Gauge Theory and the Geometric Langlands Program

Edward Witten

August 10th, 2005

Talk at the Third Simons Workshop in Mathematics and Physics
SUNY at Stony Brook, July 25 - August 26, 2005

Based on notes by Ram Sriharsha

Introduction

The Langlands program of number theory, or what we might call Langlands duality, was proposed in more or less its present form by Robert Langlands, in the late 1960s. It is a kind of unified scheme for many results in number theory ranging from quadratic reciprocity, which is hundreds of years old, to modern results such as Andrew Wiles' proof of Fermat's last theorem, which involved a sort of special case of the Langlands program. For today, however, I will not assume any prior knowledge of the Langlands program.

Langlands duality was originally formulated for number fields, but geometers have also developed analogs involving curves over a finite field and ordinary complex Riemann surfaces. For a very brief introduction to all this, see [1]. We will focus today on the geometric Langlands program for complex Riemann surfaces. Some mathematicians regard this as an important test case for the Langlands program, and some don't. Once again, however, for today's talk, you do not need to know about the geometric Langlands program. The talk will start with $\mathcal{N} = 4$ super Yang-Mills and branes, and using standard physical reasoning about gauge theories and branes, we will see the appearance of the ingredients that go into the geometric Langlands program.

The starting point therefore is $\mathcal{N} = 4$ supersymmetric Yang-Mills theory, but before we get there, there is some standard lore that needs to be stated here. Goddard, Nuyts, and Olive

(GNO) [2], observed that in gauge theories, electric charge takes values in the weight lattice of the gauge group G, while magnetic charge takes values in the root lattice. They also observed that for any simple Lie group G, there is a "dual" group, the GNO or Langlands dual group ${}^L G$, for which the two lattices are exchanged. Montonen and Olive went farther [3] and proposed that in some gauge theories, there is an electric-magnetic duality symmetry that inverts the gauge coupling, while also replacing G with ${}^L G$.

There has always been something a bit strange about the GNO picture, as it involved talking about monopole states that generically are unstable. (They can be stabilized by turning on a Higgs field expectation value, but this spoils other parts of the arguments.) An improved understanding has been given recently by Kapustin [4] (with whom I have been collaborating on matters related to the geometric Langlands program), who restated the GNO construction in terms of operators – the Wilson and 't Hooft loop operators – rather than states. Under electric-magnetic duality, the 't Hooft operators and Wilson operators get mapped into each other. A Wilson operator is given by a representation of G that has a highest weight, so it is determined by a vector in the weight lattice. The 't Hooft operator is given by an embedding of $U(1)$ into G. One constructs a singular Dirac monopole using such an embedding, and each such embedding lets us define an 't Hooft operator (which is characterized by saying that its matrix elements are computed by a path integral over fields that agree with the given singular monopole solution near the singularity). So the 't Hooft operators are classified by embeddings of $U(1)$ into G and therefore labeled by points in the root lattice of G. If we are going to exchange 't Hooft and Wilson loops, we will have to exchange the root and weight lattices and exchange G with ${}^L G$.

Physicists have tended to think of the Wilson operators only as criteria for confinement. From that point of view, the adjoint representation is considered trivial. But in quantum theory, we can define Wilson operators for every representation, whether they give a useful quantum criterion for confinement or not. (Though Wilson lines in representations such as the adjoint representation, on which the center of the gauge group acts trivially, are irrelevant for confinement, there are quantum field theories – like Chern-Simons gauge theory in three dimensions – in which they are most definitely useful and significant.) Likewise, one can define an 't Hooft operator for every embedding of $U(1)$ into G whether or not the corresponding Dirac monopoles are stable. Most of them are not.

Nowadays, we know that the simplest situation where Montonen-Olive duality is realized is in $\mathcal{N} = 4$ supersymmetric gauge theories. From the modern point of view, it was discovered in the 1980's that the original Montonen-Olive duality is only part of a more general $SL(2, \mathbb{Z})$ duality which takes τ into $(a\tau + b)/(c\tau + d)$, where a, b, c, and d are integers with $ad - bc = 1$, and $\tau = \theta/2\pi + 4\pi i/e^2$. (For simplicity, I am stating this in the simply-laced case; for non-simply-laced G, things are a little more complicated.) However, it turns out that it is the original Montonen-Olive duality, which in this language is $\tau \to -1/\tau$ (plus exchange of G and ${}^LG$), that is relevant for the Langlands program. So we will focus on this duality here. (The rest of $SL(2,\mathbb{Z})$ is actually useful in an extension of the Langlands program that we will not discuss today.)

