Smooth four-manifolds

T. Perutz

Cambridge University Part III course,
Lent Term 2006

www.dpmms.cam.ac.uk/∼tp214/PartIII

References

[1] M. Atiyah, *Collected works, Vol. 5: Gauge theories,* Oxford University Press, 1988.

The geometry of Yang-Mills fields gives a gentle account of the theory (pre-Donaldson). *Self-duality in four-dimensional Riemannian geometry* (with Hitchin and Singer) is a reference for the dimension formula for the instanton moduli space.

[2] S. Donaldson, *An application of gauge theory to four-dimensional topology,* J. Differential Geom. 18 (1983), no. 2, 279–315.

Proves the diagonalisability theorem. Briliant and concise.

[3] S. Donaldson and P. Kronheimer, *The geometry of four-manifolds,* Oxford Mathematical Monographs, Oxford University Press, 1990.

We'll cover parts of chapters 1, 2, 4 and 8. Chapter 7 is also relevant.

[4] D. Freed and K. Uhlenbeck, *Instantons and four-manifolds,* second edition, Mathematical Sciences Research Institute Publications, Springer-Verlag, 1991.

Expanded account of [2]. I'd rather you didn't recall the library copy until the end of term, as I'm using it to help write the course!

[5] J. Milnor and J. Stasheff, *Characteristic classes,* Annals of Mathematics Studies, No. 76, Princeton University Press, 1974.

Further reading on vector bundles.

[6] A. Scorpan, *The wild world of four-manifolds*, American Mathematical Society, 2005.

Survey of the state of the art.

[7] R. Wells, *Differential analysis on complex manifolds,* second edition, Graduate Texts in Mathematics, No. 65, Springer-Verlag, 1980.

Further reading on Sobolev spaces and elliptic operators.

1 Why four-manifolds?

Four-dimensional manifolds behave in a different way to manifolds of any other dimension. The difference lies is in the behaviour of *smooth structures*. We know enough about these to be sure that they really do behave differently in dimension four, but there are basic questions which we can answer in every dimension *except* four. For example:

How many non-diffeomorphic smooth structures are there on the n-sphere S^n?

Remember that to give a smooth structure (sometimes called a *differentiable structure*) is to give a preferred class of coordinate charts for which the transition functions are all infinitely differentiable. Two such classes are considered identical if their union also has infinitely differentiable transition functions.

If $\Phi\colon M_1 \to M_2$ is a homeomorphism between manifolds, and M_2 has a given smooth structure, then M_1 has an induced smooth structure. In this way, the group of self-homeomorphisms of a manifold M acts (on the right) on the set $\mathcal{S}(M)$ of smooth structures. So, starting from an initial smooth structure, we can produce a great many new ones which are *distinct* from it (the identity map is not a diffeomorphism) but *diffeomorphic* to it. The question is to understand smooth structures up to diffeomorphism—that is, to study the set of orbits in $\mathcal{S}(M)$ under the action of the homeomorphism group.[1]

There is a formula for the number $s(n)$ of non-diffeomorphic smooth structures on S^n which is valid in every dimension except 4. In low dimensions, these numbers are

n	1	2	3	4	5	6	7	8	9	10	11	12
$s(n)$	1	1	1	?	1	1	28	2	8	6	992	1

One interesting feature is that the numbers which appear are *finite*. This is a general phenomenon: there are only finitely many non-diffeomorphic smooth structures on a fixed compact topological manifold of dimension $\neq 4$. However, the quartic surface

$$\{(z_1 : z_2 : z_3 : z_4) \in \mathbb{C}P^3 : z_1^4 + z_2^4 + z_3^4 + z_4^4 = 0\}$$

is a four-manifold which admits a countable infinity of different smooth structures.

Open question: Is true that any topological four-manifold which has at least one smooth structure has infinitely many non-diffeomorphic smooth structures?

[1] Let us dispose of one point right away: the answer would be the same if one used some weaker notion of differentiability. For instance, every manifold with C^1-differentiable structure is C^1-diffeomorphic to a unique smooth (i.e. C^∞) structure.

A good (if in general rather abstract) answer to the following question was established in the 1960s—but the answer is only valid in dimensions higher than four:

> Let M be a compact topological manifold. How can we tell when M admits at least one smooth structure?

When the dimension is ≤ 3, it is known for different reasons that M always admits a unique smooth structure. We should acknowledge at this point that three-manifold theory has its own subtleties, but they have a different origin to those in four dimensions. A central problem is to understand to what extent the fundamental group of a three-manifold controls its global topology.

1.1 Existence of smooth structures in higher dimensions—sketch

It seems appropriate to say a little about what does happen in higher dimensions, though this inevitably brings in some quite sophisticated algebraic topology. If some of the terms are unfamiliar, don't worry—this material is included only to put the course into context, and we won't touch it again.[2]

Let's start with a space M which is a candidate to admit smooth structures. It will simplify matters if we suppose M is simply connected. Rather than assuming M is a compact topological n-manifold, it is better to suppose that M is a space which has the right sort of algebraic topology. So we just assume that M is a simply connected space, built out of cells of dimension $\leq n$, and that there is a 'fundamental class $[M] \in H_n(M;\mathbb{Z})$ which sets up a Poincaré duality isomorphism $c \mapsto c \cap [M]$ from $H^*(M;\mathbb{Z})$ to $H_{n-*}(M;\mathbb{Z})$. We now look for a smooth manifold M' and a homotopy equivalence $M' \simeq M$.

We can embed M in some Euclidean space $\mathbb{R}^k$, where $k \gg 0$. There is then a 'regular neighbourhood' N of M in $\mathbb{R}^k$ which retracts onto M, i.e. there is a continuous map $r \colon N \to M$ which is the identity map on N itself. It turns out that, because of the Poincaré duality property, we can arrange that the map

$$p = r|_{\partial N} : \partial N \to M$$

is a *spherical fibration*. This means that every fibre $p^{-1}(x)$ is homotopy-equivalent to the sphere S^{k-n-1}, and that p is, in a precise sense, a 'coherent family' of such homotopy-spheres.

When k is very large, the spherical fibration $p \colon \partial N \to M$ (considered up to homotopy equivalence for maps between spaces) may be regarded as an invariant of the homotopy type of M. It is called the **Spivak normal invariant**.

The Spivak normal invariant of a smooth manifold M' is special. In that case, we can embed M' smoothly in $\mathbb{R}^k$. The 'normal vector bundle' $\nu_{M'}$ is the space of pairs (x, v), where $x \in X$ and $v \in \mathbb{R}^k$ is orthogonal to the tangent space

[2] If you follow Dr Feldman's course on cobordism you'll be well-equipped to understand the theory sketched here.

T_xM'. A neighbourhood N of M' can be identified with the disc-subbundle $\{(x,v) \in \nu_{M'} : |v| \leq \epsilon\}$, for some $\epsilon > 0$. The retraction r sends (x,v) to x. Thus the spherical fibration $\partial N \to M'$ has the special feature that it is (in a canonical way) a sphere sub-bundle in a *vector bundle* over M.

A property that our candidate space M must therefore possess if it is to be homotopy equivalent to a smooth manifold is that *its Spivak normal invariant must arise from a vector bundle.* This is a property that can be studied using algebraic topology. Examples are known of topological manifolds for which the Spivak normal invariant does not arise from any vector bundle—hence the manifold has no smooth structure.

If the Spivak normal invariant arises from a vector bundle then we can at least find a smooth compact n-manifold M' and a map $f\colon M' \to M$ such that on homology we have $f_*[M'] = [M]$. The next step is to try to modify the pair (M', f) so as to make f a homotopy equivalence. The modification procedure uses 'surgery theory'. In dimensions ≥ 5, there is a 'surgery obstruction' which lies in a group $\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z}$ or 0 (depending on the dimension modulo 4), and (M', f) can be modified to a homotopy equivalence through surgery precisely when this obstruction vanishes.

1.2 The four-dimensional case

It is easy to construct examples of simply connected four-dimensional spaces with Poincaré duality—in fact, all one needs to specify such a space M_Q is a square matrix Q with integer entries and determinant ± 1. The Spivak normal invariant always arises from a (unique) vector bundle, but the surgery procedure for modifying the map breaks down; there just isn't 'enough space' to carry it out. A difficult theorem of M. Freedman says that M_Q is always homotopy equivalent to a *topological* four-manifold. However, we shall see in this course that

Theorem 1 (Donaldson). *When the matrix $Q \in \mathrm{GL}(n, \mathbb{Z})$ is positive- or negative-definite, and there is no $A \in \mathrm{GL}(n,\mathbb{Z})$ with $AQA^\top = \pm I$, the space M_Q is not homotopy-equivalent to a four-manifold which admits a smooth structure.*

Gauge theory. The methods used to prove the non-existence theorem for smooth structures, and to distinguish infinitely many smooth structures on (e.g.) the quartic surface, are quite different from those that were sketched in the previous paragraphs. They involve differential geometry, and the analysis of certain partial differential equations with 'gauge symmetry'—the **instanton** or **anti-self-duality** equations. These equations were first considered by theoretical physicists working on quantum chromodynamics; they are close cousins of Maxwell's equations in electrodynamics. In the late 1970s, geometers (among them Atiyah, Hitchin, Taubes and Uhlenbeck) began the rigorous study of the instanton equations on smooth four-manifolds. The breakthrough came in 1981, when Donaldson used these equations to prove Theorem 1. The goal of this course is to study this proof. We won't give a complete account of it; our aim is to break it up into parts, giving statements only for some parts, and rigorous proofs for others. This should make the remainder of the proof, and other results in gauge theory, quite accessible.

We round off this discussion by stating a long-standing conjecture in knot theory (the 'Property-P conjecture') which has recently been proved using gauge theory. The proof is a synthesis of several ideas, but analysis of the instanton equations lies at its heart. To give the statement, we need to say what it means to perform 'surgery' on a smooth knot $K \subset S^3$. The knot has a closed tubular neighbourhood $N_K \subset S^3$ diffeomorphic to a solid torus: $N_K \cong S^1 \times D^2$. For any self-diffeomorphism f of the boundary $\partial N_K \cong S^1 \times S^1$, one obtains a smooth three-manifold

$$S^3_{K,f} := (S^3 \setminus \mathrm{int}(N_K)) \cup_f N_K,$$

which is said to be the result of surgery on K.

Theorem 2 (Kronheimer-Mrowka, 2003). *Suppose that $S^3_{K,f}$ is simply connected. Then f is homotopic to the identity map.*

Notes. We should mention that there is another set of equations, the Seiberg-Witten equations, discovered in 1994, which for many purposes are more efficient than the instanton equations, and gives different proofs of the theorems on smooth structures. There is also a more recent method (Ozsváth-Szabó, 2002), apparently a recasting of Seiberg-Witten theory, involving handlebody decompositions and techniques from symplectic geometry. However, Donaldson's original proof is still compelling, because it demonstrates very clearly the interaction between the the structure of the space of solutions to the instanton equations and the geometry of the four-manifold.

If you would like to read more about high-dimensional manifolds, I recommend Milnor's short paper *On manifolds homeomorphic to the 7-sphere*, Annals of Mathematics 64 (1956), 399–405. You could also try M. Kervaire and J. Milnor, *Groups of homotopy spheres. I,* Ann. of Math. 77 (1963), 504–537. Both available at www.jstor.ac.uk.

2 Using PDEs to study topology

A mathematical physicist, modelling some physical situation using a PDE, needs to understand how the topology and geometry of the situation influences the solutions to the equation. Thus the solutions to the wave equation in a cavity depend on its topology and shape; magnetic fields in a superconducting magnet exhibit 'flux quantisation' phenomena when the magnet is, say, ring-shaped.

The focus of early work on the instanton equations was on a problem of this sort: that of finding the most general solution over simple spaces such as the four-sphere or four-torus. (For more information on that aspect of the theory, Dr Dunajski's Easter term course, 'Instantons and Solitons', is recommended.)

It is possible to turn this point of view on its head. The idea is that

by looking at all possible solutions to the equations, we can deduce something about the topology of the space.

This is what we will do for the instanton equations.

2.1 Example: domains in $\mathbb{R}^3$

Take a bounded, path-connected open set $\Omega \subset \mathbb{R}^3$, with smooth boundary, and a vector field $\mathbf{v}\colon \Omega \to \mathbb{R}^3$ which extends smoothly over the boundary. We ask whether it is the gradient of some function (or 'potential') $A\colon \Omega \to \mathbb{R}$, so that

$$\mathbf{v} = \mathrm{grad}(A).$$

A necessary condition is that

$$\mathrm{curl}(\mathbf{v}) = \mathbf{0}.$$

Locally, this is also a sufficient condition: if the curl vanishes then near any point we can find a potential function. *Globally*, this is not always so; it is true if Ω is simply connected, but not in general otherwise. If $\gamma\colon S^1 \to \Omega$ is a loop, then the integral $\int_\gamma (\mathbf{v}\cdot\mathbf{l})dt$ (where $\mathbf{l}$ is the unit-length tangent vector field) depends only on the homotopy class of γ. This sets up a map from the quotient vector space

$$\frac{\{\mathbf{v} : \mathrm{curl}(\mathbf{v}) = \mathbf{0}\}}{\{\mathbf{v} : \mathbf{v} = \mathrm{grad}(A) \text{ for some } A\}} \tag{1}$$

to

$$\mathrm{Hom}(\pi_1(\Omega), \mathbb{R}) = H^1(\Omega; \mathbb{R})$$

which is in fact an isomorphism. So: instead of looking at the properties of some particular vector field, we consider *all* vector fields for which the curl vanishes. By considering how these vector fields can fail to be gradients, we can deduce something about the fundamental group.

A noticeable feature of the expression (1) is that it is a quotient of infinite-dimensional vector spaces which is itself finite-dimensional. An alternative to taking quotients is to cut down to finite dimensions by another equation. For any $\mathbf{v}$, there exists an A such that

$$\mathrm{div}(\mathbf{v} + \mathrm{grad}(A)) = 0 :$$

A is the solution to Poisson's equation

$$\nabla^2 A = -\mathrm{div}(\mathbf{v}).$$

The solution to this equation is unique up to the addition of a constant function. Hence

$$\{\mathbf{v} : \mathrm{curl}(\mathbf{v}) = 0,\ \mathrm{div}(\mathbf{v}) = 0\} \cong \frac{\{\mathbf{v} : \mathrm{curl}(\mathbf{v}) = \mathbf{0}\}}{\{\mathbf{v} : \mathbf{v} = \mathrm{grad}(A) \text{ for some } A\}}.$$

2.2 Harmonic forms and the Hodge theorem

The discussion above can be carried over to a general smooth compact manifold M (without boundary). The objects of interest are now the space of differential p-forms $\Omega^p(M)$, for $p = 0, 1, 2, \ldots$. The generalisation of the quotient (1) is the pth **de Rham cohomology** of M, which is the real vector space

$$H^p_{dR}(M) := \frac{\{\alpha \in \Omega^p(M) : d\alpha = 0\}}{\{\alpha : \alpha = d\beta \text{ for some } \beta \in \Omega^{p-1}(M)\}}.$$

The definition makes sense because $d^2 = 0$. These spaces are purely topological objects, for the de Rham theorem says that there is a natural isomorphism

$$H^p_{dR}(M) \cong H^p(M; \mathbb{R})$$

between the de Rham cohomology and the singular cohomology with real coefficients.

In the case of a compact manifold, the de Rham cohomology can also be realised as a subspace (rather than a quotient) of the closed forms. This follows from the *Hodge theorem*. To state it, we need to introduce a Riemannian metric g on the manifold M. The metric, plus a choice of orientation for M, determine a volume form vol_g:

$$\mathrm{vol}_g(x) = e_1 \wedge \cdots \wedge e_n$$

where $(e_1, \ldots, e_n)$ is an oriented orthonormal basis for $T^*_x M$. The exterior derivative

$$d \colon \Omega^p(M) \to \Omega^{p+1}(M)$$

has a 'formal adjoint'

$$d^* \colon \Omega^{p+1}(M) \to \Omega^p(M)$$

characterised by the equation

$$\int_M g(d\alpha, \beta)\mathrm{vol}_g = \int_M g(\alpha, d^*\beta)\mathrm{vol}_g, \quad \alpha \in \Omega^p(M),\ \beta \in \Omega^{p+1}(M).$$

Note that $(d^*)^2 = 0$. The operator d^* plays the role of the divergence in our discussion on $\mathbb{R}^3$.

Theorem 3 (the Hodge theorem). *Let (M, g) be a compact oriented Riemannian manifold.*

1. *For any $\alpha \in \Omega^p(M)$, there exists $\gamma \in \Omega^{p-1}(M)$ with $d^*(\alpha - d\gamma) = 0$. If γ' also has this property then $d(\gamma - \gamma') = 0$.*
2. *For any $\alpha \in \Omega^p(M)$, there exists $\beta \in \Omega^{p+1}(M)$ with $d(\alpha - d^*\beta) = 0$. If β' also has this property then $d^*(\beta - \beta') = 0$.*

A more convenient way to express the theorem is this:

Any $\alpha \in \Omega^p(M)$ has a decomposition

$$\alpha = \alpha_{\text{harm}} + d^*\beta + d\gamma,$$

where

$$d\alpha_{\text{harm}} = d^*\alpha_{\text{harm}} = 0.$$

The components α_{harm}, $d^*\beta$ and $d\gamma$ are uniquely determined.

From time to time we will draw on the Hodge theorem in this formulation. *Exercise: show that the two versions are equivalent.* A differential form η with $d\eta = 0$ and $d^*\eta = 0$ is called a **harmonic form** for the metric g.

Corollary 4. *There is a canonical isomorphism*

$$\{\eta \in \Omega^p(M) : d\eta = 0,\ d^*\eta = 0\} \cong H^p_{dR}(M).$$

The proof of the Hodge theorem involves analysis in so-called 'Sobolev spaces'—techniques which are further elaborated in the study of instantons on a four-manifold. Indeed, there are strong parallels between the equations defining harmonic forms and those defining instantons.

There is also an important difference: the instanton equations are *non-linear*. So while the harmonic forms are a vector space, instantons are not. However, once one divides out by the natural symmetry group of the instanton equations (the 'gauge group') the space of solutions is in a precise sense *finite-dimensional*. For 'generic' Riemannian metrics it is a smooth manifold, apart from certain well-understood singular points. In Hodge theory, one studies the topology of a manifold through the associated vector space of harmonic differential forms. We will study the topology of a four-manifold through the topology of the space of solutions to the instanton equations.

3 The intersection form of a four-manifold

3.1 The intersection form

The key definition in this lecture is from algebraic topology (the methods should be familiar from the Michaelmas Term course, or elsewhere; you'll find a summary in the handout 'Review of homology and cohomology'). However, the emphasis is on its interpretation in differential geometry.

Definition 5. The **intersection form** of a compact, oriented four-manifold X is the *symmetric* bilinear form

$$H^2(X;\mathbb{Z}) \times H^2(X;\mathbb{Z}) \to \mathbb{Z}, \quad (a,b) \mapsto a \cdot b := \langle a \cup b, [X] \rangle.$$

Here the angle-brackets denote the 'evaluation' pairing between cohomology and homology in the same degree (four in this case).

By Poincaré duality, the induced form on $H^2(X;\mathbb{Z})/\text{torsion}$ is non-degenerate. We sometimes denote this non-degenerate form by Q_X.

Let's look at the additive structure of the (co)homology of X. First, we have $H^4 = H_0 = \mathbb{Z}$, and $H_4 = H^0 = \mathbb{Z}$. Next, $H^3 = H_1 = \pi_1^{\text{ab}}$, and $H_3 = H^1 = \text{Hom}(\pi_1, \mathbb{Z})$. Finally, $H^2 = H_2$, and there is a non-canonical 'universal coefficients' isomorphism

$$H^2(X;\mathbb{Z})_{\text{tors.}} \cong H_1(X;\mathbb{Z})_{\text{tors.}}$$

It is common practice to use the same letter to denote a class in H_2 and its Poincaré dual in H^2. So, for instance, we might write $\alpha \cdot \beta$ when α and β are in H_2.

If we suppose that X is *simply-connected*, i.e. that $\pi_1(X)$ is trivial, then all this simplifies, and we have the following.

On a simply-connected, compact, oriented four-manifold, (i) the odd-dimensional (co)homology is zero, and (ii) $H_2 = H^2$ is torsion-free.

3.2 Intersections of embedded surfaces

How should we understand the intersection form geometrically? The most appealing interpretation is in terms of intersections of embedded surfaces. As stated in the review of homology theory, when we have smooth, closed, oriented surfaces S_1 and S_2, and *transverse* embeddings $S_k \hookrightarrow X$, the cup product of their dual classes is calculated by

$$[S_1] \cdot [S_2] = \sum_{x \in S_1 \cap S_2} \varepsilon(x)$$

where $\varepsilon(x) \in \{\pm 1\}$ is $+1$ if the orientations for $T_x S_1$ and $T_x S_2$, when put together, give the orientation for $T_x X$, and -1 if they don't. To apply this to a particular pair of homology classes, we need them to be representable by a transverse pair of surfaces. In concrete examples, this can often be seen directly. However, it is reassuring to know the following two facts:

- On a smooth n-manifold M, any homology class in $H_{n-2}(M;\mathbb{Z})$ is the fundamental class of a smoothly embedded, oriented submanifold.
- If $i\colon N \to M$ and $i'\colon N' \to M$ are embeddings of submanifolds of complementary dimension in M, there is a homotopy $h_t\colon N \to M$, $t \in [0,1]$, with $h_0 = i$, such that $h_1(N)$ is transverse to $i'(N')$.

Computing self-intersections. As an example of the second point, consider the problem of calculating the *self-intersection* $[S]\cdot[S]$ of an embedded surface $S \subset X$ in a four-manifold.

