$ is the natural N-dimensional representation of $\mathrm{SL}(N, \mathbb{C})$, and $\mathcal{N}(\rho)$ (which needs no compactification, as $R^\vee$ is minuscule) is $\mathbb{CP}^{N-1}$. Not coincidentally, the cohomology of $\mathbb{CP}^{N-1}$ is of rank N, the dimension of $R^\vee$.

The generators of the cohomology of $\mathbb{CP}^{N-1}$ are in degrees $0, 2, 4, \ldots, 2N-2$. Let us shift the degrees by $-(N-1)$ so that they are symmetrically spaced around zero. Then we have the following diagonal matrix whose eigenvalues are the appropriate degrees:

$$h = \begin{pmatrix} N-1 & & \cdots & \\ & N-3 & \cdots & \\ & & \ddots & \\ & & \cdots & -(N-1) \end{pmatrix}. \tag{1.3}$$

One may recognize this matrix; it is an element of the Lie algebra of $G^\vee_\mathbb{C} = \mathrm{SL}(N, \mathbb{C})$ that, in the language of Kostant, generates the maximal torus of a 'principal SL_2 subgroup' of $G^\vee_\mathbb{C}$.

This is the general state of affairs. In the correspondence between a representation $R^\vee$ and the cohomology of the corresponding moduli space $\bar{\mathcal{N}}(\rho)$, the grading of the cohomology by degree corresponds to the action on $R^\vee$ of a generator of the maximal torus of a principal SL_2.

[1] In one important respect, our example is misleadingly simple. In our example, every possible Hecke modification can be made using a clutching function associated with a homomorphism $\tilde{\rho} : \mathbb{C}^* \to G_\mathbb{C}$ (which is conjugate to the original homomorphism $\rho : \mathbb{C}^* \to T_\mathbb{C} \subset G_\mathbb{C}$). Accordingly, in our example, $\mathcal{N}(\rho)$ is a homogeneous space for an obvious action of $G_\mathbb{C}$. In general, this is only so if the representation $R^\vee$ is 'minuscule', as it is in our example.

[2] This phenonomenon was investigated in the 1980s in unpublished work by P. Kronheimer, and more recently in [8, 15].

1.4. Characteristic classes in gauge theory

The nilpotent 'raising operator' of the principal SL_2 also plays a role. To understand this, first recall that Atiyah and Bott [2] used gauge theory to define certain universal cohomology classes over any family of $G_{\mathbb{C}}$-bundles over a Riemann surface C. The definition applies immediately to $\bar{\mathcal{N}}(\rho)$, which parametrizes a family of holomorphic $G_{\mathbb{C}}$-bundles over $\mathbb{CP}^1$ (Hecke modifications of a trivial bundle).

If G is of rank r, then the ring of invariant polynomials on the Lie algebra $\mathfrak{g}$ of G is itself a polynomial ring with r generators, say $P_1, \ldots, P_r$, which we can take to be homogeneous of degrees $d_1, \ldots, d_r$. The relation of the d_i to a principal SL_2 subgroup of $G_{\mathbb{C}}$ is as follows: the Lie algebra $\mathfrak{g}$ decomposes under the principal SL_2 as a direct sum

$$\mathfrak{g} = \bigoplus_{i=1}^{r} \mathcal{J}_i \tag{1.4}$$

of irreducible modules $\mathcal{J}_i$ of dimensions $2d_i - 1$; in particular, therefore,

$$\sum_{i=1}^{r}(2d_i - 1) = \dim G.$$

For example, if $G = \mathrm{SU}(N)$, then $r = N-1$, letting Tr denote an invariant quadratic form on $\mathfrak{g}$, we can take the P_i to be the polynomials $P_i(\sigma) = (i+1)^{-1}\operatorname{Tr}\sigma^{i+1}$ for $\sigma \in \mathfrak{g}$ and $i = 1, 2, 3, \ldots, N-1$. Thus, P_i is homogeneous of degree $i+1$. As in this example, if G is simple, the smallest value of the degrees d_i is always 2 and this value occurs precisely once. The corresponding polynomial P is simply an invariant quadratic form on the Lie algebra $\mathfrak{g}$.

If F is the curvature of a G-bundle over any space $\mathcal{M}$, then $P_i(F)$ is a $2d_i$-dimensional characteristic class, taking values in $H^{2i}(\mathcal{M})$. (For topological purposes, it does not matter if we consider G-bundles or $G_{\mathbb{C}}$-bundles.) Atiyah and Bott consider the case that $\mathcal{M}$ parametrizes a family of G-bundles over a Riemann surface C. We let $\mathcal{E} \to \mathcal{M} \times C$ be the corresponding universal G bundle. (If necessary, we consider the associated G_{ad} bundle and define the $P_i(F)$ as rational characteristic classes.) From the class $P_i(F) \in H^{2i}(\mathcal{M} \times C)$ we can construct two families of cohomology classes over $\mathcal{M}$. Fixing a point $c \in C$, and writing π for the projection $\mathcal{M} \times C \to \mathcal{M}$, we set v_i to be the restriction of $P_i(F)$ to $\mathcal{M} \times c$. We also set $x_i = \pi_*(P_i(F))$. Thus, $v_i \in H^{2d_i}(\mathcal{M})$, and $x_i \in H^{2d_i-2}(\mathcal{M})$. To summarize,

$$\left.\begin{aligned} v_i &= P_i(F)|_{\mathcal{M}\times c}, \\ x_i &= \pi_*(P_i(F)). \end{aligned}\right\} \tag{1.5}$$

For our present purposes, we want $\mathcal{M}$ to be one of the families $\bar{\mathcal{N}}(\rho)$ of Hecke modifications of the trivial bundle $U \to \mathbb{CP}^1$ at a specified point $p \in \mathbb{CP}^1$. Taking c to be disjoint from p, it is clear that the classes v_i vanish for $\mathcal{M} = \bar{\mathcal{N}}(\rho)$ (since a Hecke modification at p has no effect at c). However, the classes x_i are non-zero and interesting.

Multiplication by x_i gives an endomorphism of $H^*(\bar{\mathcal{N}}(\rho))$ that increases the degree by $2d_i - 2$. It must map under duality to an endomorphism f_i of $R^\vee$ that increases the eigenvalue of h (the generator of a Cartan subalgebra of a principal

SL_2) by $2d_i - 2$. Thus, we expect $[h, f_i] = (2d_i - 2)f_i$. Moreover, the f_i must commute, since the x_i do.

As noted above, the smallest value of the degrees d_i is 2, which occurs precisely once. So this construction gives an essentially unique class[3] x of degree 2. It turns out that duality maps x to the nilpotent raising operator of the principal SL_2 subgroup of G that we have already encountered (the action of whose maximal torus is dual to the grading of $H^*(\bar{\mathcal{N}}(\rho))$ by degree). This being so, since the x_i all commute with x, duality must map them to elements of $\mathfrak{g}$ that commute with the raising operator of the principal SL_2. These are precisely the highest weight vectors in the SL_2 modules $\mathcal{J}_i$ of (1.4).

For example, for $\mathrm{SL}(N, \mathbb{C})$, the raising operator of the principal SL_2 is the matrix

$$f = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{1.6}$$

with 1s just above the main diagonal. The image of the two-dimensional class x under duality is precisely f. Indeed, in this example, the cohomology of $\bar{\mathcal{N}}(\rho) = \mathbb{CP}^{N-1}$ is spanned by the classes $1, x, x^2, \ldots, x^{N-1}$, and in this basis (which we have used in writing the degree operator as in (1.3)), x coincides with the matrix f. The traceless matrices that commute with f (in other words, the highest weight vectors of the $\mathcal{J}_i$) are the matrices f^k, $k = 1, \ldots, N-1$. The invariant polynomial $P_i = (1/(i+1)) \operatorname{Tr} \sigma^{i+1}$, with $d_i = i + 1$, is associated with a class x_i of dimension $2(d_i - 1) = 2i$. This class maps under duality to f^i.

Let $\mathcal{T}$ be the subgroup of $G^\vee_{\mathbb{C}}$ generated by a maximal torus in a principal SL_2 subgroup together with the highest weight vectors in the decomposition (1.4). A summary of part of what we have said is that the action of $\mathcal{T}$ on the representation $R^\vee$ corresponds to a natural $\mathcal{T}$ action on the cohomology of $\bar{\mathcal{N}}(\rho)$. The rest of the $G^\vee$ action on $R^\vee$ does not have any equally direct meaning in terms of the cohomology of $\bar{\mathcal{N}}(\rho)$.

There are additional facts of a similar nature. Our goal here is not to describe all such facts but to explain how such facts can emerge from gauge theory.

1.5. Convolution and the operator product expansion

One obvious gap in what we have said so far is that we have treated independently each representation $R^\vee$ of the dual group $G^\vee$. For example, in § 1.4, the characteristic class x was defined uniformly for all $\bar{\mathcal{N}}(\rho)$ by a universal gauge theory construction. But it might appear from the analysis that x could map under duality to a multiple of the Lie algebra element f of equation (1.6) (or, more exactly, a multiple of the linear transformation by which f acts in the representation $R^\vee$), with a different multiple for every representation.

Actually, the different irreducible representations of $G^\vee$ are linked by the classical operation of taking a tensor product of two representations and decomposing it in a

[3]For $G = \mathrm{SU}(N)$, x can also be constructed as the first Chern class of the 'determinant line bundle' associated to the family $\mathcal{M}$ of vector bundles over C. For $G = \mathrm{SO}(N)$ or $\mathrm{Sp}(2N)$, x can similarly be constructed as the first Chern class of a Pfaffian line bundle.

direct sum of irreducibles. This operation is dual to a certain natural 'convolution' operation [13, 17] on the cohomology of the moduli spaces $\bar{\mathcal{N}}(\rho)$. This operation also has a gauge theory interpretation, as we recall shortly.

Suppose that $R^\vee_\alpha$ and $R^\vee_\beta$ are two irreducible representations of $G^\vee$, and that the decomposition of their tensor product is

$$R^\vee_\alpha \otimes R^\vee_\beta = \bigoplus_\gamma N^\gamma_{\alpha\beta} \otimes R^\vee_\gamma, \tag{1.7}$$

where $R^\vee_\gamma$ are inequivalent irreducible representations of $G^\vee$, and $N^\gamma_{\alpha\beta}$ are vector spaces (with trivial action of $G^\vee$). For each α, let $\rho_\alpha : U(1) \to T \subset G$ be the homomorphism corresponding to $R^\vee_\alpha$. Under the duality maps $R^\vee_\alpha \leftrightarrow H^*(\bar{\mathcal{N}}(\rho_\alpha))$, equation (1.7) must correspond to a decomposition

$$H^*(\bar{\mathcal{N}}(\rho_\alpha)) \otimes H^*(\bar{\mathcal{N}}(\rho_\beta)) = \bigoplus_\gamma N^\gamma_{\alpha\beta} H^*(\bar{\mathcal{N}}(\rho_\gamma)), \tag{1.8}$$

with the same vector spaces $N^\gamma_{\alpha\beta}$ as before. Indeed [13,17], the appropriate decomposition can be described directly in terms of the affine Grassmannian of G without reference to duality. This decomposition is compatible with the action of the group $\mathcal{T}$ described in § 1.4. In other words, $\mathcal{T}$ acts on each factor on the left of (1.8) and hence on the tensor product; it likewise acts on each summand on the right of (1.8) and hence on the direct sum; and these actions agree.

In gauge theory terms, the classical tensor product of representations (1.7) corresponds to the operator product expansion for Wilson operators, and the corresponding decomposition (1.8) corresponds to the operator product expansion for 't Hooft operators. This has been explained in [15, § 10.4], and that story will not be repeated here. However, in § 2.12, we will explain why the operator product expansions (or, in other words, the above decompositions) are compatible with the action of $\mathcal{T}$.

2. Gauge theory

2.1. The $\hat{A}$ and $\hat{B}$ models

Let M be a 4-manifold. We will be studying gauge theory on M: more specifically, the twisted version of $\mathfrak{N} = 4$ super Yang–Mills theory that is related to geometric Langlands duality. We write A for the gauge field, which is a connection on a bundle $E \to M$, and F for its curvature. Another important ingredient in the theory is a 1-form ϕ that is valued in $\mathrm{ad}(E)$.

As explained in [15], the twisting introduces an asymmetry between $G^\vee$ and G. A and ϕ combine together in quite different ways in the two cases.

In the $G^\vee$ theory, which we will loosely call the $\hat{B}$ model (because, on compactification to two dimensions, it reduces to an ordinary B model), A and ϕ combine to form a complexified connection $\mathcal{A} = A + \mathrm{i}\phi$. As explained in [15], supersymmetry in the $\hat{B}$ model requires the connection $\mathcal{A}$ to be flat. So the $\hat{B}$ model involves the study of representations of the fundamental group of M in $G^\vee_{\mathbb{C}}$. As long as the flat connection $\mathcal{A}$ is irreducible, it is the only important variable in the $\hat{B}$ model. However, we will later analyse a situation in which the condition of irreducibility is not satisfied, so we will encounter other variables (also described in [15]).

In the G theory, the pair (A, ϕ) instead obey a nonlinear elliptic equation

$$F - \phi \wedge \phi = \star \mathrm{d}_A \phi, \tag{2.1}$$

where $\star$ is the Hodge star operator and d_A is the gauge-covariant extension of the exterior derivative. This equation is analogous to the instanton equations of two-dimensional A models (as well as to other familiar equations such as Hitchin's equations in two dimensions), so we will call this the $\hat{A}$ model. Equation (2.1) may be unfamiliar, but it has various specializations that are more familiar. For example, suppose that $M = W \times \mathbb{R}$, and that the solution is invariant under rigid motions of $\mathbb{R}$, including those that reverse orientation. (To get a symmetry of (2.1), orientation reversal must be accompanied by a sign change of ϕ.) Parametrizing $\mathbb{R}$ by a real coordinate t, the conditions imply that A is pulled back from W and that $\phi = \phi_0 \, \mathrm{d}t$, where the section ϕ_0 of $\mathrm{ad}(E)$ is also pulled back from W. Then (2.1) specializes to the three-dimensional Bogomolny equations

$$F = \star \mathrm{d}_A \phi_0. \tag{2.2}$$

Here $\star$ is now the Hodge star operator in three dimensions. Similarly, (2.1) can be reduced to Hitchin's equations in two dimensions. (For this, we take $M = \Sigma \times C$, where Σ and C are two Riemann surfaces, and assume that A and ϕ are pulled back from C.) The Bogomolny equations have been extensively studied (see, for example, [4]).

2.2. Wilson and 't Hooft operators

Let $L \subset M$ be an embedded oriented 1-manifold. We want to make some modification along L of gauge theory on M.

Starting with the $\hat{B}$ model, one 'classical' modification is to suppose that L is the trajectory of a 'charged particle' in the representation $R^\vee$ of the gauge group, which we take to be $G^\vee$. Mathematically, we achieve this by including in the 'path integral' of the theory a factor consisting of the trace, in the $R^\vee$ representation, of the holonomy around L of the complexified connection $\mathcal{A}$. This trace might be denoted as $\mathrm{Tr}_{R^\vee} \mathrm{Hol}(\mathcal{A}, L)$; physicists usually write it as $\mathrm{Tr}_{R^\vee} P \exp\{-\oint_L \mathcal{A}\}$. Since it is just a function of $\mathcal{A}$, this operator preserves the topological invariance of the $\hat{B}$ model. (The $\hat{B}$ model condition that $\mathcal{A}$ is flat means that the holonomy only depends on the homotopy class of L.) When included as a factor in a quantum path integral, the holonomy is known as a Wilson operator.

In taking the trace of the holonomy, we have assumed that L is a closed 1-manifold, that is, a circle. If M is compact, this is the only relevant case. More generally, if M has boundaries or ends, one also considers the case that L is an open 1-manifold that connects boundaries or ends of M. Then instead of a trace, one considers the matrix elements of the holonomy between prescribed initial and final states: that is, prescribed initial and final vectors in $R^\vee$. This is actually the situation that we will consider momentarily.

What is the dual in G gauge theory of including the holonomy factor in $G^\vee$ gauge theory? The dual is the 't Hooft operator. It was essentially shown by 't Hooft nearly 30 years ago that the dual operation to including a holonomy factor or Wilson operator is to modify the theory by requiring the fields to have a certain type of singularity along L. This singularity gives a way to study, using gauge theory, the Hecke modifications of a G bundle on a Riemann surface. The required singularity and its interpretation in terms of Hecke modifications have been described in [15, §§ 9,10]. A few relevant points are summarized in § 2.4.

