GAUGE THEORY IN THREE AND FOUR DIMENSIONS

PAUL NORBURY

ABSTRACT. Following Kronheimer and Mrowka, we prove minimal genus bounds for surfaces embedded in manifolds of dimensions three and four.

INTRODUCTION

These notes are written for the December 1998 short courses at Melbourne University. The course consists of eight one hour lectures. The level is aimed at students who have just completed an undergraduate degree.

The aim of this course is to introduce gauge theory techniques into the study of low-dimensional topology. Given the restricted time, the background will be brief and we will satisfy ourselves with just a few applications. The main sources for this course will be John Morgan's book "The Seiberg-Witten invariants and applications to the topology of smooth four-manifolds" and John Moore's book "Lectures on Seiberg-Witten invariants" for the background and Kronheimer's paper "Embedded surfaces and gauge theory in three and four dimensions" (http://www.math.harvard.edu/ kronheim/) for the application.

The general philosophy will be to start with a topological problem and describe a topological approach to its solution. By a topological approach, I mean that using a topological object with very interesting properties, we can tackle the problem. Gauge theory will come in when trying to prove the existence of such a freakish topological object.

1. MINIMAL GENUS

Let $K \subset S^3$ be a knot and consider $Y = S^3 - N(K)$, the manifold obtained by removing a neighbourhood of the knot from S^3. A Seifert surface for K is an embedded surface $\Sigma \hookrightarrow Y$ such that $\partial\Sigma \subset \partial Y$ is given by a longitude. In fact, this property uniquely determines the longitude.

Example Let K be the unknot. Then $Y = S^3 - N(K)$ gives the solid torus. It is easy to see a spanning disk for K. In fact we can add handles to get many surfaces with the same boundary. Why can't we get any other curve on the boundary torus this way? Any two such surfaces will intersect in 1-dimensional manifolds once we ensure the intersection is transversal. The intersection will actually be an oriented 1-manifold with boundary. A $(1,n)$ curve will intersect the $(0,1)$ curve exactly once so there can be only one boundary component, which is impossible. More generally, the intersection points of two curves in the torus will induce the same orientation.

The proof for the general case is identical. The following issue naturally arises.

Question 1.1. *What is the minimal genus of a Seifert surface for a given knot K?*

Example The trefoil has a genus 1 Seifert surface. Since it doesn't bound an embedded disk its genus is 1. Why doesn't it bound an embedded disk? Any immersed disk with embedded boundary that bounds the trefoil must intersect itself an odd number of times which is not satisfied by an embedded disk.

There is a more general question we can ask.

Question 1.2. *What is the minimal genus of an embedded surface representing a given homology class in a three-manifold Y?*

By homology class, we mean to start with an embedded surface and move it around the manifold, perhaps so that it is not embedded at some places, and, if we can, squeezing holes until they disappear. We would like to know what is the minimal genus of an embedded surface in the family. (Equivalently, we might start with an immersed surface and move it as described and ask if we can ever get an embedded surface in that family.)

Here is a wishful approach to this problem. Consider a collection of oriented embedded circles $S \subset Y$. We can arrange that these intersect any embedded surface Σ transversally by moving Σ or S slightly. We can count the points of intersection (with sign) in $\Sigma \cap S$. Imagine if we could find a collection of circles S such that the number of points in the intersection $\Sigma \cap S$ is a lower bound for the genus of Σ for any Σ. Such an intersection only depends on the homology class of Σ so it could give an answer to the question.

A more sophisticated way of expressing this approach is to ask if there is a line bundle L over Y such that the restriction of L to any embedded surface Σ is "less twisted" than the tangent bundle of Σ.

Gauge theory, via the Seiberg-Witten equations, supplies us with exactly such a freakish set of circles. The lower bounds that such a collection S might give can sometimes be quite bad for a particular surface Σ. Fortunately, gauge theory gives us a finite number of sets S_i each with the property described above. For any embedded surface Σ, we can use the maximum of $\Sigma \cdot S_i$ over the i and this can give a good lower bound.

This should look similar to the way that foliations give lower bounds to the genus of an embedded surface. The theory of foliations ties in quite closely with the gauge theory. We will get to that closer to the end of the course.

2. Differential geometry

Bundles, connections and the Dirac operator. The easy version.

We will think of a bundle as a sub-bundle of a trivial vector bundle and connections as projections on to the sub-bundle. All manifolds here will be oriented.

