Calibrated Geometry and Gauge Theory

Jason D. Lotay
University of Oxford

December 8, 2022

Abstract

Calibrated geometry was inspired by ideas in complex geometry and provides a key tool in the study of submanifolds which minimize volume. Gauge theory in higher dimensions provides a potential means to define new invariants for manifolds with special holonomy, motivated by Chern–Simons and Donaldson theory. Both topics have important links with symplectic and Kähler geometry, geometric analysis and theoretical physics. We provide an overview of calibrated geometry and gauge theory in higher dimensions and discuss relationships between them. We focus on the special holonomy settings of Calabi–Yau, G_2 and Spin(7) manifolds. We describe fundamental results and techniques in the field and discuss open problems.

Contents

1 Introduction

The course has three key aims.

- Provide an overview of calibrated geometry and gauge theory in higher dimensions.
- Discuss relationships between these seemingly distinct topics.
- Describe the main results and open problems in the field.

Along the way, we shall see some of the key tools which are used in these subjects and touch on a wide variety of different areas.

1.1 Questions

Before we get started, I want to discuss some questions. The first (seemingly basic) question is the following.

Question 1.1. Given two n-planes P_1, P_2 in $\mathbb{R}^{2n}$ such that $P_1 \cap P_2 = \{0\}$, when is $P_1 \cup P_2$ *volume-minimizing*?

Here, asking for $P_1 \cup P_2$ to be volume-minimizing is that for any oriented n-dimensional submanifold N in $\mathbb{R}^{2n}$ which is equal to $P_1 \cup P_2$ outside some compact subset S or $\mathbb{R}^{2n}$ (which is allowed to depend on N), we have that

$$\mathrm{Vol}\big((P_1 \cup P_2) \cap S\big) \leq \mathrm{Vol}(N \cap S).$$

Let's think about Question 1.1 in the case $n = 1$. Here, we see that the answer is easy and it is *never*! However, it turns out that for $n > 1$ that there is a more interesting answer to Question 1.1 which, roughly speaking, is that $P_1 \cup P_2$ is volume-minimizing whenever the angles between the planes P_1, P_2 are "sufficiently large". As we shall see, one needs to be more precise about the statement and the orientations of the planes, and this is the Angle Theorem. It does however make intuitive sense even if we think about the 1-dimensional case, where we see that the pair of lines gets closed and closer to length minimizing as we move the angle between the lines to be larger and larger (until they agree).

Underlying Question 1.1 is the following "real" question.

Question 1.2. How do we know when a submanifold is volume-minimizing?

This is a very difficult question in general, but an answer to this question is provided by *calibrated geometry.* In particular, it leads to the solution of Question 1.1 in the Angle Theorem.

My second (again, seemingly basic) question is the following.

Question 1.3. Given an n-dimensional manifold M which is homeomorphic to $\mathbb{R}^n$, must M be diffeomorphic to $\mathbb{R}^n$?

The answer to Question 1.3 for $n \neq 4$ is yes but, surprisingly, the answer for $n = 4$ is no! This leads to the notion of "exotic" $\mathbb{R}^4$s which follows by work of Donaldson and Freedman.

Again, underlying Question 1.3 is really the following question.

Question 1.4. Can we construct invariants that detect different geometric structures?

Here an answer is provided by *gauge theory* and in particular yields the negative answer to Question 1.3 for $n = 4$.

1.2 Summary

Now we know that the topics of the course are interesting, but what are they?

- Calibrated geometry is a Riemannian submanifold theory inspired by complex geometry. It is a key tool in geometric analysis.
- Gauge theory was inspired by mathematical physics and concerns connections on Riemannian manifolds. Gauge theory is a key tool in low-dimensional topology.

Both of these topics have links to a wide variety of topics including: symplectic and Kähler geometry; special holonomy; spin geometry; variational problems; geometric measure theory; nonlinear elliptic PDE; and String Theory.