2.3. Choice of M

In this paper, our interest is in the representation $R^\vee$, not the 4-manifold M. So we simply want to choose M and the embedded 1-manifold L to be as simple as possible. It is convenient to take $M = W \times \mathbb{R}$, where W is a 3-manifold and $\mathbb{R}$ parametrizes the 'time'. We similarly take $L = w \times \mathbb{R}$, where w is a point in W.

Henceforth, we adopt a 'Hamiltonian' point of view in which, in effect, we work at time zero and only talk about W. So instead of a 4-manifold M with an embedded 1-manifold L labelled by a representation $R^\vee$, we consider a 3-manifold W with an embedded point w labelled by that representation. The presence of this special point means that, in the quantization, we must include an 'external charge' in the representation $R^\vee$,

Moreover, we want to make a simple choice of W so as to study the representation $R^\vee$ and its dual, keeping away from the wonders of 3-manifolds.

What is the simplest 3-manifold? S^3 comes to mind right away, but there is a snag. Suppose that we study the $G^\vee$ gauge theory on $W = S^3$, with a marked point w that is labelled by the representation $R^\vee$. What will the quantum Hilbert space turn out to be? A flat connection on S^3 is necessarily trivial, so there is no moduli space of flat connections to quantize. If the trivial flat connection on S^3 had no automorphisms, the quantum Hilbert space of the $\hat{B}$ model would be simply $R^\vee$, as there is nothing else to quantize. However, the trivial flat connection on S^3 actually has a group $G_{\mathbb{C}}$ of automorphisms and, in quantization, one is supposed to impose invariance under the group of gauge transformations. Because of this, the quantum Hilbert space is not $R^\vee$ but the $G^\vee_{\mathbb{C}}$ invariant subspace of $R^\vee$: namely, 0. Thus, simply taking $W = S^3$ with a marked point labelled by the representation $R^\vee$ will not give us a way to use the $\hat{B}$ model to study the representation $R^\vee$.

Our requirement, of a 3-manifold on which the trivial flat connection is unique and irreducible, does not exist. However, we can pick W to be a 3-manifold with boundary, provided that we endow the boundary with a supersymmetric boundary condition. For example, suppose that W is a three-dimensional ball B^3. We may pick Dirichlet boundary conditions on the boundary of B^3. In the $\hat{B}$ model, Dirichlet boundary conditions mean that $\mathcal{A}$ is trivial on the boundary ∂B^3, and that only gauge transformations that are trivial on the boundary are allowed.

If we formulate the $\hat{B}$ model on B^3 with Dirichlet boundary conditions and a marked point labelled by $R^\vee$, then as there are no non-trivial flat connections and the trivial one has no gauge symmetries, the physical Hilbert space is a copy of $R^\vee$. So this does give a way to study the representation $R^\vee$ as a space of physical states in the $\hat{B}$ model. The only trouble is that the dual of Dirichlet boundary conditions is rather complicated [11, 12], and the resulting $\hat{A}$-model picture is not very transparent.

There is another choice that turns out to be more useful because it gives something that is tractable in both the $\hat{A}$ model and the $\hat{B}$ model. This is to take $W = S^2 \times I$, where $I \subset \mathbb{R}$ is a closed interval. Of course, W has two ends, since I

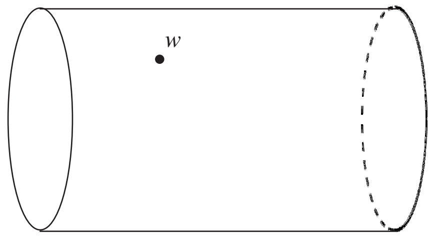

Figure 1. $W = S^2 \times I$ in G gauge theory with a marked point w at which an 't Hooft operator is inserted. Dirichlet boundary conditions are imposed at the right boundary and Neumann boundary conditions at the left boundary.

has two boundary points. Suppose that we pick Dirichlet boundary conditions at one end of $S^2 \times I$ and Neumann boundary conditions at the other. (Neumann boundary conditions in gauge theory mean that the gauge field and the gauge transformations are arbitrary on the boundary; instead, there is a condition on the normal derivative of the gauge field, though we will not have to consider it explicitly because it is a consequence of the equations that we will be solving anyway.) With these boundary conditions, the trivial flat connection on W is unique and irreducible.

By contrast, if we were to place Dirichlet boundary conditions at both ends of $S^2 \times I$, there would be non-trivial flat connections classified by the holonomy along a path from one end to the other. With Dirichlet boundary conditions at both ends, this holonomy is gauge-invariant. If we were to place Neumann boundary conditions at both ends, every flat connection would be gauge-equivalent to the trivial one, but (since there would be no restriction on the boundary values of a gauge transformation) the trivial flat connection would have a group $G_{\mathbb{C}}$ of automorphisms, coming from constant gauge transformations.

The case that works well is therefore the case of mixed boundary conditions: Dirichlet at one end and Neumann at the other. So we could study the representation $R^\vee$ in the $\hat{B}$ model by working on $W = S^2 \times I$ with mixed boundary conditions. This may even be an interesting thing to do.

Instead, here, we will do something that turns out to be simpler. We will study the $\hat{A}$ model, not the $\hat{B}$ model, on $W = S^2 \times I$, with mixed Dirichlet and Neumann boundary conditions, and one marked point labelled by an 't Hooft singularity (see figure 1). Since they make the trivial solution of the Bogomolny equations isolated and irreducible, mixed boundary conditions simplify the $\hat{A}$ model just as they simplify the $\hat{B}$ model.

In fact, the $\hat{A}$ model on $S^2 \times I$ with mixed boundary conditions was studied in [15, § 10.4] in order to investigate the operator product expansion for 't Hooft operators. At the time, it was not possible to compare with a $\hat{B}$-model description, since the duals of Dirichlet and Neumann boundary conditions in supersymmetric non-abelian gauge theory were not sufficiently clear. Here, we will complete the analysis using more recent results [11,12] on duality of boundary conditions. This will enable us to understand, using gauge theory, the results that were surveyed in § 1.

2.4. Bogomolny equations with a singularity

In the $\hat{A}$ model on $W = S^2 \times I$, we must solve the Bogomolny equations (2.2), with singularities at the positions of 't Hooft operators. For the moment, suppose that there is a single such singularity, located at $w = c \times r$, where c and r are points in S^2 and I, respectively.

If E is any G bundle with connection over $C \times I$, where C is a Riemann surface (in our case, $C = S^2$), we can restrict E to $C \times \{y\}$, for $y \in I$, to obtain a G bundle with connection $E_y \to C$. Since any connection on a bundle over a Riemann surface defines an integrable $\bar{\partial}$ operator, the bundles E_y are holomorphic $G_{\mathbb{C}}$ bundles in a natural way.

One of the many special properties of the Bogomolny equations is that if the pair (A, ϕ_0) obeys these equations, then, as a holomorphic bundle, E_y is independent of y, up to a natural isomorphism. This is proved by a very short computation. Writing z for a local holomorphic coordinate on C, a linear combination of the Bogomolny equations gives $F_{y\bar{z}} = -\mathrm{i}D_{\bar{z}}\phi_0$, or $[\partial_y + A_y + \mathrm{i}\phi_0, \bar{\partial}_A] = 0$, where $\bar{\partial}_A$ is the $\bar{\partial}$ operator on E_y determined by the connection A. Thus, $\bar{\partial}_A$ is independent of y, up to a complex gauge transformation, and integrating the modified connection $A_y + \mathrm{i}\phi_0$ in the y direction gives a natural isomorphism between the E_y of different y.

In the presence of an 't Hooft operator at $w = c \times r$, the Bogomolny equations fail (because there is a singularity) at the point w, and as a result, the holomorphic type of E_y may jump when we cross $y = r$. However, if we delete from C the point c, then we do not see the singularity and no jumping occurs. In other words, if we write E'_y for the restriction of E_y to $C \setminus c$, then E'_y is independent of y, as a holomorphic bundle over $C \setminus c$. (Moreover, there is a natural isomorphism between the E'_y of different y, by parallel transport with the connection $A_y + \mathrm{i}\phi_0$.) Thus, the jump in E_y in crossing $y = r$ is a Hecke modification at the point $c \in C$.

Suppose that $R^\vee$ is an irreducible representation of the dual group $G^\vee$. Using ideas described in § 1.1, let $\rho : U(1) \to G$ be the homomorphism corresponding to $R^\vee$, and let $E_\rho \to \mathbb{CP}^1$ be the corresponding $G_{\mathbb{C}}$ bundle. Then the 't Hooft operator dual to $R^\vee$ in G gauge theory is defined so that the Hecke modification found in the last paragraph is of type ρ. This is accomplished by specifying a suitable singularity type in the solution of the Bogomolny equations. Roughly, one arranges that the solution (A, ϕ_0) of the Bogomolny equations has the property that, when restricted to a small 2-sphere S that encloses the point w, the connection A determines a holomorphic $G_{\mathbb{C}}$ bundle over S that is equivalent holomorphically to E_ρ [15].

2.5. The space of physical states

Now let us determine the space of physical states of the $\hat{A}$ model on $W = S^2 \times I$, with mixed boundary conditions. On general grounds, this is the cohomology of the moduli space of solutions of the Bogomolny equations, with the chosen boundary conditions.

Dirichlet boundary conditions at one end of W means that E_y is trivial at that end. Neumann boundary conditions mean that, at the other end, any E_y that is produced by solving the Bogomolny equations is allowed. In the presence of a single 't Hooft operator dual to $R^\vee$, any Hecke modification of type ρ can occur. So the moduli space of solutions of the Bogomolny equations is our friend the moduli space $\mathcal{N}(\rho)$ of Hecke modifications of type ρ. This moduli space has a natural compactification by allowing monopole bubbling [8, 15], the shrinking to a point of a 'lump' of energy in a solution of the Bogomolny equations.[4] This compactification is the compactified space $\bar{\mathcal{N}}(\rho)$ of Hecke modifications.

Therefore, the space $\mathcal{H}$ of physical states of the $\hat{A}$ model is the cohomology $H^*(\bar{\mathcal{N}}(\rho))$. Together with the fact that we will find $\mathcal{H} = R^\vee$ in the $\hat{B}$ model, this is the basic reason that electric–magnetic duality establishes a map between the cohomology $H^*(\bar{\mathcal{N}}(\rho))$ and the representation $R^\vee$ of $G^\vee$.

2.6. Nahm's equations

To learn more about $\mathcal{H}$, we need to analyse its dual description in $G^\vee$ gauge theory. Some things are simpler than what we have met so far, and some things are less simple.

First of all, there are no Bogomolny equations to worry about. The supersymmetric equations of the $\hat{B}$ model are quite different. As formulated in [15], these equations involve a connection A on a $G^\vee$ bundle $E^\vee \to M$, a 1-form ϕ valued in the adjoint bundle $\mathrm{ad}(E^\vee)$ and a 0-form $\sigma^\vee$ taking values in the complexification $\mathrm{ad}(E^\vee) \otimes \mathbb{C}$. (We write $\sigma^\vee$ for this field, a slight departure from the notation in [15], as we will later introduce an analogous field σ in the $\hat{A}$ model.) It is convenient to combine A and ϕ to a complex connection $\mathcal{A} = A + \mathrm{i}\phi$ on the $G^\vee_{\mathbb{C}}$-bundle $E^\vee_{\mathbb{C}} \to M$ obtained by complexifying $E^\vee$. Moreover, we write $\mathcal{F}$ for the curvature of $\mathcal{A}$, and $\mathrm{d}_{\mathcal{A}}$, d_A for the exterior derivatives with respect to $\mathcal{A}$ and A, respectively. The supersymmetric conditions read

$$\left.\begin{aligned} \mathcal{F}_{\mathcal{A}} &= 0, \\ \mathrm{d}_{\mathcal{A}}\sigma^\vee &= 0, \\ \mathrm{d}^*_A\phi + \mathrm{i}[\sigma^\vee, \bar{\sigma}^\vee] &= 0. \end{aligned}\right\} \tag{2.3}$$

Here $\mathrm{d}^*_A = \star\mathrm{d}_A\star$ is the adjoint of d_A. The first condition says that $E^\vee_{\mathbb{C}}$ is flat, and the second condition says that $\sigma^\vee$ generates an automorphism of this flat bundle. Therefore, if $E^\vee_{\mathbb{C}}$ is irreducible, then $\sigma^\vee$ must vanish. This is the case most often considered in the geometric Langlands correspondence, but we will be in a rather different situation because, for $W = S^2 \times I$ and with the boundary conditions we have introduced, there are no non-trivial flat connections. While the first two equations are invariant under $G^\vee_{\mathbb{C}}$-valued gauge transformations, the third one is only invariant under $G^\vee$-valued gauge transformations. For a certain natural symplectic structure on the data $(A, \phi, \sigma^\vee)$, the expression $\mathrm{d}^\star_A\phi + \mathrm{i}[\sigma^\vee, \bar{\sigma}^\vee]$ is the moment map for the action of $G^\vee$ gauge transformations on this data. As this interpretation suggests, the third equation is a stability condition; the moduli space of solutions of the three equations, modulo $G^\vee$-valued gauge transformations, is the moduli space of stable pairs $(\mathcal{A}, \sigma^\vee)$ obeying the first two equations, modulo $G^\vee_{\mathbb{C}}$-valued gauge transformations. A pair is considered strictly stable if it cannot be put in a triangular form

$$\begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix}, \tag{2.4}$$

[4] Monopole bubbling is somewhat analogous to instanton bubbling in four dimensions, which involves the shrinking of an instanton. An important difference is that instanton bubbling can occur anywhere, while monopole bubbling can only occur at the position of an 't Hooft operator. Monopole bubbling involves a reduction of the weight ρ associated to the 't Hooft singularity.

and semistable if it can be put in such a form. (There are no strictly unstable pairs.) Two semistable pairs are considered equivalent if the diagonal blocks α and γ coincide. For the case when $\sigma^\vee = 0$, this interpretation of the third equation was obtained in [9].

Rather surprisingly, the system of equations (2.3) can be truncated to give a system of equations in mathematical physics that are familiar but are not usually studied in relation to complex flat connections. These are Nahm's equations. They were originally obtained [18] as the result of applying an ADHM-like transform to the Bogomolny equations on $\mathbb{R}^3$. Subsequently, they have turned out to have a wide range of mathematical applications (see, for instance, [1,16]).

To reduce (2.3) to Nahm's equations, suppose that $A = 0$ and that $\phi = \phi_y \, \mathrm{d}y$, where y is one of the coordinates on M. In our application, we have $M = W \times \mathbb{R}$, $W = S^2 \times I$, and we take y to be a coordinate on I, so that $y = 0$ is one end of I. Furthermore, write $\sigma^\vee = (X_1 + \mathrm{i}X_2)/\sqrt{2}$, where X_1 and X_2 take values in the *real* adjoint bundle $\mathrm{ad}(E^\vee)$, and set

$$\phi_y = X_3. \tag{2.5}$$

Then the equations (2.3) reduce unexpectedly to Nahm's equations $\mathrm{d}X_1/\,\mathrm{d}y + [X_2, X_3] = 0$, and cyclic permutations of indices $1, 2, 3$. Alternatively, combining X_1, X_2 and X_3 to a section $\boldsymbol{X}$ of $\mathrm{ad}(E^\vee) \otimes \mathbb{R}^3$, the equations can be written

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}y} + \boldsymbol{X} \times \boldsymbol{X} = 0. \tag{2.6}$$

Here $(\boldsymbol{X} \times \boldsymbol{X})_1 = [X_2, X_3]$, etc.

Nahm's equations (2.6) have an obvious SO(3) symmetry acting on $\boldsymbol{X}$. In the way we have derived these equations from (2.3), this symmetry is rather mysterious. Its origin is more obvious in the underlying four-dimensional gauge theory, as we explain in § 3.