A manifold is given by $Y \subset \mathbb{R}^N$. In fact, this induces a Riemannian metric. Consider the tangent bundle TY of Y consisting of the tangent spaces of Y inside $\mathbb{R}^N$. Let ν_Y be the normal bundle of Y. Notice that $TY \oplus \nu_y \cong Y \times \mathbb{R}^N$. In general, we can embed the tangent bundle $TY \subset Y \times \mathbb{R}^M$ into other trivial bundles, and there are many embeddings into each trivial bundle. (So the canonical embedding is a bit misleading.) More generally, we can define a bundle over Y to be any smoothly varying family of subspaces $E_x \subset \{x\} \times R^M$ for $x \in Y$. A section s of a bundle is a smooth map $s : Y \to \mathbb{R}^M$ such that $s(x) \in E_x$ for all x. We also call a section of the tangent bundle a vector field. The space of sections of a bundle E is notated by $\Gamma(E)$.

Example Consider $S^2 \subset \mathbb{R}^3$ given by $S^2 = \{x \in \mathbb{R}^3 | \; |x| = 1\}$ and take its tangent bundle

$$TS^2 = \{(x, v) \in S^2 \times \mathbb{R}^3 | x \cdot v = 0\}.$$

Here's a couple of sections:

$$s(x_1, x_2, x_3) = (-x_1 x_3, -x_2 x_3, x_1^2 + x_2^2), \;\; r(x_1, x_2, x_3) = (-x_2, x_1, 0).$$

Often we would like to understand how a section changes as we move across the manifold. Notice from the previous example that this notion of change is ambiguous. It seems that as we travel around the equator the vectors do not change and as we travel north they change in a way that is independent of where we start from on the equator. But that would mean that they don't change near the north pole and it is clear that they do. Perhaps I went about it the wrong way and I should start with the north pole and in fact the whole northern hemisphere. Comparing the two hemispheres, we still get a contradiction. Surely this has to do with the "twisting" (i.e. non-triviality) of the tangent bundle?

A connection allows us to differentiate sections. It is given by a projection $P_x : \mathbb{R}^M \to E_x$. Why does this help? We can differentiate vectors in a trivial bundle, so for $v \in T_xY$ and $s \in \Gamma(E)$, we have $(\partial_v s)(x) \in \mathbb{R}^M$. Define $\partial_v^A s = P_x \partial_v s$. If ξ is a vector field, then $\partial_\xi^A s \in \Gamma(E)$ as desired. We use A to denote the connection (even though P would suffice). Another way to express a connection is $\partial_v^A = \partial_v - \partial_v P$ where the extra term acts as a zero order operator.

In the example above, there is an obvious family of projections from $\mathbb{R}^3$ to T_xS^2, given by $P_x v = v - (x \cdot v)x$. Look at the point $(1,0,0)$ on the equator. Then we have

$$\partial^A_{(0,1,0)} s = P_{(1,0,0)}(0,0,0) = (0,0,0), \partial^A_{(0,0,1)} s = P_{(1,0,0)}(-1,0,0) = (0,0,0).$$

At a general point (except for a pole), let v be the unit vector parallel to the equator and let w be the unit vector pointing north. Then

$$\partial_v^A s = P_{(x_1.x_2,x_3)} Df(v) = x_3 v,\ \ \partial_w^A s = P_{(x_1.x_2,x_3)} Df(w) = -x_3 w.$$

So the vectors behave as expected as we travel north but they seem to twist as we traverse the globe. Notice that this contradicts an argument we gave above. What we see here is that covariant derivatives don't commute. This brings us to the curvature of a connection.

Define $F_A = [dP, dP]$ where we mean that for $v, w \in T_xY$, $F_A(v,w) = [\partial_v P, \partial_w P]$. The curvature acts on sections by (natrix) multiplication. That's quite remarkable since the curvature arises from the commutator $F_A(v,w) = [\partial_v^A, \partial_w^A]$ so we might expect it to be a second order operator instead of a zero order operator. What we see is that although the covariant derivative is not commutative, it is somewhere in between.