The important fact is that calibrated geometry and gauge theory in higher dimensions are *intimately connected*. It is this relationship which we wish to explore by the end of this course. In particular, we shall see that even to construct solutions to the gauge theory equations in higher dimensions will use calibrated submanifolds. Moreover, we will see that to have a hope to construct invariants from gauge theory in higher dimensions as we might wish to do based on our earlier question, one needs to know a lot about calibrated submanifolds. In fact, motivated by Gromov–Witten and Floer theory, one might (perhaps naively) hope to construct invariants just from calibrated submanifolds, but it turns out that one almost certainly needs to use gauge theory again to have a chance of obtaining such an invariant. Finally, there is an idea from String Theory called Mirror Symmetry which suggests a further relation between calibrated geometry and gauge theory that we hope to explore at the very end of the course.

Disclaimer. The relationship between calibrated geometry and gauge theory is *not* fully understood and the invariants do *not* (currently) exist! However, important progress in these topics have been made recently and so it seems an opportune time to discuss them. In particular, there are many interesting open problems which makes this an exciting research topic.

We end this introduction by providing a brief summary of the intended course content.

- **Introduction to calibrations.** Minimal submanifolds: definition and examples; first variation of volume; minimal graphs. Calibrations and calibrated submanifolds; calibrated submanifolds are volume-minimizing; calibrations and holonomy.
- **Complex and special Lagrangian submanifolds; the angle theorem.** Wirtinger's inequality; complex submanifolds in Kähler manifolds are calibrated. Special Lagrangian calibration; Calabi–Yau manifolds; examples. The angle theorem: Lawlor necks and Nance calibrations.
- **Calibrated submanifolds and exceptional holonomy.** Associative, coassociative and Cayley calibrations; G_2 and Spin(7) manifolds; examples; relations to complex and special Lagrangian geometry.
- **Constructing calibrated submanifolds and moduli problems.** Construction methods via reductions to ODEs. Deformation theory of calibrated submanifolds; links to elliptic PDE and spin geometry. Gluing methods for nonlinear PDE to construct compact examples.

- **Introduction to gauge theory in higher dimensions.** Yang–Mills functional and connections. Discussion of gauge theory in low dimensions; instantons and monopoles. Hermitian-Yang–Mills and stability.
- **Gauge theory and exceptional holonomy.** Instantons and monopoles on manifolds with special holonomy; instantons minimize Yang–Mills functional.
- **Constructing solutions to gauge theoretic equations and moduli problems.** Construction methods via reductions to ODEs. Deformation theory of instantons; links to elliptic PDE and spin geometry.
- **Links between calibrated geometry and gauge theory.** Calibrated submanifolds and limits of instantons and monopoles; Fueter sections. Examples of instantons on compact manifolds via gluing. Donaldson-Thomas/Donaldson-Segal program; conjectured links to enumerative invariants and Floer theory. Mirror symmetry and new gauge theories from calibrated geometry (time permitting).
- **Open problems.** Discussion of key problems in the field; graduate-level problems.

2 Introduction to calibrations

2.1 Minimal submanifolds

Before we begin, let us recall some facts about *geodesics* γ in a Riemannian manifold (M, g).

- Geodesics are critical points for the length functional on curves: $\gamma \mapsto \int |\gamma'|$.
- Geodesics are defined by a second order differential equation: $\nabla_{\gamma'}\gamma' = 0$.
- Geodesics locally minimize length. They are not necessarily globally minimizing, even with fixed endpoints: take a curve along the equator in the unit 2-sphere $\mathcal{S}^2$ with length greater than π. Moreover, the equator in $\mathcal{S}^2$ is minimal but not length minimizing since we can deform it to a shorter line of latitude.
- Geodesics reflect and encode aspects of the ambient geometry. For example, the way geodesics emanating from a fixed point "spread" is intimately connected to the curvature of the ambient manifold, as one sees when studying Jacobi fields.
- Geodesics are related to topology. For example, each non-trivial free homotopy class of curves in a compact Riemannian manifold is represented by a closed geodesic (by a theorem due to Cartan).