The **normal bundle** $N_{S/X} \to S$ is the rank two vector bundle $TX|_S/TS$. If we choose a Riemannian metric g on X then we can identify the normal bundle with the orthogonal complement of the tangent bundle to S:

$$N_{S/X} = (TS)^{\perp}$$

The metric determines an exponential map

$$\exp\colon N_{S/X} \to X,$$

and its derivative at $(x,0)$, $D_{(x,0)}\exp\colon T_{x,0}N_{S/X} \to T_xX$, is an isomorphism. It follows from the inverse function theorem that there is a subset

$$N_{S/X}^{\leq\epsilon} := \{(x,v) \in N_{S/X} : g_x(v,v) \leq \epsilon\}$$

such that

$$\exp\colon N_{S/X}^{\leq\epsilon} \to X$$

is an open embedding.

So a neighbourhood of S in X is diffeomorphic to a disc-subbundle of the normal bundle to S.

Lemma 6. *Let M be a smooth manifold, $\pi\colon \xi \to M$ a smooth vector bundle. Then ξ has a smooth section $\sigma\colon M \to \xi$ which vanishes transversely. That is, $\pi\circ\sigma = \mathrm{id}_M$, and whenever $\sigma(x) = 0_x \in \xi_x$, the tangent space $T_{0_x}\xi$ is spanned by the images of $D_x\sigma$ and $D_x\zeta$, where ζ is the zero-section $x' \mapsto 0_{x'}$.*

The proof will be a guided exercise on the first example sheet.

By the lemma, we can find a section of the normal bundle, $\sigma\colon S \to N_{S/X}$, which vanishes transversely. For small $t \neq 0$, the image of $t\sigma$ lies in $N_{S/X}^{\leq\epsilon}$. So we can compose with the open embedding exp to obtain an embedding of S

$$x \mapsto h_t(x) = \exp_x(t\sigma(x))$$

transverse to the original embedding. Of, course, $(h_t)_*[S] = [S]$, so we can calculate $[S]\cdot[S]$ by counting intersection points of S with $h_t(S)$. These intersection points are just the points $x \in S$ where $\sigma(x) = 0$. The sign $\varepsilon(x)$ is ± 1 according to whether the composite map

$$T_x S \xrightarrow{D_x\sigma} T_x X \xrightarrow{\text{orthog. proj.}} (N_{X/S})_x$$

is orientation-preserving or orientation-reversing.

3.3 Examples

The four-sphere S^4. This has $H^2 = 0$.

The projective plane $\mathbb{C}P^2$. We have $H^2(\mathbb{C}P^2;\mathbb{Z}) = \mathbb{Z}$, generated by a class U with $U \cdot U = 1$ (here we use the complex orientation). The Poincaré dual of U is represented by the fundamental class of a projective line $L = \mathbb{C}P^1 \subset \mathbb{C}P^2$. For any other such line L', distinct from L, the intersection $L \cap L'$ is a single point. The intersection takes place in some region $\mathbb{C}^2 \subset \mathbb{C}P^2$, where it can be seen as the transverse intersection of two complex lines. Hence the intersection number is $+1$. Thus, in homology, $[L]\cdot[L] = 1$.

The product $S^2 \times S^2$. An easy calculation in cellular homology (or recourse to the Künneth formula) shows that $H_2(S^2 \times S^2;\mathbb{Z}) = \mathbb{Z}^2$, with basis given by the classes

$$S_1 = [S^2 \times \{x\}], \quad S_2 = [\{x\} \times S^2].$$

We give $S^2 = \mathbb{C}P^1$ its usual orientation, and $S^2 \times S^2$ the product orientation. Then $S_1 \cdot S_2 = +1$. To compute $S_1 \cdot S_1$, we perturb $S^2 \times \{x\}$ to a homologous surface which meets it transversely. An obvious choice is $S^2 \times \{x'\}$, where $x \neq x'$. Then the intersection is empty, so $S_1 \cdot S_1 = 0$. Similarly for $S_2 \cdot S_2$, so the the intersection matrix is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A non-trivial S^2-bundle over S^2. We'll construct a smooth four-manifold X with a smooth map $\pi\colon X \to S^2$ such that the derivative $D_x\pi$ is surjective for all $x \in X$. This implies (by the inverse function theorem) that each fibre $\pi^{-1}(s)$ is a smooth two-dimensional submanifold of X; here, the fibres will all be diffeomorphic to S^2.

Regard the base S^2 as $\mathbb{C} \cup \{\infty\}$, and cover it by two overlapping discs,

$$D_0 = \{z : |z|^2 \leq 2\}, \quad D_\infty = \{z : |z|^2 \geq 1\} \cup \{\infty\}.$$

Let $g\colon S^1 \to \mathrm{SO}(3)$ be the loop

$$g(e^{2\pi i t}) = \text{rotation by } 2\pi t \text{ about } (0,0,1).$$

Really only the homotopy class of g matters; our choice represents the generator of $\pi_1(\mathrm{SO}(3)) = \mathbb{Z}/2$. Define

$$X = (S^2 \times D_0) \cup (S^2 \times D_\infty)/\sim, \text{ where } (g_{z/|z|}(x), z) \sim (x, z) \text{ for } z \in D_0 \cap D_\infty.$$

Then X is naturally a smooth manifold (because the gluing map $(x, z) \mapsto (g_{z/|z|}(x), z)$ is smooth). The map $\pi\colon X \to S^2$ is given by projection to the D_0- or D_∞-factor. It is easy to check that this has the properties claimed above.

Now, X is the union of two pieces, $S^2 \times D_0$ and $S^2 \times D_\infty$, which intersect along a region homeomorphic to $S^2 \times [1, 2] \times S^1$. Applying Mayer-Vietoris to this decomposition, we find that $H_2(X; \mathbb{Z}) \cong \mathbb{Z}^2$. One generator for H_2 is represented by a fibre F. This has $F \cdot F = 0$. Another is represented by a section, i.e. a smooth map $\sigma\colon S^2 \to X$ with $\pi \circ \sigma = \mathrm{id}_{S^2}$. Any such section is linearly independent from F, since $\sigma_*[S^2] \cdot F = \pm 1$. Let's fix the orientation making the sign $+1$.

An example of a section is σ_0, where $\sigma_0(z) = (0, 0, 1)$ for any z. More generally, there is a path of sections $\{\sigma_s\}_{s \in [0, \frac{1}{2}]}$ defined as follows:

$$\sigma_s(z) = \begin{cases} (s, 0, \sqrt{1 - s^2}), & z \in D_\infty, \\ (\frac{s \operatorname{Re} z}{\chi(|z|)}, \frac{s \operatorname{Im} z}{\chi(|z|)}, \sqrt{1 - \frac{|z|^2 s^2}{\chi(|z|)^2}}), & z \in D_0. \end{cases}$$

Here $\chi\colon [0, 2] \to [\frac{1}{2}, 2]$ is a smooth function such that $\chi(x) = x$ when $x \geq 1$ and $\chi(x) = 1$ when $x \leq 1/2$. All these sections represent a single homology class G. To compute $G \cdot G$, let's consider σ_0 and $\sigma_{\frac{1}{2}}$. Their images intersect only at the point $((1, 0, 0); 0) \in S^2 \times D_0$. Writing $z = x + \mathrm{i}y$, we have

$$\partial_x \sigma_{\frac{1}{2}}(0) = (\frac{1}{4}, 0, 0) \in \mathbb{R}^3, \quad \partial_y \sigma_{\frac{1}{2}}(0) = (0, \frac{1}{4}, 0) \in \mathbb{R}^3.$$

It follows that $\sigma_{\frac{1}{2}}$ and σ_0 intersect transversely, with sign $+1$ (think through why this is so!). So the intersection matrix with respect to the basis (F, G) is

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

With respect to the basis $(G, F - G)$, it is therefore

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

4 The intersection form via differential forms

4.1 The wedge product form

We saw in the last lecture that the intersection form can be interpreted in terms of intersections of embedded surfaces. Today we give a different interpretation, in terms of closed two-forms. (There is a third interpretation of H^2, as the group of complex line bundles, which we relegate to the example sheets for lack of time.)

The second de Rham cohomology group $H^2_{dR}(X)$ carries the symmetric bilinear form

$$([\alpha], [\beta]) \mapsto \int_X \alpha \wedge \beta.$$

Here $[\alpha]$ is the cohomology class of the closed two-form α. This form is well-defined because if we change α to $\alpha + d\gamma$ then the value of the form changes by

$$\begin{aligned} \int_X d\gamma \wedge \beta &= \int_X d\,(\gamma \wedge \beta) && \text{as } d\beta = 0 \\ &= 0 && \text{by Stokes' theorem.} \end{aligned}$$

Thom forms. To any embedded surface we can associate a closed two-form (which depends on some additional choices, but only up to addition by exact forms).

Choose a smooth, decreasing function $\chi\colon [0,\infty) \to [0,\infty)$ such that (i) $\chi(t) = 0$ when $t \geq 2$; and (ii) $\int_0^\infty \chi(t)dt = 1$. Notice that, on $\mathbb{R}^2$, with coordinates $y = (y_1, y_2)$, the two-form

$$\chi(|y|)dy_1 \wedge dy_2$$

is invariant under SO(2).

Let $\xi \to \Sigma$ be a rank two vector bundle over a compact surface. Equip it with a Euclidean metric, i.e. a smoothly-varying choice of inner product on each fibre ξ_x. Now, Σ is covered by neighbourhoods U over which ξ is trivial. Over U, we can choose oriented orthonormal bases $(e_{1,x}, e_{2,x})$ for each fibre ξ_x, $x \in U$. We can then write $v_x \in \xi_x$ uniquely as $v_x = y_1(x)e_{1,x} + y_2(x)e_{2,x}$, and this defines coordinate functions y_1 and y_2. On $\xi|_U$, we have the two-form

$$4\epsilon^{-2}\chi(\frac{2|v_x|}{\epsilon})dy_1 \wedge dy_2,$$

which is *closed* and *non-zero only at points of length* $\leq \epsilon$. Because of its invariance under SO(2), this form does not depend on the choice of basis $(e_{1,x}, e_{2,x})$. We can therefore patch together forms defined in different local trivialisations and obtain a globally-defined two-form

$$\tau_{\xi,\epsilon} \in \Omega^2(\xi).$$

Notice that

$$d\tau_{\xi,\epsilon} = 0; \quad \int_{\xi_x} \tau_{\xi,\epsilon} = 1.$$

Now choose a Riemannian metric g on our compact four-manifold X. Given an embedded surface $\Sigma \subset X$, with normal bundle $N_{\Sigma/X} \to \Sigma$, the metric determines an embedding

$$\exp\colon N_{\Sigma/X}^{\leq\epsilon} \hookrightarrow X$$

for sufficiently small ϵ, as in Lecture 3. We define the **Thom form** $\tau_\Sigma \in \Omega^2(X)$ by

$$\tau_\Sigma(p) = \begin{cases} 0, & p \notin \exp(N_{\Sigma/X}^{\leq\epsilon}), \\ \tau_{N_{X/\Sigma},\epsilon}(\exp^{-1}(p)), & p \in \exp(N_{\Sigma/X}^{\leq\epsilon}). \end{cases}$$

We should really notate the choice of g and ϵ. However, the cohomology class of τ_Σ does not depend on these choices.

Lemma 7. *For two transverse surfaces Σ_1 and Σ_2 in X, we have*

$$\int_X \tau_{\Sigma_1} \wedge \tau_{\Sigma_2} = [\Sigma_1] \cdot [\Sigma_2].$$

The map $[\Sigma] \to [\tau_\Sigma]$ is a version of the Poincaré duality map. It extends $\mathbb{R}$-linearly to an isomorphism $H_2(X;\mathbb{R}) \cong H^2_{dR}(X)$. So *the wedge-product form is isomorphic to the $\mathbb{R}$-linear extension of the intersection form on $H_2(X;\mathbb{R}) = H_2(X;\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{R}$.*

4.2 Hodge theory and self-duality

Self-dual and anti-self-dual two-forms. Let $(V, \langle \cdot, \cdot \rangle)$ be a real inner product space, and $(e_1, \dots, e_n)$ an orthonormal basis. There is an induced inner product on $\Lambda^k V$, with orthonormal basis $\{e_{i_1} \wedge \cdots \wedge e_{i_k} : i_1 < \cdots < i_k\}$. The **Hodge star** $*\colon \Lambda^k V \to \Lambda^{n-k} V$ is the linear map characterised by the identity

$$*a \wedge b = \langle a, b \rangle e_1 \wedge \cdots \wedge e_n, \quad a, b \in \Lambda^k V.$$

Of particular interest to us is the case $(n,k) = (4,2)$, where we have an endomorphism $* \in \mathrm{End}(\Lambda^2)$. Thus

$$*(e_{\sigma(1)} \wedge e_{\sigma(2)}) = \mathrm{sign}(\sigma) e_{\sigma(3)} \wedge e_{\sigma(4)}$$

for permutations σ of $\{1,2,3,4\}$, which shows that $** = \mathrm{id}_{\Lambda^2 V}$. This, with the identity

$$a = \frac{1}{2}(a + *a) + \frac{1}{2}(a - *a),$$

implies that

$$\Lambda^2 V = \Lambda^+ V \oplus \Lambda^- V$$

where $\Lambda^{\pm}V$ is the (± 1)-eigenspace of $*$. An orthonormal basis for $\Lambda^{\pm}V$ is given by

$$\frac{1}{\sqrt{2}}(e_1 \wedge e_2 \pm e_3 \wedge e_4), \quad \frac{1}{\sqrt{2}}(e_1 \wedge e_3 \pm e_4 \wedge e_2), \quad \frac{1}{\sqrt{2}}(e_1 \wedge e_4 \pm e_2 \wedge e_3).$$

Now consider a Riemannian n-manifold (X, g). At each point $x \in X$, we have a Hodge star operation

$$*_x \in \mathrm{Hom}(\Lambda^k T_x^* X, \Lambda^{n-k} T_x^* X), \quad *a \wedge b = \langle a, b \rangle \, \mathrm{vol}_{g,x}.$$

When $(n, k) = (4, 2)$ we get a splitting

$$\Lambda^2 T_x^* X = \Lambda_x^+ \oplus \Lambda_x^-.$$

The subspaces $\Lambda_x^{\pm}$ sweep out sub-vector bundles $\Lambda^{\pm}(X) \subset \Lambda^2 T^* X$ (*why?*).

Similarly, there is a Hodge star operation on forms,

$$* \in \mathrm{Hom}(\Omega^k(X), \Omega^{n-k}(X)), \quad (*\eta)(x) = *_x(\eta(x)),$$

and when $(n, k) = (4, 2)$ a decomposition (of vector spaces, or of $C^\infty(X)$-modules)

$$\Omega^2(X) = \Omega^+(X) \oplus \Omega^-(X),$$

where $\Omega^{\pm}(X)$ is the space of sections of $\Lambda^{\pm}(X)$.

Definition 8. On a Riemannian four-manifold, sections of $\Omega^+(X)$ are called **self-dual two-forms**. Sections of $\Omega^-(X)$ are called **anti-self-dual two-forms**.

Notice that if $\alpha \in \Omega^2(X)$ has components $\alpha^{\pm} = (\alpha \pm *\alpha)/2$ then

$$\int_X \alpha \wedge \alpha = \int_X (|\alpha^+|^2 - |\alpha^-|^2) \mathrm{vol}_g.$$

Self-dual and anti-self-dual harmonic two-forms. Recall from Lecture 2 that on a closed Riemannian manifold (X, g), the d-operator on forms has a formal adjoint operator $d^* \colon \Omega^k(X) \to \Omega^{k-1}(X)$, characterised by

$$\int_X g(d^*\alpha, \beta) \mathrm{vol}_g = \int_X g(\alpha, d\beta) \mathrm{vol}_g, \quad \alpha \in \Omega^k(X),\ \beta \in \Omega^{k-1}(X).$$

Lemma 9. *For any k-form α, we have $d^*\alpha = (-1)^{(k+1)(n-k)+1} * d(*\alpha)$.*

Proof.

$$\begin{aligned}
\int_X g(\alpha, d\beta) \mathrm{vol}_g &= \int_X *\alpha \wedge d\beta \\
&= (-1)^{(n-k+1)} \int_X d(*\alpha) \wedge \beta && \text{(Stokes' theorem)} \\
&= (-1)^{(k+1)(n-k)+1} \int_X (** \, d(*\alpha)) \wedge \beta && \text{(as } * \circ * = (-1)^{k(n-k)}\text{)} \\
&= (-1)^{(k+1)(n-k)+1} \int_X g(*d * \alpha, \beta) \mathrm{vol}_g.
\end{aligned}$$

∎

By the Hodge theorem, there is a unique *harmonic* representative η for each de Rham cohomology class:

$$d\eta = 0, \quad d^*\eta = 0.$$

This inclusion of the vector space $\mathcal{H}_g^k$ of harmonic forms into the closed forms gives an isomorphism $\mathcal{H}_g^k \to H_{dR}^k(X)$.

By the lemma, when η is harmonic, so is $*\eta$. In particular, if η is a harmonic two-form on a Riemannian four-manifold, its self-dual and anti-self-dual components, $(\eta \pm *\eta)/2$, are also harmonic. Hence $\mathcal{H}_g^2$ decomposes as

$$\mathcal{H}_g = \mathcal{H}_g^+ \oplus \mathcal{H}_g^-$$

into spaces of (anti-)self-dual harmonic two forms.

When $0 \neq \alpha \in \mathcal{H}_g^+$, we have

$$\int_X \alpha \wedge \alpha = \int_X *\alpha \wedge \alpha = \int_X |\alpha|^2 \mathrm{vol}_g > 0.$$

Thus the wedge-product form is positive-definite on $\mathcal{H}_g^+$. Similarly, it is negative-definite on $\mathcal{H}_g^-$. Since these two subspaces span $\mathcal{H}_g^2$, they are *maximal* definite subspaces for the wedge-product form.

5 Simply-connected four-dimensional complexes

In this lecture we will prove that two closed, oriented four-manifolds which are simply connected and have isomorphic intersection forms are homotopy-equivalent:

Theorem 10 (Milnor). *Suppose that X, X' are compact, connected, simply connected topological four-manifolds. Any homomorphism $H_2(X;\mathbb{Z}) \to H_2(X';\mathbb{Z})$ which preserves the intersection forms is induced by an orientation-preserving homotopy equivalence $X \to X'$.*

Remark 11. *If X and X' are smooth (and not necessarily simply connected) then any orientation-preserving homotopy equivalence $F\colon X \to X'$ satisfies $F^*(TX') \cong TX$. So, in the simply connected case, Q_X determines both the homotopy type and the tangent bundle.*

Here's an explanation for readers who know some vector bundle theory. Rank-four oriented vector bundles over a finite cell complex are classified by the characteristic classes w_2, w_3, p_1 and e. On a smooth four-manifold, w_2 is characterised by the identity $w_2 \cup x = x \cup x$ ($x \in H^2(X;\mathbb{Z}/2)$), while $w_3 = \beta w_2$, where β is the Bockstein operation; these are two instances of the Wu formula*. A remarkable result of Thom/Rokhlin says that the cobordism class of X is determined by $\sigma(X)$, the signature of Q_X. From this follows* Hirzebruch's signature theorem*: $\langle p_1(TX), [X]\rangle = 3\sigma(X)$. Hence $w_2(F^*TX') = w_2(TX)$, $w_3(F^*TX') = w_3(TX)$, and $p_1(F^*TX') = p_1(TX)$. Finally, $\langle e(TX), [X]\rangle$ is the Euler characteristic of X, so $e(F^*TX') = e(TX)$.*

5.1 Attaching maps

We begin our discussion with the following construction of simply connected four-dimensional cell-complexes: start with $\bigvee_{i=1}^n S^2$, the wedge sum (one-point union) of n copies of S^2. Let X_f be the space obtained by attaching a four-cell by a map $f\colon S^3 \to \bigvee S^2$:

$$X_f = \left(\bigvee S^2\right) \cup_f e^4.$$

Since X_f can be built from cells of dimensions 0, 2 and 4 only, its (co)homology is zero except in these degrees. We have $H_4(X_f;\mathbb{Z}) = \mathbb{Z}$, generated by e_4, and $H_2(X_f;\mathbb{Z}) = \mathbb{Z}^n$. Similarly, $H^4(X_f;\mathbb{Z}) = \mathbb{Z}$ and $H^2(X_f;\mathbb{Z}) = \mathbb{Z}^n$. The cup product

$$H^2(X_f;\mathbb{Z}) \times H^2(X_f;\mathbb{Z}) \to H^4(X_f;\mathbb{Z})$$

is therefore given by an $n \times n$ symmetric matrix Q_f.

The homotopy type of X_f depends only on n and the homotopy class of the attaching map f. It turns out that the available homotopy classes are exactly classified by the matrices Q_f :

Lemma 12. *The map $f \mapsto Q_f$ gives an isomorphism of abelian groups*

$$\pi_3(\bigvee S^2) \to \{\textit{symmetric } n \times n \textit{ matrices over } \mathbb{Z}\}.$$

This generalises the isomorphism $\pi_3(S^2) \cong \mathbb{Z}$. In that case, the generator is the Hopf map $h\colon S^3 \to S^2$. Attaching a four-cell to S^2 via h gives the space $X_h = \mathbb{C}P^2$.