2.7. The dual boundary conditions

Nahm's equations admit certain singular solutions that are important in many of their applications [1,16,18]. Let $\vartheta : \mathfrak{su}(2) \to \mathfrak{g}^\vee$ be any homomorphism from the SU(2) Lie algebra to that of $G^\vee$. It is given by elements $\boldsymbol{t} = (t_1, t_2, t_3) \in \mathfrak{g}^\vee$ that obey the $\mathfrak{su}(2)$ commutation relations $[t_1, t_2] = t_3$, and cyclic permutations. Then Nahm's equations on the half-line $y > 0$ are obeyed by

$$\boldsymbol{X} = \frac{\boldsymbol{t}}{y}. \tag{2.7}$$

Consider $G^\vee$ gauge theory on a half-space $y \geqslant 0$. Dirichlet boundary conditions on $G^\vee$ gauge fields can be extended to the full $\mathfrak{N} = 4$ super Yang–Mills theory in a supersymmetric (half-BPS) fashion. When this is done in the most obvious way, the fields $\boldsymbol{X}$ actually obey free (or Neumann) boundary conditions and thus are unconstrained, but non-singular, at the boundary. With the aid of the singular solutions (2.7) of Nahm's equations, one can describe boundary conditions [11] in $G^\vee$ gauge theory that generalize the most obvious Dirichlet boundary conditions in that they preserve the same supersymmetry. To do this, instead of saying that $\boldsymbol{X}$ is regular at $y = 0$, we say that it should have precisely the singular behaviour

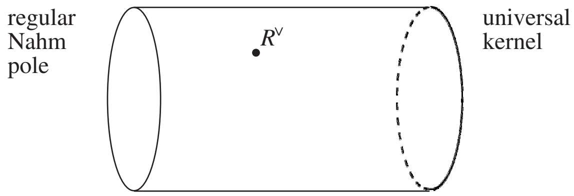

Figure 2. $W = S^2 \times I$ in $G^\vee$ gauge theory with a marked point at which an external charge in the representation $R^\vee$ is included. The boundary conditions are dual to those of figure 1. Dirichlet boundary conditions modified with a regular Nahm pole are shown on the left, while more complicated boundary conditions associated with the universal kernel of geometric Langlands are shown on the right.

of (2.7) near $y = 0$. This condition can be uniquely extended to the full $\mathfrak{N} = 4$ theory in a supersymmetric fashion. This use of a classical singularity to define a boundary condition in quantum theory is somewhat analogous to the definition of the 't Hooft operator via a classical singularity (in that case, a singularity along a codimension-3 submanifold of space-time).

The most important case for us will be what we call a regular Nahm pole. This is the case that $\vartheta : \mathfrak{su}(2) \to \mathfrak{g}^\vee$ is a principal embedding. (Usually the principal embedding is defined as a homomorphism $\mathfrak{sl}(2,\mathbb{C}) \to \mathfrak{g}^\vee_{\mathbb{C}}$. The complexification of ϑ is such a homomorphism.) For $G^\vee_{\mathbb{C}} = \mathrm{SL}(N,\mathbb{C})$, a principal $\mathfrak{sl}(2)$ embedding (or at least the images of two of the three $\mathfrak{sl}(2)$ generators) was described explicitly in (1.3) and (1.6). As in this example, a principal $\mathfrak{sl}(2)$ embedding ϑ is always irreducible in the sense that the subalgebra of $\mathfrak{g}^\vee_{\mathbb{C}}$ that commutes with the image of ϑ is zero. Conversely, an irreducible $\mathfrak{sl}(2)$ embedding is always conjugate to a principal one.

2.7.1. *The dual picture*

We finally have the tools to discuss the dual of the $\hat{A}$ model picture that was analysed in § 2.5. In our study of G gauge theory, we imposed mixed Dirichlet–Neumann boundary conditions: say, Neumann at $y = 0$ and Dirichlet at $y = L$. To compare this with a description in $G^\vee$ gauge theory we need to know what happens to Neumann and Dirichlet boundary conditions under duality.

For $G = G^\vee = \mathrm{U}(1)$, electric–magnetic duality simply exchanges Dirichlet and Neumann boundary conditions. One of the main results of [11,12] is that this is not true for non-abelian gauge groups. Rather, electric–magnetic duality maps Neumann boundary conditions to Dirichlet boundary conditions modified by a regular Nahm pole. Furthermore, it maps Dirichlet boundary conditions to something that is very interesting (and related to the 'universal kernel' of geometric Langlands, as explained in § 3.5), but more difficult to describe.

For our purposes, all we really need to know about the dual of Dirichlet boundary conditions is that in the $\hat{B}$ model on $W = S^2 \times I$ (times $\mathbb{R}$), with boundary conditions at $y = 0$ given by the regular Nahm pole, and the appropriate boundary conditions at $y = L$, the solution of Nahm's equations is unique. In fact, the relevant solution is precisely $\boldsymbol{X} = \boldsymbol{t}/y$. (The boundary conditions at $y = 0$ require that the solution should take this form, modulo regular terms; the regular terms are fixed by the boundary condition at $y = L$.) How this comes about is described in § 3.4.

The dual picture, therefore, is as described in figure 2.

2.8. The space of physical states in the $\hat{B}$ model

Now we can describe the space of physical states in the $\hat{B}$ model on $S^2 \times I$, with these boundary conditions, and with a marked point $w = c \times r$ labelled by the representation $R^\vee$.

The analysis is easy because, with the boundary conditions, Nahm's equations have a unique and irreducible solution, with no gauge automorphisms and no moduli that must be quantized. Moreover, no moduli appear when Nahm's equations are embedded in the more complete system (2.3). This follows from the irreducibility of the solution of Nahm's equations with a regular pole.

In the absence of marked points, the physical Hilbert space $\mathcal{H}$ would be a copy of $\mathbb{C}$, from quantizing a space of solutions of Nahm's equations that consists of only one point. However, we must take into account the marked point. In general, in the presence of the marked point, $\mathcal{H}$ would be computed as the $\bar{\partial}$ cohomology of a certain holomorphic vector bundle[5] with fibre $R^\vee$ over the moduli space $\mathcal{M}$ of solutions of (2.3). (If a generic point in $\mathcal{M}$ has an automorphism group H, then one takes the H-invariant part of the $\bar{\partial}$ cohomology.) In the present case, as $\mathcal{M}$ is a single point (with no automorphisms), the physical Hilbert space is simply $\mathcal{H} = R^\vee$.

Therefore, electric–magnetic duality gives a natural map from $H^*(\bar{\mathcal{N}}(\rho))$, which is the space $\mathcal{H}$ of physical states computed in the $\hat{A}$ model, to $R^\vee$. Now, in the $\hat{B}$ model, we can try to identify the grading of $\mathcal{H}$ that, in the $\hat{A}$ model, corresponds to the grading of the cohomology $H^*(\bar{\mathcal{N}}(\rho))$ by degree.

In the underlying $\mathfrak{N} = 4$ super Yang–Mills theory, there is a Spin(6) group of global symmetries (these symmetries act non-trivially on the supersymmetries and are hence usually called R-symmetries). The twisting that leads eventually to geometric Langlands breaks this Spin(6) symmetry down to Spin(2). (This remark and related remarks in the next paragraph are explained more fully in § 3.) In the context of topological field theory, this Spin(2) symmetry is usually called 'ghost number' symmetry. The action of this Spin(2) symmetry on the $\hat{A}$ model gives the grading of $H^*(\bar{\mathcal{N}}(\rho))$ by degree.

So we must consider the action of the Spin(2) or ghost number symmetry in the $\hat{B}$ model. In the $\hat{B}$ model, Spin(2) acts by rotation of $\sigma^\vee$, that is, by rotation of the $X_1 - X_2$ plane. To be more precise, $\sigma^\vee$ has ghost number 2. Equivalently, the Spin(2) generator acts on the $X_1 - X_2$ plane as $2(X_1\partial/\partial X_2 - X_2\partial/\partial X_1)$.

In quantizing the $\hat{B}$ model on $S^2 \times I$ with the boundary conditions that we have chosen, X_1 and X_2 are not zero; in fact, they appear in the solution of Nahm's equations with the regular pole, $\boldsymbol{X} = \boldsymbol{t}/y$. So this solution is not invariant under a rotation of the $X_1 - X_2$ plane, understood naively. Why, therefore, is there a Spin(2) grading of the physical Hilbert space $\mathcal{H}$ in the $\hat{B}$ model?

The answer to this question is that we must accompany a Spin(2) rotation of the $X_1 - X_2$ plane with a gauge transformation. The regular Nahm pole $\boldsymbol{X} = \boldsymbol{t}/y$ is invariant under the combination of a rotation of the $X_1 - X_2$ plane and a gauge

[5] Let $\mathcal{E} \to \mathcal{M} \times W$ be the universal bundle. We construct the desired bundle $\mathcal{E}_{R^\vee} \to \mathcal{M}$ by restricting $\mathcal{E}$ to $\mathcal{M} \times w$ and taking the associated bundle in the representation $R^\vee$. (In this construction, in general $\mathcal{E}$ must be understood as a twisted bundle, twisted by a certain gerbe.)

transformation generated by t_3. The rotation of the $X_1 - X_2$ plane does not act on the representation $R^\vee$, but the gauge transformation does. So, on the $\hat{B}$ model side, the grading of $\mathcal{H}$ comes from the action of t_3. But since the boundary condition involves a regular Nahm pole, t_3 generates the maximal torus of a principal SL_2 subgroup of $G^\vee$.

So electric–magnetic duality maps the grading of $H^*(\bar{\mathcal{N}}(\rho))$ by degree to the action on $R^\vee$ of the maximal torus of a principal SL_2 subgroup. This fact was described in § 1.3. Now we understand it via gauge theory.

2.9. Universal characteristic classes in the $\hat{A}$ model

It remains to understand, using gauge theory, an additional fact described in § 1.4: under duality, certain natural cohomology classes of $\bar{\mathcal{N}}(\rho)$ map to elements of $\mathfrak{g}^\vee$ acting on $R^\vee$. There are three steps to understanding this:

(i) interpreting these cohomology classes as local quantum field operators in the $\hat{A}$ model;

(ii) determining their image under electric–magnetic duality;

(iii) computing the action of the dual operators in the $\hat{B}$ model.

We consider step (i) here and steps (ii) and (iii) in § 2.11.

The method for carrying out the first step is known from experience with Donaldson theory. In defining polynomial invariants of 4-manifolds [10], Donaldson adapted to four dimensions the universal gauge theory cohomology classes that were described in two dimensions in [2] (and reviewed in § 1.4). Donaldson's construction was interpreted in quantum field theory in [24]. One of the main steps in doing so was to interpret the universal characteristic classes in terms of quantum field theory operators. The resulting formulae were understood geometrically by Atiyah and Jeffrey [5]. Formally, the construction of Donaldson theory by twisting of $\mathfrak{N} = 2$ super Yang–Mills theory is just analogous to the construction of the $\hat{A}$ model relevant to geometric Langlands by twisting of $\mathfrak{N} = 4$ super Yang–Mills theory. (The instanton equation plays the same formal role in Donaldson theory that the equation $F - \phi \wedge \phi = \star \mathrm{d}_A \phi$ plays in the $\hat{A}$ model related to geometric Langlands.) As a result, we can carry out the first step by simply borrowing the construction of [24].

As in § 1.4, the starting point is an invariant polynomial P_i on the Lie algebra $\mathfrak{g}$ of G. Using this polynomial, one constructs corresponding supersymmetric operators in the $\hat{A}$ model (or in Donaldson theory). The construction uses the existence of a field σ of degree or ghost number 2, taking values in the adjoint bundle $\mathrm{ad}(E)_{\mathbb{C}}$ associated to a G bundle, E. (There is also an analogous field $\sigma^\vee$ in the $\hat{B}$ model; it has already appeared in (2.3), and will reappear in § 2.11.) σ is invariant under the topological supersymmetry of the $\hat{A}$ model, so it can be used to define operators that preserve the topological invariance of that model.

The most obvious way to do this is simply to define $\mathcal{P}_i(z) = P_i(\sigma(z))$. This commutes with the topological supersymmetry of the $\hat{A}$ model, since it is a function only of σ, which has this property. Here, z is a point in a 4-manifold M, and we have made the z-dependence explicit to emphasize that $\mathcal{P}_i$ is supposed to be a local operator in quantum field theory. We will usually not write the z-dependence explicitly.

Suppose that $P_i(\sigma)$ is homogeneous of degree d_i. Then, as σ has degree 2, $\mathcal{P}_i$ is an operator of degree $2d_i$. It corresponds to the cohomology class v_i of degree $2d_i$ that was defined from a more topological point of view in (1.5). (The link between the two points of view depends on the fact that σ can be interpreted as part of the Cartan model of the equivariant cohomology of the gauge group acting on the space of connections and other data; see [5] for related ideas.) This has an important generalization, which physicists call the descent procedure. It is possible to derive from the invariant polynomial P_i a family of r-form-valued supersymmetric operators of degree $2d_i - r$, for $r = 1, \ldots, 4$. (The definition stops at $r = 4$ since we are in four dimensions.) Let us write

$$\hat{\mathcal{P}}_i = \mathcal{P}_i^{(0)} + \mathcal{P}_i^{(1)} + \cdots + \mathcal{P}_i^{(4)},$$

where $\mathcal{P}_i^{(0)} = \mathcal{P}_i = P_i(\sigma)$, and $\mathcal{P}_i^{(r)}$ will be a local operator with values in r-forms on M. We define the $\mathcal{P}_i^{(r)}$ for $r > 0$ by requiring that

$$(\mathrm{d} + [Q, \cdot])\hat{\mathcal{P}}_i = 0, \tag{2.8}$$

where d is the ordinary exterior derivative on M, and Q is the generator of the topological supersymmetry.

$\hat{\mathcal{P}}_i$ is uniquely determined by the condition (2.8) plus the choice of $\mathcal{P}_i^{(0)}$ and the fact that $\hat{\mathcal{P}}_i$ is supposed to be a locally defined quantum field operator (in other words, a universally defined local expression in the fields of the underlying super Yang–Mills theory). For example, both in Donaldson theory and for our purposes, the most important component is the 2-form component $\mathcal{P}_i^{(2)}$. It turns out to be

$$\mathcal{P}_i^{(2)} = \left\langle \frac{\partial P_i}{\partial \sigma}, F \right\rangle + \left\langle \frac{\partial^2 P_i}{\partial \sigma^2}, \psi \wedge \psi \right\rangle. \tag{2.9}$$

The notation here means the following. As P_i is an invariant polynomial on the Lie algebra $\mathfrak{g}$, we can regard $\partial P_i / \partial \sigma$ as an element of the dual space $\mathfrak{g}^*$. Hence, it can be paired with the $\mathfrak{g}$-valued 2-form F (the curvature of the gauge connection A) to make a gauge-invariant 2-form-valued field that appears as the first term on the right-hand side of (2.9). Similarly, we can consider $\partial^2 P_i / \partial \sigma^2$ as an element of $\mathfrak{g}^* \otimes \mathfrak{g}^*$. On the other hand, ψ is a $\mathfrak{g}$-valued fermionic 1-form (of degree or ghost number 1) that is part of the twisted super Yang–Mills theory under consideration here (either twisted $\mathfrak{N} = 2$ relevant to Donaldson theory, or twisted $\mathfrak{N} = 4$ relevant to geometric Langlands). So $\psi \wedge \psi$ is a 2-form valued in $\mathfrak{g} \otimes \mathfrak{g}$; it can be paired with $\partial^2 P_i / \partial \sigma^2$ to give the second term on the right-hand side of (2.9).

From (2.8), we have $[Q, \mathcal{P}_i^{(2)}] = -\mathrm{d}\mathcal{P}_i^{(1)}$; thus $[Q, \mathcal{P}_i^{(2)}]$ is an exact form. So the integral of $\mathcal{P}_i^{(2)}$ over a 2-cycle $\mathcal{S} \subset M$, that is,

$$x_i(\mathcal{S}) = \int_{\mathcal{S}} \mathcal{P}_i^{(2)}, \tag{2.10}$$

commutes with the generator Q of the topological supersymmetry. Thus, $x_i(\mathcal{S})$ is an observable of the $\hat{A}$ model. Since $\mathrm{d}\mathcal{P}_i^{(2)} = -\{Q, \mathcal{P}_i^{(3)}\}$, this observable element

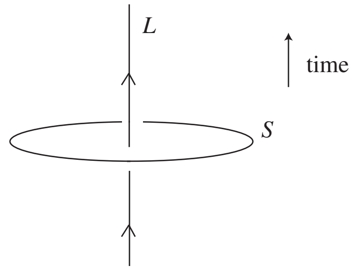

Figure 3. An 't Hooft or Wilson line operator that runs in the time direction (shown vertically) at a fixed position in W. A small two-surface S (sketched here as a circle) is supported at a fixed time and is linked with L.

only depends on the homology class of $\mathcal{S}$. Concretely, $x_i(\mathcal{S})$ will correspond to a cohomology class on the relevant moduli spaces.

In our problem with $M = W \times \mathbb{R}$, $W = S^2 \times I$ and an 't Hooft operator supported on $w \times \mathbb{R}$ with $w \in W$, what choice do we wish to make for $\mathcal{S}$? Part of the answer is that we will take $\mathcal{S}$ to be supported at a particular time. In other words, we take it to be the product of a point $t_0 \in \mathbb{R}$ and a 2-cycle in W that we will call S.

Remark 2.1. The fact that $\mathcal{S}$ is localized in time means that the corresponding quantum field theory expression $x_i(\mathcal{S})$ is an operator that acts on the quantum state at a particular time. (In topological field theory, the precise time does not matter, but, in general, as operators may not commute, their ordering does.) By contrast, the 't Hooft operator in this problem is present for all time, as its support is $w \times \mathbb{R}$. Being present for all time, it is part of the definition of the quantum state, rather than being an operator that acts on this state.