A year ago, in a conference at the IAS on the geometric Langlands program, David Ben-Zvi explained some things that made it clear for me that the geometric Langlands program has to do with the compactification of $\mathcal{N} = 4$ super Yang-Mills from four dimensions to two dimensions on a Riemann surface Σ. This dimensional reduction had been studied around 1995 in several papers [5], [6].

In compactifying on Σ, one needs to topologically twist the theory in the compact directions, in order to preserve supersymmetry. If we compactify, we break all supersymmetries because Σ is curved. However, it is possible to twist two of the scalars in such a manner that the compactification preserves half of the original supersymmetries (this is the case considered in [5], while the authors of [6] worked in genus one, where the twisting had no effect). As this twist preserves half of the supersymmetry of the $\mathcal{N} = 4$ super Yang-Mills theory, it leads to a two dimensional theory with eight supercharges, and gives rise to a $\mathcal{N} = (4,4)$ supersymmetric sigma model whose target space turns out to be the moduli space $\mathcal{M}_H$ of "Higgs bundles," as defined by Hitchin. $\mathcal{M}_H$ is a hyper-Kahler manifold, as we would expect for the target space of a two-dimensional sigma model with $(4,4)$ supersymmetry.

The sigma model with target has a T-duality, because of the "Hitchin fibration" of $\mathcal{M}_H$. This is a map from $\mathcal{M}_H$ to a vector space, in which the generic fiber is an abelian variety. One can carry out a fiber-wise T-duality on the toroidal fibers. Loosely speaking, what was shown in [6] and [5] was that the Montonen-Olive duality in the gauge theory corresponds to T-duality in the Hitchin moduli space. To be precise, it is actually T-duality together with a rotation of the two twisted scalars that corresponds to the $\tau \to -1/\tau$ transformation in the gauge theory. So the T-duality comes from the S-duality together with a classical symmetry of the action.

This story was pointed out in the physics literature in 1995, but was not really developed after that. However from Ben-Zvi's lecture and some clues about what mathematicians were doing, it became clear what one must do: one must consider branes in the sigma model. The two dimensional sigma model is supersymmetric, and we can consider branes in this model. Considering a brane just means that we formulate the sigma model on a half-space $\mathbb{R}^2_+$ and we impose supersymmetric boundary conditions. The bulk theory has $\mathcal{N} = (4,4)$ supersymmetry, with four supersymmetries carried by left-movers and four by right-movers, and the boundary conditions reflect the left movers into right movers, so it is not normally possible for a boundary condition to preserve more than half the supersymmetries of the bulk theory.

But there are many ways for a boundary condition to preserve half of the supersymmetry. Many of these choices are of interest in the Langlands program. The T-duality of the Hitchin moduli space has actually been used mathematically in elucidating some features of the geometric Langlands program [7]. So if mathematicians at least have a glimmering that the geometric Langlands program has to do with this particular T-duality, why have they not gone farther? One partial answer is that it is very hard to understand everything just from the two-dimensional sigma model. In particular, the existence and properties of the Wilson and 't Hooft operators are more natural from the four-dimensional gauge theory. These operators do not correspond to anything typical in two dimensional sigma models. Their existence makes this particular sigma model rather special.

N=4 super Yang-Mills and topological twisting

The starting point in our analysis, therefore, is a four dimensional topological field theory that arises from twisting four dimensional $\mathcal{N} = 4$ super Yang-Mills theory. There are several possible twists. Recall that $\mathcal{N} = 4$ super Yang-Mills theory in four dimensions can be obtained by dimensional reduction from $\mathcal{N} = 1$ super Yang-Mills theory in ten dimensions. The dimensional reduction produces an $SO(6)$ symmetry in four dimensions, often called $SO(6)_R$. The ten dimensional theory has a gauge field A_I, $I = 1, \ldots, 10$ and a fermionic superpartner χ in the adjoint representation, this being the unique supersymmetric gauge theory in ten dimensions. Dimensional reduction of the ten dimensional gauge field gives rise to six scalars ϕ_I (I runs from say 5 to 10) and a four dimensional gauge field A_μ (μ runs from 1 to 4) upon reduction to four dimensions. The $SO(6)$ symmetry acts on ϕ_I.