An important object in differential geometry is a differential form. Let's start with usual integration. We can make sense of $\int_a^b f(t)dt$ ofr any function f defined on $[a,b]$. This can be thought of as integration over a manifold. We would like to integrate over submanifolds. Consider a curve in the plane. If we have a function in the plane can we integrate it over the curve? Really, integration needs $\mathbb{R}^n$. So, parametrise the curve. Notice, however, that the integral depends on the parametrisation:

$$\int f(g(s))dg = \int f(s)g'(s)ds \neq \int f(s)ds$$

for a change of parametrisation g. It ends up that we cannot integrate functions over submanifolds but we can integrate differential forms like $f(s)ds = f(g)dg/g'(s)$. In three dimensions, a 1-form is given by three functions

$$f_1(x_1,x_2,x_3)dx_1 + f_2(x_1,x_2,x_3)dx_2 + f_1(x_1,x_2,x_3)dx_3.$$

Similarly, a 2-form (in three dimensions) is given by three functions

$$f_{12}(x_1,x_2,x_3)dx_1 \wedge dx_2 + f_{13}(x_1,x_2,x_3)dx_1 \wedge dx_3 + f_{23}(x_1,x_2,x_3)dx_2 \wedge dx_3$$

and it can be integrated over surfaces in the manifold. Notice that a 2-form becomes a well-defined function when restricted to a surface. Thus, another way to think of a 2-form on a three-manifold is as a function that makes sense when you specify a surface (or even a plane of tangent vectors at a point). The expression $\wedge$ is used to show that the definition is orientation sensitive which is necessary in changes of coordinates.

The curvature defined above is a 2-form..

Perhaps we will leave the Dirac operator to another lecture. Instead, let's look at the global topology that a connection captures, independently of the connection!

Let L be a complex line bundle over Y. So $L_x \subset Y \times \mathbb{C}^N$. Let s be a section of L. Then the zero set of s is a one-dimensional (oriented) submanifold $S \subset Y$, once we choose s appropriately—transversal to the zero section. A one-dimensional submanifold must be a collection of circles in Y. Take any (oriented) surface $\Sigma \subset Y$. The intersection of S and Σ gives a number by counting the points in the intersection with appropriate sign—+1 if S points in the positive normal direction of Σ and -1 if S points in the negative normal direction of Σ. We know that this number only depends on the homology classes of S and Σ. What is interesting is that the homology class of S is independent of the section s so depends only on the line bundle L. It is probably easiest to see this if we restrict L to the surface Σ. The zero set of any section is an isolated set of points in Σ (transversality). Any two sections s_1, s_2 can be joined by a path of sections $s = (1-t)s_1 + ts_2$. Since $s : \Sigma \times I \to L$ is a section its zero set is a 1-manifold (perhaps we perturb the homotopy to get transversality) with boundary given by the zero sets of s_1 and s_2. This gives an oriented cobordism between the two sets so they are counted the same way.

The set S can be thought of as an element of the second cohomology group $H^2(Y)$. Is there a convenient 2-form representing this class? Yes—the curvature of the connection gives us what we want. First notice that if A_1 and A_2 are two connections, then $A_1 - A_2 = a$ is a 1-form and $F_{A_2} = F_{A_1} + da$ os the cohomology class is

These are Dirac magnetic monopoles.

I would like to mention that bundles can be defined intrinsically where we use a structure group to keep track of the twisting rather than an ambient trivial bundle that has the ability to see twisting.

3. Spinors and vectors

This will make the second of the Seiberg-Witten equations friendlier. Also introduce the Dirac operator and vectors and Hermitian matrices. The advantage of doing it this way and then referring to the spin representation is that are forced to see the matrices rather than settle for mere existence. (More the physicists way.)

Simply understand the isomorphism $\mathbb{CP}^1 \cong \{v \in \mathbb{R}^3 |\ |v| = 1\}$.

The two-sphere can be realised as the unit vectors in $\mathbb{R}^3$ or, equivalently, the oriented projective space of lines in $\mathbb{R}^3$ so we can label this realisation $\tilde{\mathbb{RP}}^2$. We can also realise the two-sphere as the projective space $\mathbb{CP}^1$ of complex lines in $\mathbb{C}^2$. These two realisations respectively make the actions of $SO(3)$ and $SU(2)$ manifest. Is there a nice way to associate to a real vector in $\mathbb{R}^3$ its complex vector in $\mathbb{C}^2$ and such that the actions of $SO(3)$ and $SU(2)$ are compatible?