To generalize to higher dimensions, we start by analysing the submanifolds which are critical points for the volume functional. Let N be a k-dimensional submanifold (without boundary) of an n-dimensional Riemannian manifold (M, g). Let $F : N \times (-\epsilon, \epsilon) \to M$ be a variation of N with compact support; i.e. $F = \mathrm{Id}$ outside a compact subset $\overline{S}$ of N with S open and $F(p, 0) = p$ for all $p \in N$. The vector field $X = \frac{\partial F}{\partial t}|_N$ is called the variation vector field (which will be zero outside of $\overline{S}$). We then have the following definition.

Definition 2.1. An oriented submanifold N of a Riemannian manifold (M, g) is *minimal* if

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Vol}(F(S,t))|_{t=0} = 0$$

for all variations F with compact support $\overline{S}$ (depending on F).

Remark. Notice that we do not ask for N to minimize volume: it is only stationary for the volume. It could even be a maximum!

Example. A plane in $\mathbb{R}^n$ is minimal since any small variation will have larger volume.

Example. Geodesics are critical points for length, so geodesics are minimal.

For simplicity let us suppose that N is compact. We wish to calculate $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Vol}(F(N,t))|_{t=0}$. Given local coordinates x_i on N we know that

$$\operatorname{Vol}(F(N,t)) = \int_N \sqrt{\det\left(g\left(\frac{\partial F}{\partial x_i}, \frac{\partial F}{\partial x_j}\right)\right)}\operatorname{vol}_N .$$

Let $p \in N$ and choose our coordinates x_i to be normal coordinates at p: i.e. so that $\frac{\partial F}{\partial x_i}(p,t) = e_i(t)$ satisfy $g(e_i(0), e_j(0)) = \delta_{ij}$. If $g_{ij}(t) = g(e_i(t), e_j(t))$ and $(g^{ij}(t))$ denotes the inverse of the matrix $(g_{ij}(t))$ then we know that

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\det(g_{ij}(t))}|_{t=0} = \frac{1}{2}\frac{\sum_{i,j} g^{ij}(t)g'_{ij}(t)}{\sqrt{\det(g_{ij}(t))}}|_{t=0} = \frac{1}{2}\sum_i g'_{ii}(0).$$

Now, if we let ∇ denote the Levi-Civita connection of g, then

$$\begin{aligned}\frac{1}{2}\sum_i g'_{ii}(0) &= \frac{1}{2}\sum_i \frac{\mathrm{d}}{\mathrm{d}t} g\left(\frac{\partial F}{\partial x_i}, \frac{\partial F}{\partial x_i}\right)|_{t=0}\\ &= \sum_i g(\nabla_X e_i, e_i)\\ &= \sum_i g(\nabla_{e_i} X, e_i) = \operatorname{div}_N(X)\end{aligned}$$

since $[X, e_i] = 0$ (i.e. the t and x_i derivatives commute). Moreover, we see that

$$\operatorname{div}_N(X) = \sum_i g(\nabla_{e_i} X, e_i) = \operatorname{div}_N(X^{\mathrm{T}}) - \sum_i g(X^\perp, \nabla_{e_i} e_i) = \operatorname{div}_N(X^{\mathrm{T}}) - g(X, H)$$

(since $\nabla_{e_i}\big(g(X^\perp, e_i)\big) = 0$) where $^{\mathrm{T}}$ and $^\perp$ denote the tangential and normal parts and

$$H = \sum_i \nabla^\perp_{e_i} e_i$$

is the *mean curvature vector*. Overall we have the following.

Theorem 2.2. *The first variation formula is*

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Vol}(F(N,t))|_{t=0} = \int_N \operatorname{div}_N(X)\operatorname{vol}_N = -\int_N g(X,H)\operatorname{vol}_N .$$

Remark. The $\operatorname{div}_N(X^{\mathrm{T}})$ term does not appear in the first variation formula because its integral vanishes by the divergence theorem as N is compact without boundary. In general, it will still vanish since we assume for our variations that there exists a compact submanifold of N with boundary which contains the support of X^{T} and so that X^{T} vanishes on the boundary.