If we naively compactify this theory on a general four manifold, we break all supersymmetries. In order to preserve some supersymmetries, we need to topologically twist the theory. The topological twist we are interested in is easily described if we keep in mind the $SO(6)_R$ symmetry. We replace the Euclidean symmetry group $SO(4)$ of $\mathbb{R}^4$ with a group $SO(4)'$ obtained as the subgroup of $SO(4) \times SO(6)_R$ of the form $(g, R(g))$, $g \in SO(4)$, where R is a homomorphism of $SO(4)$ to $SO(6)$. For any such choice of R, we get a new Lorentz group in the four dimensional theory. We want to pick R so that one or more of the supercharges of the original theory are $SO(4)'$ singlets. If R is trivial, there are no supersymmetries that are invariant, because in that case the new $SO(4)$ is the same as the old one, and none of the supersymmetries were invariant under the old $SO(4)$ – they transform as spinors. However, for a non-trivial choice of R, there may be some invariant supersymmetries, and then it turns out that it is possible, when formulating the theory on a curved four-manifold, to preserve those supersymmetries and get a topological field theory.

There are three choices that work [8], two of which have been exploited before. The third twist has not had much application until now, though it has been investigated in the literature, originally by Marcus [9]. It can best be described by saying that the **6** of $SO(6)$ decomposes under $SO(4)'$ as $\mathbf{4}+\mathbf{1}+\mathbf{1}$. After analyzing how the spinors transform under $SO(4)'$, one learns that two of the four supercharges are $SO(4)'$ invariant, so the twisted theory preserves two supersymmetries in the four dimensional sense. The supercharges that are preserved are those that satisfy $(\gamma_\mu\gamma_\nu + \gamma_{\mu+4}\gamma_{\nu+4})\epsilon = 0$ (where μ runs from 1 to 4) and this gives us the two singlets. So on a general four manifold M, we have two supersymmetries.

This construction actually leads to a one complex parameter family of twisted topological theories, parameterized by $\mathbb{CP}^1$. The reason is simple. When performing the topological twisting, the topological symmetry can be identified with an arbitrary linear combination $Q = uQ_L + vQ_R$ of the two supercharges Q_L and Q_R that are preserved by the twisting (the L and R refer to the four dimensional chirality of the supercharges). Since the supersymmetry conditions are invariant under an overall scaling of the supercharges, (u, v) parameterize $\mathbb{CP}^1$. It is also convenient to introduce an affine coordinate $t = v/u$, allowing for t to become zero or ∞. It is important that the twisted theory has no symmetry that acts nontrivially on t. As a result, there are no trivial equivalences among this family of topological field theories, only interesting equivalences that come from dualities.

How do we identify the field configurations that are invariant under supersymmetry? We write down the general equation $\delta\chi = \gamma^{IJ}F_{IJ}\epsilon$ for how the fermions χ transform under supersymmetry. (F_{IJ} includes the four-dimensional field strength $F_{\mu\nu}$ for $\mu, \nu = 1\ldots4$, the covariant derivaties of the scalars $F_{\mu J} = D_\mu\Phi_J$ for $\mu = 1\ldots4$ and $J = 5\ldots10$, and the commutators of the scalars $F_{IJ} = [\Phi_I, \Phi_J]$ for $I, J = 5\ldots10$.) The condition for a field to be invariant under the supersymmetry generated by a particular ϵ is that this variation vanishes:

$$\gamma^{IJ}F_{IJ}\epsilon = 0 \quad I = 1, \ldots, 10. \tag{1}$$

In practice, we take $\epsilon = u\epsilon_L + v\epsilon_R$, where ϵ_L and ϵ_R are the singlets of $SO(4)'$. Imposing the condition that (1) be solved for this ϵ gives us a condition on $F_{\mu\nu}$.