Remark 13. *The group $\pi_3(\bigvee S^2)$ parametrises* based *maps $S^3 \to \bigvee S^2$, whilst the attaching maps (and their homotopies) are not based. This makes no difference here, as every free homotopy can be deformed to a based homotopy. This is because $\bigvee S^2$ is simply connected.*

Relative homotopy groups. In the proof we will need to make use of relative homotopy groups. Suppose (X, A) is a pair of spaces (i.e. $A \subset X$) and $x \in A$ a basepoint. Then the homotopy classes of maps $(D^n, \partial D^n, *) \to (X, A, \{x\})$ form a group $\pi_n(X, A, x)$ when $n \geq 2$ (usually we drop the basepoint from the notation). This group is abelian when $n \geq 3$.[3] When $\{x\} = A$ we get the usual homotopy group $\pi_n(X, x)$. There is a long exact sequence

$$\cdots \to \pi_{n+1}(X, A) \overset{\partial_{n+1}}{\to} \pi_n(A) \to \pi_n(X) \to \pi_n(X, A) \overset{\partial_n}{\to} \pi_{n-1}(X, A) \to \cdots$$

where ∂ is restriction from D^n to ∂D^n. The *Hurewicz homomorphism*

$$h\colon \pi_n(X, A) \to H_n(X, A)$$

sends the homotopy class of $f\colon (D^n, \partial D^n) \to (X, A)$ to $f_*[D^n, \partial D^n]$, where $[D^n, \partial D^n] \in H_n(D^n, \partial D^n; \mathbb{Z})$ is the orientation class. The Hurewicz homomorphisms form a map of complexes from the exact sequence of relative homotopy groups to the exact sequence of relative homology groups. The *Hurewicz theorem* is as follows:

Suppose (X, A) is a CW pair, with X path-connected and simply connected, and A path connected. Let $x \in A$. If $H_k(X, A; \mathbb{Z}) = 0$ for $k < n$ then $\pi_k(X, A) = 0$ for $k < n$ and the Hurewicz map $h\colon \pi_n(X, A) \to H_n(X, A; \mathbb{Z})$ is an isomorphism.

Proof of the lemma. Think of the wedge $\bigvee S^2$ as embedded into the n-fold product,

$$\bigvee S^2 \subset S^2 \times \cdots \times S^2,$$

as the subspace where all but one coordinate is equal to a chosen basepoint $* \in S^2$. Embed $S^2 = \mathbb{C}P^1$ into $\mathbb{C}P^2$ in the usual way (that is, by attaching e_4 via h), and hence embed $\bigvee S^2$ into the product $P = \mathbb{C}P^2 \times \cdots \times \mathbb{C}P^2$.

[3] In fact $\pi_n(X, A, x) \cong \pi_{n-1}P(X; x, A)$, the $(n-1)$th absolute homotopy group of the space of paths in X originating at x and terminating in A.

Now, $\pi_3(\mathbb{C}P^2) = \pi_4(\mathbb{C}P^2) = 0$. Indeed, the circle-bundle $S^5 = S(\mathbb{C}^3) \to \mathbb{C}P^2$ induces injective maps on homotopy groups π_k, for all k, and $\pi_k(S^5) = 0$ when $k < 5$. (If you know about the long exact sequence of homotopy groups of a fibration, you can see this injectivity property immediately. If not, you could think about why it should be true.) Hence $\pi_3(P) = \pi_4(P) = 0$.

The homotopy and homology exact sequences of the pair $(P, \bigvee S^2)$, give a diagram with exact rows

$$\begin{array}{ccccccccc} & & 0 & \longrightarrow & \pi_4(P, \bigvee S^2) & \longrightarrow & \pi_3(\bigvee S^2) & \longrightarrow & 0 \\ & & & & {\scriptstyle h}\big\downarrow & & & & \\ 0 & \longrightarrow & H_4(P;\mathbb{Z}) & \longrightarrow & H_4(P, \bigvee S^2;\mathbb{Z}) & \longrightarrow & 0 & & \end{array}$$

$H_k(P, \bigvee S^2) = 0$ for $k < 4$ (*check this!*). The Hurewicz theorem therefore tells us that h is an isomorphism. Hence $\pi_3(\bigvee S^2) \cong H_4(P;\mathbb{Z})$. Let e^i denote the generator of $H^2(P;\mathbb{Z})$ obtained by pulling back the standard generator of $H^2(\mathbb{C}P^2;\mathbb{Z})$ by the ith projection map $P \to \mathbb{C}P^2$. The Künneth formula, or better, a direct calculation in cellular cohomology (*do it!*), shows that

$$H^4(P;\mathbb{Z}) = \bigoplus_{1\leq i\leq j\leq n} \mathbb{Z}(e^i \cup e^j).$$

This is a free abelian group (of rank $n(n+1)/2$), and so is identified with $\mathrm{Hom}(H_4(P;\mathbb{Z}), \mathbb{Z})$. We obtain an isomorphism

$$H_4(P;\mathbb{Z}) \to \{\text{symmetric } n \times n \text{ matrices over } \mathbb{Z}\}$$

by sending the class $\tilde{f} \in H_4(P;\mathbb{Z})$ corresponding to $f \in \pi_3(\bigvee S^2)$ to the matrix $\tilde{Q}_f$ with entries

$$(\tilde{Q}_f)_{ij} = \langle e^i \cup e^j, \tilde{f}\rangle.$$

All that remains is to see that $\tilde{Q}_f$ is equal to Q_f, the matrix of the cup product form on X_f. By the surjectivity of $\pi_4(P, \bigvee S^2) \to \pi_3(\bigvee S^2)$, the inclusion of $\bigvee S^2$ in P extends to a map $\eta\colon X_f \to P$. The fundamental class $[X_f] \in H_4(X_f;\mathbb{Z})$ maps to $\eta_*[X_f] = \tilde{f}$. Hence

$$\begin{aligned} (\tilde{Q}_f)_{ij} = \langle e^i \cup e^j, \tilde{f}\rangle &= \langle e^i \cup e^j, \eta_*[X_f]\rangle \\ &= \langle \eta^*(e^i \cup e^j), [X_f]\rangle \\ &= \langle \eta^* e^i \cup \eta^* e^j, [X_f]\rangle = (Q_f)_{ij}. \end{aligned}$$

□

Remark 14. *(i) In view of the lemma, we can sensibly write X_Q instead of X_f, where $Q = Q_f$ is the matrix corresponding to $f \in \pi_3(\bigvee S^2)$.*

(ii) A necessary condition for X_Q to be homotopy-equivalent to a topological four-manifold is that Poincaré duality should hold. This is just the condition that Q is unimodular, i.e. that $\det(Q)$ *is invertible in* $\mathbb{Z}$.

5.2 Proof of Milnor's theorem

In the following proof we will quote the theorem that *any compact topological manifold M, possibly with boundary, has the homotopy type of a CW complex with finitely many cells.* This basic result is unfortunately not easy to prove. It *is* straightforward in the smooth case, where one obtains such a cell decomposition by choosing a Morse function.

Proof of Theorem 10. The proof is in three steps.

1. Choose a basis $(e_1, \dots, e_n)$ for $H_2(X;\mathbb{Z})$. Since X is simply-connected, the Hurewicz theorem applies; it tells us that the Hurewicz map $\pi_2(X) \to H_2(X;\mathbb{Z})$ is an isomorphism. Hence $e_i = s_{i*}[S^2]$ for maps $s_i \colon S^2 \to X$ which send the basepoint $* \in S^2$ to a basepoint $x \in X$. We can conflate these maps into a single one,

$$s \colon \bigvee_{i=1}^{n} S^2 \to X.$$

Applying a homotopy if necessary, we may assume that s is not surjective (for example, s is homotopic to a map which is cellular, so its image lies in the 2-skeleton of X). Let $B \subset X$ be a closed four-ball disjoint from $\operatorname{im}(s)$, and let $X_0 = X \setminus \operatorname{int}(B)$.

2. We now observe (by using the Mayer-Vietoris sequence, for instance) that X_0 has non-zero homology only in degrees 0 and 2. In those degrees the homology is the same as that of X. Hence the map

$$s \colon \bigvee S^2 \to X_0$$

induces an isomorphism on H_*. Here we invoke a well-known theorem of J. H. C. Whitehead: *Suppose A, B are connected and simply connected spaces homotopy-equivalent to CW complexes. Then any map $f \colon A \to B$ which induces an isomorphism on H_* is a homotopy-equivalence.* Since X_0 (a compact manifold with boundary) has the homotopy type of a CW complex, $s \colon \bigvee S^2 \to X_0$ is a homotopy equivalence.

3. X is obtained by attaching a four-cell to X_0. Hence it is homotopy equivalent to X_f, for some $f \in \pi_3(\bigvee S^2)$. Moreover, there is a homotopy equivalence $X \simeq X_f$ which maps $[X]$ to $[X_f]$ and which sends the basis e_i for $H_2(X;\mathbb{Z})$ to the standard basis for $H_2(X_f)$. By the lemma, f is determined up to homotopy by the intersection matrix Q of X in the basis e_i. Hence $X_f \simeq X_Q$.

Since we can apply the same argument to X', Milnor's theorem follows.

□

References for this lecture. The main argument is from Milnor and Husemoller, *Symmetric bilinear forms*, Springer 1973. Background in homotopy theory is covered in e.g. A. Hatcher, *Algebraic topology*, CUP, 2002; P. May, *A concise course in algebraic topology*, University of Chicago Press, 1999.

6 Further results on intersection forms

6.1 Indefinite forms

For a few four-manifolds it is feasible to calculate the intersection form directly. Usually, however, one has to appeal to an algebraic fact, the *Hasse-Minkowski theorem*. To state it, we need to give another example of a unimodular form, the rank 8 form E_8. This is the lattice in $\mathbb{R}^8$ generated by the vectors

$$e_2 - e_3, e_3 - e_4, \ldots, e_7 - e_8, \frac{1}{2}(e_1 + e_8 - e_2 - \cdots - e_7), e_7 + e_8.$$

(This is actually a root lattice, and the eight vectors form a set of 'simple roots'.) It has the following properties.

- It is positive-definite (since $\mathbb{R}^8$ is!).
- Its matrix entries are integers: the non-zero entries are

$$E_8 = \begin{pmatrix} 2 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & -1 & 2 & -1 & & & \\ & & & -1 & 2 & -1 & -1 & \\ & & & & -1 & 2 & & \\ & & & & -1 & & 2 & -1 \\ & & & & & & -1 & 2 \end{pmatrix}$$

- It is *even*, i.e. the squared length of any vector in E_8 is in $2\mathbb{Z}$: indeed, the eight basis vectors all have squared length 2.
- It is *unimodular*, i.e. its matrix is invertible over $\mathbb{Z}$: indeed, wedging together the basis vectors one finds $\det(E_8) = 1$.

We also introduce the 'hyperbolic' form

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which is even and unimodular, and the rank one unimodular forms $\langle \pm 1 \rangle$.

Theorem 15 (Hasse-Minkowski). *Suppose Q is a quadratic form over $\mathbb{Z}$, represented by a symmetric matrix of integers, which is unimodular (*$\det(Q) = \pm 1$*) and indefinite ($Q(x)$ takes both positive and negative values). Then*

- *If Q is even, $Q \simeq m(-E_8) + nH$ for some $m, n \geq 0$.*
- *If Q is odd (i.e. not even) then $Q \simeq m\langle 1 \rangle + n\langle -1 \rangle$.*

Here $\simeq$ denotes equivalence of quadratic forms over $\mathbb{Z}$, and the additive notation refers to direct sum of forms. Hence Q is determined, up to equivalence, by its rank, signature, and type (odd/even).

The situation for *definite* unimodular forms is quite different. The number of equivalence classes of indecomposable, positive-definite, unimodular forms of rank n is an unbounded function of n. However, the theorem of Donaldson, to be discussed in due course, tells us that only the very simplest definite forms arise as intersection forms of smooth four-manifolds.

6.2 The signature is a cobordism-invariant

The dimension of maximal positive-definite (resp. negative-definite) subspace of $H^2(X;\mathbb{R})$ is an invariant of Q_X, hence of X. It is denoted by b^+ (resp. b^-). The **signature** $\sigma(X)$ of X is the signature $\sigma(Q_X)$ of the intersection form on $H^2(X;\mathbb{R})$:

$$\sigma(X) = b^+ - b^-.$$

Proposition 16. *Let X be a closed four-manifold (smooth and oriented but not necessarily connected) which bounds a compact oriented five-manifold (again smooth and orientable but not necessarily connected). Then $\sigma(X) = 0$.*

We need an algebraic lemma, whose proof is left as an exercise.

Lemma 17. *Let q be a non-degenerate quadratic form on a vector space V of dimension $2n$. If there is an n-dimensional subspace U such that $q|_U = 0$ then q has signature zero.*

Proof of the proposition. Say $X = \partial W^5$. Write i for the inclusion of X in W. The intersection form is identically zero on $i^*H^2(W;\mathbb{R})$, for

$$i^*\alpha \cdot i^*\beta = \langle i^*\alpha \cup i^*\beta, [X]\rangle = \langle i^*(\alpha \cup \beta), [X]\rangle = \langle \alpha \cup \beta, i_*[X]\rangle = 0,$$

where in the last step we use the fact that $i_*[X] = 0$ since $X = \partial W$. We claim that

$$\dim H^2(X;\mathbb{R}) = 2 \dim i^*H^2(W;\mathbb{R}).$$

There is a commutative diagram with exact rows

$$\begin{array}{ccccc}
H^2(W;\mathbb{R}) & \xrightarrow{i^*} & H^2(X;\mathbb{R}) & \xrightarrow{\delta} & H^3(W,X;\mathbb{R}) \\
\cong\downarrow & & \cong\downarrow & & \cong\downarrow \\
H_3(W,X;\mathbb{R}) & \xrightarrow{\partial} & H_2(X;\mathbb{R}) & \xrightarrow{i_*} & H_2(W;\mathbb{R}).
\end{array}$$

The vertical arrows are the Poincaré-Lefschetz duality isomorphisms. By exactness of the upper row, $H^2(X) \cong \mathrm{im}(i^*) \oplus \mathrm{im}(\delta)$. But $\mathrm{im}(\delta) \cong \mathrm{im}(i_*)$, and i_* is the dual homomorphism to i^*, and so has the same rank. Hence $H^2(X) \cong \mathrm{im}(i^*) \oplus \mathrm{im}(i^*)$.

The result now follows from the algebraic lemma. □

Corollary 18. *If X_1 is cobordant to X_2 (i.e. there exists a compact, smooth oriented five-manifold W with $\overline{X_1} \cup X_2$ as its oriented boundary) then $\sigma(X_1) = \sigma(X_2)$.*

That's the end of this lecture as far as course-material is concerned. The following notes survey some further results; they are included only to put the results of the course into their proper perspective.

6.3 (*) The theorems of Freedman, Rokhlin and Wall

Theorem 19 (Freedman, 1980). *Fix a unimodular matrix Q.*

- *There exists a compact, simply-connected, oriented topological four-manifold M homotopy-equivalent to X_Q.*
- *If Q is even as a quadratic form then M is unique up to homeomorphism.*
- *If Q is odd then there are precisely two such manifolds M_1, M_2, up to homeomorphism. They are distinguished by the property that $M_1 \times S^1$ admits a smooth structure but $M_2 \times S^1$ does not.*

Of course, M_2 cannot admit a smooth structure either (but in general we cannot say whether or not M_1 does). Hence we have the

Corollary 20. *Two compact, oriented, simply-connected four-manifolds which admit smooth structures and which have isomorphic intersection forms are homeomorphic.*

We should also mention here two classical results about *smooth* four-manifolds.

Theorem 21 (Rokhlin). *Let X be a smooth, compact, simply connected, oriented four-manifold. If X has even intersection form Q_X then its signature $\sigma(X)$ is divisible by 16.*

The signature of an even unimodular form is always divisible by 8 (example sheet 2). Freedman tells us that there is a topological four-manifold with intersection form E_8, but since this has signature 8, the manifold carries no smooth structure.[4]

Theorem 22 (Wall). *If X, X' are closed, oriented, simply-connected smooth four-manifolds with isomorphic intersection forms then there exists an integer k and an orientation-preserving diffeomorphism*

$$X \# k(S^2 \times S^2) \cong X' \# k(S^2 \times S^2)$$

between the connected sums of X, X' with k copies of $S^2 \times S^2$.

[4] Rokhlin's theorem is easy to prove if you know about Dirac operators and the Atiyah-Singer index theorem. The hypotheses imply that X admits a spin-structure. The Dirac operator, sending positive spinors to negative spinors, is quaternion-linear. The number $-\sigma(X)/16$ is its quaternionic index.

The proof uses surgery theory, and the notion of 'h-cobordism'. It is an open problem to determine whether one can always take $k = 1$. The invariants which are used to distinguish X from X' always vanish upon taking the connected sum with $S^2 \times S^2$.

6.4 (*) Complex hypersurfaces in $\mathbb{C}P^3$

Since in this course we assume knowledge neither of complex manifolds, nor of characteristic classes, we do not really have the tools available to construct interesting four-manifolds and compute their intersection forms. For example, Chern classes are essential in computing invariants of complex hypersurfaces in $\mathbb{C}P^3$. Here are the answers in that case. Choose a homogeneous polynomial $f_d(z_0, z_1, z_2, z_3)$ of degree d, and let

$$X_d = \{z \in \mathbb{C}P^3 : f_d(z) = 0\}.$$

For suitable choices of f_d (e.g. $f_d = \sum z_i^d$) this is a smooth four-manifold. The intersection form can be calculated in a two-step process. The first step is to calculate the rank, signature and type; the second is to appeal to the Hasse-Minkowski theorem.

A standard result in complex geometry, the Lefschetz hyperplane theorem, says that X_d is simply-connected. The Euler characteristic $\chi(X_d)$ is also the second Chern number of the tangent bundle, which can be calculated using the Whitney sum formula and the fact that the total Chern class of $\mathbb{C}P^3$ is $(1+u)^4$ where $u \in H^2$ is the hyperplane class. The result is

$$\chi(X_d) = d(d^2 - 4d + 6).$$

Hence $b_2(X_d) = d(d^2 - 4d + 6) - 2$. The signature $\sigma(X_d)$of the intersection form can also be computed using characteristic classes, via the Hirzebruch signature theorem, which says that $\sigma(X) = \frac{1}{3}p_1(X)[X]$. The result is

$$\sigma(X_d) = \frac{1}{3}d(4 - d^2).$$

The type of Q_X is determined by whether or not the second Stiefel-Whitney class w_2 vanishes. This follows from the Wu formula, which says that $w_2 \cup x = x \cup x$ for $x \in H^2(X; \mathbb{Z}/2)$. One deduces that the type of Q_{X_d} is the parity of d.

One can then read off using Hasse-Minkowski that, for example, the intersection form of X_4 (a K3 surface) is

$$2(-E_8) + 3H.$$

7 SU(2)-bundles over four-manifolds

The group $\mathrm{SU}(n)$ can be thought of as the subgroup of $\mathrm{GL}(n, \mathbb{C})$ consisting of matrices A with $|Ax| = |x|$ (where $|\cdot|$ is the standard hermitian inner product) and $Ax_1 \wedge \cdots \wedge Ax_n = x_1 \wedge \cdots \wedge x_n$ for $x_i \in \mathbb{C}^n$.

Definition 23. An $\mathrm{SU}(n)$-bundle $(E, |\cdot|, o)$ over a space X is a rank n complex vector bundle E over X equipped with a hermitian metric $|\cdot|$ and a determinant form, that is, a unit-length section o of $\Lambda^n E$. An isomorphism of $\mathrm{SU}(n)$-bundles $(E, |\cdot|, o)$ and $(E', |\cdot|', o')$ over X is an isomorphism of complex vector bundles $\theta \colon E \to E'$ such that $|\theta(v)|' = |v|$ and $(\Lambda^n\theta)o = o'$.

Often we will simply write E for an $\mathrm{SU}(n)$-bundle. To define $\mathrm{U}(n)$-bundles, delete each occurrence of o.

Gluing construction. We immediately specialise to SU(2). Let X be a compact, oriented four-manifold, and $B^4 \subset X$ be a smoothly embedded closed four-ball. Given any smooth map $g \colon S^3 \to \mathrm{SU}(2)$, we construct an SU(2)-bundle $E_g \to X$:

Think of X as the space obtained from $X_0 := X \setminus \mathrm{int}(B^4)$ and B^4 by gluing together their common boundary 3-sphere. Thus

$$X = X_0 \cup_f B^4$$

for some diffeomorphism $f \colon \partial X_0 \to \partial B^4$. We will glue together the trivial bundles

$$\mathbb{C}^2 \times X_0 \to X_0, \quad \mathbb{C}^2 \times B^4 \to B^4.$$

over the boundaries. Choose a smooth map $g \colon S^3 \to \mathrm{SU}(2)$, and let E_g be the quotient space of the disjoint union $(\mathbb{C}^2 \times X_0) \cup (\mathbb{C}^2 \times B^4)$ under the equivalence relation

$$(v, x) \sim ((g \circ f(x))v, f(x)), \quad x \in \partial X_0.$$

There is an obvious map $E_g \to X_0 \cup_f B^4 = X$. Convince yourself that $E_g \to X$ is locally trivial, hence a rank two vector bundle.

$\mathbb{C}^2$ has its standard hermitian inner product $|\cdot|$ and volume element $o = e_1 \wedge e_2 \in \Lambda^2\mathbb{C}^2$, and these induce a hermitian metric and volume form for the trivial $\mathbb{C}^2$-bundles over B^4 and X_0. These are compatible with the gluing map g, since it has values in SU(2). Hence they jointly define a hermitian metric and volume form on E_g.

We interpolate an observation about vector bundles. In the following statement, 'bundle' should be read as 'real vector bundle', 'complex vector bundle', '$\mathrm{SU}(n)$-bundle' or '$\mathrm{U}(n)$-bundle.'