What will we choose for $S \subset W$? (In what follows, we will not distinguish in the notation between $S \subset W$ and $\mathcal{S} = S \times t_0 \in W \times \mathbb{R}$.) One obvious choice is to let S be the left or right boundary of W. If S is the right boundary, where we have imposed Dirichlet boundary conditions, so that the G-bundle is trivialized, then $x_i(S)$ vanishes. (From a quantum field theory point of view, the supersymmetric extension of Dirichlet boundary conditions actually says that A, ψ and σ all vanish, so $\hat{\mathcal{P}}_i$ certainly does.) On the other hand, if S is the left boundary, with Neumann boundary conditions, there is no reason for $x_i(S)$ to vanish. The difference between the left and right boundaries of $S^2 \times I$ is homologous to a small 2-sphere that 'links' the point $w = c \times r \in W$ at which an 't Hooft operator is present. This is the most illuminating choice of S (figure 3). At any rate, whether we make this choice of S or take S to simply be the left boundary of W, $x_i(S)$ coincides with the class $x_i \in H^{2d_i-2}(\bar{\mathcal{N}}(\rho))$ that was defined in (1.5). In view of remark 2.1, we should think of $x_i(S)$ not just as an element of $H^{2d_i-2}(\bar{\mathcal{N}}(\rho))$ but as an operator acting on this space (by cup product, as follows from general properties of the $\hat{A}$ model).

In § 1.4, we also used the invariant polynomial P_i to define gauge theory characteristic classes v_i of degree $2d_i$. As we have already mentioned, in the quantum field theory language, these classes simply correspond to the quantum field operator $\mathcal{P}_i(z)$, evaluated at an arbitrary point $z \in M$. In § 1.4, we noted that the v_i vanish as elements of $H^{2d_i}(\bar{\mathcal{N}}(\rho))$ (though, of course, they are non-zero in other gauge theory moduli spaces). We can prove this in the quantum field theory approach by taking z to approach the Dirichlet boundary of W; on this boundary, $\sigma = 0$ so $\mathcal{P}_i$ vanishes.

2.10. A group theory interlude

Before describing the dual picture, we need a small group theory interlude.

Let T and $T^\vee$ be the maximal tori of G and $G^\vee$ and let $\mathfrak{t}$ and $\mathfrak{t}^\vee$ be their Lie algebras. Because $\mathfrak{t}$ and $\mathfrak{t}^\vee$ are dual vector spaces, and G and $G^\vee$ have the same Weyl group, Weyl-invariant and non-degenerate quadratic forms on $\mathfrak{t}$ correspond in a natural way to Weyl-invariant and non-degenerate quadratic forms on $\mathfrak{t}^\vee$. Indeed, thinking of an invariant quadratic form γ on $\mathfrak{t}$ as a Weyl-invariant map from $\mathfrak{t}$ to $\mathfrak{t}^\vee$, its inverse γ^{-1} is a Weyl-invariant map in the opposite direction, or, equivalently, a quadratic form on $\mathfrak{t}^\vee$. If γ and $\gamma^\vee$ are invariant quadratic forms on the Lie algebras $\mathfrak{g}$ and $\mathfrak{g}^\vee$ whose restrictions to $\mathfrak{t}$ and $\mathfrak{t}^\vee$ are inverse matrices, then we formally write $\gamma^\vee = \gamma^{-1}$ even without restricting to $\mathfrak{t}$ and $\mathfrak{t}^\vee$. (As we state more fully later, the most natural relation between quadratic forms on the two sides that comes from duality really contains an extra factor of $n_{\mathfrak{g}}$, the ratio of length squared of long and short roots.)

G-invariant polynomials on the Lie algebra $\mathfrak{g}$ are in natural correspondence with Weyl-invariant polynomials on $\mathfrak{t}$. Similarly, $G^\vee$-invariant polynomials on $\mathfrak{g}^\vee$ correspond naturally to Weyl-invariant polynomials on $\mathfrak{t}^\vee$.

Combining the above statements, once an invariant quadratic form $\gamma^\vee$ or γ is picked on $\mathfrak{g}^\vee$ or equivalently on $\mathfrak{g}$, we get a natural map from homogeneous invariant polynomials on $\mathfrak{g}$ to homogeneous invariant polynomials on $\mathfrak{g}^\vee$ of the same degree. Given an invariant polynomial on $\mathfrak{g}$, we restrict to a Weyl-invariant polynomial on $\mathfrak{t}$, multiply by a suitable power of $\gamma^\vee$ so it can be interpreted as a Weyl-invariant polynomial on $\mathfrak{t}^\vee$ and then associate it to a $G^\vee$-invariant polynomial on $\mathfrak{g}^\vee$. To restate this, let $(\mathrm{Sym}^{d_i}(\mathfrak{g}))^G$ and $(\mathrm{Sym}^{d_i}(\mathfrak{g}^\vee))^{G^\vee}$ be the spaces of homogeneous and invariant polynomials of the indicated degrees, and let Θ and $\Theta^\vee$ be the spaces of invariant quadratic forms on $\mathfrak{g}$ and $\mathfrak{g}^\vee$. For brevity, we suppose that G and $G^\vee$ are simple. Then Θ and $\Theta^\vee$ are one dimensional and $(\mathrm{Sym}^{d_i}(\mathfrak{g}))^G = \Theta^{d_i} \otimes (\mathrm{Sym}^{d_i}(\mathfrak{g}^\vee))^{G^\vee}$. Hence, if $\gamma^\vee \in \Theta^\vee$ is picked, we get a correspondence

$$P_i \leftrightarrow (\gamma^\vee)^{d_i} P_i^\vee \tag{2.11}$$

between homogeneous polynomials $P_i \in (\mathrm{Sym}^{d_i}(\mathfrak{g}))^G$ and $P_i^\vee \in (\mathrm{Sym}^{d_i}(\mathfrak{g}^\vee))^{G^\vee}$.

The main reason that these considerations are relevant to gauge theory is that an invariant quadratic form appears in defining the Lagrangian. For example, the kinetic energy of the gauge fields is commonly written

$$-\frac{1}{2e^2}\int_M \mathrm{Tr}\, F \wedge \star F, \tag{2.12}$$

where $-\mathrm{Tr}$ is usually regarded as an invariant quadratic form on $\mathfrak{g}$ that is defined *a priori*, and $1/e^2$ is a real number. However, as the theory does not depend separately on the quadratic form $-\mathrm{Tr}$ and the real number $1/e^2$ but rather only on their product, we may as well combine them[6] to $\gamma = -(4\pi/e^2)\,\mathrm{Tr}$ and say that the theory simply depends on an arbitrary choice of a positive definite invariant quadratic form on $\mathfrak{g}$. The $G^\vee$ theory similarly depends on a quadratic form $\gamma^\vee = -(4\pi/e^{\vee 2})\mathrm{Tr}$. The relation between the two that follows from electric–magnetic duality is

$$\gamma^\vee = \frac{1}{n_{\mathfrak{g}}}\gamma^{-1}, \tag{2.13}$$

where $n_{\mathfrak{g}}$ is the ratio of length squared of long and short roots of G or $G^\vee$.

2.11. Remaining steps

The procedure that was started in § 2.9 has two remaining steps: to find the $\hat{B}$ model duals of the operators $\mathcal{P}_i^{(r)}$ of the $\hat{A}$ model, and to determine their action on the space $\mathcal{H}$ of physical states.

The $\hat{B}$ model of $G^\vee$ has a complex adjoint-valued scalar field $\sigma^\vee$ whose role is somewhat similar to that of σ in the $\hat{A}$ model. We have already encountered this field in (2.3).

For $G = G^\vee = U(1)$, the action of electric–magnetic duality on these fields is very simple: σ maps to a multiple of $\sigma^\vee$. For non-abelian G and $G^\vee$, the relation cannot be as simple as that, since σ and $\sigma^\vee$ take values in different spaces; they are valued in the complexified Lie algebras of G and $G^\vee$, respectively. However, G-invariant polynomials in σ do transform into $G^\vee$-invariant polynomials in $\sigma^\vee$ in a way that one would guess from (2.11):

$$P_i(\sigma) = (\sqrt{n_{\mathfrak{g}}}\gamma^\vee)^{d_i} P_i^\vee(\sigma^\vee). \tag{2.14}$$

So $\mathcal{P}_i = P_i(\sigma)$ maps to a multiple of $\mathcal{P}_i^\vee = P_i^\vee(\sigma^\vee)$.

We can apply this right away to our familiar example of quantization on $W = S^2 \times I$ with mixed Dirichlet and Neumann boundary conditions in the $\hat{A}$ model, and the corresponding dual boundary conditions in the $\hat{B}$ model. Picking a point $z \in W$ (or, more accurately, $z \in M = W \times \mathbb{R}$), the operator $\mathcal{P}_i(z) = P_i(\sigma(z))$ corresponds in general to a natural cohomology class of degree $2d_i$ on the $\hat{A}$-model moduli space. However, for the specific case of $W = S^2 \times I$ with our chosen boundary conditions, $\mathcal{P}_i(z)$ vanishes, as we explained at the end of § 2.9. To see the equivalent vanishing in the $\hat{B}$ model on W, we note that $\sigma^\vee$, being the raising operator of a principal SL_2 subgroup of $G^\vee$, is nilpotent. Hence, $P_i^\vee(\sigma^\vee)$ vanishes for every invariant polynomial $P_i^\vee$. This is the dual of the vanishing seen in the $\hat{A}$ model.

It is probably more interesting to understand the $\hat{B}$-model duals of those $\hat{A}$ model operators that are non-vanishing. For this, we must understand the duals of the other operators $\mathcal{P}_i^{(r)}$ introduced in § 2.9. As these operators were obtained by a descent procedure starting with $\mathcal{P}_i^{(0)} = \mathcal{P}_i$, we can find their duals by applying the descent procedure starting with $\mathcal{P}_i^{\vee(0)} = \mathcal{P}_i^\vee$. In other words, we look for a family of r-form valued operators $\mathcal{P}_i^{\vee(r)}$, $r = 0, \ldots, 4$, with $\mathcal{P}_i^{\vee(0)} = \mathcal{P}_i^\vee$ and such that

[6] The factor of 4π is convenient here. Actually, a more complete description involves the gauge theory θ angle as well. Then the theory really depends on a complex-valued invariant quadratic form $\tau = (\theta/2\pi + 4\pi\mathrm{i}/e^2)(-\mathrm{Tr})$, whose imaginary part is positive definite. For our purposes, we omit θ and set $\gamma = \mathrm{Im}\,\tau$.

$(\mathrm{d} + \{Q, \ \cdot\ \})\hat{\mathcal{P}}_i^\vee = 0$, where

$$\hat{\mathcal{P}}_i^\vee = \mathcal{P}_i^{\vee(0)} + \mathcal{P}_i^{\vee(1)} + \cdots + \mathcal{P}_i^{\vee(4)}.$$

The $\mathcal{P}_i^{\vee(r)}$ are uniquely determined by those conditions and must be the duals of the $\mathcal{P}_i^{(r)}$.

For our application, the important case is $r = 2$. The explicit formula for $\mathcal{P}_i^{\vee(2)}$ is very similar to the formula (2.9) for $\mathcal{P}_i^{(2)}$, except that the curvature F must be replaced by $\star F$, as follows:

$$\mathcal{P}_i^{(2)} = \left\langle \frac{\partial P_i}{\partial \sigma^\vee}, \star F \right\rangle + \left\langle \frac{\partial^2 P_i}{\partial \sigma^{\vee 2}}, \psi \wedge \psi \right\rangle. \tag{2.15}$$

Of course, F and ψ are now fields in $G^\vee$ rather than G gauge theory, though we do not indicate this in the notation.

We can now identify the $\hat{B}$-model dual of the classes $x_i \in H^{2d_i-2}(\bar{\mathcal{N}}(\rho))$ that were defined in (2.10). We simply replace $\mathcal{P}_i^{(2)}$ by $(\sqrt{n_\mathfrak{g}}\gamma^\vee)^{d_i}\mathcal{P}_i^{\vee(2)}$ in the definition of these classes (the power of $\sqrt{n_\mathfrak{g}}\gamma^\vee$ is from (2.14)), so the dual formula is

$$x_i(S) = (\sqrt{n_\mathfrak{g}}\gamma^\vee)^{d_i} \int_S \mathcal{P}_i^{\vee(2)}. \tag{2.16}$$

As in § 2.9, S is a small 2-sphere that links the marked point $w = c \times r \in W$. We recall that, in the $\hat{B}$ model, an external charge in the representation $R^\vee$ is present at the point w.

All we have to do, then, is to evaluate the integral on the right-hand side of (2.16). Since S is a small 2-cycle around the point $w = c \times r$, a non-zero integral can arise only if the 2-form $\mathcal{P}_i^{\vee(2)}$ has a singularity at w. The reason that there is such a singularity is that the external charge in the representation $R^\vee$ produces an electric field, or, in other words, a contribution to $\star F$. In keeping with Coulomb's law, the electric field is proportional to the inverse of the square of the distance from the location w of the external charge. As a result, $\star F$ has a non-zero integral over S. The electric field due to the external point charge is proportional to $e^{\vee 2}$, or, in other words, to $(\gamma^\vee)^{-1}$. It is also proportional to the charge generators, that is, to the matrices that represent the $G^\vee$ action on $R^\vee$. Taking this into account, we find that

$$\int_S \mathcal{P}_i^{\vee(2)} = (\gamma^\vee)^{-1}\frac{\partial P_i^\vee}{\partial \sigma^\vee}. \tag{2.17}$$

To understand this formula, observe that as P_i is an invariant polynomial on $\mathfrak{g}^\vee$ its derivative $\partial P_i/\partial\sigma^\vee$ can be understood as an element of the dual space $(\mathfrak{g}^\vee)^*$; understanding $(\gamma^\vee)^{-1}$ as a map from $(\mathfrak{g}^\vee)^*$ to $\mathfrak{g}^\vee$, the right-hand side of (2.17) is an element of $\mathfrak{g}^\vee$, or, in other words, an operator that acts on the space $\mathcal{H} = R^\vee$ of physical states.

So at last, the $\hat{A}$-model cohomology class

$$x_i = \int_S P_i^{(2)}$$

can be written in the $\hat{B}$ model as

$$x_i = n_\mathfrak{g}^{d_i/2}(\gamma^\vee)^{d_i-1}\frac{\partial P_i^\vee(\sigma^\vee)}{\partial\sigma^\vee}. \tag{2.18}$$

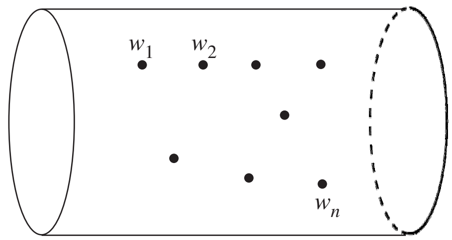

Figure 4. $S^2 \times I$ with n marked points (only a few of which have been labelled) at which 't Hooft or Wilson operators have been inserted.

An illuminating special case of this result is the case that we pick P_i to be of degree 2, corresponding to an invariant quadratic form on $\mathfrak{g}$ and to a two-dimensional class $x \in H^2(\bar{\mathcal{N}}(\rho))$. In this case, $\partial P^\vee/\partial\sigma^\vee$ is a Lie algebra element that is linear in $\sigma^\vee$, and is in fact simply a multiple of $\sigma^\vee$. In the relevant solution of Nahm's equations, $\sigma^\vee$ is the raising operator of a principal SL_2. So in other words, the class $x \in H^2(\bar{\mathcal{N}}(\rho))$ maps to the raising operator of a principal SL_2, acting on $R^\vee$. This is a typical fact described in § 1.4.

Finally, we can understand the sense in which this result is independent of the choice of $\gamma^\vee$ (which should be irrelevant in the $\hat{B}$ model). The raising operator of a principal SL_2 is well defined only up to a scalar multiple. As the right-hand side of (2.18) is homogeneous in $\sigma^\vee$ of degree $d_i - 1$, a change in $\gamma^\vee$ can be absorbed in a rescaling of $\sigma^\vee$. The same rescaling works for all i.

2.12. Compatibility with fusion

For simplicity, we have considered the case of a single marked point $w \in W = S^2 \times I$. However, there is an immediate generalization to the case of several distinct marked points $w_\alpha \in W$, labelled by representations $R_\alpha^\vee$ of $G^\vee$. At these points there is an 't Hooft singularity in the $\hat{A}$ model, or an external charge in the given representation in the $\hat{B}$ model (see figure 4).