The equation (1) leads when worked out more explicitly to the following two equations:

$$(F_{\mu\nu} - [\Phi, \Phi]_{\mu\nu} + t(D\Phi)_{\mu\nu})^+ = 0 \tag{2}$$

$$\left(F_{\mu\nu} - [\Phi, \Phi]_{\mu\nu} - t^{-1}(D\Phi)_{\mu\nu}\right)^- = 0 \tag{3}$$

For a two-form Ω, I write Ω^+ or Ω^- for its self-dual or anti-self-dual projection. If t is real, the above equations are elliptic PDEs. As written, they look bad for $t = 0$ or $t = \infty$, but this is misleading. By multiplying one equation by t or the other by t^{-1}, the family can be extended over $t = 0, \infty$. So we get a family of elliptic PDE's parameterized by $\mathbb{RP}^1$.

Like many systems of elliptic PDE's derived from supersymmetry, these are subject to unusual vanishing theorems. These state that (among other things) that on a closed four-manifold M with a bundle E of non-vanishing first Pontryagin class, there are no solutions to (2) and (3) except at $t = 0$ and $t = \infty$ (each of which is possible for one sign of the Pontryagin class). Moreover, when the first Pontryagin class vanishes, then any fields A, ϕ that obey these equations for some value of t obeys then for all t, and is described by a flat connection $\mathcal{A} = A + i\phi$ with values in the complexification $G_{\mathbb{C}}$ of the original gauge group.

For real t, we get a topological field theory which counts solutions of these elliptic PDEs in a suitable sense. This is analogous to an A-model in two dimensions or to Donaldson theory in four dimensions. What happens for complex t? The equations, understood for fields A, ϕ valued in the real Lie algebra of G, are overdetermined. This is much more like a B-model in two dimensions. A more careful study shows that this analogy is quite close and becomes precise at $t = \pm i$. For t not real and not equal to $\pm i$, the closest analogy is to two-dimensional models based on generalized complex geometry, which are not yet so well understood.

Upon dimensional reduction of (2) and (3) to two dimensions, we get Hitchin's equations. To be more precise, this goes as follows. We take our four-manifold to be $\mathbb{R}^2 \times \Sigma$, with Σ a two-dimensional surface. Then we prove that any translationally invariant solution of the four-dimensional equations is actually a pullback from Σ. (Translational invariance minimizes the energy or action.) The equations for the fields on Σ are equivalent to Hitchin's equations.

If instead we make a dimensional reduction to *three* dimensions, we get a sort of complex version of the Bogomoln'yi equations. These equations, which appear to be new, are important in understanding the 't Hooft operators.

So we have answered a few questions. I have told you at least partly how the theory depends on t, and I have told you how the Hitchin equations come in. The next question to answer is how does the S-duality symmetry $\tau \to -1/\tau$ act on t. One can answer this question by first determining how S-duality acts on the four-dimensional supersymmetries. The result is that under S-duality (when $\mathrm{Re}(\tau) = 0$, for simplicity), the supercharges transform as:

$$\epsilon \to \frac{1 + i\bar{\Gamma}}{\sqrt{2}}\epsilon, \tag{4}$$

where $\bar{\Gamma}$ measures the four-dimensional chirality. So $\epsilon_L \to i^{1/2}\epsilon_L$, $\epsilon_R \to i^{-1/2}\epsilon_R$, and $t \to -it$.

Hence, in particular, S-duality takes the A-model at $t = \pm 1$ into the B-model at $\mp i$, while $t = 0$ and $t = \infty$ are fixed points. For the geometric Langlands program, the most important values of t are $t = \pm 1$ and $t = \pm i$.