Lemma 24. *Let ξ_0 and ξ_1 be two bundles over a paracompact Hausdorff space Y. Then ξ_0 is isomorphic to ξ_1 iff there is a bundle ξ over $Y \times [0,1]$ such that, for $i \in \{0,1\}$, the restriction of ξ to $Y \times \{i\} = Y$ is equal to ξ_i.*

Briefly, the two bundles are isomorphic iff they are homotopic. In our gluing construction, we observe that if we have a homotopy g_t then the bundles E_{g_t} together form a bundle over $X \times [0,1]$. By the lemma, the SU(2) bundles E_{g_0} and E_{g_1} are therefore isomorphic.

To prove the lemma in one direction, observe that if $\theta\colon \xi_0 \to \xi_1$ is an isomorphism then we can glue $\xi_0 \times [0, \frac{1}{2}]$ to $\xi_1 \times [\frac{1}{2}, 1]$ over $\{\frac{1}{2}\}$ using θ to obtain the required vector bundle ξ. For a proof of the converse, see for example Prop. 1.7 in Allen Hatcher, *Vector bundles and K-theory*, available at `www.math.cornell.edu/∼hatcher`. In the case of smooth vector bundles over a smooth manifold, there is a simple proof using connections.

Remark 25. *We can make some particular choices for g so as to enumerate the possible homotopy classes. The homotopy class of $g\colon S^3 \to \mathrm{SU}(2)$ is detected by the* degree $\deg(g) \in \mathbb{Z}$. *Recall that the degree of a map $\phi\colon M \to M'$ between closed oriented n-manifolds is given by the formula*

$$\phi_*[M] = \deg(\phi)[M'] \in H_n(M'; \mathbb{Z}),$$

or, upon choosing a volume form ω on M', by

$$\int_M \phi^*\omega = \deg(\phi) \int_{M'} \omega.$$

Think of S^3 as $S(\mathbb{H})$, the unit sphere in the quaternions. Quaternions act on $\mathbb{C}^2 = \mathbb{H}$ by left quaternion multiplication, and this defines the diffeomorphism $S^3 \to \mathrm{SU}(2) \subset \mathrm{GL}(2, \mathbb{C})$. Define g_k to be the composite

$$S^3 \xrightarrow{\cong} S(\mathbb{H}) \xrightarrow{q \mapsto q^k} S(\mathbb{H}) \xrightarrow{\cong} S^3.$$

Then $\deg(g_k) = k$.

SU(2) *is naturally isomorphic to* Sp(1)*, the group of unit-length quaternions. It sometimes helpful to view* SU(2)*-bundle as* Sp(1)*-bundles, i.e. rank 1 quaternionic vector bundles with quaternionic metric.*

Degree versus Euler number. Here is one way to compute degrees of maps between spheres. Let B^{n+1} be the closed unit ball in $\mathbb{R}^{n+1}$.

Lemma 26. *Suppose that $s\colon B^{n+1} \to \mathbb{R}^{n+1}$ is a smooth map such that $s(x) \neq 0$ when $x \in \partial B^{n+1}$ and which vanishes transversely. Then the degree of the map*

$$\frac{s}{|s|}\colon \partial B^{n+1} = S^n \to S^n$$

is the signed count of points in $s^{-1}(0)$.

Proof. Write t for the S^n-valued function $s/|s|$, defined on $B^{n+1} \setminus s^{-1}(0)$. Suppose that B and B' are two smaller balls contained in the interior of B^{n+1}, such that $B \cap B' = \emptyset$ and $s^{-1}(0) \subset \operatorname{int}(B) \cup \operatorname{int}(B')$. Then

$$\deg t|_{\partial B^{n+1}} = \deg t|_{\partial B} + \deg t|_{\partial B'}. \tag{2}$$

This follows from the following observation: there is a homotopy of maps $h_t \colon S^n \to B^{n+1}$ such that $h_0 = t|_{\partial B^{n+1}}$ and h_1 is as follows. Let $(x_0, \dots, x_n)$ be coordinates on $\mathbb{R}^{n+1}$:

$$h_1(x) = \begin{cases} t \circ g(x), & x_0 < -\frac{1}{2} \\ t \circ \gamma(x_0), & -\frac{1}{2} \leq x_0 \leq \frac{1}{2} \\ t \circ g'(x), & x_0 > \frac{1}{2}. \end{cases}$$

Here $\gamma \colon [-\frac{1}{2}, \frac{1}{2}] \to B^{n+1}$ is a path connecting a point $p \in \partial B$ to $p' \in \partial B'$; g is a diffeomorphism from $\{x \in S^n : x_0 < -\frac{1}{2}\}$ to $\partial B \setminus \{p\}$; and g' a diffeomorphism from $\{x \in S^n : x_0 > \frac{1}{2}\}$ to $\partial B \setminus \{p'\}$.

Applying (2) successively to nested balls, we see that it suffices to show that the degree of $t|_{\partial U}$ is equal to $\varepsilon(z) \in \{\pm 1\}$ when U is a very small ball containing a single zero $z \in s^{-1}(0)$, of sign $\varepsilon(z)$. Since the zero is transverse, the inverse function theorem implies that $s|_U$ is a diffeomorphism onto its image, for small enough U; thus $s|_U \colon U \to s(U)$ has degree $\varepsilon(z)$ (the sign of $\det(D_z s)$). Shrinking U further if necessary, the map $s(U) \to S^n$, $x \mapsto x/|x|$, will be an orientation-preserving diffeomorphism. (What is needed for this is that the line segment from x to $x/|x|$ contains no points of $s(U)$ besides x. This would certainly be true if $s|_U$ were a linear map; in reality, $s|_U$ is a small perturbation of a linear map.) Hence $\deg(t|_{\partial U}) = \deg(s|_{\partial U}) = \varepsilon(z)$, as required.

□

Definition 27. Let ξ be a real vector bundle of rank n over a closed n-manifold M. The Euler number $e(\xi) \in \mathbb{Z}$ is defined to be the signed count of zeros of a transversely-vanishing section s of ξ.

Lemma 28. *$e(\xi)$ does not depend on the choice of section s.*

Proof. Any two sections s_0, s_1 are homotopic; there is a path of sections $t \mapsto s_t$ connecting them. We can think of $\{s_t\}_{t \in [0,1]}$ as a section $s_\bullet$ of $p_1^* \xi \to X \times [0,1]$, where $p_1 \colon X \times [0,1] \to X$ is the projection. If s_0 and s_1 are transverse then we can perturb $s_\bullet$ so as to make it transverse also. Then $s_\bullet^{-1}(0)$ is an oriented 1-manifold with boundary in $X \times [0,1]$. The boundary points lie in $X \times \{0,1\}$; they are the points of $s_0^{-1}(0)$ and of $s_1^{-1}(0)$. The signed count of boundary points in an oriented 1-manifold with boundary is always zero. But this count is the difference between $e(\xi)$ as calculated using s_0 and $e(\xi)$ as calculated using s_1.

□

Now we return to our $\mathrm{SU}(2)$-bundles over X^4 (and their underlying rank 4 real vector bundles).

Proposition 29. *We have* $e(E_g) = \deg(g)$. *Hence if* $E_g \cong E_{g'}$ *then* $\deg(g) = \deg(g')$. *Conversely, any* $\mathrm{SU}(2)$*-bundle* E' *over* X *is isomorphic to* E_g *for any* g *of degree* $e(E)$.

In a nutshell, the Euler number gives a bijection between isomorphism classes of $\mathrm{SU}(2)$-bundles over X and the integers.

Proof. We have $E_g|_{X_0} = \mathbb{C}^2 \times X_0$. Let σ_0 denote the constant section $x \mapsto ((1,0),x)$ of $E_g|_{X_0}$. Under the gluing relation $\sim$, $\sigma_0|_{\partial X_0}$ is transformed to a section of the trivial bundle $\mathbb{C}^2 \times \partial B^4 \to \partial B^4$, namely $x \mapsto (g(x)(1,0),x)$. Extend this to a section σ_1 of $\mathbb{C}^2 \times B^4 \to B^4$ with transverse zeros. Glue σ_0 and σ_1 together to obtain a section σ of $E \to X$. By the lemma, $e(E_g)$—the signed count of points in $\sigma^{-1}(0)$—is the degree of the map $S^3 \to S^3$, $x \mapsto g(x)(1,0)$, i.e. with the composite

$$S^3 \xrightarrow{g} \mathrm{SU}(2) \xrightarrow{\alpha \mapsto \alpha(1,0)} S^3.$$

Since the second map is an orientation-preserving diffeomorphism, the composite has degree $\deg(g)$. Hence $e(E_g) = \deg(g)$.

Now let E be any $\mathrm{SU}(2)$-bundle. Choose a section s with transverse zeros. There is a closed neighbourhood U of $s^{-1}(0)$ which is diffeomorphic to a closed ball (*why?*).

Over $X \setminus U$, there is a unique trivialisation $\tau\colon E|_{X\setminus U} \to \mathbb{C}^2 \times (X \setminus U)$ which respects the hermitian metric and determinant form and which sends $s(x)$ to $((|s(x)|,0),x)$ (*why?*).

Over U, E is also trivial as an $\mathrm{SU}(2)$-bundle—like any bundle over a ball. Hence E is obtained by gluing together trivial bundles over $X \setminus U$ and U by some map $g\colon \partial(X \setminus U) \to \mathrm{GL}(2,\mathbb{C})$. The gluing must respect the standard hermitian metric on $\mathbb{C}^2$ and the standard volume element $e_1 \wedge e_2 \in \Lambda^2\mathbb{C}^2$; hence g takes values in $\mathrm{SU}(2)$, and $E \cong E_g$.

□

8 Connections and curvature

There are a number of points of view one can take on connections and their curvature. We will focus on just one of these: for us, a connection in a vector bundle is synonymous with a covariant derivative. This is probably the most straightforward method, though perhaps not the most geometrically transparent. In a more leisurely course one would develop the different methods and the relations between them—in particular, the interpretation of curvature as the failure of integrability of a tangent distribution on E.

Definition 30. Let X be a smooth manifold and $E \to X$ a smooth complex vector bundle. A **covariant derivative** or **connection** in E is a $\mathbb{C}$-linear map

$$\nabla\colon C^\infty(E) \to C^\infty(E \otimes_{\mathbb{R}} T^*X)$$

such that

$$\nabla(fs) = df \otimes s + f\nabla s$$

for all $s \in C^\infty(E)$ and all smooth real-valued (resp. complex valued) functions f on X.

Remark 31. *You might be more used to thinking of a covariant derivative as a collection of linear maps*

$$\nabla_v\colon C^\infty(E) \to C^\infty(E),$$

one for each vector field $v \in C^\infty(TX)$. This is the same thing: one obtains $\nabla_v s$ from ∇s by contracting with v:

$$E \otimes T^*_x X \ni e \otimes \alpha \mapsto \alpha(v)e.$$

Lemma 32. *Covariant derivatives on E form an affine space modelled on the complex vector space $C^\infty(T^*X \otimes \mathrm{End}_{\mathbb{C}}(E))$.*

Proof. The first point is that covariant derivatives exist. To see this one observes first that they exist locally. Indeed, on a trivial vector bundle $U \times \mathbb{R}^n \to U$, a section takes the form of a map $x \mapsto (x, s(x))$, so it is specified by the function $s\colon U \to \mathbb{R}^n$. The derivative of s at x is a linear map $(ds)_x \in \mathrm{Hom}(T_xU, \mathbb{R}^n) = T^*_xU \otimes \mathbb{R}^n$. The Leibnitz rule implies that $s \mapsto ds$ is indeed a covariant derivative. One can now patch together locally-defined covariant derivatives using a partition of unity. *Exercise: why does this work?*

If ∇ is a covariant derivative and $\alpha \in C^\infty(T^*X \otimes \mathrm{End}_{\mathbb{C}}(E))$ then the operator $\nabla + \alpha$ given by

$$(\nabla + \alpha)(s)(x) = (\nabla s)(x) + \alpha(s(x))$$

is another covariant derivative. Conversely, the difference $\nabla - \nabla'$ of covariant derivatives is an endomorphism-valued one-form α. To see this, we recall a general principle in differential geometry. If we have two vector bundles $E_1 \to E_2$ over X, and an operator

$$A\colon C^\infty(E_1) \to C^\infty(E_2)$$

which is linear over functions (i.e. $A(fs) = fA(s)$ for $f \in C^\infty(X)$) then there exists a homomorphism $a\colon E_1 \to E_2$ such that $(As)(x) = a(s(x))$. You should remind yourself (or work out) how to prove this. The difference $\nabla - \nabla'$ of covariant derivatives is linear over functions, hence arises from an algebraic operator in this way, i.e.

$$\nabla - \nabla' \in C^\infty(T^*X \otimes \mathrm{End}_{\mathbb{C}}(E)).$$

□

Local description. In a local trivialisation, $E|_U \cong U \times \mathbb{R}^n$, we can compare a general covariant derivative ∇ with the trivial one d. Thus

$$(\nabla s)(x) = (ds)_x + \omega_x s(x), \quad \omega \in C^\infty(T^*U \otimes \mathrm{End}(E|_U)).$$

More briefly,

$$\nabla = d + \omega.$$

We should stress that ω depends on the particular choice of trivialisation. We can write

$$s = \sum_{i=1}^n s_i e_i, \quad s_i \in C^\infty(U)$$

where $(e_1, \ldots, e_n)$ is the standard basis for $\mathbb{R}^n$. We think of ω as an $n \times n$ matrix whose entries are one-forms on U, so $\omega e_i = \sum_j \omega_{ji} e_j$. Then

$$(\nabla s)(x) = \sum_i \left(ds_i + \sum_j \omega_{ij} s_j \right) e_i.$$

So a covariant derivative is represented locally by a first-order differential operator.

In a $\mathrm{U}(n)$ or $\mathrm{SU}(n)$-bundle, there are connections compatible with the extra structure:

Definition 33. A covariant derivative ∇ in a $\mathrm{U}(n)$-bundle $(E, (\cdot,\cdot))$ is called a $\mathrm{U}(n)$-connection if

$$d(s_1, s_2) = (\nabla s_1, s_2) + (s_1, \nabla s_2)$$

for all smooth sections s_i. In an $\mathrm{SU}(n)$-bundle $(E, (\cdot,\cdot))$, an $\mathrm{SU}(n)$-connection is a $\mathrm{U}(n)$-connection which also satisfies

$$\nabla^{\Lambda^n E} o = 0.$$

Here $\nabla^{\Lambda^n E}$ is the induced covariant derivative on $\Lambda^n E$, given by

$$\nabla(s_1 \wedge \cdots \wedge s_n) = \sum_{i=1}^{n} s_1 \wedge \cdots \wedge \nabla s_i \wedge \cdots \wedge s_n.$$

The $\mathrm{SU}(n)$-connections form an affine space modelled on $C^\infty(T^*X \otimes \mathfrak{su}(E))$, where $\mathfrak{su}(E)$ is the subbundle of $\mathrm{End}_{\mathbb{C}}(E)$ of endomorphisms T satisfying

$$T + T^* = 0, \quad \mathrm{tr}(T) = 0,$$

where T^* is the hermitian adjoint endomorphism.

Pullbacks. Covariant derivatives behave well under pullback of vector bundles. If ∇ is a covariant derivative on $E \to X$, and $\phi\colon Y \to X$ is a smooth map, then there is an induced covariant derivative $\phi^*\nabla$ on ϕ^*E, characterised by the relation

$$(\phi^*\nabla)(f\phi^*s) = df \otimes \phi^*s + f\phi^*(\nabla s), \quad f \in C^\infty(Y),\ s \in C^\infty(E).$$

Curvature. Write Λ^p_X for $\Lambda^p T^*X$ (so $\Lambda^1_X = T^*X$). Given a covariant derivative $\nabla\colon C^\infty(E) \to C^\infty(\Lambda^1_X \otimes E)$ we obtain an $\mathbb{R}$-linear operator, also denoted ∇,

$$\nabla\colon C^\infty(\Lambda^1_X \otimes E) \to C^\infty(\Lambda^2_X \otimes E),$$
$$\nabla(\eta \otimes e) = d\eta \otimes e - \eta \wedge \nabla e.$$

The composite $\nabla \circ \nabla\colon C^\infty(E) \to C^\infty(\Lambda^2_X \otimes E)$ is linear over functions:

$$\nabla \circ \nabla(fs) = d^2 f \otimes s - df \otimes \nabla s + df \otimes \nabla s + f\nabla \circ \nabla s = f\nabla \circ \nabla(s).$$

Hence there exists a bundle map

$$F(\nabla)\colon E \to \Lambda^2_X \otimes E$$

such that

$$(\nabla \circ \nabla s)(x) = F(\nabla)(s(x)).$$

Since $\mathrm{Hom}_{\mathbb{C}}(E, \Lambda^2_X \otimes E) = \Lambda^2_X \otimes \mathrm{End}_{\mathbb{C}}(E)$, we can think of $F(\nabla)$ as an endomorphism-valued two-form:

$$F(\nabla) \in C^\infty(\Lambda^2_X \otimes \mathrm{End}_{\mathbb{C}}(E)).$$

This two-form is the **curvature** of ∇. In the $\mathrm{SU}(n)$ case, the curvature is a section of $\Lambda^2_X \otimes \mathfrak{su}(E)$.

Lemma 34. *If u and v are vector fields then the contraction $F(\nabla)(u,v) \in C^\infty(\mathrm{End}(E))$ is given by*

$$F(\nabla)(u,v) = \nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]}.$$

Proof. Write $\nabla s = \sum_i \eta_i \otimes \sigma_i$ for one-forms η_i and sections σ_i. Then

$$\begin{aligned} F(\nabla)(s)(u,v) &= \sum_i (d\eta_i \otimes \sigma_i - \eta_i \wedge \nabla \sigma_i)(u,v) \\ &= \sum_i (d\eta_i(u,v)\sigma_i + \eta_i(u)\nabla_v \sigma_i - \eta_i(v)\nabla_u \sigma_i); \end{aligned}$$

$$\begin{aligned} (\nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]})(s) &= \sum_i \left(u(\eta_i(v)) - v(\eta_i(u)) - \eta_i([u,v])\right) \sigma_i \\ &\quad + \sum_i \eta_i(u) \otimes \nabla_v \sigma_i - \eta_i(v) \otimes \nabla_u \sigma_i. \end{aligned}$$

So the result follows from the formula for the exterior derivative,

$$d\eta_i(u,v) = u(\eta_i(v)) - v(\eta_i(u)) - \eta_i([u,v]).$$

□

In our local trivialisation, with $\nabla = d + \omega$, we have

$$\begin{aligned} F(\nabla)(e_i) &= \nabla \left(\sum_j \omega_{ji} \otimes e_j \right) \\ &= \sum_j \left((d\omega_{ji}) \otimes e_j - \sum_k \omega_{ji} \wedge \omega_{kj} \otimes e_k \right) \\ &= \sum_j \left((d\omega_{ji}) \otimes e_j + \sum_k \omega_{kj} \wedge \omega_{ji} \otimes e_k \right), \end{aligned}$$

that is,

$$F(\nabla) = d\omega + \omega \wedge \omega$$

where $\omega \wedge \omega$ denotes the wedge product of *endomorphism-valued* one-forms, i.e. the combination of wedge product with matrix multiplication (unlike the wedge-square of ordinary one-forms, this need not vanish).

There is a similar formula for the change in the curvature when one changes the connection by an endomorphism-valued one-form:

$$F(\nabla + a) = F(\nabla) + \nabla^{\mathrm{End}(E)} a + a \wedge a.$$

Here $\nabla^{\mathrm{End}(E)} a = [\nabla, a]$, i.e. $(\nabla^{\mathrm{End}(E)} a)(s) = \nabla a(s) - a \wedge \nabla s$. This brings out the important point that the curvature varies non-linearly with the connection.

Note that curvature pulls back via the usual pullback on two-forms:

$$F(\phi^* \nabla) = \phi^* F(\nabla).$$

9 Holonomy and flat connections

9.1 Holonomy

Let $E \to X$ be an $\mathrm{SU}(n)$-bundle, ∇ a covariant derivative in E. Fix a basepoint $x \in X$. The holonomy function

$$\mathrm{hol}_\nabla \colon \Omega_x X \to \mathrm{SU}(E_x)$$

maps the set $\Omega_x X$ of piecewise-smooth loops based at x to the group of automorphisms of the fibre E_x respecting the metric and determinant form. It is compatible with concatenation in that

$$\mathrm{hol}_\nabla(\gamma_1 * \gamma_2) = \mathrm{hol}_\nabla(\gamma_1) \circ \mathrm{hol}_\nabla(\gamma_2),$$

where $\gamma_1 * \gamma_2$ means go round γ_2 then γ_1. It is defined as follows.

Given a path $\gamma \colon [0,1] \to X$, there is a pullback vector bundle $\gamma^* E \to [0,1]$, and a pullback covariant derivative $\gamma^*\nabla$. The existence and uniqueness theorem for first-order linear ODEs implies that, given $e \in E_x = E_{\gamma(0)}$, there exists a unique path $t \mapsto e(t)$ in E such that (i) $e(0) = e$, (ii) $e(t) \in E_{\gamma(t)}$, and (iii) $e(t)$ is horizontal with respect to ∇, in the sense that

$$(\gamma^*\nabla)_{\partial_t}(e(t)) = 0.$$

These paths vary linearly with the initial condition e, so one obtains *linear* maps

$$E_{\gamma(0)} \to E_{\gamma(t)}$$

by sending e to $e(t)$. These are called parallel transport maps. They are invertible—the inverse is parallel transport along the reversed path. They also respect the hermitian metrics and determinant forms (*exercise*).

In the particular case where γ is a loop, one gets at $t = 1$ an invertible linear map $E_x \to E_x$. This, by definition, is $\mathrm{hol}_\nabla(\gamma)$.

Definition 35. The **holonomy group** $\mathcal{H}_{\nabla,x}$ is the image of hol_∇ in $\mathrm{SU}(E_x)$.