Take a non-zero vector $s = (z_1, z_2) \in \mathbb{C}^2$. So s represents a point in $\mathbb{CP}^1$. From s we can get a vector

$$v = \sigma(s) = ((|z_1|^2 - |z_2|^2, 2z_1\bar{z}_2) \in \mathbb{R}^3$$

where we have used $\mathbb{R}^3 \cong \mathbb{R}\times\mathbb{C}$. Then $|v| = |s|^2$ and we claim that this induces the desired map $\tilde{\sigma} : \mathbb{CP}^1 \to \tilde{\mathbb{RP}}^2$. Really there isn't a canonical way to associate a point of $\mathbb{CP}^1$ with a point of $\mathbb{RP}^2$. So after making a choice (in this case the isomorphism $\mathbb{R}^3 \cong \mathbb{R} \times \mathbb{C}$) we can ask at least that the $SU(2)$ and $SO(3)$ actions are compatible.

The map from $SU(2)$ to $SO(3)$

$$\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \mapsto \begin{pmatrix} 1 & re(a\bar{b} & -im(a\bar{b}) \\ -re(a\bar{b}) & * & * \\ -im(ab) & * & * \end{pmatrix}$$

shows how the respective actions on $\mathbb{CP}^1$ and $\tilde{\mathbb{RP}}^2$ agree.

Some further features of this map:

- The map $\sigma : \mathbb{C}^2 \to \mathbb{R}^3$ induces the Hopf fibration $S^3 \to S^2$ when restricted to any sphere of constant norm vectors in $\mathbb{C}^2$.
- A slicker way to see σ is to use the fact that we know that the action of $SU(2)$ on its Lie algebra gives the standard $SO(3)$ action. Here we are identifying the Lie algebra $su(2)$ with $\mathbb{R}^3$. This identification is really encoding the Clifford action of vectors on spinors. Then notice that $\sigma(s) = s\bar{s}^T - \frac{1}{2}|s|^2$. For $u \in SU(2)$, we get $\sigma(us) = u\sigma(s)baru^T = u\sigma(s)u^{-1}$, the adjoint action of

$SU(2)$ on its Lie algebra (times i). Notice that s is an eigenvector of $\sigma(s)$. In fact this property determines $\sigma(s)$ uniquely up to oriented scaling.

- We can say that $\mathbb{CP}^1$ are isomorphic $\tilde{\mathbb{RP}}^2$ since an eigenspace of a Hermitian matrix is (almost) the same as a Hermitian matrix.
- Or, we say that "two" spinors is in some sense the same as a vector. This is the twistor theory of Penrose.

Back to connections on a bundle on a three-manifold and the Dirac operator. Define the Dirac operator by

$$Ds = \sum_j \Gamma(e_j)\nabla_{e_j} s$$

and more generally

$$D_A s = \sum_j \Gamma(e_j)\nabla^A_{e_j} s.$$

Here $\{e_1, e_2, e_3\}$ is an orthonormal set of vectors in the tangent bundle of Y and $\Gamma(e_j)$ is the Hermitian matrix associated to the vector e_j.

Look at the Dirac operator in two dimensions to get a feel for this.

$$\begin{aligned} Ds &= \Gamma(e_1)\nabla_{e_1} s + \Gamma(e_2)\nabla_{e_2} s \\ &= \begin{pmatrix} \partial_x & \partial_y \\ -\partial_y & \partial_x \end{pmatrix} s \end{aligned}$$

Notice that the algebra of matrices agrees with the algebra of complex numbers. The Dirac equation is the Cauchy-Riemann equation. More generally, solutions of the twisted Dirac operator correspond to holomorphic sections of a (complex) line bundle, and moreover we can use the Dirac operator to define the holomorphic structure on a complex line bundle.

4. The Seiberg-Witten equations

The Seiberg-Witten equations are given by

$$D_A\Phi = 0 \tag{1}$$

$$\rho(F_A) = \sigma(\Phi) \tag{2}$$

where A is a connection on L, Φ is a section of $S \otimes L$, $D_A : \Gamma(S\otimes L) \to \Gamma(S \otimes L)$ is the Dirac operator twisted by the connection A, F_A is the curvature 2-form of A and $\rho(F_A)$ associates a Hermitian matrix valued function to the curvature, and $\sigma(\Phi)\Phi\bar{\Phi}^T - \frac{1}{2}|\Phi|^2 I$ is described in the previous section.

A Weitzenbock formula for the Dirac operator helps to justify these equations. Before describing this, we will quickly define the scalar curvature of a metric. For a surface Σ, the scalar curvature can be defined in many equivalent ways:

(i) Say $\Sigma \subset \mathbb{R}^3$. At each point $x \in \Sigma$ we can find local coordinates (x_1, x_2, x_3) in $\mathbb{R}^3$ around x so that Σ is locally given by $x_3 = f(x_1, x_2) = \sum_{i,j} Q_{ij} x_i x_j + ...$ where there is no constant or degree one term and we have included the second order term and omitted the higher order terms. Then $Q_{i,j}$ is a symmetric matrix that is almost an invariant of Σ at the point x. Actually, the invariant quantities are tr$Q = 2H$, the mean curvature, and det$Q = s$, the scalar, or Gaussian, curvature.