On the $\hat{A}$-model side, the moduli space with our usual mixed boundary conditions is

$$\mathcal{M} = \prod_\alpha \bar{\mathcal{N}}(\rho_\alpha),$$

where ρ_α is related to $R_\alpha^\vee$ as described in § 1.1. This follows from the relation of the Bogomolny equations to Hecke modifications. The space of physical states is the cohomology of $\mathcal{M}$ or

$$\mathcal{H} = \bigotimes_\alpha H^*(\bar{\mathcal{N}}(\rho_\alpha)). \tag{2.19}$$

On the $\hat{B}$-model side, since the solution of Nahm's equations is unique and irreducible, with a regular pole at one end, the physical Hilbert space is simply the tensor product of the representations $R_\alpha^\vee$ associated with the marked points,

$$\mathcal{H} = \bigotimes_\alpha R_\alpha^\vee. \tag{2.20}$$

The duality map between (2.19) and (2.20) is simply induced from the individual isomorphisms $H^*(\bar{\mathcal{N}}(\rho_\alpha)) \leftrightarrow R_\alpha^\vee$.

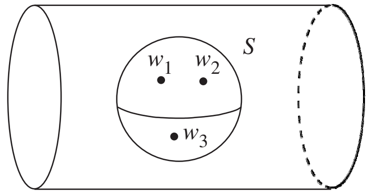

Figure 5. A 2-sphere S (at fixed time) surrounding all of the marked points $w_\alpha \in S^2 \times I$. (In this example, there are three marked points). S is homologous to a sum of 2-spheres S_α, each of them linking just one of the w_α.

As discussed in § 1.5, we can also let some of the points w_α coalesce. This leads to an operator product expansion of 't Hooft operators in the $\hat{A}$ model, or of Wilson operators in the $\hat{B}$ model. On the $\hat{B}$-model side, the operator product expansion for Wilson operators corresponds to the classical tensor product $R^\vee_\alpha \otimes R^\vee_\beta = \bigoplus_\gamma N^\gamma_{\alpha\beta} R^\vee_\gamma$. The corresponding $\hat{A}$-model picture is more complicated and is described in gauge theory terms in [15, § 10.4].

The only observation that we will add here is that the operator product expansion for Wilson or 't Hooft operators commutes with the action of the group $\mathcal{T}$ described at the end of § 1.4. We recall that $\mathcal{T}$ is generated on the $\hat{A}$-model side by the grading of the cohomology by degree and the action of the cohomology classes $x_i(S)$. For example, consider the grading of the $\hat{A}$-model cohomology by degree. With

$$\mathcal{M} = \prod_\alpha \bar{\mathcal{N}}(\rho_\alpha),$$

the operator that grades $\mathcal{H} = H^*(\mathcal{M})$ by the degree of a cohomology class is the sum of the corresponding operators on the individual factors $H^*(\bar{\mathcal{N}}(\rho_\alpha))$. The operator product expansion of 't Hooft operators commutes with the degree or ghost number symmetry, which, after all, originates as a symmetry group (a group of R-symmetries) of the full $\mathfrak{N} = 4$ super Yang–Mills theory. So after fusing some of the 't Hooft operators together, the action of the ghost number symmetry is unchanged. Similarly, the dual $\hat{B}$-model grading is by the generator t_3 of a maximal torus of a principal SL_2 subgroup of $G^\vee$. Again, the linear transformation by which t_3 acts on $\mathcal{H} = \bigotimes_\alpha R^\vee_\alpha$ is the sum of the corresponding linear transformations for the individual $R^\vee_\alpha$. This linear transformation is unchanged if some of the points are fused together, since it originates as a combination of an R-symmetry and a gauge transformation, both of which are symmetries of the full theory and therefore of the operator product expansion of Wilson operators.

A similar story holds for the linear transformations that correspond to the gauge theory cohomology classes x_i introduced in § 1.4. Let S be a 2-cycle that encloses all of the marked points w_α, as indicated in figure 5. In the $\hat{A}$ model, we have

$$x_i = \int_S \mathcal{P}^{(2)}_i,$$

while in the $\hat{B}$ model the analogue is

$$x_i = \int_S \mathcal{P}^{\vee(2)}_i.$$

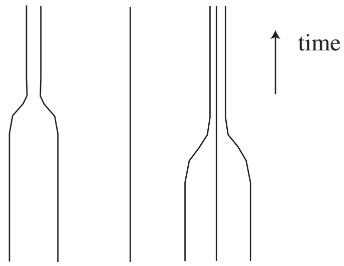

Figure 6. The operator product expansion as a time-dependent process. Time runs vertically in the figure. In the past, there are n distinct marked points with insertions of Wilson or 't Hooft operators. In the future, the points fuse together in various groups. (Complete fusion is not shown in the figure.) In this example, $n = 6$ and the groups are of sizes 2, 1 and 3. If, as in figure 5, we add a 2-sphere S that surrounds all of the points, then, by topological invariance we could move it to the past, where it acts on n distinct line operators, or to the future, where it acts on a smaller number of line operators created by fusion. Hence, the action of the operators $x_i(S)$ commutes with the operator product expansion of Wilson or 't Hooft operators.

These definitions make it clear that nothing happens to x_i if we fuse together some of the points w_α that are contained inside S.

It is illuminating here to think of the fusion as a time-dependent process. We go back to a four-dimensional picture on $M = W \times \mathbb{R}$, where $\mathbb{R}$ parametrizes the time, and instead of thinking of the marked points as having time-independent positions (as we have done so far in this paper) we take them to be separate in the past and to possibly fuse together (in arbitrary subsets) in the distant future, as in figure 6. The surface S is located at a fixed time, but topological invariance means that we can place it in the distant past, acting on the Hilbert space of a collection of isolated points w_α, or in the far future, after some fusing may have occurred. So fusion commutes[7] with the action of $x_i(S)$.

Finally, we want to see that, for each i, the linear transformation by which x_i acts on the physical Hilbert space $\mathcal{H} = \bigotimes_\alpha H^*(\bar{\mathcal{N}}(\rho_\alpha)) = \bigotimes_\alpha R_\alpha^\vee$ can be written as a sum of the linear transformations by which x_i would act on the individual factors $H^*(\bar{\mathcal{N}}(\rho_\alpha))$ or $R_\alpha^\vee$. For each α, let S_α be a 2-cycle that encloses only the single marked point w_α. Then S is homologous to the sum of the S_α. Hence,

$$\int_S \mathcal{P}_i^{(2)} = \sum_\alpha \int_{S_\alpha} \mathcal{P}_i^{(2)},$$

and similarly

$$\int_S \mathcal{P}_i^{\vee(2)} = \sum_\alpha \int_{S_\alpha} \mathcal{P}_i^{\vee(2)}.$$

So in either the $\hat{A}$ model or the $\hat{B}$ model, x_i acts on $\mathcal{H}$ by the sum of the linear transformations by which x_i would act on a single factor $H^*(\bar{\mathcal{N}}(\rho_\alpha))$ or $R_\alpha^\vee$.

[7]The ability to move the $x_i(S)$ backwards or forwards in time also means that they are central: they commute with any other operators that may act on the Wilson or the 't Hooft operators.

3. From physical Yang–Mills theory to topological field theory

In §§ 3.1 and 3.2, we describe some details of the relation between supersymmetric Yang–Mills theory and topological field theory in four dimensions that were omitted in § 2.

In § 3.3, we discuss the compactification (not reduction) of the theory to three dimensions, hopefully shedding light on some recent mathematical work [6].

In § 3.4, we explain the claim of § 2.7 that, with the boundary conditions that we chose on $W = S^2 \times I$, Nahm's equations have a unique solution. Finally, in § 3.5, we explain the relation of the dual of Dirichlet boundary conditions to the universal kernel of geometric Langlands.

3.1. Twisting

We begin by reviewing the 'twisting' procedure by which topological field theories can be constructed, starting from supersymmetric Yang–Mills theory in four dimensions. The original example involved starting with $\mathfrak{N} = 2$ super Yang–Mills theory. The twisted theory is then essentially unique and is related to Donaldson theory [24]. Starting from $\mathfrak{N} = 4$ super Yang–Mills theory, there are three choices [22], one of which is related to geometric Langlands [15].

We begin by considering $\mathfrak{N} = 4$ super Yang–Mills theory on the Euclidean space $\mathbb{R}^4$. The rotation group is SO(4). We denote the positive and negative spin representations of the double cover Spin(4) as V_+ and V_-, respectively; they are both two dimensional. One important point is that, although $\mathfrak{N} = 4$ super Yang–Mills theory is conformally invariant, both classically and quantum mechanically, the twisting procedure does not use this conformal invariance. (A close analogy of the construction with the twisting of $\mathfrak{N} = 2$ super Yang–Mills theory would not be possible if we had to make use of conformal invariance, since $\mathfrak{N} = 2$ super Yang–Mills theory is not conformally invariant quantum mechanically.)

$\mathfrak{N} = 4$ super Yang–Mills theory also has an R-symmetry group Spin(6). An R-symmetry group is simply a group of symmetries that acts by automorphisms of the supersymmetries, while acting trivially on space-time. The group Spin(6) has positive and negative spin representations that we will call U_+ and U_-. They are both of dimension 4. The supersymmetries of $\mathfrak{N} = 4$ super Yang–Mills theory transform under $\mathrm{Spin}(4) \times \mathrm{Spin}(6)$ as

$$\mathcal{Y} = V_+ \otimes U_+ \oplus V_- \otimes U_-. \tag{3.1}$$

Classically [7], it is possible to construct $\mathfrak{N} = 4$ super-Yang Mills theory by dimensional reduction from 10 dimensions, that is, from $\mathbb{R}^{10}$. This entails an embedding $(\mathrm{Spin}(4) \times \mathrm{Spin}(6))/\mathbb{Z}_2 \subset \mathrm{Spin}(10)$. In this way of constructing the $\mathfrak{N} = 4$ theory, $\mathcal{Y}$ simply corresponds to one of the irreducible spin representations of Spin(10). The supersymmetry algebra in 10 dimensions reads

$$\{Q_\gamma, Q_\delta\} = \sum_{I=1}^{10} \Gamma^I_{\gamma\delta} P_I, \tag{3.2}$$

where the notation is as follows. Q_γ and Q_δ are two supersymmetry charges, corresponding to elements of $\mathcal{Y}$. The P_I generate the translation symmetries of $\mathbb{R}^{10}$, while the Γ_I are the generators of the Clifford algebra, understood as bilinear maps $\mathrm{Sym}^2\,\mathcal{Y} \to V_{10}$, where V_{10} is the 10-dimensional representation of Spin(10). Reduction to four dimensions is achieved by requiring the fields to be independent of the last six coordinates of $\mathbb{R}^{10}$. This reduces Spin(10) symmetry to the subgroup $(\mathrm{Spin}(4)\times\mathrm{Spin}(6))/\mathbb{Z}_2$ considered in the last paragraph. In the reduced theory,[8] the P_I vanish in (3.2) for $5 \leqslant I \leqslant 10$. Therefore, in the reduced theory, the right-hand side of (3.2) contains precisely the four operators P_I, $I = 1, \dots, 4$.

Remark 3.1. In particular, in the theory reduced to four dimensions, there is no Spin(4)-invariant operator on the right-hand side of (3.2). On the other hand, the right-hand side of (3.2) is Spin(6)-invariant.

The idea of twisting is to replace Spin(4) by another subgroup of $(\mathrm{Spin}(4) \times \mathrm{Spin}(6))/\mathbb{Z}_2$ that acts in the same way on space-time, but has some convenient properties that will be described. This is accomplished by picking a homomorphism $\lambda : \mathrm{Spin}(4) \to \mathrm{Spin}(6)$. Then we extend this to an embedding $(1 \times \lambda) : \mathrm{Spin}(4) \to \mathrm{Spin}(4) \times \mathrm{Spin}(6)$ and we define $\mathrm{Spin}'(4) = (1 \times \lambda)(\mathrm{Spin}(4))$. The twisted theory is one in which the ordinary rotation group Spin(4) is replaced by $\mathrm{Spin}'(4)$. In other words, whenever we make a rotation of $\mathbb{R}^4$ by an element $f \in \mathrm{Spin}(4)$, we accompany this by a Spin(6) transformation $\lambda(f)$.

We want to pick λ so that the $\mathrm{Spin}(4) \times \mathrm{Spin}(6)$ module $\mathcal{Y}$ contains a non-zero $\mathrm{Spin}'(4)$ invariant. Supposing that this is the case, pick such an invariant and write Q for the corresponding supersymmetry. Q automatically satisfies the fundamental condition $Q^2 = 0$. The reason for this is that Q^2 is $\mathrm{Spin}'(4)$ invariant and (since Q is a linear combination of the Q_γ) can be computed from (3.2). But, in view of remark 3.1, this is no $\mathrm{Spin}'(4)$ invariant on the right-hand side of (3.2).

Since $Q^2 = 0$, one can pass from $\mathfrak{N} = 4$ super Yang–Mills theory to a much 'smaller' theory by taking the cohomology of Q. One considers only operators (or states) that commute with Q (or are annihilated by Q) modulo operators of the form $\{Q, \dots\}$ (or states in the image of Q).

It is possible to state a simple condition under which the small theory can be extended to a topological field theory. The condition is that the stress tensor T of the theory, which measures the response of the theory to a change in the metric of $\mathbb{R}^4$, must be trivial in the cohomology of Q; that is, it must be of the form $T = \{Q, \Lambda\}$ for some Λ. In practice, this condition is always satisfied in four dimensions. Given this, one can promote the 'local' construction on $\mathbb{R}^4$ sketched in the last few paragraphs to a 'global' construction that makes sense on a rather general smooth 4-manifold M. (Depending on λ, M may require some additional structure such as an orientation or a spin structure; however, for the choice of λ that leads to geometric Langlands, no such additional structure is required.)

Three possible twists of $\mathfrak{N} = 4$ super Yang–Mills theory lead to topological field theories. Two of these are close cousins of Donaldson theory, and the third is related to geometric Langlands.

[8]It is possible [20] to pick boundary conditions such that the P_I, $I \geqslant 5$, survive in the reduced theory as central charges (electric charges) that commute with all local operators (and in this case magnetic charges appear in the algebra as additional central charges). We are not interested here in such boundary conditions. In any event, the automorphism of the algebra of local operators generated by the P_I always vanishes; this is what we need in the following arguments.

The twist which leads to geometric Langlands is easily described as follows: $\mathrm{SO}(6) = \mathrm{Spin}(6)/\mathbb{Z}_2$ has an obvious $\mathrm{SO}(4) \times \mathrm{SO}(2)$ subgroup

$$\begin{pmatrix} * & * & * & * & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * & * \end{pmatrix}. \tag{3.3}$$

Taking the double cover, Spin(6) has commuting Spin(4) and Spin(2) subgroups whose centres coincide, and hence a global embedding

$$\frac{\mathrm{Spin}(4) \times \mathrm{Spin}(2)}{\mathbb{Z}_2} \subset \mathrm{Spin}(6). \tag{3.4}$$

We simply take $\lambda : \mathrm{Spin}(4) \to \mathrm{Spin}(6)$ to be an isomorphism onto this Spin(4) subgroup of Spin(6). Since Spin(2) commutes with the image of λ, it becomes a global symmetry of the model. This actually is the group Spin(2) that played an important role in § 2.8.

The spin representations $U_\pm$ of Spin(6) decompose under $\mathrm{Spin}(4) \times \mathrm{Spin}(2)$ as

$$\left.\begin{aligned} U_+ &= V_+^1 \oplus V_-^{-1}, \\ U_- &= V_+^{-1} \oplus V_-^1. \end{aligned}\right\} \tag{3.5}$$

Here the notation is as follows. As before, V_+ and V_- are the two spin representations of Spin(4). As for Spin(2), it is abelian and isomorphic to U(1). Its spin representations are one-dimensional representations of U(1) of 'charge' 1 and -1; the charge is indicated by the superscripts ± 1 in (3.5).

Now, in view of equation (3.1), the supersymmetries of the theory transform under $\mathrm{Spin}'(4) \times \mathrm{Spin}(2)$ as

$$V_+ \otimes (V_+^1 \oplus V_-^{-1}) \oplus V_- \otimes (V_+^{-1} \oplus V_-^1). \tag{3.6}$$

We want to find the $\mathrm{Spin}'(4)$ invariants. Decomposing (3.6) into a direct sum of irreducibles, both $V_+ \otimes V_+^1$ and $V_- \otimes V_-^1$ contain a one-dimensional $\mathrm{Spin}'(4)$-invariant subspace, while there are no invariants in $V_\pm \otimes V_\mp^{-1}$.