If we choose an orthonormal, oriented basis for E_x, then $\mathrm{SU}(E_x)$ is identified with $\mathrm{SU}(n)$. The holonomy group $\mathcal{H}_{\nabla,x} \subset \mathrm{SU}(n)$ depends on the basis, but of course its conjugacy class does not. Moreover, we can vary the basepoint x without affecting this conjugacy class either. So we think of the holonomy group as a *conjugacy class of subgroups*

$$\mathcal{H}_\nabla \subset \mathrm{SU}(n).$$

Example 36. If E_i ($i = 1, 2$) are $\mathrm{SU}(n_i)$-bundles over X, carrying covariant derivatives ∇_i, then $E_1 \oplus E_2$ carries the covariant derivative $\nabla_1 + \nabla_2$. Its holonomy group is contained in $\mathrm{SU}(n_1) \times \mathrm{SU}(n_2) \subset \mathrm{SU}(n_1 + n_2)$.

Remark 37. *In $\mathbb{R}^n$, let u, v be linearly independent tangent vectors at the origin. Let $\gamma_{u,v,\epsilon}$ be the path starting at 0 which traces out a rhombus of side ϵ in the directions u then v. Then*

$$F_\nabla(u,v) = \lim_{\epsilon\to 0} \epsilon^{-2}\,\mathrm{hol}_\nabla(\gamma_{u,v,\epsilon}).$$

Thus curvature is an infinitesimal form of holonomy.

9.2 Gauge transformations

Definition 38. Let $p\colon E \to X$ be an $\mathrm{SU}(n)$-bundle. A **gauge transformation** is a bundle automorphism of E, i.e. a smooth map $u\colon E \to E$ such that $p\circ u = p$, which maps each fibre E_x to itself by an element of $\mathrm{SU}(E_x)$.

The gauge transformations form a group $\mathcal{G}_E$. When X is compact, it is an infinite-dimensional Lie group. The gauge group $\mathcal{G}_E$ acts on covariant derivatives by $(u,\nabla) \mapsto u\cdot\nabla$, where

$$(u\cdot\nabla)(s) = u\nabla(u^{-1}s).$$

Covariant derivatives in the same orbit under the gauge-group action are called **gauge-equivalent**. Now, u is a section of $\mathrm{SU}(E) \subset \mathrm{End}(E)$, and as such its covariant derivative $\nabla^{\mathrm{End}(E)}u$ makes sense: $\nabla^{\mathrm{End}(E)}u = [\nabla, u]$. We have

$$\nabla(u^{-1}s) = u^{-1}\nabla s + \nabla^{\mathrm{End}(E)}(u^{-1})(s) = u^{-1}\nabla s - u^{-1}(\nabla u)(u^{-1}s).$$

Therefore

$$u\cdot\nabla = \nabla - (\nabla^{\mathrm{End}(E)}u)u^{-1}.$$

In general, the curvature pulls back in the obvious way under bundle isomorphisms. Thus

$$F(u\cdot\nabla) = u^*F(\nabla) = uF(\nabla)u^{-1}.$$

Lemma 39. *Fix a basepoint $x \in X$. The map $u \mapsto u(x)$ gives an isomorphism*

$$\mathrm{stab}_{\mathcal{G}_E}(\nabla) \cong Z_{\mathrm{SU}(E_x)}(\mathcal{H}_{\nabla,x})$$

between the gauge transformations which preserve ∇ and the centraliser of the holonomy group, i.e. the elements of $\mathrm{SU}(E_x)$ which commute with the holonomies of loops based at x.

Proof. Suppose that $u\cdot\nabla = \nabla$. Recall that $u\cdot\nabla = \nabla - (\nabla^{\mathrm{End}(E)}u)u^{-1}$, hence $\nabla^{\mathrm{End}(E)}u = 0$. This implies that $u \in \mathrm{stab}_{\mathcal{G}}(\nabla)$ is uniquely determined by its value at x (*why?*). Moreover,

$$\mathrm{hol}_\nabla(\gamma) = \mathrm{hol}_{u^*\nabla}(\gamma) = u(x)\mathrm{hol}_\nabla(\gamma)u(x)^{-1}$$

(*why?*), so $u(x) \in Z_{\mathrm{SU}(E_x)}(\mathcal{H}_{\nabla,x})$. Conversely, given $u(x) \in Z_{\mathrm{SU}(E_x)}(\mathcal{H}_{\nabla,x})$, define $u(y)$ by choosing a path $t \mapsto \delta(t)$ from x to y and solving the linear ODE $(\delta^*\nabla)u(\delta(t)) = 0$. Since $u(x)$ commutes with holonomy, this does not depend on the choice of path. $\square$

Definition 40. An $\mathrm{SU}(n)$-connection is **irreducible** if, for some (hence all) $x \in X$, the holonomy group $\mathcal{H}_{\nabla,x}$ is equal to $\mathrm{SU}(E_x)$. Otherwise it is **reducible**.

We have proved that the stabiliser of ∇ under the gauge group action is isomorphic to the centraliser $Z_{\mathrm{SU}(E_x)}(\mathcal{H}_{\nabla,x})$. So for an irreducible connection, the stabiliser is as small as it can be: it is the centre $Z(\mathrm{SU}(E_x))$ of the automorphism group. In a suitable topology on the space of connections, the irreducible connections form a dense open subset [3, p.132].

9.3 Flat connections

A **flat connection** is a covariant derivative ∇ such that $F(\nabla) = 0$.

Theorem 41. *A flat* $\mathrm{SU}(n)$*-connection* ∇ *in the trivial bundle* $\mathbb{C}^n \times (-1,1)^m \to (-1,1)^m$ *is gauge-equivalent to the trivial connection.*

Proof. This proof is from Donaldson and Kronheimer [3, p.48]. What we shall do is, in effect, to use parallel transport to define a new trivialisation of the bundle, and prove that in this trivialisation the connection is trivial.

The connection can be written as $\nabla = \sum_{i=1}^m \nabla_i \otimes dx_i$, where

$$\nabla_i = \frac{\partial}{\partial x_i} + A_i, \quad A_i \colon (-1,1)^m \to \mathfrak{su}(n).$$

The curvature coefficients $F_{ij} = F(\nabla)(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j})$ are given by

$$F_{ij} = [\nabla_i, \nabla_j],$$

so the flatness condition is that the 'partial covariant derivatives' ∇_i commute with one another. We seek a gauge transformation $u \colon (-1,1)^m \to \mathrm{SU}(n)$ such that $u^*\nabla = d$, i.e. such that $u\nabla_i u^{-1} = \frac{\partial}{\partial x_i}$ for $i = 1, \dots, m$. Expanding the left-hand side, we can write these equations as

$$\frac{\partial u}{\partial x_i} = uA_i.$$

We proceed by induction on m. The case $m = 0$ is trivial. So we suppose that we have found a trivialisation such that $\nabla_i = \partial_i$ (i.e. $A_i = 0$) for $i = 1, \dots, m-1$. We regard the equation

$$\frac{\partial u}{\partial x_m} = uA_m(x_1, \dots, x_{m-1}, \cdot) \tag{3}$$

as a first-order linear ODE in x_m. It has a unique solution $u(x_1, \dots, x_{m-1}, \cdot)$ which satisfies the initial condition

$$u(x_1, \dots, x_{m-1}, 0) = \mathrm{id}.$$

Since the solutions to linear ODEs vary smoothly with parameters, u is a smooth function $(-1,1)^m \to \mathrm{GL}(n,\mathbb{C})$. We claim that the gauge-transformation u does the job. To show this we must prove that we have not spoilt the condition $\nabla_i = \partial_i$ for $i < m$. That is, we must prove that

$$\frac{\partial u}{\partial x_i} = 0, \quad i = 1, \ldots, m-1. \tag{4}$$

The flatness condition says that

$$0 = \left[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_m} + A_m\right] = \frac{\partial A_m}{\partial x_i}, \quad i < m.$$

Hence A_m is a function of x_m alone, independent of $x_1, \ldots, x_{m-1}$. Therefore the solution $u(x_1, \ldots, x_{m-1}, \cdot)$ to (3) with its initial condition is also independent of $x_1, \ldots, x_{m-1}$, i.e. (4) holds.

We must check that u takes values in $\mathrm{SU}(n)$. Clearly $\frac{\partial}{\partial x_i} u^* u = 0$ for $i < m$; and

$$\frac{\partial}{\partial x_m}(u^* u) = \frac{\partial u^*}{\partial x_m} u + u^* \frac{\partial u}{\partial x_m} = A_m^* u^* u + u^* u A_m = A_m^* + A_m = 0.$$

Hence $u^* u = \mathrm{id}$. Similarly, one checks $\det(u) = 1$. $\square$

Corollary 42. *Let ∇ be a flat $\mathrm{SU}(n)$ connection in an $\mathrm{SU}(n)$-bundle $E \to X$. Then the holonomy of loops based at a point $x \in X$ is invariant under homotopy, and so induces a homomorphism*

$$\mathrm{hol}_\nabla \colon \pi_1(X, x) \to \mathrm{SU}(E_x).$$

Proof. Any homotopy of loops is a product of homotopies supported in balls. The theorem shows that holonomy is invariant under these 'small' homotopies. $\square$

Theorem 43. *Holonomy sets up a bijection between*

1. *isomorphism classes of pairs (E, ∇) of $\mathrm{SU}(n)$-bundle with flat connection, and*
2. *representations $\pi_1(X, x) \to \mathrm{SU}(n)$, modulo $\mathrm{SU}(n)$ acting by conjugation.*

Proof. We have seen how a flat connection gives a representation of π_1. To go in the opposite direction, take $\rho \colon \pi_1(X, x) \to \mathrm{SU}(n)$. The fundamental group acts on the right on the universal cover $\widehat{X}$ of X. Define

$$E_\rho = \widehat{X} \times_{\pi_1(X)} \mathbb{C}^n = (\widehat{X} \times \mathbb{C}^n)/(x \cdot \gamma, z) \sim (x, \rho(\gamma) z), \quad \gamma \in \pi_1(X, x).$$

Then E_ρ is naturally a bundle over X, and the trivial connection in $\widehat{X} \times \mathbb{C}^n \to \widehat{X}$ descends to give a flat connection in E_ρ. *Exercise*: fill in the details. $\square$

10 Anti-self-dual connections

Recall (Lecture 4) that on a Riemannian four-manifold (X, g) we have an involution $*$ on the bundle of two-forms, and that its ± 1-eigenspaces effect a decomposition

$$\Lambda^2_X = \Lambda^+ \oplus \Lambda^-$$

into two 3-dimensional subbundles. Their sections are called self-dual or anti-self-dual two-forms. We write ω^+ for the projection of a two-form ω to Λ^+, i.e. $\omega^+ = (\omega + *\omega)/2$; similarly for ω^-. A two-form with values in some auxiliary bundle, say $F \in C^\infty(\Lambda^2_X \otimes \xi)$, can be projected to its self-dual part (a section of $\Lambda^+ \otimes \xi$) in just the same way.

Definition 44. Let (X, g) be an oriented Riemannian four-manifold and $E \to X$ a $\mathrm{U}(n)$ or $\mathrm{SU}(n)$-bundle. An **anti-self-dual connection** or **instanton** is a $\mathrm{U}(n)$- (resp. $\mathrm{SU}(n)$-) connection ∇ on E such that

$$F(\nabla)^+ = 0.$$

We will be concerned only with SU(2)-bundles and (briefly) U(1)-bundles.[5]

The ASD equation $F(\nabla)^+ = 0$ has an enormous symmetry group. If u is an $\mathrm{SU}(n)$ gauge transformation then

$$F(u \cdot \nabla)^+ = (uF(\nabla)u^{-1})^+ = uF(\nabla)^+u^{-1},$$

since conjugation by u affects the endomorphism part of $F(\nabla)$ but not the two-form part. Hence the action of the gauge group on connections preserves the ASD connections.

Remark 45. *The Hodge star does not change if one makes a conformal change in the metric, i.e. rescales it by a positive function. Hence anti-self-duality is also a conformally-invariant condition.*

10.1 The Yang-Mills functional

Let $E \to X$ be an $\mathrm{SU}(n)$-bundle over a compact Riemannian manifold (X, g). The **Yang-Mills functional** on the space of unitary connections is the function which sends a connection to the L^2-norm of its curvature:

$$\|F(\nabla)\|_{L^2} = \int_X |F(\nabla)|^2 \mathrm{vol}_g.$$

Here $F(\nabla)$ is a section of $\Lambda^2_X \otimes \mathfrak{su}(E)$, and we should make clear which pointwise norm is being used. It comes from a hermitian metric on $\Lambda^2_X \otimes \mathfrak{su}(E)$. The chosen

[5] In a systematic account one would consider other compact groups of rank 1, i.e. U(2) and $\mathrm{PU}(2) = \mathrm{SO}(3)$. Higher rank bundles bring additional difficulties, arising from reducible connections, but so far no new topological information.

hermitian metric on E induces one on $\mathrm{End}(E)$: for an endomorphism A with matrix entries A_{ij}, we put

$$|A|^2 = \mathrm{tr}(AA^*) = \sum_{i,j} |A_{ij}|^2.$$

This restricts to $\mathfrak{su}(E)$ as $|A|^2 = -\,\mathrm{tr}(A^2)$. We combine it with the Riemannian metric on two-forms and put

$$|\eta \otimes A|^2 = g(\eta,\eta)|A|^2.$$

Let us work out the Euler-Lagrange equations for this functional—the condition for a connection ∇ to be a stationary point for Yang-Mills. Let $D = \nabla^{\mathrm{End}(E)}$. A stationary connection satisfies

$$\begin{aligned} 0 &= \left.\frac{d}{dt}\right|_{t=0} \|F(\nabla + ta)\|^2 = \left.\frac{d}{dt}\right|_{t=0} \|F(\nabla) + tDa + t^2 a \wedge a\|^2 \\ &= 2\int_X (F(\nabla), Da)\mathrm{vol}_g \\ &= 2\int_X (D^*F(\nabla), a)\mathrm{vol}_g. \end{aligned}$$

Since this must hold for every a, we find that

$$D^*F(\nabla) = 0.$$

This is the **Yang-Mills equation**. It is a second order PDE (that is, in local coordinates, it is a second-order PDE in the coefficients of the connection matrix).

Remark 46. *For any connection one has the Bianchi identity $DF(\nabla) = 0$. This brings out the formal similarity of the Yang-Mills equation to the equations defining harmonic forms, $d\alpha = 0$, $d^*\alpha = 0$.*

Yang-Mills on a four-manifold. Although the Yang-Mills functional makes sense over manifolds of arbitrary dimension, it has a special property in the four-dimensional case: *it is invariant under conformal changes to the metric*. A conformal change is a rescaling, $g(v_1,v_2) \rightsquigarrow \lambda g(v_1,v_2)$, where $\lambda \colon X \to (0,\infty)$. If $(e^1,\dots,e^n)$ is an g-orthonormal basis for T_x^*X, then $(\lambda^{-1/2}e^1,\dots,\lambda^{-1/2}e^n)$ is a λg-orthonormal basis. So the volume element and the norm on sections of $T^*X \otimes \mathrm{End}(E)$ scale as

$$\mathrm{vol}_{\lambda g} = \lambda^{-n/2}\mathrm{vol}_g, \quad |\cdot|^2_{\lambda g} = \lambda^2 |\cdot|^2_g.$$

Hence

$$\|F(\nabla)\|^2_{L^2(\lambda g)} = \int_X \lambda^{2-n/2}|F(\nabla)|^2_g \mathrm{vol}_g.$$

So when $n = 4$, $\|F(\nabla)\|_{L^2(\lambda g)} = \|F(\nabla)\|_{L^2(g)}$.

Another key feature of four-dimensional Yang-Mills theory is its interaction with anti-self-duality. The decomposition of two-forms into self-dual and anti-self-dual parts is orthogonal in the sense that $\alpha \wedge \beta = 0$ when $\alpha \in \Omega^+_X$ and $\beta \in \Omega^-_X$. Hence for any two-form ω

$$\omega^2 = (\omega^+)^2 + (\omega^-)^2 = (|\omega^+|^2 - |\omega^-|^2)\mathrm{vol}_g.$$

The decomposition is *also* orthogonal with respect to the Riemannian metric:

$$|\omega|^2 = |\omega^+|^2 + |\omega^-|^2.$$

This gives us an interpretation of anti-self-duality:

$$\omega^+ = 0 \quad \Leftrightarrow \quad -\omega^2 = |\omega|^2\mathrm{vol}_g.$$

We now apply this to the curvature $F = F(\nabla)$ of a connection. Taking traces of the equation

$$F^2 = (F^+)^2 + (F^-)^2,$$

we find

$$-\operatorname{tr} F^2 = (|F^+|^2 - |F^-|^2)\mathrm{vol}_g.$$

There are two conclusions we can draw from this. The first is a pointwise criterion for anti-self-duality of a connection ∇:

$$\nabla \text{ is ASD} \quad \Leftrightarrow \quad \operatorname{tr}(F(\nabla)^2) = |F(\nabla)|^2\mathrm{vol}_g.$$

The second is an integral criterion. We have an inequality

$$\|F(\nabla)\|^2_{L^2} \geq \int_X \operatorname{tr} F(\nabla)^2, \tag{5}$$

and equality holds iff ∇ is ASD. The integral on the right is actually independent of the connection by Stokes' theorem, since we have the following identity:

$$\operatorname{tr} F(\nabla + a)^2 - \operatorname{tr} F(\nabla)^2 = d \operatorname{tr}\left[\left(F_{\nabla+a} + F_\nabla - \frac{1}{3} a \wedge a\right) \wedge a\right],$$

This gives the

Proposition 47. *An ASD connection is a global minimum for the Yang-Mills functional, and in particular, satisfies the Yang-Mills equations.*

10.2 ASD connections in U(1)-bundles

Proposition 48. *Let L be a* U(1)*-bundle over a closed, simply connected four-manifold. Then (i) there is at most one gauge-orbit of ASD connections in L; (ii) if Q_X is negative-definite, L admits an ASD connection.*

Proof. Since $\mathrm{End}(L)$ is canonically trivial, the curvature of a U(1)-connection is a two-form with values in $\mathfrak{u}(1) = \mathrm{i}\mathbb{R}$. When one changes the connection from ∇ to $\nabla + a$, the curvature changes by $[\nabla, a] + a \wedge a = da + 0 = da$. Hence $\omega \in \mathrm{i}\Omega^2_X$ is the curvature form of a U(1)-connection iff it lies in the same cohomology class as $\mathrm{i}F(\nabla)$. An imaginary one-form a is therefore a global minimum for $a \mapsto \|F(\nabla + a)\|^2_{L^2}$ iff $F(\nabla + a)$ is a harmonic form.

(i) If ∇ is ASD, then $F(\nabla)$ is harmonic. The uniqueness of harmonic forms tells us that if ∇ and $\nabla + a$ are both ASD then $da = 0$. Thus the ASD connections (if any exist) form an affine space modelled on the vector space of closed one-forms. A gauge transformation $u \in \mathcal{G}_L$ sends the ASD connection $\nabla + a$ to $\nabla + a + (du)u^{-1}$, using commutativity of U(1). If u has a logarithm, say $u = e^{\mathrm{i}f}$, then $u^{-1}du = \mathrm{i}df$. When X is simply connected, every closed one-form is exact, and every gauge transformation has a logarithm. This gives the claimed uniqueness.

(ii) If Q_X is negative-definite then any harmonic form is ASD. Hence the Hodge theorem implies existence of ASD connections. $\square$

10.3 ASD connections in SU(2)-bundles

As pointed out above, the integral of $\mathrm{tr}\, F(\nabla)^2$ is a topological quantity, depending on X and E but not on ∇. In an SU(2)-bundle, we can choose a connection which is trivial away from a finite set of balls B_i, and standard over each of the B_i. Hence it must be some universal multiple of the Euler number. So to prove the following result one just has to compute one example:

Theorem 49. *For any* SU(2)*-connection E over a compact four-manifold X, we have*

$$e(E)[X] = \frac{1}{8\pi^2} \int_X \mathrm{tr}\, F(\nabla)^2.$$

Combining this with (5), we arrive at the following principle:

Theorem 50. *Any* SU(2)*-connection ∇ satisfies*

$$\|F(\nabla)\|^2_{L^2} \geq 8\pi^2 e(E).$$

Equality holds iff ∇ is ASD.

Corollary 51. *There are no ASD connections in an* SU(2)*-bundle with $e < 0$. In a bundle with $e = 0$, a connection is ASD iff it is flat: $F(\nabla) = 0$. In particular, in a trivial bundle over a simply connected four-manifold, any ASD connection is gauge-equivalent to the trivial connection.*

11 Donaldson's diagonalisability theorem

We now state the main theorem of the course.

Theorem 52. *Let X be a compact, oriented, simply connected, smooth four-manifold such that the intersection form Q_X is positive-definite. Then Q_X is equivalent to the diagonal form $\langle 1, \dots, 1\rangle$. That is, there exists a basis $(e_1, \dots, e_{b_2(X)})$ for the free abelian group $H_2(X;\mathbb{Z})$ such that $Q_X(e_i, e_j) = \delta_{ij}$.*

Remark 53. *(i) It is known that the theorem holds for arbitrary fundamental groups (of course, one must replace $H_2(X;\mathbb{Z})$ by $H_2(X;\mathbb{Z})/\text{torsion}$).*

(ii) If Q_X is negative-definite then it is equivalent to $\langle -1, \dots, -1\rangle$, as we see by applying the theorem to X with the opposite orientation.