How do the Wilson and 't Hooft operators depend on t? To analyze 't Hooft and Wilson loops, it is better to combine (2) and (3) into the equation:

$$F + \frac{(t - t^{-1})}{2} D\phi + \frac{(t + t^{-1})}{2} * D\phi - [\phi, \phi] = 0 \tag{5}$$

To get a hint about 't Hooft and Wilson loops, we will look for abelian, time independent solutions, where we drop the last term in (5). If $t = 1$ we get $F = - * D\phi$, and this is the equation for a Bogomoln'yi monopole. In particular, the field created by a supersymmetric 't Hooft operator obeys this equation. That is one indication of the fact that these operators exist at this value of t. By duality, it follows that the $t = i$ theory should allow Wilson loops. By checking (5) substituting $t = i$ we do indeed find solutions corresponding to Wilson operators where $\vec{E} = -iD\phi$. These are nothing but the supersymmetric Wilson operators that are familiar in the context of AdS/CFT. The standard Wilson loop operator $W(C) \sim \mathrm{Tr}\,\mathrm{P}\exp\int A$ is not supersymmetric. However, as first pointed out by Maldacena [10] in the AdS/CFT correspondence, we can make a supersymmetric Wilson operator which schematically looks like $W(C) \sim \mathrm{Tr}\,\mathrm{P}\ \exp\int(A + i\phi)$.

Now, when we compactify on a Riemann surface Σ, we get a sigma model on $\mathbb{R}^2$. In order to consider branes in this sigma model, we work on the half-space $\mathbb{R}^2_+$, which from the point of view of the gauge theory means working on $\mathbb{R}^2_+ \times \Sigma$. What is really special about this sigma model in comparison to the standard sigma models is that this theory has 't Hooft and Wilson loop operators. All sigma models have point-like operators analogous to those that arise from wrapping Wilson loops on Σ. The fun really comes when we consider Wilson loops extended along $\mathbb{R}^2_+$. Consider a 't Hooft or Wilson operator near the boundary of the half-plane. What is its physical interpretation? If we look at it from afar, it just looks like part of the boundary condition. In fact we can let the perturbed theory run under RG flow and the resulting infra-red fixed point would possess a conformally invariant boundary condition that depends on the Wilson-'t Hooft operator (as well as the actual, microscopic boundary condition that we started with). Let us denote by $T_{R'}$ and W_R the 't Hooft and Wilson operators, where R is a representation of G and R' is a representation of the dual group ${}^LG$. These operators map one brane to another brane, changing boundary conditions, whereas local operators in the quantum field theory map a state in the quantum field theory with a given brane to another state with the same brane.

The geometric Langlands program is all about operators that map branes to different branes.

Every now and then we might find what I will call an eigenbrane (corresponding to what is called an eigensheaf in the geometric Langlands program). The most naive notion of an eigenbrane for an operator W_R would be a brane B such that $W_R B = B$, in other words, we get back the same brane when we bring the operator to the boundary. We should allow a more general notion, that $W_R B = B \otimes V_R$ where V_R is a fixed vector space (which may depend on R).

What do I mean by tensoring a brane with a vector space V? In geometry, a brane is constructed using a vector bundle, called the Chan-Paton bundle, on a manifold (which usually is a submanifold of the target space of the sigma model). Tensoring with a vector space V means we tensor the Chan-Paton bundle with V. This also makes sense in abstract conformal field theory. An electric eigenbrane gives back itself, tensored with some V_R, when multiplied by W_R. A magnetic eigenbrane likewise gives back itself, tensored with some $V_{R'}$, when multiplied by $T_{R'}$.

How are we going to find eigenbranes? The electric eigenbranes are actually trivial to construct in this picture and the goal of the geometric Langlands program is to construct magnetic eigenbranes. The electric eigenbranes are just zerobranes supported at a point in the Hitchin moduli space $\mathcal{M}_H$. The dual of an electric eigenbrane is going to be a magnetic eigenbrane, and we can get the dual magnetic brane by performing T-duality on the fibers of $\mathcal{M}_H$. From standard T-duality, we know what is going to be the dual of a zerobrane: it is a brane that wraps on a fiber of the Hitchin fibration and is endowed with a flat line bundle.

It is easy to see that the zerobrane is an electric eigenbrane, so the brane wrapped on a fiber will be a magnetic eigenbrane. This can also be demonstrated explicitly, using the properties of the 't Hooft operators. The 't Hooft operators can be shown to correspond to the Hecke operators of the geometric Langlands program.