Let us write Q_+ and Q_- for $\mathrm{Spin}'(4)$-invariant supersymmetries derived from the invariant part of $V_+ \otimes V_+^1$ and $V_- \otimes V_-^1$, respectively. Note that they both transform under Spin(2) with charge 1. A general complex linear combination

$$Q = uQ_+ + vQ_- \tag{3.7}$$

is $\mathrm{Spin}'(4)$ invariant and also has charge 1. It also turns out that any such Q (with u and v not both zero) obeys the condition for defining a topological field theory: the stress tensor can be written as $T = \{Q, \dots\}$. The topological field theory that we get by passing to the cohomology of Q is invariant under rescaling Q by a non-zero complex number. So we should think of u and v as homogeneous coordinates on a copy of $\mathbb{CP}^1$ that parametrizes a family of topological field theories.

Because $\mathrm{Spin}(2) \cong U(1)$, its representations are labelled by integers, corresponding to the characters $\exp(\mathrm{i}\theta) \to \exp(\mathrm{i}n\theta)$, $n \in \mathbb{Z}$. The action of Spin(2) gives a $\mathbb{Z}$ grading of the full physical Hilbert space $\hat{\mathcal{H}}$ of $\mathfrak{N} = 4$ super Yang–Mills theory.

In topological field theory, we want a $\mathbb{Z}$ grading not of $\hat{\mathcal{H}}$, but of a vastly smaller space $\mathcal{H}$: the cohomology of Q. In order for the cohomology of Q to be $\mathbb{Z}$-graded, we require that Q should transform in a definite character of Spin(2). This is true for any choice of u and v because both Q_+ and Q_- transform with the same character of Spin(2): what we have called charge 1. So any complex linear combination $Q = uQ_+ + vQ_-$ also has charge or degree 1, and the cohomology of Q is $\mathbb{Z}$-graded.

If it were the case, for example, that Q_+ and Q_- had charge 1 and -1, respectively, then a generic complex linear combination $Q = uQ_+ + vQ_-$ would not have definite charge, and its cohomology would be only $\mathbb{Z}_2$-graded. In § 3.3, we describe a situation in which something similar to that occurs.

3.1.1. *A slight complication*

Roughly speaking, the $\hat{A}$ model and the $\hat{B}$ model correspond to different values of the ratio v/u. The full details are a little more complicated, and involve also the coupling parameter $\tau = \theta/2\pi + 4\pi\mathrm{i}/e^2$ of the gauge theory, as explained in [15, § 3.5].

The complication arises because the Lagrangian of the theory cannot be written in the form $\{Q, \dots\}$, but is of this form only modulo a multiple of the topological invariant

$$\int_M \operatorname{Tr} F \wedge F.$$

Consequently, the topological field theory depends not only on the twisting parameter v/u, but also on τ. Actually, the topological field theory depends on τ and the twisting parameter only via a single parameter defined in [15, (3.50)]; as a result, it is true that twisting leads to a family of topological field theories parametrized by $\mathbb{CP}^1$ and that the $\hat{A}$ model and the $\hat{B}$ model correspond to two points in this space.

The expression

$$\int_M \operatorname{Tr} F \wedge F$$

actually has another interpretation. Let $P(\sigma)$ be an invariant quadratic polynomial on the Lie algebra $\mathfrak{g}$. Applying the construction of § 2.9 to P, we construct a sequence of r-form-valued operators $\mathcal{P}^{(r)}$, $r = 0, \dots, 4$, with $\mathcal{P}^{(0)} = P(\sigma)$ and

$$(\mathrm{d} + \{Q, \cdot\}) \sum_r \mathcal{P}^{(r)} = 0.$$

If we pick P correctly, then $\mathcal{P}^{(4)} = (1/8\pi^2) \operatorname{Tr} F \wedge F$. The integral

$$\int_M \mathcal{P}^{(4)} = \frac{1}{8\pi^2} \int_M \operatorname{Tr} F \wedge F$$

(which is none other than the instanton number) is Q-invariant but is non-trivial in the cohomology of Q. It is this fact that causes the coupling parameter τ to be relevant in the topological field theory.

In a superficially similar situation that will be considered in § 3.3, $\int_M \mathcal{P}^{(4)}$ will disappear from the Q-cohomology (by 'cancelling' a certain integral of $\mathcal{P}^{(3)}$, which will also disappear at the same time). This being so, τ will be irrelevant in the topological field theory, which will depend only on the choice of Q.

3.2. Scalar fields in twisted theory

Now we want to describe the bosonic fields of $\mathfrak{N} = 4$ super Yang–Mills theory before and after twisting. In 10 dimensions, the only bosonic field is the connection $\hat{A}$. Writing

$$\hat{A} = \sum_{I=1}^{4} A_I \, \mathrm{d}x^I + \sum_{J=5}^{10} A_J \, \mathrm{d}x^J,$$

we can parametrize $\hat{A}$ by the four-dimensional connection

$$A = \sum_{I=1}^{4} A_I \, \mathrm{d}x^I$$

and six scalar fields $\Phi_I = A_{4+I}$, $I = 1, \dots, 6$, that are valued in the adjoint representation of the gauge group G.

In particular, the six scalar fields Φ transform in the 'vector' representation of $\mathrm{SO}(6) = \mathrm{Spin}(6)/\mathbb{Z}_2$. Under the embedding $\mathrm{SO}(4) \times \mathrm{SO}(2) \subset \mathrm{SO}(6)$ sketched in (3.3), Φ splits into 'upper' components that transform under SO(4) and 'lower' components that transform under SO(2).

In the twisted theory on a general 4-manifold M, the upper components are interpreted as an $\mathrm{ad}(E)$-valued 1-form ϕ. Twisting transforms ϕ from a collection of four scalar fields (or 0-forms) into a 1-form. In other words, the upper components of ϕ are invariant under Spin(4), but transform under $\mathrm{Spin}'(4)$ in such a way that it is natural to interpret

$$\phi = \sum_{I=1}^{4} \Phi_I \, \mathrm{d}x^I$$

as a 1-form. This 1-form entered prominently in § 2. In the $\hat{B}$ model it combines with A to the complex connection $\mathcal{A} = A + \mathrm{i}\phi$, and in the $\hat{A}$ model it appears with A in the elliptic differential equations $F - \phi \wedge \phi = \star \, \mathrm{d}_A \phi$.

The lower components of Φ are a pair of $\mathrm{ad}(E)$-valued scalar fields that transform trivially under $\mathrm{Spin}'(4)$ but in a real two-dimensional representation of Spin(2). In § 2.6, these fields were called X_1 and X_2 and were combined into a complex field $\sigma = (X_1 + \mathrm{i}X_2)/\sqrt{2}$. The field σ has charge or degree 2 for the following reason. We defined the charge so that the fundamental representation of Spin(2) has charge 1, so the fundamental representation of $\mathrm{SO}(2) = \mathrm{Spin}(2)/\mathbb{Z}_2$ has charge 2. The fields X_1 and X_2 transform in the fundamental representation of SO(2), as is clear from the embedding $\mathrm{SO}(4) \times \mathrm{SO}(2) \subset \mathrm{SO}(6)$. In the $\hat{A}$ model, σ can be viewed as part of the Cartan model of the equivariant cohomology of the gauge group acting on the fields (A, ϕ). In the $\hat{B}$ model, its role was described in § 2.6.

All of this holds on a generic 4-manifold M. However, matters simplify if M is the product of a 3-manifold M_3 with a 1-manifold M_1. Here, M_1 may be either $\mathbb{R}$

or S^1 or a compact interval I with some boundary conditions chosen. Topological field theory on M does not really depend on what metric is chosen on M, but if M is a product, it is simplest to do the computations with a product metric. The cotangent bundle of M then splits metrically (as well as topologically) as a direct sum $T^*M = T^*M_3 \oplus T^*M_1$, where the connection on T^*M_1 is trivial.

We now should re-examine the four 'upper' components of Φ that are interpreted for generic M after twisting as a 1-form ϕ. In the case of the $3 \oplus 1$ split of the last paragraph, only three components of Φ are twisted. They can be interpreted as a 1-form on M_3. As for the fourth 'upper' component, it is a 1-form on M_1, but the cotangent bundle of M_1 is completely trivial, topologically, metrically and from the point of view of the Riemannian connection. So, in this particular situation, twisting has done nothing at all to this scalar field. Since it is unaffected by the twisting, just like the 'lower' components X_1 and X_2, we may as well combine it with them and call it X_3.

The Spin(2) global symmetry of $\mathfrak{N} = 4$ twisted super Yang–Mills theory on a generic M is now promoted to Spin(3), rotating X_1, X_2 and X_3. This is the Spin(3) symmetry that mysteriously appeared when we derived Nahm's equations in § 2.6. The example in that section and in most of § 2 was $M = \mathbb{R} \times S^2 \times I$, which can be decomposed as $M_3 \times M_1$ in more than one way. The decomposition that is relevant for understanding § 2.6 is $M_3 = \mathbb{R} \times S^2$, $M_1 = I$. Indeed, the formula $X_3 = \phi_y$ of equation (2.5) shows that X_3 is the component of ϕ in the I direction.

Although physical Yang–Mills theory on $M_3 \times M_1$ (after twisting but before passing to the Q cohomology) has Spin(3) symmetry, the topological field theory that we get by taking the Q cohomology does not. That is because Q does not transform in a one-dimensional representation of Spin(3). In fact, it lies in a two-dimensional representation of Spin(3).

3.3. More general construction in three dimensions

We will now make a digression aimed at making contact with some recent mathematical work [6]. At the end of § 3.2, we considered a four-dimensional topological field theory specialized to a 4-manifold with a product structure. Henceforth, we take this to be specifically $M = M_3 \times S^1$. Keeping S^1 fixed and letting M_3 vary, the four-dimensional topological field theory reduces to a three-dimensional one.

Starting with $\mathfrak{N} = 4$ super Yang–Mills in four dimensions, it is possible to modify the construction slightly to obtain a three-dimensional topological field theory that does *not* quite come in this way from a four-dimensional topological field theory. Roughly speaking, to do this, we require Q to have only $\mathrm{Spin}'(3)$ invariance, not $\mathrm{Spin}'(4)$ invariance.

To explain the construction in more detail, begin with $\mathfrak{N} = 4$ super Yang–Mills theory on $\mathbb{R}^3 \times S^1$. The spin group of $\mathbb{R}^3$ is Spin(3), and of course the R-symmetry group of the theory is still Spin(6). Now we want to pick a homomorphism $\tilde{\lambda}$: Spin(3) $\to$ Spin(6) and to define $\mathrm{Spin}'(3)$ as the image of

$$(1 \times \tilde{\lambda}) : \mathrm{Spin}(3) \to \mathrm{Spin}(3) \times \mathrm{Spin}(6).$$

We simply define $\tilde{\lambda}$ to be the restriction to Spin(3) of the homomorphism λ : Spin(4) $\to$ Spin(6) that we used before. In other words, we now begin with a subgroup $(\mathrm{Spin}_1(3) \times \mathrm{Spin}_2(3))/\mathbb{Z}_2 \subset \mathrm{Spin}(6)$ (here $\mathrm{Spin}_i(3)$, $i = 1, 2$, are two commuting copies of Spin(3)). We define $\mathrm{Spin}'(3)$ as the diagonal product of $\mathrm{Spin}(3) \times \mathrm{Spin}_1(3) \subset \mathrm{Spin}(3) \times \mathrm{Spin}(6)$.

Clearly, $\mathrm{Spin}'(3)$ commutes with the group $F = \mathrm{Spin}_2(3)$, which is yet another copy of Spin(3). F will play the role that was played in §§ 3.1 and 3.2 by Spin(2). The reason for the extension of Spin(2) to Spin(3) is the same as in § 3.2: only three scalar fields have been twisted, not four. We will also be interested in the complexification of F, which is $F_{\mathbb{C}} = \mathrm{Spin}(3, \mathbb{C}) \cong \mathrm{SL}(2, \mathbb{C})$.

To construct a three-dimensional topological field theory we must pick a $\mathrm{Spin}'(3)$-invariant supercharge. So let us determine how the supercharges transform under $\mathrm{Spin}'(3) \times F$. We write V, V_1 and V_2 for the spin representations of Spin(3), $\mathrm{Spin}_1(3)$ and $\mathrm{Spin}_2(3)$. The two spin representations $V_\pm$ of Spin(4) are both equivalent to V when restricted to Spin(3). Similarly, the two spin representations $U_\pm$ of Spin(6) are both equivalent under $(\mathrm{Spin}_1(3) \times \mathrm{Spin}_2(3))/\mathbb{Z}_2$ to $V_1 \otimes V_2$. So, as a $\mathrm{Spin}(3) \times \mathrm{Spin}_1(3) \times \mathrm{Spin}_2(3)$ module, the space of supersymmetries is

$$\mathcal{Y} = V \otimes V_1 \otimes V_2 \otimes \mathbb{C}^2. \tag{3.8}$$

We restrict to $\mathrm{Spin}'(3) \times \mathrm{Spin}_2(3)$ by setting $V_1 = V$, giving $\mathcal{Y} = V \otimes V \otimes V_2 \otimes \mathbb{C}^2$ The first step in constructing three-dimensional supersymmetric field theories is to extract the $\mathrm{Spin}'(3)$-invariant subspace. The $\mathrm{Spin}'(3)$-invariant subspace of $V \otimes V$ is one-dimensional, so the $\mathrm{Spin}'(3)$-invariant subspace of $\mathcal{Y}$ is four-dimensional. We call this subspace J. As an F-module, J is isomorphic to $V_2 \otimes \mathbb{C}^2$, where V_2 is a two-dimensional module for $F_{\mathbb{C}} \cong \mathrm{SL}(2, \mathbb{C})$.

If Q is the supersymmetry corresponding to a generic point in J, it is *not* true that $Q^2 = 0$. We can see this from (3.2). Though there is no $\mathrm{Spin}'(4)$ invariant on the right-hand side of (3.2), there is an essentially unique $\mathrm{Spin}'(3)$ invariant. It is the generator of the rotation of S^1, the second factor of $\mathbb{R}^3 \times S^1$. Let us call this generator $\mathcal{V}$. A generic $\mathrm{Spin}'(3)$-invariant supersymmetry squares not to zero but to a multiple of $\mathcal{V}$. On the $\mathrm{Spin}'(3)$ invariant subspace J, (3.2) reduces to something that in coordinates looks like

$$\{Q_\alpha, Q_\beta\} = \delta_{\alpha\beta}\mathcal{V}. \tag{3.9}$$

Intrinsically, $\delta_{\alpha\beta}$ is a quadratic form $(\cdot, \cdot)$ on the four-dimensional vector space J. This quadratic form is obviously $F_{\mathbb{C}}$ invariant, and this is actually enough to ensure that it is non-degenerate, given that $J \cong V_2 \otimes \mathbb{C}^2$. Indeed, the quadratic form is the tensor product of an $F_{\mathbb{C}}$-invariant skew form on V_2 and a non-zero (and therefore non-degenerate) skew form on $\mathbb{C}^2$. The skew form on $\mathbb{C}^2$ is invariant under a group $\tilde{F}_{\mathbb{C}}$ that is another copy of $\mathrm{SL}(2, \mathbb{C})$. $\tilde{F}_{\mathbb{C}}$ is therefore a group of symmetries of the quadratic form, though there is no natural way to make it act on the states and operators of the full theory.

Suppose that Q is a $\mathrm{Spin}'(3)$-invariant supersymmetry with $Q^2 = \mathcal{V}$ (or equivalently, Q^2 a non-zero multiple of $\mathcal{V}$). Can we use Q as a differential to construct a topological field theory? Superficially, the answer is 'no', since Q^2 is non-zero. However, $\mathcal{V}$ generates a symmetry (a compact group of rotations of $\mathbb{R}^3 \times S^1$) and we can restrict to $\mathcal{V}$-invariant operators and states. In this smaller space, $Q^2 = 0$ and we can pass to the cohomology of Q. In fact, similar constructions have been made previously [19, 21]. These constructions, respectively, involve non-free S^1 actions on

$\mathbb{R}^4$ or S^4. The construction we describe here is similar but simpler as it involves a free S^1 action.

The relation $Q^2 = \mathcal{V}$ is reminiscent of equivariant cohomology. Consider a U(1) action on a manifold B generated by a vector field V. Localized equivariant cohomology can be described by the operator $\mathrm{d}_V = \mathrm{d} + \iota_V$ acting on differential forms on B; here ι_V is the operator of contraction with V. One has $\mathrm{d}_V^2 = \mathcal{L}_V$, where $\mathcal{L}_V$ is the Lie derivative with respect to V. The operator d_V was related to supersymmetric nonlinear sigma models in [23] and was interpreted in equivariant cohomology in [3]. In our problem, since $\mathcal{V}$ generates the natural S^1 action on $M_4 = M_3 \times S^1$, the relation $Q^2 = \mathcal{V}$ is suggestive of localized equivariant cohomology for this action. This connection is made much more precise in [19, 21].