The theorem is deduced as a corollary of the following

Theorem 54. *Let X be a compact, oriented, simply connected four-manifold such that the intersection form Q_X is positive-definite. Then there exists an oriented cobordism from X to the disjoint union of $q(X)$ copies of $\mathbb{C}P^2$ (perhaps with different orientations), where*

$$q(X) = \frac{1}{2}\#\{x \in H_2(X;\mathbb{Z}) : Q_X(x,x) = 1\}.$$

Note that $q(X) \in \mathbb{Z}$ because if $Q_X(x,x) = 1$ then $Q_X(-x,-x) = 1$ also.

We need a simple algebraic lemma.

Lemma 55. *Let $(\cdot,\cdot)$ be a positive-definite symmetric bilinear form on a free abelian group $A \cong \mathbb{Z}^n$. Then there are at most n unordered pairs $\{a, -a\}$ where $a \in A$ satisfies $(a,a) = 1$. Equality holds if and only if $(\cdot,\cdot)$ is equivalent to the standard form $\langle 1\rangle^n$.*

Proof. Take $a \in A$ with $(a,a) = 1$. Then the identity

$$x = (x,a)a + \big(x - (x,a)a\big)$$

shows that we have an orthogonal decomposition

$$A = \langle 1\rangle \oplus a^{\perp}.$$

The lemma follows by induction on n. □

Proof of the diagonalisability theorem. Since X is cobordant to $q = q(X)$ copies of $\mathbb{C}P^2$, and signature is a cobordism invariant, we have

$$\sigma(X) = \varepsilon_1 + \cdots + \varepsilon_q$$

where $\varepsilon_i = \pm 1$ is the signature of the ith copy of $\mathbb{C}P^2$ (the sign depends on its orientation). Since Q_X is positive-definite, its signature $\sigma(X)$ is equal to its rank, so

$$\operatorname{rank}(Q_X) = \sigma(X) \leq q.$$

By the lemma, $\operatorname{rank}(Q_X) = q$ and Q_X is equivalent to the standard form $\langle 1\rangle^{\operatorname{rank}(Q_X)}$. □

Remark 56. *We can see in retrospect that the q copies of $\mathbb{C}P^2$ in Theorem 54 must all be oriented so as to have signature $+1$.*

Deducing Theorem 54 from the structure of the moduli space. We need to produce a cobordism Z from our four-manifold X to the disjoint union of q $\mathbb{C}P^2$s. In the proof it is convenient to reverse the orientation of X, so that Q_X is now *negative*-definite.

The idea is to consider the **ASD moduli space** space $\mathcal{M}_{X,E,g}$ of g-ASD: connections modulo gauge transformations in an SU(2)-bundle E with $e(E) = 1$. The topology is inherited from an (affine) topological vector space structure on the space of connections.

Theorem 57. *Let X be negative-definite and simply connected. There is a C^0-dense set of metrics g such that $\mathcal{M}_{X,g}$ has the following properties:*

1. *The subspace $\mathcal{R} \subset \mathcal{M}_{X,E,g}$ of gauge-equivalence classes of reducible connections contains precisely q points. Its complement $\mathcal{M}^*_{X,E,g} = \mathcal{M}_{X,E,g} \setminus \mathcal{R}$ has a natural structure of smooth five-dimensional manifold.*

2. *For each $s \in \mathcal{R}$, there is a neighbourhood U_s and a homeomorphism*

$$(U_s, s) \to (\mathbb{C}^3/S^1, [0]),$$

smooth on $U_s \setminus \{s\}$. Here $S^1 \subset \mathbb{C}$ acts on $\mathbb{C}^3$ by scalar multiplication.

3. *There is an open subset $C \subset \mathcal{M}_{X,E,g}$ such that the complement $\mathcal{M}_{X,E,g} \setminus C$ is compact and contains $\mathcal{R}$, and a diffeomorphism*

$$C \cong X \times (0, 1).$$

4. *$\mathcal{M}_{X,E,g} \setminus \mathcal{R}$ is orientable.*

To deduce Theorem 54 from this result we modify $\mathcal{M}_{X,E,g}$ in two ways:

(i) Since $C \cong X \times [0, 1)$, we can obtain a compact space $\overline{\mathcal{M}}_{X,g}$ by slicing off the end. That is, we remove $X \times (0, 1)$ and replace it by $X \times (0, \frac{1}{2}]$ so as to produce a boundary component X. (Alternatively, we can compactify $X \times (0, 1)$ to $X \times (0, 1]$, adjoining a copy of X.)

(ii) We chop off the q singular points $s \in \mathcal{R}$. That is, we excise small neighbourhoods $B_\epsilon/S^1 \subset \mathbb{C}^3/S^1 = U_s \subset \overline{\mathcal{M}}_{X,g}$, where $B_\epsilon \subset \mathbb{C}^3$ is the open ball of radius ϵ.

After modifying $\mathcal{M}_{X,E,g}$ in these two ways, we obtain a space Z which is a compact, orientable manifold with boundary. The boundary is the union of X and q copies of $S^5/S^1 = \mathbb{C}P^2$. This proves Theorem 54.

Remark 58. *In this course, points (1) and (2) will be proved, modulo various background results on elliptic operators and Coulomb gauge slices. In particular, we will show that if $q(X) > 0$, $\mathcal{M}_{X,E,g}$ is non-empty.*

For (3), we will identify the open set C, prove (modulo Uhlenbeck's analytic results on connections) that its complement is compact, and write down a suitable map $C \to X \times (0,1)$. This explains the geometric significance of compactifying the moduli space by adjoining a copy of X. The proof that the map is a diffeomorphism is the technical heart of Donaldson's argument, and we will not give it. We will not even prove that C (or indeed $\mathcal{M}_{X,E,g}$) is non-empty.

The orientability property (4) could be covered in a lecture or two, but we do not have time to explain it.

We can immediately explain the first point of the theorem.

Proposition 59. *Under the hypotheses of the structure theorem (57), there are exactly $q(X)$ gauge-orbits of reducible ASD connections.*

Recall that a connection ∇ is reducible if its holonomy group $\mathcal{H}_\nabla$ is properly contained in $\mathrm{SU}(2)$.

Proof of the proposition. We construct a map

$$f\colon \mathcal{R} \to \{c \in H^2(X;\mathbb{Z}) : c^2 = -1\}/\pm 1,$$

where $\mathcal{R}$ is the set of reducible ASD connections modulo gauge-equivalence.

We have to quote a general property of connections in (say) $\mathrm{U}(n)$-bundles (Ambrose-Singer, 1953): *The holonomy group $\mathcal{H}_{\nabla,x} \subset \mathrm{Aut}(E_x)$ is a closed subgroup. Its Lie algebra is the sub-Lie algebra of $\mathfrak{u}(E_x)$ generated by the curvatures $F_\nabla(u,v)$, $u, v \in T_xX$.*

Since X is simply connected, the holonomy group is path connected. So the holonomy of a reducible connection ∇ is a connected Lie subgroup of $\mathrm{SU}(2)$. A connected Lie subgroup in a Lie group is determined by its Lie algebra, hence the only possibilities are

1. The Lie algebra is 0, and the holonomy group is trivial.
2. The Lie algebra is one-dimensional, and the holonomy group is conjugate to the diagonal subgroup $S^1 \subset \mathrm{SU}(2)$.

A connection which has trivial holonomy is flat, by the Ambrose-Singer theorem. But an $\mathrm{SU}(2)$-bundle which carries a flat connection necessarily has $e = 0$ (since $-8\pi^2 e$ is the integral of $\operatorname{tr} F^2$) whereas we have assumed that $e = 1$. So the reducible connections in our bundle have holonomy $S^1 \subset \mathrm{SU}(2)$. The stabiliser of ∇ in $\mathcal{G}_E$ (consisting of gauge transformations u with $\nabla u = 0$) is isomorphic to $Z_{\mathrm{SU}(2)}(S^1) = S^1$.

We can give a geometric interpretation of the fact that $\mathcal{H}_{\nabla,x} \subset \mathrm{Aut}(E_x)$ is isomorphic to $S^1 \subset \mathrm{SU}(2)$:

At every point $x \in X$, there exist complex lines $L_x, L'_x \subset E_x$, such that $E_x = L_x \oplus L'_x$, each preserved by the holonomy group $\mathcal{H}_{\nabla,x}$. These lines are unique (up to reordering).

Globally, these lines trace out *line bundles* L, L' with $E = L \oplus L'$. We can decide which line is L_x and which L'_x in a coherent way, because the possible choices make a two-fold covering space of X. Since X is simply-connected, this must be trivial. Because E is an SU(2)-bundle, L' is isomorphic to the conjugate line bundle $\bar{L}$.

Choose transverse sections s of L and s' of L'. Their zero-sets represent classes c and $-c$. The Euler number of $E = L \oplus L'$ is given by the count of zeros of $s \oplus s'$; thus $e(E) = -c \cdot c$. By hypothesis, $e(E) = 1$; hence $c^2 = -1$. The map f sends a reducible ASD connection to the class c arising in this way.

We now construct a map

$$g \colon \{c \in H^2(X; \mathbb{Z}) : c^2 = -1\}/\pm 1 \to \mathcal{R}.$$

For this we need to know that isomorphism classes of line bundles over X (as over any CW complex[6]) are in bijection with $H^2(X; \mathbb{Z})$. The bijection sends L to its Euler class $e(L) \in H^2(X; \mathbb{Z})$, which is Poincaré dual to the zero-set of a transverse section.

Given c with $c^2 = -1$, there exists a line bundle L, unique up to isomorphism, with $e(L) = c$. We proved in lecture 10 that (X being simply connected and negative-definite) L carries an ASD connection, unique up to the action of $\mathcal{G}_L$. The conjugate line bundle $\bar{L}$ likewise carries an ASD connection ∇'. The direct sum $L \oplus \bar{L}$ is an SU(2)-bundle with Euler number 1, hence it is isomorphic to E. The isomorphism carries $\nabla \oplus \nabla'$ to an ASD connection. Its equivalence class is $g(c)$.

It is easy to see that g is inverse to f. Hence f is bijective.

$\square$

[6]U(1)-bundles over Y (resp. classes in $H^2(Y; \mathbb{Z})$) correspond to homotopy classes of maps from Y into a classifying space $B\,\mathrm{U}(1)$ (resp. into a $K(\mathbb{Z}, 2)$). The space $\mathbb{C}P^\infty$ carries a U(1)-bundle with contractible total space, and is therefore a model for $B\,\mathrm{U}(1)$. It also has trivial homotopy groups, with exception of $\pi_2 = \mathbb{Z}$, which makes it a $K(\mathbb{Z}, 2)$.

12 Instantons over $\mathbb{R}^4$

In this lecture we shall write down a five-parameter family of instantons in the $\mathrm{SU}(2)$-bundle with Euler number 1 over $\mathbb{R}^4$. An open set in the parameter space gives rise to instantons whose curvature density is concentrated near a point in $\mathbb{R}^4 \subset S^4$. Such a solution can be transplanted into a chosen manifold X. The resulting connection will not be precisely ASD, but its self-dual curvature is sufficiently small that the connection can be perturbed so as to become ASD (via a careful application of the implicit function theorem).

This has the important consequence that *the moduli space* $\mathcal{M}_{X,E,g}$ *is non-empty.* The instantons which arise this way make up a collar neighbourhood of the boundary component X in the compactified moduli space.

Remark 60. *In this lecture we use quaternionic methods. The entire discussion (construction of connections, action of the group of conformal transformation) has a precise analogue with* $\mathbb{C}$ *replacing* $\mathbb{H}$*, and* $\mathrm{U}(1)$*-bundles over* $\mathbb{C}$ *replacing* $\mathrm{SU}(2)$*-bundles over* $\mathbb{R}^4$*, and it is instructive to work this through. In the quaternionic case, the bundles and connections extend over the one-point compactification* S^4 *of* $\mathbb{R}^4$*. In the complex case, they extend over* S^2*.*

12.0.1 The basic instanton.

(i) We use the quaternions $\mathbb{H}$, i.e. the $\mathbb{R}$-algebra generated by $\{i, j, k\}$ subject to the relations $i^2 = j^2 = k^2 = ijk = -1$. Recall the standard notation

$$q = q_0 + q_1 i + q_2 j + q_3 k, \qquad \bar{q} = q_0 - q_1 i - q_2 j - q_3 k,$$
$$|q|^2 = q\bar{q}, \qquad \mathrm{Im}(q) = \frac{1}{2}(q - \bar{q}).$$

$\mathbb{H}$ is a Lie algebra with bracket $[p, q] = pq - qp$. It splits as the direct sum of two sub-Lie algebras, the reals $\mathbb{R}$ and the imaginary quaternions $\mathrm{Im}(\mathbb{H})$. The latter is isomorphic, via its adjoint action on $\mathbb{H} = \mathbb{C}\langle 1, j\rangle$, to $\mathfrak{su}(2)$.

(ii) Writing $dq = dq_0 + dq_1 i + dq_2 j + dq_3 k$, we note that the expression

$$\begin{aligned} d\bar{q} \wedge dq/2 = &(dq_0 \wedge dq_1 - dq_2 \wedge dq_3)i \\ &+(dq_0 \wedge dq_2 - dq_3 \wedge dq_1)j \\ &+(dq_0 \wedge dq_3 - dq_1 \wedge dq_2)k \end{aligned}$$

is an ASD two-form with values in $\mathrm{Im}\,\mathbb{H} = \mathfrak{su}(2)$.

(iii) We can define $\mathrm{SU}(2)$-connections in the trivial bundle over $\mathbb{R}^4 = \mathbb{H}$ by writing $\nabla_A = d + A$, where

$$A = \mathrm{Im}(f\,dq) \in \Omega^1(\mathbb{H}; \mathfrak{su}(2)), \quad f \colon \mathbb{H} \to \mathbb{H}.$$

Expanding $f dq$ as $\sum a_n dq_n$, we have $A = \sum \mathrm{Im}(a_n) dq_n$. The curvature is

$$\begin{aligned} F(\nabla_A) &= dA + A \wedge A \\ &= \sum_m d(\mathrm{Im}\, a_m) \wedge dq_m + \frac{1}{2} \sum_{m,n} [\mathrm{Im}\, a_m, \mathrm{Im}\, a_n]\, dq_m \wedge dq_n \\ &= \sum_m \mathrm{Im}(da_m) \wedge dq_m + \frac{1}{2} \sum_{m,n} [a_m, a_n]\, dq_m \wedge dq_n \\ &= \mathrm{Im}(df \wedge dq + f dq \wedge f dq). \end{aligned}$$

With these preliminaries in place, define the **basic instanton**

$$A(q) = \mathrm{Im}\left(\frac{\bar{q}\, dq}{1 + |q|^2} \right) \in \Omega^1(\mathbb{H}; \mathfrak{su}(2)).$$

Its curvature is

$$\mathrm{Im}\left(\frac{d\bar{q} \wedge dq}{1 + |q|^2} - \frac{\bar{q}\, dq + d\bar{q}\, q}{(1 + |q|^2)^2} \wedge \bar{q}\, dq + \frac{\bar{q}\, dq \wedge \bar{q}\, dq}{(1 + |q|^2)^2} \right) = \mathrm{Im}\left(\frac{d\bar{q} \wedge dq}{(1 + |q|^2)^2} \right).$$

The expression inside the brackets is itself imaginary, hence

$$F(\nabla_A) = \frac{d\bar{q} \wedge dq}{(1 + |q|^2)^2}.$$

Hence $F(\nabla_A)$ is ASD, i.e., ∇_A is an ASD connection.

Remark 61. *The formula for $A(q)$ is fairly simple, but it would take a stroke of inspiration to write it down directly. Its proper geometric home is actually in a bundle over over S^4. That setting reveals that the basic instanton has a high degree of symmetry, which is (partially) broken by restricting to $\mathbb{R}^4$. In the language of principal bundles, ∇_A corresponds to a tangent distribution in S^7. The latter is a principal $\mathrm{SU}(2)$-bundle over S^4 (by the quaternionic analogue of the Hopf map). The distribution consists of tangent vectors which are orthogonal to tangents to the fibres of $S^7 \to S^4$ (with respect to the standard metric on S^7, restricted from $\mathbb{R}^8$).*

12.0.2 Conformal transformations

If we take a conformal automorphism of standard Euclidean four-space $(\mathbb{R}^4, g)$, i.e. a diffeomorphism $f \colon \mathbb{R}^4 \to \mathbb{R}^4$ such that $f^* g = \lambda g$ for some positive function λ, then the pullback of a g-ASD connection ∇ by f is ASD with respect to λg. But then, by the conformal invariance of the ASD condition, it is also g-ASD.

We apply this to the basic instanton. Two basic examples of conformal maps are translations and dilations. Thus we consider the pullback of ∇ by the map

$$m \colon q \mapsto \lambda(q - c), \quad c \in \mathbb{H}, \quad \lambda > 0,$$

i.e. the map

$$\nabla_{m^*A} = d + m^*A.$$

Exercise. Calculate m^*A and $F(\nabla_{m^*A})$. Show that the curvature density of the transformed instanton is

$$|F(m^*A)|^2 = |m^*F(A)|^2 = \frac{48\lambda^2}{(1+\lambda^2|q-c|^2)^4}.$$

Show that

$$\int_{\mathbb{R}^4} (|F(\nabla_{m^*A})|^2)\mathrm{vol}_g = 8\pi^2.$$

(Note that the volume of the unit 3-sphere is $2\pi^2$).

The curvature density $|F|^2$ is a gauge-invariant quantity. But $|F(\nabla_{m^*A})|^2$ has its maximum at c; the maximum is $48\lambda^2$. Hence these instantons are all inequivalent under the gauge-group action.

Remark 62. *The fact that the integral of the curvature density is $8\pi^2$ shows that, if these instantons extend to connections in a bundle E over S^4 (the one-point compactification of $\mathbb{R}^4$), then E has Euler number $+1$. These connections do indeed extend.*

In this way we obtain a family of gauge-inequivalent instantons parametrised by $\mathbb{H} \times \mathbb{R}^+$. The $\mathbb{H}$-parameter determines the 'centre' of the instanton, i.e. the point where the curvature density achieves its maximum; the $\mathbb{R}^+$-parameter determines the 'scale', i.e. how concentrated the curvature is. As $\lambda \to \infty$, the curvature density approaches a delta-function centred at $c \in \mathbb{H}$.

So this open end of the parameter space, in which the parameter λ goes to infinity, has a definite meaning in terms of instantons.

13 The compactness theorem

Suppose X is a closed, oriented, smooth four-manifold. Let $E \to X$ be the $\mathrm{SU}(2)$-bundle with Euler number $+1$.

Theorem 63 (Uhlenbeck-Donaldson). *Let ∇_n be a sequence of ASD connections in E. Then there is a subsequence $\nabla_{n'}$ such that one of the following two things occurs:*

- *There are gauge transformations u_n' such that $u_n(\nabla_{n'})$ converges in C^∞ to an ASD connection.*
- *There is a point $x \in X$ with the following property. There are trivialisations $\rho_{n'} \colon (X \setminus \{x\}) \times \mathbb{C}^2 \to E|_{X\setminus\{x\}}$, such that (i) $\rho^*_{n'}(\nabla_{n'}|_{X\setminus\{x\}})$ converges, in C^∞ over compact subsets, to a flat connection ∇ on the trivial bundle; and (ii) the curvature densities $|F(\nabla_n')|^2$ converge as measures to $8\pi^2\delta_x$. That is, for any $f \in C^0(X)$, we have*

$$\frac{1}{8\pi^2}\int_X f|F(\nabla_{n'})|^2 \mathrm{vol}_g \to f(x).$$

Convergence in C^∞ of a sequence of connections ∇_n over a compact set K can be defined as follows. Fix a reference connection ∇, and let $A_n = \nabla_n - \nabla$. Then A_n is required to converge uniformly to a limiting one-form A_∞ (as measured using the norm on forms supplied by the Riemannian metric); and for any sequence of vector fields $v_1, \dots, v_k$, the iterated covariant derivative $\nabla_{v_1} \dots \nabla_{v_k} A_n$ should converge uniformly to $\nabla_{v_1} \dots \nabla_{v_k} A_\infty$.

Remark 64. *When X is simply connected, every flat connection is equivalent to the trivial connection. Thus we can take the limiting connection ∇ in the statement of the theorem to be the trivial connection.*

We will conceal the analytic difficulty of the proof by quoting the following two results:

Theorem 65 (Uhlenbeck's removable singularities theorem). *Consider the four-ball B^4, equipped with a metric g. If ∇ is an ASD connection in the trivial* $\mathrm{SU}(2)$*-bundle over $B^4 \setminus \{0\}$, satisfying*

$$\int_{B^4\setminus\{0\}} |F(\nabla)|^2 \mathrm{vol}_g < \infty,$$

then there is a smooth ASD connection over B^4, gauge-equivalent to ∇ over $B^4 \setminus \{0\}$.

Lemma 66. *There exists a universal constant $\epsilon > 0$ such that the following holds. Given an oriented Riemannian four-manifold (M, g) (not necessarily* *compact), and a sequence ∇_n of ASD connections in an* $\mathrm{SU}(2)$*-bundle E such that every $x \in M$ is contained in a geodesic ball D_x with*

$$\int_{D_x} |F(\nabla_n)|^2 \mathrm{vol}_g \leq \epsilon^2,$$

there exists a subsequence $\nabla_{n'}$ and gauge transformations $u_{n'}$ such that $u_{n'}(\nabla_{n'})$ converges in C^∞ over compact subsets of M.