We deduce the following.

Definition 2.3. N is a *minimal submanifold* if and only if $H = 0$.

Example. A plane in $\mathbb{R}^n$ is trivially minimal because if X, Y are any vector fields on the plane then $\nabla^\perp_X Y = 0$ as the second fundamental form of a plane is zero.

Example. For curves γ, $H = 0$ is equivalent to the geodesic equation $\nabla_{\dot\gamma}\dot\gamma = 0$.

The most studied minimal submanifolds (other than geodesics) are minimal surfaces in $\mathbb{R}^3$, since here the equation $H = 0$ becomes a scalar equation on a surface, which is the simplest to analyse. In general we would have a system of equations, which is more difficult to study.

Example. The helicoid $M = \{(t\cos s, t\sin s, s) \in \mathbb{R}^3 : s, t \in \mathbb{R}\}$ is a complete embedded minimal surface, discovered by Meusnier in 1776.

Example. The catenoid $M = \{(\cosh t\cos s, \cosh t\sin s, t) \in \mathbb{R}^3 : s, t, \in \mathbb{R}\}$ is a complete embedded minimal surface, discovered by Euler in 1744 and shown to be minimal by Meusnier in 1776. The catenoid is another explicit example which is a critical point for volume but not minimizing.

In fact the helicoid and the catenoid are locally isometric, and there is a 1-parameter family of locally isometric minimal surfaces deforming between the catenoid and helicoid: see, for example, [18, Theorem 16.5] for details.

It took about 70 years to find the next minimal surface, but now we know many examples of minimal surfaces in $\mathbb{R}^3$, as well as in other spaces by studying the nonlinear elliptic PDE given by the minimal surface equation. The amount of literature in the area is vast, with key results including the proofs of the Lawson [1], Willmore [63] and Yau [29, 64, 77] Conjectures, and minimal surfaces have applications to major problems in geometry including the Positive Mass Theorem [75, 76], Penrose Inequality [24] and Poincaré Conjecture [74].

2.2 Minimal graphs

The equation $H = 0$ is a *second order nonlinear PDE.* We can see this explicitly in the following simple case. For a function $f : U \subseteq \mathbb{R}^{n-1} \to \mathbb{R}$ where $\overline{U}$ is compact, we see that if $N = \mathrm{Graph}(f) \subseteq \mathbb{R}^n$ then the volume of N is given by

$$\mathrm{Vol}(N) = \int_U \sqrt{1 + |\nabla f|^2}\,\mathrm{vol}_U\,.$$

Any sufficiently small variation can be written $F(N, t) = \mathrm{Graph}(f + th)$ for some $h : U \to \mathbb{R}$, so we can compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Vol}(F(N,t))|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\int_U \sqrt{1+|\nabla f + t\nabla h|^2}\,\mathrm{vol}_U \\
&= \int_U \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\sqrt{1+|\nabla f|^2 + 2t\langle\nabla f, \nabla h\rangle + t^2|\nabla h|^2}\,\mathrm{vol}_U \\
&= \int_U \frac{\langle\nabla f, \nabla h\rangle}{\sqrt{1+|\nabla f|^2}}\,\mathrm{vol}_U \\
&= -\int_U h\,\mathrm{div}\left(\frac{\nabla f}{\sqrt{1+|\nabla f|^2}}\right)\mathrm{vol}_U\,.
\end{aligned}$$

We therefore see that N is minimal if and only if this vanishes for all h.