Up to scaling by a non-zero complex number, Q corresponds *a priori* to an arbitrary point in the projective space $\mathbb{P}(J) \cong \mathbb{CP}^3$. But it is not true that $\mathbb{CP}^3$ parametrizes a family of inequivalent topological field theories. If f is any invertible operator acting on the Hilbert space $\mathcal{H}$ of $\mathfrak{N} = 4$ super Yang–Mills theory, then Q and fQf^{-1} lead to equivalent topological field theories. In particular, picking $f \in F_\mathbb{C} \cong \mathrm{SL}(2,\mathbb{C})$, we see that, to classify the three-dimensional topological field theories that emerge from this construction, we must divide by the action of $F_\mathbb{C}$ on $\mathbb{CP}^3$.

Let us first classify those topological field theories for which $Q^2 = 0$. These correspond to the zeros of the non-degenerate quadratic form $(\cdot,\cdot)$ on $\mathbb{P}(J)$. They form a non-degenerate quadric $\mathcal{Q}$, which is a copy of $\mathbb{CP}^1 \times \mathbb{CP}^1$. This particular copy of $\mathbb{CP}^1 \times \mathbb{CP}^1$ is a homogeneous space for the group $\mathrm{SO}(4,\mathbb{C}) \cong (F_\mathbb{C} \times \tilde{F}_\mathbb{C})/\mathbb{Z}_2$ that acts on $\mathbb{P}(J)$ preserving the quadric, so we write it as $\mathbb{CP}^1 \times \widetilde{\mathbb{CP}}^1$. Here $\mathbb{CP}^1$ is a homogeneous space for $F_\mathbb{C}$, and $\widetilde{\mathbb{CP}}^1$ is a homogeneous space for $\tilde{F}_\mathbb{C}$. The quotient $(\mathbb{CP}^1 \times \widetilde{\mathbb{CP}}^1)/F_\mathbb{C}$ is just a copy of $\widetilde{\mathbb{CP}}^1$. However, $F_\mathbb{C}$ does not act freely on $\widetilde{\mathbb{CP}}^1$. Each point in $\widetilde{\mathbb{CP}}^1$ is left fixed by a Borel subgroup $\mathcal{B}$ of $F_\mathbb{C}$, isomorphic to

$$\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}. \tag{3.10}$$

In the topological field theory associated to a particular choice of Q, this Borel group acts as a group of symmetries. In particular, the cohomology of Q is $\mathbb{Z}$-graded by the action of the diagonal matrices in $\mathcal{B}$.

So we have a family of $\mathbb{Z}$-graded three-dimensional topological field theories, parametrized by a copy of $\mathbb{CP}^1$ (at this point we drop the tilde), with the property that $Q^2 = 0$. Actually, these are simply the examples that come by compactification on S^1 of a four-dimensional topological field theory.

To obtain something new we consider the examples for which Q^2 is a non-zero multiple of $\mathcal{V}$. Note that $\mathbb{P}(J)$ is a complex manifold of complex dimension 3, as is $F_\mathbb{C}$. This makes it possible for the complement of the quadric $\mathcal{Q} \subset \mathbb{P}(J)$ to consist of a single $F_\mathbb{C}$ orbit. This is, in fact, the situation. Bearing in mind the decomposition $J \cong V_2 \otimes \mathbb{C}^2$, where $F_\mathbb{C}$ acts on the first factor and $\tilde{F}_\mathbb{C}$ on the second, we can think of an element of J as a 2×2 matrix $K_{A\dot{A}}$, $A, \dot{A} = 1, 2$, with $F_\mathbb{C}$ and $\tilde{F}_\mathbb{C}$ acting on K respectively on the left and right. In this representation, the $F_\mathbb{C} \times \tilde{F}_\mathbb{C}$-invariant quadratic form is $K \to \det(K)$ and the condition for K not to be a null vector for the quadratic form is that it should be an invertible matrix. But any two invertible matrices are equivalent under the action of $F_\mathbb{C} \times \mathbb{C}^*$ ($F_\mathbb{C}$ acts on the 2×2 matrix K

by left multiplication, while $\mathbb{C}^*$ acts by scaling $K \to \lambda K$, $\lambda \in \mathbb{C}^*$: we must divide by $\mathbb{C}^*$ since we view K as an element of the projective space $\mathbb{P}(J)$). So, as claimed, the complement of the quadric in $\mathbb{P}(J)$ is a single $F_{\mathbb{C}}$ orbit.

Although the left action of $F_{\mathbb{C}}$ on the space of invertible 2×2 matrices is free, when we project to $\mathbb{P}(J)$, the action becomes only semi-free (that is, the stabilizer of a point is a finite group). In fact, $F_{\mathbb{C}} \cong \mathrm{SL}(2, \mathbb{C})$ contains a central subgroup $\mathbb{Z}_2$ consisting of the matrices -1 and 1. These matrices act trivially on $\mathbb{P}(J)$, and the subgroup of $F_{\mathbb{C}}$ that leaves a fixed point in $\mathbb{P}(J)$ that is not on the quadric is $\mathbb{Z}_2$. So if Q corresponds to a point that is not on the quadric, then its cohomology is $\mathbb{Z}_2$-graded, but not $\mathbb{Z}$-graded.

We can thus summarize what three-dimensional topological field theories arise from this construction. There is the usual $\mathbb{CP}^1$ family of theories that arise by compactification from four dimensions. Two points in this family are the $\hat{A}$ model and $\hat{B}$ model of G (which are equivalent, respectively, to the $\hat{B}$ model and $\hat{A}$ model of $G^\vee$). The generic point in this family corresponds to what is sometimes called quantum geometric Langlands (of G or equivalently of $G^\vee$). There is one more theory that does *not* arise by compactification of a four-dimensional theory. It is only $\mathbb{Z}_2$ graded and, as we explain momentarily, does not distinguish G from $G^\vee$.

What we have established so far is really that, by varying Q at a fixed value of the coupling parameter τ of the theory, we can construct only one new theory. In § 3.3.1 we will show that because of vanishing of a certain element of cohomology, the parameter τ is irrelevant in the new theory. This means that the new theory is really unique.

This new $\mathbb{Z}_2$-graded theory appears to be a candidate for the one studied in [6]. Electric–magnetic duality acts non-trivially on the $\mathbb{CP}^1$ that parametrizes theories that come from four dimensions. But the new theory, being unique, must be invariant under duality. In particular, as duality exchanges G and $G^\vee$, the new three-dimensional theory defined for G is equivalent to the same theory defined for $G^\vee$.

Starting with any point on the quadric $\mathcal{Q}$, corresponding to one of the usual theories studied in (ordinary or quantum) geometric Langlands, and making an infinitesimal perturbation away from $\mathcal{Q}$, one lands on the same generic $F_{\mathbb{C}}$ orbit. So the same theory (the one that is symmetrical between G and $G^\vee$) can be reached (after compactification to three dimensions) by an infinitesimal perturbation of any of the theories of four-dimensional origin. The required perturbation reduces the $\mathbb{Z}$-grading to a $\mathbb{Z}_2$-grading.

3.3.1. *Vanishing of a certain element of cohomology*

As explained in § 3.1.1, the reason that the gauge coupling parameter τ is not completely irrelevant in the twisted four-dimensional theories that lead to geometric Langlands is that the instanton number

$$\nu = \int_M \mathcal{P}^{(4)} = \frac{1}{8\pi^2} \int_M \mathrm{Tr}\, F \wedge F$$

is Q-invariant and not of the form $\{Q, \dots\}$; that is, it represents a non-trivial cohomology class of Q. Adding a multiple of ν to the Lagrangian gives a non-trivial deformation of the theory.

It turns out that when we perturb slightly away from the quadric, this cohomology class disappears. As a result, the parameter τ becomes irrelevant, completing the justification of the claim that, after compactification to three dimensions on a circle, there is precisely one new $\mathbb{Z}_2$-graded topological field theory that we can make.

REMARK 3.2. The fact that the cohomology class disappears under perturbation away from the quadric can be anticipated as follows. As shown in [15], the deformation by the cohomology class ν is equivalent to the deformation associated with a change in the linear combination $Q = uQ_+ + vQ_-$. We have already seen that once we move away from $\mathcal{Q}$, the deformation by changing Q becomes trivial, so the deformation by ν must also become trivial. Instead of relying on this sort of argument, we prefer to be more explicit.

In general, for a cohomology class to disappear under an infinitesimal perturbation, it must annihilate another cohomology class whose $\mathbb{Z}$-grading differs by ± 1 (if the perturbation preserves a $\mathbb{Z}$-grading, as in the case usually considered), or at least one that has the opposite $\mathbb{Z}_2$ grading (if the perturbation preserves only a $\mathbb{Z}_2$-grading, as in the case considered here). In the four-dimensional topological field theories related to geometric Langlands, there is no 4-form-valued cohomology class of Q with an odd grading that could possibly cancel $\int_M \mathcal{P}^{(4)}$ in the cohomology. However, once we compactify to three dimensions, there is such a class. Our construction on $M = M_3 \times S^1$ made use of a vector field $\mathcal{V}$ that generates the rotation of S^1. There is a natural $\mathcal{V}$-invariant 1-form $\mathrm{d}y$ on S^1 with $\int_{S^1} \mathrm{d}y = 1$. This enables us to consider the expression

$$\tilde{\nu} = \int_M \mathcal{P}^{(3)} \wedge \mathrm{d}y,$$

which is a Q cohomology class of degree 1. If the four-dimensional $\mathbb{Z}$-graded theory is restricted to 4-manifolds of the form $M_3 \times S^1$, then, in addition to the usual complex modulus corresponding to the cohomology class ν (this modulus is tangent to the usual $\mathbb{CP}^1$ family), there is an odd modulus corresponding to $\tilde{\nu}$.

But, when one perturbs away from the quadric $\mathcal{Q}$ to a $\mathbb{Z}_2$-graded theory, the cohomology classes ν and $\tilde{\nu}$ both disappear, as we will now argue. Let Q be the topological supersymmetry generator corresponding to a point in $\mathcal{Q}$, so that $Q^2 = 0$ and Q descends from four dimensions. Pick a one-parameter deformation $Q_\varepsilon = Q + \varepsilon Q'$, where Q' corresponds to another point in $\mathbb{P}(J)$ and $Q_\varepsilon^2 \neq 0$. After possibly replacing Q' by a linear combination of Q and Q', we can assume that $(Q')^2 = 0$ and that

$$\{Q, Q'\} = \mathcal{V}. \tag{3.11}$$

Let

$$\mathrm{CS}(A) = \frac{1}{8\pi^2} \operatorname{Tr}(A \wedge \mathrm{d}A + \tfrac{2}{3} A \wedge A \wedge A)$$

be the Chern–Simons 3-form. Its periods are not well defined as real numbers, but rather take values in $\mathbb{R}/\mathbb{Z}$. Let

$$\Theta = \int_M \mathrm{CS}(A) \wedge \mathrm{d}y.$$

In defining Θ, we pick a point $y_0 \in S^1$ and, at that point, we pick an $\mathbb{R}$-valued lift of

$$\int_{M_3 \times y_0} \mathrm{CS}(A).$$

Then we pick an $\mathbb{R}$-valued lift of

$$f(y) = \int_{M_3 \times y} \mathrm{CS}(A)$$

so that this function is continuous for $y > y_0$, and define

$$\Theta = \int_{S^1} \mathrm{d}y\, f(y).$$

Once we go all the way around the circle, $f(y)$ will jump by ν, the instanton number, so the definition of Θ depends on both the choice of y_0 and the real lift chosen for $\int_{M_3 \times y_0} \mathrm{CS}(A)$. But the indeterminacy of Θ is independent of A, and hence it makes sense to compute the commutator $[\mathcal{V}, \Theta]$, where $\mathcal{V}$ acts on A by generating the rotation of the circle. Since

$$\int_{S^1} \mathrm{d}y \left(\frac{\mathrm{d}f}{\mathrm{d}y} \right)$$

(which is the change in f in going around the circle) equals the instanton number ν, the commutator is

$$[\mathcal{V}, \Theta] = \nu. \tag{3.12}$$

(Physicists would usually describe this computation by saying that $[\mathcal{V}, A_i] = F_{yi}$, where A_i is a component of the connection tangent to M_3 and F_{yi} is a corresponding curvature component. Using this, a formal evaluation of the commutator gives (3.12).)

Another useful calculation gives

$$[Q, \Theta] = \int_M \mathcal{P}^{(3)} \wedge \mathrm{d}y = \tilde{\nu}. \tag{3.13}$$

Again, the commutator makes sense because Θ is well defined modulo an additive constant. To compute this commutator, one needs to know that $[Q, A] = \psi$, where ψ is an adjoint-valued fermion field such that $\mathcal{P}^{(3)} = (1/4\pi^2)\,\mathrm{Tr}\, F \wedge \psi$. The formula (3.13) does not make $\tilde{\nu} = \int_M \mathcal{P}^{(3)}$ trivial in the cohomology of Q, since Θ is not a well-defined real-valued function.

However, now we find $\{Q', \tilde{\nu}\} = \{Q', [Q, \Theta]\} = -\{Q, [Q', \Theta]\} + \{\mathcal{V}, \Theta\}$, where (3.11) has been used along with the Jacobi identity. Also using (3.12), we obtain

$$\{Q', \tilde{\nu}\} = \nu - \{Q, [Q', \Theta]\}. \tag{3.14}$$

Again, the commutator $[Q', \Theta]$ is well defined despite the uncertainty of Θ by a real constant (an explicit local quantum field theory expression can be written for this commutator). So when we pass to the cohomology of Q, the last term in (3.14) is trivial and this equation reduces to $\{Q', \tilde{\nu}\} = \nu$. This implies that when we perturb Q to $Q_\varepsilon = Q + \varepsilon Q'$, both $\tilde{\nu}$ and ν disappear from the cohomology.

Remark 3.3. Going back to four dimensions, we can select an invariant polynomial P_i of degree d_i and perturb the topological field theories related to geometric Langlands by the Q-invariant interaction $\int_M \mathcal{P}_i^{(4)}$. This perturbation has degree $2d_i - 4$, so, for $d_i > 2$, it gives a $\mathbb{Z}_{2d_i-4}$-graded theory. If we include a linear combination of such perturbations with all possible values of i, we will get a family of four-dimensional topological field theories that (for most simple Lie groups G) are generically only $\mathbb{Z}_2$-graded. These theories have similar behaviour under electric–magnetic duality to the $\mathbb{Z}$-graded theories that are usually considered in geometric Langlands. It is not clear to the author whether they contain any essentially new information.

3.4. Uniqueness of the solution of Nahm's equation

An important point in § 2.7 was that, with the appropriate boundary conditions at the two ends, the solution of Nahm's equations on the half-open interval $(0, L]$ is unique. The boundary condition for $y \to 0$ was described in equation (2.7): $\boldsymbol{X}$ should have a regular pole at $y = 0$, the singular part being

$$\boldsymbol{X} = \frac{\boldsymbol{t}}{y}, \tag{3.15}$$

where $\boldsymbol{t}$ are the images of the $\mathfrak{su}(2)$ generators under a principal embedding $\vartheta : \mathfrak{su}(2) \to \mathfrak{g}^\vee$.

The boundary condition at $y = L$ was not explained in § 2 but, as we will explain, its effect is that the solutions of Nahm's equations on $(0, L]$ with the conditions we will want at $y = L$ are tautologically the same as the solutions of Nahm's equations on the open half-line $(0, \infty)$ with a requirement that $\boldsymbol{X} \to 0$ at infinity.

Kronheimer [16] investigated Nahm's equations on the open half-line with these conditions[9] (including the regular Nahm pole at $y = 0$) and showed that the solution is unique. So once we have explained how our problem on the half-open interval $(0, L]$ is related to Kronheimer's problem on the half-line $(0, \infty)$, the uniqueness claimed in § 2.7 will follow.