Proof of the compactness theorem. **1.** The first observation is that, after passing to a subsequence, the curvature densities converge to a limiting measure ν:

The curvature densities of the ∇_n define positive Borel measures $\nu_n \in C^0(X)^*$ (so $\nu(f) \geq 0$ when $f \geq 0$):

$$\nu_n(f) = \int_X f|F(\nabla_n)|^2 \mathrm{vol}_g.$$

A suitable collection of geodesic balls gives a countable basis $U_1, U_2, \dots$ for the topology of X. For each k, there is a subsequence of $\{n'_k\}$ of $\{n\}$ such that $\nu_{n'_k}(U_k)$ has a limit. (Here the measure of an open subset means the supremum of the measures of continuous functions bounded above by its characteristic function.) By the diagonal argument, we can arrange that the same subsequence $\{n'\}$ works for all k. There is then a unique measure ν such that $\nu_{n'}(U_k) \to \nu(U_k)$ (straightforward check). Since $\nu_n(X) = 8\pi^2$, we also have $\nu(X) = 8\pi^2$.

2. The next step is to use the measure ν to identify a finite set $\mathbf{x}$ of 'bad points'. It will emerge later that $\mathbf{x}$ contains at most one point. Using the universal constant ϵ as above, define

$$\mathbf{x} = \{x \in X : \text{ every geodesic ball containing } x \text{ has measure } \geq \epsilon^2\}.$$

We know that $\int_X d\nu = 8\pi^2$; it follows that $|\mathbf{x}| \leq 8\pi^2\epsilon^{-2}$, since if not, we could find disjoint geodesic balls D_i in X such that $\bigcup_i D_i$ has measure exceeding $\epsilon^2 \cdot 8\pi^2/\epsilon^2 = 8\pi^2$. So $\mathbf{x}$ is a finite set.

3. Using the lemma, we see that there is a subsequence $\{n'\}$ and gauge transformations $u_{n'}$ over $X \setminus \mathbf{x}$, such that $u_{n'}(\nabla_{n'}|_{X\setminus\mathbf{x}})$ converges to an ASD connection ∇. In particular, if $\mathbf{x} = \emptyset$, the first of the two possibilities allowed by the theorem holds.

We now suppose $\mathbf{x} \neq \emptyset$. Since $\int_{X\setminus S} |F(\nabla)|^2 \mathrm{vol}_g \leq 8\pi^2$, the removable singularities theorem implies that, over a neighbourhood D_x of $x \in \mathbf{x}$, there is a trivialisation of $E|_{D_x \setminus \{x\}}$ in which ∇ extends to a smooth ASD connection on D_x. This means that, globally, ∇ extends to a smooth ASD connection in *some new bundle $E' \to X$.*

4. Suppose $\mathbf{x} = \{x_1, \dots, x_p\}$. We observe that the limiting measure is of the form

$$\nu = |F(\nabla)|^2 + \sum_{i=1}^{p} c_i \delta_{x_i}, \quad c_i > 0.$$

Indeed, this is certainly true when applied to a continuous function supported in $X \setminus \mathbf{x}$, or to (a continuous approximation to) the characteristic function of a ball centred on x_i; and an arbitrary continuous function can be approximated by sums of functions of these two types.

In particular, $\int_X |F(\nabla)|^2 \mathrm{vol}_g < 8\pi^2$. On the other hand, since ∇ is ASD, the curvature integral must be a non-negative multiple of $8\pi^2$. Hence the only possibility is that $|F(\nabla)|^2 = 0$, i.e. ∇ is flat.

5. We claim that $c_i/8\pi^2$ is an integer for each i. The only possibility is then that $p = 1$ and $c_1 = 8\pi^2$.

To prove this, choose small closed balls Z_i centred on x_i, with boundaries $Y_i = \partial Z_i$, and consider the connection $\nabla' = \nabla|_{Y_i}$. The *Chern-Simons invariant* of (Y_i, ∇'),

$$\mathrm{CS}_{Y_i}(\nabla') \in \mathbb{R}/\mathbb{Z},$$

is defined as follows: choose *any* compact four-manifold W with boundary $\partial W = Y_i$, and *any* SU(2)-connection ∇'' in a bundle over W extending $E|_{Y_i}$, and set

$$\mathrm{CS}_{\partial Z_i}(\nabla') = \frac{1}{8\pi^2} \int_W \operatorname{tr} F(\nabla'')^2 \mod \mathbb{Z}.$$

This is well-defined because given two such extensions, say W and Z_i, we can glue $-W$ to Z_i along Y_i to obtain an SU(2)-connection in a bundle over a closed manifold $-W \cup_{Y_i} Z_i$. The difference between the Z_i-integral and the W-integral is then given by an integral over $-W \cup_{Y_i} Z_i$, which is an integer by the Chern-Weil formula (it it the Euler number of the resulting bundle).

It is easy to see that $\mathrm{CS}_{Y_i}(\nabla')$ depends continuously on ∇'. In particular, the convergence of $u_{n'}\nabla_{n'}$ to ∇ gives

$$\mathrm{CS}_{Y_i}(\nabla') = \lim_{n'} \mathrm{CS}_{Y_i}(u_{n'}(\nabla_{n'}|_{X \setminus S})).$$

The LHS is zero, since ∇ is flat. We have

$$\begin{aligned} c_i = \int_{Z_i} d\nu &= \lim_{n'} \int_{Z_i} d\nu_{n'} \\ &= \lim_{n'} \int_{Z_i} |F(\nabla_{n'}|^2 \mathrm{vol}_g \\ &= \lim_{n'} \int_{Z_i} \operatorname{tr} F(\nabla_{n'}^2) \\ &= 8\pi^2 \lim_{n'} \mathrm{CS}_{Y_i}(\nabla_{n'}|_{Y_i}) \\ &= 8\pi^2 \mathrm{CS}_{Y_i}(\nabla') = [0] \in \mathbb{R}/8\pi^2\mathbb{Z}. \end{aligned}$$

In the last line we use the gauge-invariance of CS. □

A compact subset. Choose $\delta > 0$, smaller than the injectivity radius of (X, g), and define

$$K_\delta = \{[\nabla] \in \mathcal{M}_{X,E,g} : \int_D |F(\nabla)|^2 \mathrm{vol}_g \leq 4\pi^2 \text{ for all geodesic balls } D \text{ of radius } \leq \delta\}.$$

By the theorem, given any sequence $[\nabla_n]$ in K_δ, there are gauge transformations u_n such that $u_n(\nabla_n)$ has a subsequence which converges in C^∞ to an ASD connection ∇_∞.

At this point it becomes intolerable that we have not set up a topology on $\mathcal{M}_{X,E,g}$, let alone a smooth structure. We will discuss this issue further in the next lecture. For now, though, we can give $\mathcal{M}_{X,E,g}$ the quotient C^∞ topology, and assert that this topology is metrizable. Then K_δ becomes a sequentially compact subset.

Concentrated connections. Now consider $C_\delta := \mathcal{M}_{X,E,g} \setminus K_\delta$. This is the open subset in the moduli space corresponding to connections whose curvature density is concentrated near a point.

An imprecise form of Donaldson's collar theorem states that, for $\delta \ll 1$, C_δ is homeomorphic to $X \times (0, \delta)$. The major part of Donaldson's paper [2] is devoted to proving a more precise version of this theorem.

To define a map $C_\delta \to X \times (0, \delta)$, one should first prove that, for small δ, the curvature density $|F(\nabla)|^2$ of any $[\nabla] \in C_\delta$ has a unique maximum. This follows from the compactness theorem, which says that as $\delta \to 0$, the curvature becomes concentrated at a single point. With that in place, one defines

$$C_\delta \to X \times (0, \delta) \quad [\nabla] \mapsto \big(\mathrm{centre}(\nabla), \mathrm{scale}(\nabla)\big),$$

where $\mathrm{centre}(\nabla)$ is the unique point x at which $|F(\nabla)|^2$ achieves its maximum, and

$$\mathrm{scale}(\nabla) = \inf\{r : \int_{D_r(x)} |F(\nabla)|^2 \mathrm{vol}_g \geq 4\pi^2\}$$

where $D_r(x)$ is the geodesic ball of radius r centred at x. So the scale is the smallest radius R such that a geodesic ball of radius R carries half of the total curvature density.

The inverse map is constructed by transferring instantons on S^4 to X. In normal coordinates in the geodesic ball $D_x(\delta)$, the metric there takes the form $g_{ij} = \delta_{ij} + O(|x|^2)$; thus, for δ small, it is close to the Euclidean metric. Identifying this ball with $\mathbb{R}^4 \subset S^4$. Given $R \in (0, \delta)$, there is a unique gauge-orbit of instantons on S^4 having scale R and centre $0 \in \mathbb{R}^4 \subset S^4$. A particular representative for this orbit is obtained by conformal rescaling of the basic instanton. Pull back this connection to $D_x(\delta)$, and plug it into the trivial connection over $X \setminus D_x(\delta)$ by means of a cutoff function. A delicate application of the implicit function theorem, due to Taubes, shows that this connection lies close to a unique instanton having the same centre and scale. This defines the inverse to the $(\mathrm{centre}, \mathrm{scale})$ map.

14 Analysing the space of connections modulo gauge

In this lecture we write down a good model for the space $\mathcal{B}_E = \mathcal{A}_E/\mathcal{G}_E$ of connections modulo gauge (the ambient space of the moduli space $\mathcal{M}_{X,E,g}$): 'good' in that it is a complete metric space, and in that the action of the gauge group admits local slices which are Hilbert spaces.

14.1 (*) Sobolev spaces of connections and gauge transformations

Definition 67. Let X be a compact Riemannian manifold, and $E \to X$ a $\mathrm{U}(n)$-bundle. For a non-negative integer $l > 0$, the **Sobolev norm** $\|\cdot\|_{L^2_l}$ on $C^\infty(E)$ is defined by

$$\|s\|^2_{L^2_l} = \int_X (|s|^2 + |Ds|^2 + \cdots + |D^{(l)}s|^2)\mathrm{vol}_g.$$

Here $D^{(k)}s$ is the kth derivative of s, a section of $(T^*X)^{\otimes k} \otimes E$. The **Sobolev space** $L^2_l(E)$ is the completion of $C^\infty(E)$ in the L^2_l topology.

So $L^2_0(E)$ is the usual space of L^2-sections, and the higher L^2_l form a nested sequence of subspaces. Note that the Sobolev norms come from hermitian inner products, so that $L^2_l(E)$ is a *Hilbert space*.

Remark 68. *An L^2 section is really an equivalence class of square-integrable sections, where sections are equivalent if they are equal outside a set of measure zero. Over a compact base, an L^2 section Ds has a 'weak derivative' Ds, which is the equivalence class of sections defined by the relation*

$$\int_X (Ds, \phi)\mathrm{vol}_g = \int_X (s, D^*\phi)\mathrm{vol}_g$$

for every smooth 'test-section' $\phi \in C^\infty(T^*X \otimes E)$. The Sobolev space $L^2_l(E)$ can be realised as the subspace of $L^2(E)$ consisting of sections with l square-integrable weak derivatives.

There are two important theorems about Sobolev spaces:

1. **Rellich's lemma.** The inclusion $i\colon L^2_{k+1}(E) \hookrightarrow L^2_k(E)$ is a compact operator. That is, it is continuous, and for any sequence s_n with $\|s_n\|_{L^2_{k+1}} \leq 1$, the sequence $i(s_n)$ has a convergent subsequence in $L^2_k(E)$.
2. **Sobolev embedding theorem.** If r is a non-negative integer such that

$$l - \dim(X)/2 > r$$

then the inclusion $C^\infty(E) \to C^r(E)$ extends to a bounded embedding $L^2_l(E) \hookrightarrow C^r(E)$.

There are further embedding theorems, of which we will state one more:

3. When $\dim(X) = 4$, there is a bounded embedding $L^2_1(E) \hookrightarrow L^4(E)$.

Definition 69. Let $E \to X$ be an $\mathrm{SU}(n)$-bundle over a compact Riemannian manifold. An **L^2_k-connection** is an operator

$$C^\infty(E) \to L^2(T^*X \otimes E)$$

of the form $\nabla + a$, where $a \in L^2_k(T^*X \otimes \mathfrak{su}(E))$. Thus an L^2_k-connection differs from a smooth one by a endomorphism-valued one-form of class L^2_k.

Lemma 70. *For* $\mathrm{SU}(n)$*-bundles over a compact Riemannian manifold four-manifold,*

1. L^2_3 *embeds into the continuous sections* C^0.
2. *There is a bounded bilinear multiplication map* $L^2_1 \times L^2_1 \to L^2$.
3. *For* $l \leq 3$*, there is a bounded bilinear multiplication map* $L^2_l \times L^2_3 \to L^2_l$.

Proof. The first point is an instance of the Sobolev embedding theorem. The second follows from the embedding $L^2_1 \hookrightarrow L^4$ and the Cauchy-Schwarz inequality. For the third, let's consider the case $l = 2$; the other possibilities are proved in the same way. Consider sections $s \in L^2_2$ and $t \in L^2_3$. We need to estimate the L^2 norms of the six terms st, $(Ds)t$, sDt, $(D^2s)t$, $(Ds)(Dt)$, and sD^2t. Now, s, Ds, t, Dt and D^2t are all of class L^2_1. So we can use the multiplication $L^2_1 \times L^2_1 \to L^2$ to estimate five of the six terms. The remaining one is $(D^2s)t$, and we can estimate this one using the embedding $L^2_3 \to C^0$. $\square$

Definition 71. In four dimensions, we denote by $\mathcal{A}_E$ the space of L^2_2-connections with its L^2_2-topology. We denote by $\mathcal{G}_E$ the space of $\mathrm{SU}(n)$-gauge transformations of class L^2_3, with its L^2_3-topology.

Note that *the definition of* $\mathcal{G}_E$ *only makes sense by virtue of the Sobolev embedding* $L^2_3 \hookrightarrow C^0$. The $\mathrm{SU}(n)$ gauge transformations u are sections of $\mathrm{End}(E)$ satisfying the pointwise conditions $u(x)^*u(x) = \mathrm{id}_{E_x}$ and $\det(u(x)) = 1$. For general L^2 sections of $\mathrm{End}(E)$, we cannot make sense of pointwise conditions such as these, because L^2 sections are really just *equivalence classes* of sections. But L^2_3 endomorphisms can be thought of as continuous endomorphisms, so it makes sense to impose pointwise conditions.

Lemma 72.
1. *The curvature of a connection in* $\mathcal{A}_E$ *lies in* $L^2_1(\Lambda^2_X \otimes \mathfrak{su}(E))$.
2. $\mathcal{G}_E$ *operates continuously on* $\mathcal{A}_E$.
3. $\mathcal{G}_E$ *is a topological group.*[7]

[7]Actually it is a Hilbert Lie group, with Lie algebra $L^2_3(\mathfrak{su}(E))$.

Proof. The curvature lies in $L^2_1(\Lambda^2_X \otimes \mathfrak{su}(E))$ because of the bounded multiplication $L^2_2 \times L^2_2 \to L^2_1$. The L^3_2 gauge group acts continuously on L^2_2-connections because of the multiplication $L^3_2 \times L^2_2 \to L^2_2$. There is a well-defined and continuous product structure in $\mathcal{G}_E$ because of the multiplication $L^2_3 \times L^2_3 \to L^2_3$. The inversion map $u \mapsto u^{-1} = u^*$ is evidently well-defined and continuous, hence $\mathcal{G}_E$ is a topological group. □

14.2 The orbit space

Since $\mathcal{G}_E$ operates on $\mathcal{A}_E$ we can form the orbit space

$$\mathcal{B}_E = \mathcal{A}_E/\mathcal{G}_E$$

with its topology inherited from the L^2_2-topology on $\mathcal{A}_E$. The orbit space is Hausdorff: indeed, it is metrizable in the coarser quotient L^2 topology.

Lemma 73. *The function $d\colon \mathcal{A}_E \times \mathcal{A}_E \to \mathbb{R}$ given by*

$$d(\nabla, \nabla') = \inf_{u \in \mathcal{G}_E} \|\nabla - u \cdot \nabla'\|_{L^2}$$

descends to give a metric on $\mathcal{B}_E$, inducing the quotient L^2 topology.

For the proof (which makes substantial use of Sobolev space techniques) see [3, p.130].

Usually when studying the orbit space of a M/G (where G is a Lie group operating smoothly on a manifold) it is convenient to be able to work upstairs in M. To do this one needs to have local slices.

Definition 74. A **local slice** at $x \in M$ for the action of G on M is a smooth submanifold $S_x \subset M$, containing x and invariant under the action of $\mathrm{stab}_G(x) \subset G$, such that the projection $M \to M/G$ induces a homeomorphism from $S_x/\mathrm{stab}_G(x)$ to a neighbourhood of $[x]$ in M/G.

Local slices for the action of $\mathcal{G}_E$ on $\mathcal{A}_E$ are also known as **local gauges**.

Definition 75. Let $\nabla \in \mathcal{A}_E$, and define the **Coulomb gauge slice**

$$S_{\nabla,\epsilon} = \{\nabla + a : \|a\|_{L^2_2} < \epsilon,\ d^*_\nabla a = 0\} \subset \mathcal{A}_E.$$

Theorem 76. *(See [3, p.56].) For small enough $\epsilon = \epsilon(\nabla)$, $S_{\nabla,\epsilon}$ is a local slice at ∇ for the action of $\mathcal{G}_E$ on $\mathcal{A}_E$. That is, the projection $S_\nabla/\Gamma_\nabla \to \mathcal{B}_E$ is an open embedding, where $\Gamma_\nabla = \mathrm{stab}_{\mathcal{G}}(\nabla)$.*

Recall that the stabiliser Γ_∇ consists of gauge transformations which are covariant-constant and therefore determined by their values at a basepoint x. The image of the resulting embedding $\Gamma_\nabla \hookrightarrow \mathrm{SU}(E_x)$ is the centraliser of the holonomy group $\mathcal{H}_{\nabla,x}$.

14.3 Local slices and ASD connections

The moduli space $\mathcal{M}_{X,E,g}$ is defined to be the subspace of $\mathcal{B}_E$ consisting of gauge-orbits of ASD connections. The theorem on Coulomb gauge slices immediately gives the following:

Proposition 77. *Let ∇ be an ASD connection. For $\epsilon > 0$, define*

$$Z_{\nabla,\epsilon} = \{\nabla + a \in S_{\nabla,\epsilon} : F(\nabla + a)^+ = 0\} \subset S_{\nabla,\epsilon}.$$

Then the projection from $Z_{\nabla,\epsilon}$ to $\mathcal{M}_{X,g}$ is a surjection onto an open set, and it induces an open embedding

$$Z_{\nabla,\epsilon}/\Gamma_{\nabla,x} \hookrightarrow \mathcal{M}_{X,E,g}.$$

Recall the formula for the variation in the curvature:

$$F(\nabla + a) = F(\nabla) + d_\nabla a + a \wedge a.$$

If ∇ is ASD then $F(\nabla + a)^+ = (d_\nabla a + a \wedge a)^+$. Hence the linearisation of the ASD equation at ∇ is the equation

$$d^+_\nabla a = 0.$$

Define the operator

$$\delta_\nabla = d^*_\nabla \oplus d^+_\nabla \colon \Omega^1_X(\mathfrak{su}(E)) \to \Omega^0_X(\mathfrak{su}(E)) \oplus \Omega^+_X(\mathfrak{su}(E)).$$

Then a lies in $\ker(\delta_\nabla)$ iff it satisfies both this equation and also the equation $d^*_\nabla a = 0$. Thus $\ker(\delta_\nabla)$ gives the linearisation of the ASD condition *together with the Coulomb gauge-fixing condition.*

If it happens that δ_∇ is surjective then it follows from the inverse function theorem that $Z_{\nabla,\epsilon}$ is a smooth submanifold of $\mathcal{A}_E$. Its tangent space at ∇ is $\ker(\delta_\nabla)$.

15 Elliptic operators and complexes

A first-order linear differential operator $D\colon C^\infty(U,\mathbb{R}^m)\to C^\infty(U,\mathbb{R}^n)$ over an open set $U\subset\mathbb{R}^N$ is one of the form

$$(Ds)_i=\sum_j\left(\sum_k a_{ij}^k\frac{\partial}{\partial x_k}+b_{ij}\right)s_j,\quad a_{ij}^k,\,b_{ij}\in\mathbb{R}. \tag{6}$$

Definition 78. (i) A **first-order linear differential operator** D between vector bundles ξ_1 and ξ_2 over a smooth manifold M is a map $D\colon C^\infty(\xi_1)\to C^\infty(\xi_2)$ which can be expressed, in local coordinates, and local trivialisations of the bundles, by a first order operator as in (6).

(ii) The **symbol** of D is the map of vector bundles

$$\sigma_D\colon T^*M\to\mathrm{Hom}(\xi_1,\xi_2)$$

given, over $x\in M$, by

$$\sigma_D(e_x)(v)=D(f\widetilde{v})(x)-f(x)D(\widetilde{v})(x),$$

where f is a function such that $df(x)=e_x$, and $\widetilde{v}$ is a section of ξ_1 such that $\widetilde{v}(x)=v$.

(iii) A first-order linear differential operator D is called **elliptic** if $\sigma_D(e_x)$ is an isomorphism, for all $x\in M$ and all non-zero $e_x\in T_x^*M$.

One checks that the symbol is well-defined and that the symbol of the operator D in (6) is given by $\sigma_D(dx_k)=(a_{ij}^k)_{i,j}$. Thus the symbol captures the leading order terms algebraically.

We will give examples presently. Meanwhile, here is the crucial result about elliptic operators.

Theorem 79. *Let $D\colon C^\infty(\xi_1)\to C^\infty(\xi_2)$ be a first-order elliptic operator over a compact manifold. Then*

- $\mathrm{im}(D)$ *is closed and* $C^\infty(\xi_2)=\mathrm{im}(D)\oplus\ker(D^*)$*;*
- D *and* D^* *are both Fredholm.*

A bounded linear map between topological vector spaces is *Fredholm* if it has closed image, and its kernel and cokernel are finite-dimensional.