Hence, $\mathrm{Graph}(f)$ is minimal in $\mathbb{R}^n$ if and only if and only if

$$\mathrm{div}\left(\frac{\nabla f}{\sqrt{1+|\nabla f|^2}}\right) = 0.$$

We see that we can write this equation as $\Delta f + Q(\nabla f, \nabla^2 f) = 0$ where Q consists of nonlinear terms (but linear in $\nabla^2 f$) and

$$\Delta f = -\,\mathrm{div}(\nabla f) = -\,\mathrm{tr}(\mathrm{Hess}\, f).$$

(The minus sign gives the "geometer's Laplacian", whereas the plus sign gives the "analyst's Laplacian"). Hence, if we linearise this equation we just get $\Delta f = 0$, so f is harmonic. More concretely, linearising the operator $Pf = 0$ (at 0) means calculating the linear operator

$$Lf = L_0 Pf = \frac{\partial}{\partial t}P(tf)|_{t=0}.$$

In other words, the minimal submanifold equation is a nonlinear equation whose linearisation is just Laplace's equation: this is an example of a nonlinear *elliptic* PDE, which we shall discuss further later. For now, to compute the symbol of a linear operator L of order k, you compute

$$\sigma_L(x, \xi) = \lim_{t\to\infty} t^{-k} e^{-itf} L(e^{itf})(x)$$

where $\xi = \mathrm{d}f(x) \in T_x^*M$ for a function f with $f(x) = 0$. Ellipticity says σ_L is an isomorphism whenever $\xi \neq 0$. In the case of the Laplacian $L = \Delta$ we get

$$\sigma_\Delta(x, \xi) = -|\xi|^2,$$

which is clearly an isomorphism for $\xi \neq 0$.

Now, as we have mentioned before, minimal submanifolds are only critical points for the volume functional, so a natural question to ask is: when is a minimal submanifold a minimizer for the volume functional?

This is very difficult to answer in general, and we see already for example that a plane is a minimizer but the catenoid is not a minimizer (simply by dilating it). We now see that minimal graphs are always volume minimizers, but even in this simple case the reason why we know that is due to calibrated geometry.

Theorem 2.4. *Suppose that $N = \mathrm{Graph}(f)$ for $f : U \subseteq \mathbb{R}^{n-1} \to \mathbb{R}$, where U is an open subset with compact closure. Let N' be a variation of N in $U \times \mathbb{R}$ with the same boundary as N. Then*

$$\mathrm{Vol}(N) \leq \mathrm{Vol}(N').$$

Proof. Since T^*N is trivial (as T^*U is trivial) we can choose a global orthonormal coframe $\xi_1, \ldots, \xi_{n-1}$ on N and form the $(n-1)$-form

$$\eta = \xi_1 \wedge \ldots \wedge \xi_{n-1}.$$

We can trivially extend η to $U \times \mathbb{R} \subseteq \mathbb{R}^n$ in a parallel fashion so that it is independent of the "vertical" x_n coordinate.

One fact is clear about η:

$$\eta(e_1, \ldots, e_{n-1}) \leq 1 \text{ for all unit tangent vectors } e_1, \ldots, e_{n-1} \text{ on } \mathbb{R}^n$$

since $\xi_1, \ldots, \xi_{n-1}$ are all unit.

The second fact is less clear: if N is minimal then

$$\mathrm{d}\eta = 0.$$

The reason is that $\mathrm{d}\eta$ is, up to a sign,

$$\mathrm{div}\left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}}\right) \mathrm{vol}_{\mathbb{R}^{n+1}}$$

which vanishes precisely when N is minimal. To show this, we notice that $\eta = \nu \lrcorner \mathrm{vol}_{\mathbb{R}^n}$ for a unit normal ν to N (then extended in a parallel manner to $U \times \mathbb{R}$. We see that on $U \times \mathbb{R}$ we have that (up to sign)

$$\nu = \frac{(-\nabla f, 1)}{\sqrt{1 + |\nabla f|^2}}.$$

The computation that $\mathrm{d}\eta = 0$ is then a straightforward exercise.