Actually, Kronheimer considered a more general problem in which $\vartheta : \mathfrak{su}(2) \to \mathfrak{g}^\vee$ is taken to be an arbitrary homomorphism, not necessarily related to a principal embedding. We will need to know about the opposite case, when $\vartheta = 0$. For this choice, there is no pole at $y = 0$, so we are studying solutions of Nahm's equations on the closed half-line $[0, \infty)$. In this case, the moduli space of solutions of Nahm's equations turns out to be a hyper-Kahler manifold $\mathcal{X}(G^\vee)$ that, in any of its complex structures, is equivalent to the nilpotent cone in the complex Lie algebra $\mathfrak{g}_{\mathbb{C}}^\vee$. The moduli space $\mathcal{X}(G^\vee)$ has $G^\vee$ symmetry for an easily understood reason: if $\vartheta = 0$, then the group $G^\vee$ acts on the solutions of Nahm's equations in the obvious fashion $\boldsymbol{X} \to g\boldsymbol{X}g^{-1}$. (For $\vartheta \neq 0$, the group that acts is the subgroup of $G^\vee$ that commutes with the image of ϑ.) The hyper-Kahler moment map for the $G^\vee$ action on $\mathcal{X}(G^\vee)$ turns out to be $\boldsymbol{\mu} = \boldsymbol{X}(0)$. All this has the following trivial generalization. If we

[9] Kronheimer also considered a generalization of the condition $\boldsymbol{X} \to 0$ at infinity, the requirement being instead that $\boldsymbol{X}$ is conjugate at infinity to a specified triple of elements of $\boldsymbol{t}^\vee$. It is possible to modify our boundary conditions on both the $\hat{A}$-model and $\hat{B}$ model sides so as to arrive at this generalization. The necessary facts are mostly presented in [11]. However, we will omit this generalization here.

solve Nahm's equations on the half-line $[L,\infty)$ (rather than on $[0,\infty)$), we obtain an isomorphic hyper-Kahler manifold, the moment map for the $G^\vee$ action now being

$$\boldsymbol{\mu} = \boldsymbol{X}(L). \tag{3.16}$$

Here we will only require the extreme cases that ϑ is either 0 or a principal embedding. The general result [16], however, for any ϑ, is that the moduli space of solutions of Nahm's equations turns out to be, as a complex manifold in any of its complex structures, the Slodowy slice transverse to the nilpotent element $t_1 + \mathrm{i}t_2$ of $\mathfrak{g}_{\mathbb{C}}^\vee$.

Now we need to describe the boundary conditions at $y = L$ in the construction of § 2.7. The relevant notion of a boundary condition is more extended than one may be accustomed to in the world of partial differential equations, for example. A boundary condition in a quantum field theory defined on d-manifolds is a choice of how to extend the definition to d-manifolds with boundary in such a way that all the usual axioms of local quantum field theory are preserved. This notion allows us to include on the boundary a $(d-1)$-dimensional quantum field theory. It is only interesting to do that, however, if the $(d-1)$-dimensional boundary theory is coupled in some way to the 'bulk' theory.

One might think that this notion of a boundary condition is too broad. However, it is shown in [11,12] that this extended notion of a boundary condition is unavoidable if one wishes electric–magnetic duality to act on boundary conditions, since the dual of a more conventional boundary condition can very well be a boundary condition in this extended sense. For example, as shown in [12], the dual of Dirichlet boundary conditions in G-gauge theory is a boundary condition in $G^\vee$ gauge theory that involves the coupling of the $G^\vee$ gauge fields to a very special superconformal field theory $T(G^\vee)$ that is supported on the boundary. For our purposes, $T(G^\vee)$ has the following important properties. It has $G^\vee \times G$ as a group of global symmetries. The Higgs branch of vacua of $T(G^\vee)$ turns out to be the Kronheimer manifold $\mathcal{X}(G^\vee)$, and the Coulomb branch of vacua is the dual Kronheimer manifold $\mathcal{X}(G)$.

As explained in [11], for a boundary condition in $G^\vee$ gauge theory that is obtained by coupling to a boundary theory with $G^\vee$ symmetry, the appropriate boundary condition in Nahm's equations is to set $\boldsymbol{X}$ equal on the boundary to $\boldsymbol{\mu}$, the moment map for the action of $G^\vee$ on the Higgs branch:

$$\boldsymbol{X}(L) = \boldsymbol{\mu}. \tag{3.17}$$

This equation looks just like (3.16), even though the two equations have a completely different meaning. In (3.17), $\boldsymbol{X}$ is a solution of Nahm's equations on the interval $(0,L]$, where the quantum field theory is defined. In equation (3.16), $\boldsymbol{X}$ is a solution of Nahm's equations on the half-line $[L,\infty)$. It defines a point in the Higgs branch of the boundary theory. Nevertheless, if we simply combine the two equations we see that, even though their interpretations are completely different, the solution of Nahm's equations on $(0,L]$ agrees at $y = L$ with the solution of Nahm's equations on $[L,\infty)$. Hence, they fit together to a single solution of Nahm's equations on the open half-line $(0,\infty)$. Nahm's equations ensure that this solution is smooth near $y = L$. It has the singular behaviour (3.15) near $y = 0$, and vanishes for $y \to \infty$, since this is a characteristic of the moduli space $\mathcal{X}(G)$. According to the first result of Kronheimer mentioned at the beginning of this subsection, Nahm's equations have a unique solution (namely $\boldsymbol{X} = \boldsymbol{t}/y$) obeying these conditions. This is the uniqueness asserted in § 2.7.

The examples that we have described here of the role of Nahm's equations in duality of boundary conditions in $\mathfrak{N} = 4$ super Yang–Mills theory are really only the tip of the iceberg. Much more can be found in [11, 12]. The full story involves, among other things, the more general moduli spaces defined by Kronheimer for an arbitrary $\vartheta : \mathfrak{su}(2) \to \mathfrak{g}$.

3.5. More on the dual of Dirichlet boundary conditions

In § 3.4 we exploited, in a rather technical way, the special properties of the dual of Dirichlet boundary conditions. We should perhaps not leave the subject without explaining that the dual of Dirichlet boundary conditions actually plays a rather basic role in the geometric Langlands correspondence.

We start by explaining intuitively why the dual of Dirichlet boundary conditions should be important. In geometric Langlands, one considers the $\hat{B}$ model of $\mathfrak{N} = 4$ super Yang–Mills theory, compactified on a Riemann surface C, for gauge group $G^\vee$. We compare it with the $\hat{A}$ model of G on the same Riemann surface. The most basic branes in the $\hat{B}$ model are branes associated with a homomomorphism $\chi : \pi_1(C) \to G^\vee_{\mathbb{C}}$. We would like to understand their duals in the $\hat{A}$ model.

Let us start with the case that χ is trivial. Let $\mathcal{B}$ be the corresponding $\hat{B}$-brane. We could modify $\mathcal{B}$ by introducing a Nahm pole, but let us not do so.

Then the brane $\mathcal{B}$ is simply the one that is defined by Dirichlet boundary conditions for the complexified gauge field $\mathcal{A} = A + \mathrm{i}\phi$ (extended to all other fields to preserve the topological supersymmetry of the $\hat{B}$ model). After all, Dirichlet boundary conditions say that $\mathcal{A}$ should be trivialized on the boundary, so that the boundary data corresponds to a trivial flat connection representing the trivial homomorphism from $\pi_1(C)$ to $G^\vee_{\mathbb{C}}$.

Dirichlet boundary conditions can be considered without any compactification, as indeed was the case in [11, 12]. Thus the brane $\mathcal{B}$ associated to the trivial flat connection without a Nahm pole has a universal meaning, independent of any choice of Riemann surface C. (This is also true for the analogous problem with a specified Nahm pole.)

Let $\mathcal{B}^*$ be the $\hat{A}$-brane that is dual to $\mathcal{B}$. Then $\mathcal{B}^*$, like $\mathcal{B}$, can be defined universally without any choice of compactification. As explained in [12], and as already stated in § 3.4, $\mathcal{B}^*$ is defined by coupling G gauge theory to a three-dimensional superconformal field theory $T(G)$ that has $G \times G^\vee$ global symmetry.[10] We use the G symmetry of $T(G)$ to couple it to G gauge fields in bulk. This leaves a $G^\vee$ global symmetry, matching the fact that $G^\vee$ is the automorphism group of Dirichlet boundary conditions (or of the trivial homomorphism $\pi_1(C) \to G^\vee_{\mathbb{C}}$) in $G^\vee$ gauge theory.

The duality between $\mathcal{B}$ and $\mathcal{B}^*$ holds before or after compactification on a Riemann surface C. However, after compactification, we can consider a twisted version of the picture in which we twist using the automorphism group $G^\vee_{\mathbb{C}}$, which $\mathcal{B}$ and

[10] For three-dimensional superconformal field theories with the relevant amount of supersymmetry, there is a notion of mirror symmetry [14], somewhat analogous to the more familiar mirror symmetry in two dimensions. The mirror of $T(G)$, in this sense, is $T(G^\vee)$. Indeed, $T(\mathrm{SU}(2))$ was one of the fundamental examples considered in [14].

$\mathcal{B}^*$ have in common. On the $\hat{B}$-model side, the twisted version of the picture simply involves a choice of homomorphism $\chi : \pi_1(C) \to G^\vee_{\mathbb{C}}$. To each choice of χ, one defines a $\hat{B}$-brane $\mathcal{B}(\chi)$ that is locally isomorphic to $\mathcal{B}$, but globally is obtained from $\mathcal{B}$ by twisting by the homomorphism χ from $\pi_1(C)$ to the automorphism group $G^\vee_{\mathbb{C}}$ of $\mathcal{B}$. (The statement that $\mathcal{B}(\chi)$ is 'locally' isomorphic to $\mathcal{B}$ means that they are locally isomorphic along C.) Let us denote by $\mathcal{B}^*(\chi)$ the dual of $\mathcal{B}(\chi)$. Then $\mathcal{B}^*(\chi)$ is obtained from $\mathcal{B}^*$ exactly as $\mathcal{B}(\chi)$ was obtained from $\mathcal{B}$: by twisting, via a homomorphism, from $\pi_1(C)$ to the automorphism group. This makes sense, since $\mathcal{B}$ and $\mathcal{B}^*$ have the same automorphism group $G^\vee$.

So the dual of any $\mathcal{B}(\chi)$ can be constructed if one understands the three-dimensional superconformal field theory $T(G)$ that is the main ingredient in describing the dual of Dirichlet boundary conditions. Thus, a knowledge of $T(G)$ gives the same sort of results that one would expect mathematically from a description of the universal kernel of geometric Langlands. This universal kernel is supposed to be a brane in the product theory $\hat{A}(G) \times \hat{B}(G^\vee)$ that has certain universal properties. In fact, $T(G)$ can be used to construct the appropriate brane. This can be done prior to compactification, and thus independently of any choice of C.

The relevant construction is quite simple and was described in [12, § 4]. We divide $\mathbb{R}^4$ into two half-spaces separated by a copy of $\mathbb{R}^3$, supported at, say, $y = 0$, where y is one of the Euclidean coordinates of $\mathbb{R}^4$. For $y < 0$, we place $\mathfrak{N} = 4$ super Yang–Mills theory with gauge group G, while, for $y > 0$, we place the same theory with gauge group $G^\vee$. At $y = 0$, one places the theory $T(G)$. Using its $G \times G^\vee$ global symmetries, it can be coupled to G gauge theory on the left and $G^\vee$ gauge theory on the right. Moreover, the coupling can be chosen so that the whole construction is supersymmetric or, to be more precise, invariant under a subgroup $\mathrm{OSp}(4|4)$ of the symmetry supergroup $\mathrm{PSU}(2,2 \mid 4)$ of $\mathfrak{N} = 4$ super Yang–Mills theory. One can pick a fermionic generator of $\mathrm{OSp}(4|4)$ that for $y < 0$ generates the topological supersymmetry of the $\hat{A}$ model of G, and for $y > 0$ generates the corresponding symmetry of the $\hat{B}$ model of $G^\vee$.

To get closer to the usual mathematical point of view, we can 'fold' $\mathbb{R}^4$ along the hypersurface $y = 0$, so that the G and $G^\vee$ gauge groups are now both supported at $y < 0$ and there is nothing for $y > 0$. In this description, then, the theoretical $T(G)$ provides a boundary condition in the product of G and $G^\vee$ gauge theories. After topological twisting, this boundary condition corresponds to a brane $\tilde{\mathcal{B}}$ in the product of the $\hat{A}$ model of G and the $\hat{B}$ model of $G^\vee$. Like the branes $\mathcal{B}$ and $\mathcal{B}^*$ discussed above, $\tilde{\mathcal{B}}$ can be defined in a universal way without any compactification. This was indeed the viewpoint in [12], where properties were discussed that correspond to the desired universal properties in geometric Langlands.

Acknowledgements

The idea of connecting the Langlands correspondence with gauge theory was first proposed in the 1970s by M. F. Atiyah (motivated by the observation that the Goddard–Nuyts–Olive dual group is the same as the Langlands dual group). I am grateful to him for introducing me to those ideas at that time. I also thank S. Cherkis and the referee for a close reading of the manuscript and careful comments, E. Frenkel for some helpful questions and D. Gaiotto for collaboration on electric–magnetic duality of boundary conditions. This paper is based on a lecture at the ICMS conference 'Geometry and Physics: Atiyah80' in Edinburgh, April 2009. Supported in part by NSF Grant no. Phy-0503584.

References

1 M. F. Atiyah and R. Bielawski. Nahm's equations, configuration spaces, and flag manifolds. *Bull. Brazil. Math. Soc.* **33** (2002), 157–176.

2 M. F. Atiyah and R. Bott. The Yang–Mills equations over Riemann surfaces. *Phil. Trans. R. Soc. Lond.* A **308** (1982), 523–615.

3 M. F. Atiyah and R. Bott. The moment map and equivariant cohomology. *Topology* **23** (1984), 1–28.

4 M. F. Atiyah and N. Hitchin. *The physics and geometry of magnetic monopoles* (Princeton University Press, 1988).

5 M. F. Atiyah and L. Jeffrey. Topological Lagrangians and cohomology. *J. Geom. Phys.* **7** (1990), 119–136.

6 D. Ben-Zvi and D. Nadler. The character group of a complex group. Preprint arXiv: 0904.1247.

7 L. Brink, J. H. Schwarz and J. Scherk. Supersymmetric Yang–Mills theories. *Nucl. Phys.* B **121** (1977), 77–92.

8 S. A. Cherkis and B. Durcan. The 't Hooft–Polyakov monopole in the presence of an 't Hooft operator. *Phys. Lett.* B **671** (2009), 123–127.

9 K. Corlette. Flat G-bundles with canonical metrics. *J. Diff. Geom.* **28** (1988), 361–382.

10 S. Donaldson. Polynomial invariants for smooth 4-manifolds. *Topology* **29** (1990), 257–315.

11 D. Gaiotto and E. Witten. Supersymmetric boundary conditions in $\mathfrak{N} = 4$ super Yang–Mills theory. Preprint arXiv:0804.2902 [hep-th].

12 D. Gaiotto and E. Witten. S-duality of boundary conditions in $\mathfrak{N} = 4$ super Yang–Mills theory. Preprint arXiv:0807.3720 [hep-th].

13 V. Ginzburg. Perverse sheaves on a loop group and geometric Langlands duality. Preprint alg-geom/9511007.

14 K. A. Intriligator and N. Seiberg. Mirror symmetry in three dimensional gauge theories. *Phys. Lett.* B **387** (1996), 513–519.

15 A. Kapustin and E. Witten. Electric–magnetic duality and the geometric Langlands program. Preprint arXiv:hep-th/0604151.

16 P. B. Kronheimer. Instantons and the geometry of the nilpotent variety. *J. Diff. Geom.* **32** (1990), 473–490.

17 I. Mirkovic and K. Vilonen. Perverse sheaves on affine Grassmannians and geometric Langlands duality. Preprint alg-geom/9911050.

18 W. Nahm. A simple formalism for the BPS monopole. *Phys. Lett.* B **90** (1980), 413–414.

19 N. A. Nekrasov. Seiberg–Witten prepotential from instanton counting. *Adv. Theor. Math. Phys.* **7** (2004), 831–864.

20 D. I. Olive and E. Witten. Supersymmetry algebras that include topological charges. *Phys. Lett.* A **78** (1978), 97–101.

21 V. Pestun. Localization of gauge theory on a four-sphere and supersymmetric Wilson loops. Preprint arXiv:0712.2824 [hep-th].

22 C. Vafa and E. Witten. A strong coupling test of S-duality. *Nucl. Phys.* B **431** (1994), 3–77.

23 E. Witten. Supersymmetry and Morse theory. *J. Diff. Geom.* **17** (1982), 661–692.

24 E. Witten. Topological quantum field theory. *Commun. Math. Phys.* **117** (1988), 353–386.

25 E. Witten. Mirror symmetry, Hitchin's equations, and Langlands duality. Preprint arXiv: 0802:0999.

26 E. Witten. Geometric Langlands from six dimensions. In *A Celebration of the mathematical legacy of Raoul Bott* (ed. R. Kotiuga *et al.*), CRM Proceedings and Lecture Notes, vol. 50 (Providence, RI: American Mathematical Society, 2010). Preprint arXiv:0905.2720.

(*Issued 6 August 2010*)