The topological vector space $C^\infty(\xi_i)$ does not carry a complete norm. However, D extends continuously to a bounded map between Sobolev spaces $L^2_{k+1}(\xi_1)\to L^2_k(\xi_2)$, $k\geq 1$ (ditto D^*). A section in $\ker(D)$ or $\ker(D^*)$ in these larger spaces is automatically C^∞ ('elliptic regularity').

If E_1, E_2 are Banach spaces, the space $\mathcal{B}(E_1,E_2)$ of bounded linear maps $E_1\to E_2$ (with its operator norm topology) contains the subspace $\mathrm{Fred}(E_1,E_2)$ of Fredholm operators. The index function

$$\mathrm{ind}\colon\mathrm{Fred}(E_1,E_2)\to\mathbb{Z},\quad F\mapsto\dim\ker(F)-\dim\mathrm{coker}(F)$$

is continuous, i.e. locally constant.

One can deduce from this that the index of an elliptic operator over a compact manifold depends only on its symbol σ.

15.1 Elliptic complexes

Definition 80. An **elliptic complex** $\mathcal{K}$ is a finite sequence of first-order linear differential operators

$$\cdots \longrightarrow C^\infty(\xi_1) \xrightarrow{D_1} C^\infty(\xi_2) \xrightarrow{D_2} C^\infty(\xi_3) \longrightarrow \cdots$$

such that, for every $x \in M$ and every non-zero $e_x \in T^*_x M$, the symbol sequence

$$\cdots \longrightarrow (\xi_1)_x \xrightarrow{\sigma_{D_1}(e_x)} (\xi_2)_x \xrightarrow{\sigma_{D_2}(e_x)} (\xi_3)_x \longrightarrow \cdots$$

is an exact sequence.

There is a simple relation between elliptic complexes and elliptic operators, at least over compact manifolds. Using the notation from the definition, let $\xi_{ev} = \bigoplus_{i\in\mathbb{Z}} \xi_{2i}$, $\xi_{odd} = \bigoplus_{i\in\mathbb{Z}} \xi_{2i+1}$. Then the operators $\{D_i\}$ give a map $D\colon C^\infty(\xi_{ev}) \to C^\infty(\xi_{odd})$. After choosing metrics on the ξ_i, the D_i have formal adjoint maps D_i^*, and these too give a map $D^*\colon C^\infty(\xi_{ev}) \to C^\infty(\xi_{odd})$. We have

Proposition 81. *The first-order operator $D \oplus D^*\colon C^\infty(\xi_{ev}) \to C^\infty(\xi_{odd})$ is elliptic. Its index is minus the Euler characteristic of the complex:*

$$\operatorname{ind}(D \oplus D^*) = \sum_{i\in\mathbb{Z}} (-1)^{i+1} \dim_{\mathbb{R}} H_i(\mathcal{K}, D).$$

Example 82. The symbol of the exterior derivative $d\colon \Omega^p_M \to \Omega^{p+1}_M$ is $\sigma_d(e_x)(\omega_x) = e_x \wedge \omega_x$, whence the de Rham complex is an elliptic complex.

Choosing a metric on TM, and using the induced metrics on the Ω^p_M, the kernel of $d \oplus d^*$ in degree p is the harmonic space $\mathcal{H}^p(M)$.

15.2 The three-term complex in Hodge theory

Here is an example of an elliptic complex which we can analyse using the Hodge theorem. This complex governs the deformation theory of U(1)-instantons.

Let (X, g) be a compact oriented Riemannian four-manifold. Define the following complex $\mathcal{K}$:

$$0 \longrightarrow \Omega^0_X \xrightarrow{d} \Omega^1_X \xrightarrow{d^+} \Omega^+_{X,g} \longrightarrow 0$$

where $d^+\alpha = (d\alpha)^+$.

Lemma 83. *The sequence $\mathcal{K}$ is an elliptic complex.*

Proof. The symbol sequence at e_x is

$$0 \longrightarrow \Lambda^0_x \xrightarrow{e_x\cdot} \Lambda^1_x \xrightarrow{(e_x\wedge\cdot)^+} \Lambda^+_x \longrightarrow 0.$$

It is clear that this is a complex, and that, when $e_x \neq 0$, the map $e_x\cdot$ is injective and $(e_x \wedge \cdot)^+$ surjective. Exactness in the middle then follows from a dimension count. $\square$

We can calculate the cohomology spaces $H_i(\mathcal{K})$ (we place the three non-zero terms in degrees 0, 1 and 2).

Proposition 84. *There are canonical isomorphisms*

$$\begin{aligned} H_0(\mathcal{K}) &\cong H^0_{dR}(X) \cong \mathbb{R}, \\ H_1(\mathcal{K}) &\cong H^1_{dR}(X), \\ H_2(\mathcal{K}) &\cong \mathcal{H}^+_{X,g}. \end{aligned}$$

Hence the index of the elliptic operator

$$\delta = d^* \oplus d^+ \colon \Omega^1_X \to \Omega^0_X \oplus \Omega^+_{X,g}$$

(which is the minus the Euler characteristic of $\mathcal{K}$) is $-(1 - b_1 + b_2^+)$.

Proof. The assertion about H_0 is trivial (a zero-form in the kernel of d is a constant function). For the middle term, we note that for any two-form η we have $\eta \wedge \eta = (\eta^+ + \eta^-)^2 = (\eta^+)^2 + (\eta^-)^2 = (|\eta^+|^2 - |\eta^-|^2)\mathrm{vol}_g$. Applying this to $d\alpha$, where $\alpha \in \Omega^1_X$, we obtain

$$0 \overset{\text{Stokes}}{=} \int_X d(\alpha \wedge d\alpha) = \int_X d\alpha \wedge d\alpha = \int_X (|d^+\alpha|^2 - |d^-\alpha|^2)\mathrm{vol}_g.$$

Hence $d^+\alpha = 0$ implies $d\alpha = 0$, and $H_1(\mathcal{K}) = H^1_{dR}(X)$.

The interesting part is the isomorphism $H_2(\mathcal{K}) \cong \Omega^+_{X,g}/d^+\Omega^1_X$. The Hodge theorem tells us that any two-form η has a unique decomposition

$$\eta = \eta_{\mathrm{harm}} + d\alpha + d^*\beta, \quad \alpha \in \Omega^1_X,\ \beta \in \Omega^3_X$$

with η_{harm} harmonic. Recall also that $d^* = - * d *$. If $\eta \in \Omega^+_{X,g}$, so $*\eta = \eta$, we find

$$\eta = *\eta_{\mathrm{harm}} + *d^*\beta + *d\alpha = *\eta_{\mathrm{harm}} - d(*\beta) + d^*(*\alpha),$$

with $*\eta_{\mathrm{harm}}$ harmonic. So by the uniqueness of the Hodge decomposition, $*\eta_{\mathrm{harm}} = \eta_{\mathrm{harm}}$ and $d^*\beta = d^*(*\alpha)$. So now

$$\eta = \eta_{\mathrm{harm}} + d\alpha + *d\alpha = \eta_{\mathrm{harm}} + 2d^+\alpha.$$

Hence the map sending η to η_{harm} induces an isomorphism from $\Omega^+_{X,g}/d^+\Omega^1_X$ to the space of harmonic self-dual two-forms $\mathcal{H}_{X,g}$. $\square$

15.3 The three-term complex for instantons

Let $E \to X$ be an SU(2)-bundle, and ∇ an ASD connection. There is then a sequence

$$0 \longrightarrow \Omega^0_X(\mathfrak{su}(E)) \xrightarrow{d_\nabla} \Omega^1_X(\mathfrak{su}(E)) \xrightarrow{d^+_\nabla} \Omega^+_{X,g}(\mathfrak{su}(E)) \longrightarrow 0.$$

Since $d^+_\nabla \circ d_\nabla s = F(\nabla)^+ s$, the sequence is a complex. The same argument as for $\mathcal{K}$ shows that its symbol sequence is exact, hence that this is an elliptic complex. The associated elliptic operator is

$$\delta_\nabla = d^*_\nabla \oplus d^+_\nabla \colon \Omega^1_X(\mathfrak{su}(E)) \to \Omega^0_X(\mathfrak{su}(E)) \oplus \Omega^+_{X,g}(\mathfrak{su}(E)).$$

This operator was introduced in Lecture 13. Its kernel gives the linearisation of the ASD condition *together with the Coulomb gauge-fixing condition.* (Note that, in the absence of a gauge-fixing condition, the ASD equation is *not* elliptic.)

The three cohomology groups of the complex are denoted by H^0_∇, H^1_∇ and H^2_∇. Note that H^0_∇ is isomorphic to the Lie algebra of $\mathrm{stab}_{\mathcal{G}}(\nabla)$.

$$\boxed{\begin{array}{l} H^1_\nabla \text{ is called the deformation space of } \nabla. \\ H^2_\nabla \text{ is called the obstruction space of } \nabla. \end{array}}$$

Remark 85. *One can show, using a 'Weitzenböck formula', that over S^4 every instanton ∇ satisfies $H^2_\nabla = 0$.*

Donaldson managed without the following, but it is a handy result:

Theorem 86 (Freed-Uhlenbeck regularity). *Fix an* SU(2)*-bundle $E \to X$. Then there is a dense set of metrics g on X such that $H^2_\nabla = 0$—that is, the operator d^+_∇ is surjective—for every g-ASD connection ∇ in E.*

16 Elliptic theory and the moduli space of instantons

In Lectures 14 and 15 we introduced the elliptic operator

$$\delta_\nabla = d_\nabla^* \oplus d_\nabla^+ \colon \Omega^1_X(\mathfrak{su}(E)) \to \Omega^0_X(\mathfrak{su}(E)) \oplus \Omega^+_{X,g}(\mathfrak{su}(E))$$

associated with an ASD connection ∇. Its kernel consists of solutions to the linearised ASD equation in the Coulomb gauge slice at ∇.

Theorem 87 (Atiyah, Hitchin and Singer, see [1]). *The index of δ_∇ is given by*

$$\operatorname{ind}(\delta_\nabla) = 8e(E) - 3(1 - b_1(X) + b^+(X)).$$

In particular, when $e(E) = 1$ and $b_1 = b^+ = 0$, the index is 5.

This is a consequence of the *Atiyah-Singer index theorem.* There is also a more direct proof [3, ch. 7] which isolates and applies the relevant part of the Atiyah-Singer theorem—an 'excision' property for indices of elliptic operators.

16.1 The Kuranishi model

The Kuranishi model concerns a differentiable map ψ between Banach spaces which is non-linear but has *Fredholm derivative.* It show that the zero-set $\psi^{-1}(0)$ can be described in finite-dimensional terms.

Let T, U be real or complex Banach spaces, $\psi \colon T \to U$ a smooth map with $\psi(0) = 0$. Suppose that $D := D_0\psi \colon T \to U$ is Fredholm. Then $\ker(D)$ and $\operatorname{im}(D)$ are both closed and so have topological complements, so we may write

$$T = \ker(D) \oplus T_0; \quad U = \operatorname{im}(D) \oplus U_0.$$

$\ker(D)$ and $U_0 \cong \operatorname{coker}(D)$ are finite-dimensional, and D carries T_0 isomorphically to $\operatorname{im}(D)$. The **Kuranishi map** κ is the map

$$\kappa \colon \ker(D) \to \operatorname{coker}(D), \quad \operatorname{pr} \circ \psi|_{\ker(D)}$$

where pr is the projection $U \to \operatorname{coker}(D)$.

Theorem 88 (Kuranishi model). *There is a smooth embedding $i \colon \ker(D) \hookrightarrow T$ which carries a neighbourhood of 0 in $\kappa^{-1}(0)$ homeomorphically onto a neighbourhood of 0 in $\psi^{-1}(0)$.*

Remark 89. *When T and U are representation spaces for a compact group, and ψ equivariant, the Kuranishi map is also equivariant; the proof will make it clear that we make i equivariant too.*

The starting point for the proof is a corollary of the

Theorem 90 (Inverse function theorem). *Let T, U be Banach spaces (i.e. complete normed spaces), $\tilde{T}$ and $\tilde{U}$ neighbourhoods of $0 \in T$ and $0 \in U$ respectively, and $\psi\colon \tilde{T} \to \tilde{U}$ a smooth map with $\psi(0) = 0$. If the derivative $D_0\psi\colon T \to U$ is a Banach space isomorphism then ψ is a diffeomorphism from a neighbourhood of $0 \in \tilde{T}$ to a neighbourhood of $0 \in \tilde{U}$.*

Corollary 91. *If $\psi\colon \tilde{T} \to \tilde{U}$ is a smooth map as above, and $D_0\psi\colon T \to U$ is surjective and has a bounded right inverse $s\colon U \to T$ (so $D_0\psi \circ s = \mathrm{id}_U$) then, after shrinking $\tilde{T}$ and $\tilde{U}$, there is a diffeomorphism $f\colon \tilde{T} \to \tilde{T}$ such that $D_0\psi = \psi \circ f$.*

Proof. Write $T = \mathrm{im}(s) \oplus \ker(D\psi)$, and define $f = \phi \times \mathrm{id}_{\ker(D\psi)}$, where $\phi\colon \tilde{U} \to \mathrm{im}(s)$ is the two-sided inverse to $\psi|_{\mathrm{im}(s)}$. □

The inverse function theorem is an application of the contraction mapping theorem—so completeness is important.

Proof. Convention: In this proof, when we write a map between Banach spaces we mean the germ of a map between small neighbourhoods of the origin.

1. Let $p_1\colon U \to \mathrm{im}(D)$, $p_2\colon U \to U_0$ be the projections. The derivative at 0 of $p_1 \circ \psi$ is $p_1 \circ D$. This is surjective, so the corollary of the inverse function theorem tells us that $p_1 \circ \psi$ has a (local) right inverse, i.e. there is a diffeomorphism $f\colon T \to T$ with $p_1 \circ \psi \circ f = p_1 \circ D$ (first instance of the convention!).

2. The map $\psi \circ f\colon T \to U$ can be expressed as

$$\ker(D) \times T_0 \to \mathrm{im}(D) \times U_0, \quad (x, y) \mapsto (D(y), \chi(x, y)),$$

where $\chi = p_2 \circ \psi \circ f$. Note that D is a linear isomorphism $T_0 \to U_0$, and that $D_0\chi = 0$.

3. The zero-set of $\psi \circ f$ coincides with that of the map

$$\ker(D) \to U_0, \quad x \mapsto \chi(x, 0).$$

But (under the natural identification $U_0 = \mathrm{coker}(D)$) this map is just $\kappa \circ f$. Letting $f = i|_{\ker(D)}$, the result follows.

□

16.2 Local models for the moduli space

The Kuranishi model allows us to describe the space

$$Z_{\nabla,\epsilon} = \{\nabla + a : F(\nabla + a)^+ = 0,\, d_\nabla^* a = 0,\, |a| < \epsilon\}, \quad \nabla \text{ ASD}.$$

Let $T = \ker(d_\nabla^*) \subset L_3^2(\Lambda_X^1 \otimes \mathfrak{su}(2))$, and $U = L_2^2(\Lambda_X^+ \otimes \mathfrak{su}(2))$, and consider the map

$$\psi\colon T \to U, \quad a \mapsto F(\nabla + a)^+.$$

We have $\psi(0) = 0$ and $D_0\psi(a) = (d_\nabla a)^+$. Thus

$$\ker(D_0\psi) = \ker(d_\nabla^*) \cap \ker d_\nabla = \ker(\delta_\nabla) = H^1_\nabla.$$

The image of $D_0\psi$ is closed; indeed, a complement for $\mathrm{im}(D_0\psi)$ is given by $\ker(d_\nabla^*)$. Note that $\mathrm{coker}(D_0\psi) \cong H^2_\nabla$. Hence $D_0\psi$ is Fredholm with index

$$\mathrm{ind}(D_0\psi) = \dim H^1_\nabla - \dim H^2_\nabla = \mathrm{ind}(\delta_\nabla) + \dim H^0_\nabla.$$

By the proposition, when ϵ is small, $Z_{\nabla,\epsilon}$ is homeomorphic to a neighbourhood of 0 in $\kappa^{-1}(0) \subset H^1_\nabla$, the zero-set of the Kuranishi map

$$\kappa\colon H^1_\nabla \to H^2_\nabla, \quad \kappa(0) = 0.$$

The ASD moduli space is modelled, locally near $[\nabla]$, on $Z_{\nabla,\epsilon}/\Gamma_\nabla$. We can understand this quotient in the Kuranishi model as follows.

$H^1_\nabla = \ker(\delta_\nabla)$ and H^2_∇ are representations of the compact group $\Gamma_\nabla \subset \mathrm{SU}(2)$. Indeed, Γ_∇ acts by conjugation on sections of $\mathfrak{su}(E)$, hence on forms with values in $\mathfrak{su}(E)$. Each $u \in \Gamma_\nabla$ satisfies $[\nabla, u] = 0$, and hence also commutes with d_∇ and d^+_∇. So Γ_∇ operates on the whole elliptic complex, and in particular on its cohomology groups.

The map $\psi\colon a \mapsto F(\nabla + a)^+ = (\nabla a + \frac{1}{2}[a,a])^+$ is also equivariant. As remarked above, in this situation the Kuranishi model applies equivariantly. We now state our conclusions:

Theorem 92. *An ASD connection ∇ determines a Γ_∇-equivariant map*

$$\kappa\colon H^1_\nabla \to H^2_\nabla, \quad \kappa(0) = 0.$$

Small neighbourhoods of $[\nabla] \in \mathcal{M}_{X,g}$ are canonically homeomorphic to small neighbourhoods of 0 in $\kappa^{-1}(0)/\Gamma_\nabla$.

Corollary 93. *If $H^2_\nabla = 0$ then $Z_{\nabla,\epsilon}$ is Γ_∇-equivariantly homeomorphic to an open ball in the finite-dimensional complex vector space H^1_∇, and small neighbourhoods of $[\nabla]$ are canonically homeomorphic to neighbourhoods of 0 in H^1_∇/Γ_∇.*

To keep things brief, we have not discussed smooth structures, but they are easily accommodated within this description.

Under the assumption that every ASD connection satisfies $H^2_\nabla = 0$, we can at last understand the local structure of our moduli space $\mathcal{M}_{X,E,g}$.

∇ irreducible, $H^2_\nabla = 0$: The stabiliser $\Gamma_\nabla = \{\pm 1\}$ is the centre of $\mathrm{SU}(2)$, and it operates trivially on the cohomology spaces. Its Lie algebra H^0_∇ is trivial. Thus the dimension of H^1_∇ is equal to $\mathrm{ind}\,\delta_\nabla$. The local model is $\mathbb{R}^d$, where $d = \mathrm{ind}\,\delta_\nabla = 8e - 3(1 - b_1 + b^+)$. In particular, when $b_1 = b^+ = 0$ and $e = 1$, the model is $\mathbb{R}^5$.

∇ reducible, $H^2_\nabla = 0$: At a reducible connection with $\Gamma_\nabla = S^1 \subset \mathrm{SU}(2)$, we have $H^0_\nabla = \mathrm{Lie}(S^1) = \mathbb{R}$. Thus $\mathrm{ind}(\delta_\nabla) = \dim H^1_\nabla - 1 - \dim H^2_\nabla$. When $b_1 = b^+ = 0$ and $e = 1$, we have $\dim H^1_\nabla = 6$.

We need to understand how Γ_∇ acts on H^1_∇. So we should consider the (conjugation) action of the diagonal subgroup $S^1 \subset \mathrm{SU}(2)$ on $\mathfrak{su}(2)$. We have

$$\mathfrak{su}(2) = \mathbb{R}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus \left\{ \begin{pmatrix} 0 & \alpha \\ -\overline{\alpha} & 0 \end{pmatrix} : \alpha \in \mathbb{C} \right\},$$

where S^1 acts trivially on the first (real) summand and as $z \mapsto z^2$ on the second summand.

Globally, if ∇ respects a line-bundle decomposition $E = L \oplus \bar{L}$, we have

$$\mathfrak{su}(E) \cong \underline{\mathbb{R}} \oplus L^{\otimes 2}.$$

Here the fibre at x of the trivial $\mathbb{R}$-bundle corresponds to the Lie algebra of the holonomy group $H_{\nabla,x}$.

The three-term complex splits as the sum of two sub-complexes: a copy of the ordinary three-term complex $\mathcal{K}$, and a complex

$$\Omega^0(L^{\otimes 2}) \to \Omega^1(L^{\otimes 2}) \to \Omega^2(L^{\otimes 2}).$$

Now, the cohomology groups of $\mathcal{K}$ have dimensions $b_0 = 1$, $b_1 = 0$, and $b^+ = 0$. Hence it only the $L^{\otimes 2}$-part which contributes to H^1_∇. As a representation of S^1, H^1_∇ is equivalent to $\mathbb{C}^3$ with the weight two action

$$z \cdot (a, b, c) = (z^2 a, z^2 b, z^2 c).$$

We have:

Theorem 94. *Let X be simply connected and negative-definite, and ∇ a reducible connection with $H^2_\nabla = 0$ in an $\mathrm{SU}(2)$-bundle with $e = 1$. Then ∇ has a neighbourhood in $\mathcal{M}_{X,g}$ modelled on $\mathbb{C}^3/S^1$, where S^1 acts diagonally (with weight two) on $\mathbb{C}^3$.*

The quotient $\mathbb{C}^3/S^1$ is homeomorphic to the cone on $\mathbb{C}P^2$ by the map

$$\mathbb{C}^3/S^1 \to \frac{\mathbb{C}P^2 \times \mathbb{R}^+}{\mathbb{C}P^2 \times \{0\}}, \quad [a, b, c] \mapsto [(a : b : c), a^2 + b^2 + c^2]$$

(sending $[0, 0, 0]$ to the cone point).