But there is something very seriously missing in what I have told you so far. In the geometric Langlands program, the Hecke eigensheaves – which for us are the magnetic eigenbranes – are supposed to be $\mathcal{D}$-modules, that is, modules for a certain algebra of differential operators. How does this come about here? There is a trick, which involves the existence of an unusual sort of A-brane. (A-branes come in because our four-dimensional gauge theory at $t = 1$ reduces to a two-dimensional A-model.)

In two-dimensional sigma models, the familiar A-branes are flat bundles over Lagrangian submanifolds of the target space. On a generic Calabi-Yau three-fold, for example, these are all the A-branes. But as shown by Kapustin and Orlov [11], there are under suitable conditions more A-branes. These are the coisotropic A-branes, which are more than middle-dimensional and carry a non-flat Chan-Paton or gauge bundle. I will give an example of a coisotropic brane instead of giving its definition. An example of such a brane in our context is a brane wrapping all of $\mathcal{M}_H$, and endowed with a $U(1)$ bundle whose curvature is a certain linear combination of the symplectic forms of $\mathcal{M}_H$. If one does this just right, one gets a coisotropic A-brane. If α is this brane, then the supersymmetric $\alpha - \alpha$ strings turn out to correspond to differential operators on $\mathcal{M}$, the moduli space of G-bundles. (This statement depends among other things on Hitchin's result that on the complement of a set of high codimension, $\mathcal{M}_H$ is the cotangent bundle of $\mathcal{M}$.)

The strategy is that we prove the Hecke eigensheaf property of an A-brane and then we interpret the A-branes as $\mathcal{D}$-modules, where the last step depends on the existence of the very special coisotropic brane α that I just described. It all works out very nicely.

Finally, let us summarize how this quantum field theory story is related to the usual picture. The standard statement of the geometric Langlands correspondence is that flat ${}^LG_{\mathbb{C}}$ bundles on our two-dimensional surface Σ are in natural one-to-one correspondence to $\mathcal{D}$-modules on the moduli space $\mathcal{M}$ of holomorphic G-bundles. A flat ${}^LG_{\mathbb{C}}$ bundle on Σ gives us a point in $\mathcal{M}_H$, corresponding to an electric eigenbrane. Its S-dual is a magnetic eigenbrane, and the existence of the coisotropic brane α means that magnetic eigenbranes actually give us $\mathcal{D}$-modules. In this way, the standard statement of the geometric Langlands correspondence arises from electric-magnetic duality of gauge theory.

References

[1] E. Frenkel, "Recent Advances In The Langlands Program," math.AG/0303074.

[2] P. Goddard, J. Nuyts and D. I. Olive, "Gauge Theories And Magnetic Charge," Nucl. Phys. B **125**, 1 (1977).

[3] C. Montonen and D. I. Olive, "Magnetic Monopoles As Gauge Particles?," Phys. Lett. B **72**, 117 (1977).

[4] A. Kapustin, "Wilson-'t Hooft Operators in Four-Dimensional Gauge Theories and S-Duality," arXiv:hep-th/0501015.

[5] M. Bershadsky, A. Johansen, V. Sadov and C. Vafa, "Topological Reduction of 4-d SYM to 2-d Sigma Models," Nucl. Phys. B **448**, 166 (1995) [arXiv:hep-th/9501096].

[6] J. A. Harvey, G. W. Moore and A. Strominger, "Reducing S Duality to T Duality," Phys. Rev. D **52**, 7161 (1995) [arXiv:hep-th/9501022].

[7] T. Hausel and M. Thaddeus, "Mirror Symmetry, Langlands Duality, and the Hitchin System," math.AG/0205236.

[8] C. Vafa and E. Witten, "A Strong Coupling Test of S Duality," Nucl. Phys. B **431**, 3 (1994) [arXiv:hep-th/9408074].

[9] N. Marcus, "The Other Topological Twisting Of $\mathcal{N} = 4$ Yang-Mills," Nucl. Phys. **B452** (1995) 331, hep-th/9506002.

[10] J. M. Maldacena, "Wilson Loops in Large N Field Theories," Phys. Rev. Lett. **80**, 4859 (1998) [arXiv:hep-th/9803002].

[11] A. Kapustin and D. Orlov, "Remarks on A-Branes, Mirror Symmetry, and the Fukaya Category," J. Geom. Phys. **48**, (2003b) [arXiv:hep-th/0109098].