(ii) The Levi-Civita connection on $T\Sigma$ has curvature $F_{LC} = s\omega$, a multiple of the area form ω. The multiplier is the scalar curvature.

(iii) Take a geodesic triangle on Σ and consider $(\alpha+\beta+\gamma-\pi)/\text{area} \to s$ as the area goes to 0.

In three dimensions, the scalar curvature is twice the sum of the scalar curvatures of a set of three planes generated by three orthogonal vectors, $s = 2\sum_{i,j} s_{i,j}$.

Now, consider

$$D_A D_A = \sum_{i,j} \Gamma(\partial_{x_i})\partial^A_{x_i}\Gamma(\partial_{x_j})\partial^A_{x_j} = \sum_i \Gamma(\partial_{x_i})^2 \partial^A_{x_i}\partial^A_{x_i} + ...$$

where the missing terms are $s + \rho(F_A)$. From this we get

$$\begin{aligned} \Delta|\Phi|^2 &= \sum \partial^2_{x_i}|\Phi|^2 \\ &= \sum 2(\partial^A_{x_i}\partial^A_{x_i}\Phi, \Phi) + \sum 2(\partial^A_{x_i}\Phi, \partial^A_{x_i}\Phi) \\ &\geq \frac{s}{2}|\Phi|^2 + (\rho(F_A)\Phi, \Phi) \\ &\geq \frac{s}{2}|\Phi|^2 + \frac{1}{2}|\Phi|^4 \end{aligned}$$

which allows us to deduce that $|\Phi|^2 \leq -s$. Thus $F_A = 0$ or $|F_A| \leq -s$ or $|F_A| \leq -s_-$ where s_- is the negative part of s and is 0 otherwise. We have,

$$|\int_\Sigma F_A| \leq \int_\Sigma |F_A| d(\text{area}) \leq -\int_\Sigma s_- d(\text{area})$$

and this last expression would be helpful if we could relate the scalar curvature of Y at $x \in \Sigma$ to the scalar curvature of Σ there and if the scalar curvature on Σ is always non-positive.

We can change the metric so that locally it gives a product $\Sigma \times I$ in a neighbourhood of Σ of a non-positively curved metric on Sigma and the flat metric in the normal direction. But perhaps the Seiberg-Witten equations don't have a solution for this quite special metric. In the next section we will study an invariant produced from the Seiberg-Witten equations that is independent of the metric and whose non-vanishing for a particular metric ensures a solution for that metric.

5. Properties of the Seiberg-Witten equations and examples

We have seen from the previous section that not only do we want a solution of the Seiberg-Witten equations but we want a solution to survive as we vary the metric.

In this section we will follow Kronheimer "Embedded surfaces and gauge theory in three and four dimensions" quite closely, adding background material.

We want to use the analogy with the critical points of a function on a compact manifold.

As for the finite dimensional case we need to ensure that we have

(i) isolated critical points—we may have to perturb the equations;

(ii) finitely many critical points—compactness of the space of solutions;

(iii) as usual, we will need to count with sign if we want the sum to be independent of the metric—spectral flow.

To isolate critical points we use Sard's theorem on a Banach space that tells us that if a family is transverse then most points in the family are too.

The space of solutions is compact because we know how connections can blow up (or bubble) and that a uniform bound will prevent this. A sequence of spinors will converge since they are uniformly bounded.

Spectral flow is best seen via an ordinary differential equation. This gives the spectral flow in terms of the index of a Fredholm operator.

6. Final comments and four dimensions

The Alexander polynomial of a knot K gives the Seiberg-Witten invariants of Y which is zero surgery on $S^3 - N(K)$. This gives a pretty ordinary bound on the genus of a Seifert surface. In fact, $g(\Sigma) \geq r$ where r is the degree of the Alxander polynomial.

It is necessary to use a deeper theory, namely Floer homology, which again gives an invariant of the Seiberg-Wittens equations that is independent of the metric. The existemce of taut foliations and associated contact structures is used, via symplectic geometry in four dimensions to get these sturdier solutions.

Symplectic geometry, Thom conjecture, unknotting number. Does our topological viewpoint help in four dimensions? And property P?