If N' is another submanifold in $U \times \mathbb{R}$ with the same boundary as N which is a variation of N, then there exists K compact and n-dimensional interpolating between N and N' and we can apply Stokes' Theorem:

$$0 = \int_K \mathrm{d}\eta = \int_{N'} \eta - \int_N \eta.$$

Why is this good? Well, $\eta|_N = \mathrm{vol}_N$ and $\eta|_{N'} \leq \mathrm{vol}_{N'}$ so

$$\mathrm{Vol}(N) = \int_N \mathrm{vol}_N = \int_N \eta = \int_{N'} \eta \leq \int_{N'} \mathrm{vol}_{N'} = \mathrm{Vol}(N').$$

Hence N is volume-minimizing. $\square$

Another question to ask is: how do we find minimal submanifolds? There are two main answers to this. The first, and most natural, is the variational approach. That is, simply minimize the volume functional. The problem with this is that the topological space of submanifolds does not have good compactness properties: i.e. it is quite easy to find a sequence of compact submanifolds with a uniform bound on their volume whose limit is not a smooth submanifold. Therefore, one has to enlarge the space of submanifolds to a weaker notion, for example integral varifolds or integral currents. We will discuss this briefly later. The point is that if you directly minimize you have no control on the minimizing sequence and you might end up with some very singular object at the end. The task is then to prove that maybe it is not as singular as you thought, and in the best case scenario that is smooth. This is precisely the method by which Hodge theory is proved. However, this is much more challenging in the case of submanifolds, and very often one has to deal with complicated singularities.

2.3 Calibrations and calibrated geometry

As we have seen, minimal submanifolds are extremely important. However there are two key issues.

- Minimal submanifolds are defined by a second order nonlinear PDE system – therefore they are hard to analyse.
- Minimal submanifolds are only critical points for the volume functional, but we are often interested in minima for the volume functional – we need a way to determine when this occurs.

We can help resolve these issues using the notion of calibration and calibrated submanifolds, introduced by Harvey–Lawson [20] in 1982.

Definition 2.5. A differential k-form η on a Riemannian manifold (M, g) is a *calibration* if

- $\mathrm{d}\eta = 0$ and
- $\eta(e_1, \ldots, e_k) \leq 1$ for all unit tangent vectors $e_1, \ldots, e_k$ on M.

Example. Any non-zero form with constant coefficients on $\mathbb{R}^n$ can be rescaled so that it is a calibration with at least one plane where equality holds.

This example shows that there are many calibrations η, but the interesting question is: for which oriented planes $P = \mathrm{Span}\{e_1, \ldots, e_k\}$ does $\eta(e_1, \ldots, e_k) = 1$? More importantly, can we find submanifolds N so that this equality holds on each tangent space? This motivates the next definition.

Definition 2.6. Let η be a calibration k-form on (M, g). An oriented k-dimensional submanifold N of (M, g) is *calibrated* by η if $\eta|_N = \mathrm{vol}_N$, i.e. if for all $p \in N$ we have $\eta(e_1, \ldots, e_k) = 1$ for an oriented orthonormal basis $e_1, \ldots, e_k$ for T_pN.

Example. Any oriented plane in $\mathbb{R}^n$ is calibrated. If we change coordinates so that the plane P is $\{x \in \mathbb{R}^n : x_{k+1} = \ldots = x_n = 0\}$ (with the obvious orientation) then $\eta = \mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_k$ is a calibration and P is calibrated by η.

Notice that the calibrated condition is now an algebraic condition on the tangent vectors to N, so being calibrated is a *first order nonlinear PDE*. We shall motivate these definitions further later, but for now we make the following observation.

Theorem 2.7. *Let N be a calibrated submanifold. Then N is minimal and, moreover, if F is any variation with compact support $\overline{S}$ then $\mathrm{Vol}(F(S,t)) \geq \mathrm{Vol}(S)$; i.e. N is volume-minimizing. In particular, if N is compact then N is volume-minimizing in its homology class.*

Proof. Suppose that N is calibrated by η and suppose for simplicity that N is compact. We will show that N is homologically volume-minimizing.

Suppose that N' is homologous to N. Then

$$\int_{N'} \eta = [\eta] \cdot [N'] = [\eta] \cdot [N] = \int_N \eta,$$

where $[\eta]\cdot[N]$ represents the pairing between the cohomology class of η in $H^k(M)$ and the homology class of N in $H_k(M)$. We deduce that, since

$$\eta|_N = \mathrm{vol}_N \quad \text{and} \quad \eta|_{N'} \leq \mathrm{vol}_{N'}$$

as η is a calibration and N is η-calibrated:

$$\mathrm{Vol}(N) = \int_N \mathrm{vol}_N = \int_N \eta = \int_{N'} \eta \leq \int_{N'} \mathrm{vol}_{N'} \,\mathrm{Vol}(N').$$

We then have the result by the definition of minimal submanifold. □

We conclude this introduction with the following elementary result.

Proposition 2.8. *There are no compact calibrated submanifolds in* $\mathbb{R}^n$*.*

Proof. Suppose that η is a calibration and N is compact and calibrated by η. Then $\mathrm{d}\eta = 0$ so by the Poincaré Lemma $\eta = \mathrm{d}\zeta$, and hence

$$\mathrm{Vol}(N) = \int_N \eta = \int_N \mathrm{d}\zeta = 0$$

by Stokes' Theorem. □

2.4 Calibrations and holonomy

Although there are many calibrations, having calibrated submanifolds greatly restricts the calibrations you want to consider. The calibrations which have calibrated submanifolds have special significance and there is a particular connection with *special holonomy*, as we now explain.

First, we have to say what the holonomy group of a Riemannian manifold is.

Definition 2.9. Let (M, g) be a connected n-dimensional Riemannian manifold. Given $p \in M$, we define $\mathrm{Hol}_p(g)$ to be the group generated by parallel transport around loops in M based at p. In fact, $\mathrm{Hol}_p(g)$ is independent of p in the sense that, if $q \in M$ and $P_{p,q}$ denotes parallel transport along some path from p to q then

$$\mathrm{Hol}_q(g) = P_{p,q}\mathrm{Hol}_p(g)P_{p,q}^{-1}.$$

Since each parallel transport map defines an orothogonal transformation, each $\mathrm{Hol}_p(g)$ can be viewed as a subgroup of $\mathrm{O}(n)$, which are the same up to conjugation.

Hence, we define the *holonomy group* $\mathrm{Hol}(g)$ of g to be subgroup of $\mathrm{O}(n)$ up to conjugation defined by any of the $\mathrm{Hol}_p(g)$.

Remark. It is straightforward to see that if M is simply connected then $\mathrm{Hol}(g)$ is connected, and so will be a subgroup of $\mathrm{SO}(n)$.

Let $G = \mathrm{Hol}(g)$ be the holonomy group of a Riemannian metric g on an n-manifold M. Then G acts on the k-forms on $\mathbb{R}^n$, so suppose that η_0 is a G-invariant k-form with constant coefficients. We can always rescale η_0 so that $\eta_0|_P \leq \mathrm{vol}_P$ for all oriented k-planes P and equality holds for at least one P. Since η_0 is G-invariant, if P is calibrated then so is $\gamma\cdot P$ for any $\gamma \in G$, which usually means we have quite a few calibrated planes. We know by the *holonomy principle* (see, for example, [42, Proposition 2.5.2]) that we then get a parallel k-form η on M which is identified with η_0 at every point. Since $\nabla\eta = 0$, we have $\mathrm{d}\eta = 0$ and hence η is a calibration. Moreover, we have a lot of calibrated tangent planes on M, so we can hope to find calibrated submanifolds.

This discussion motivates the following question.

Question 2.10. Which groups can be holonomy groups of a Riemannian manifold?

To make this question more tractable and make sure our classification doesn't have some obvious redundancies, we make some observations.

We first observe that if (M, g) is isometric to a product of two Riemannian manifolds (M_1, g_1), (M_2, g_2), then

$$\mathrm{Hol}(g) = \mathrm{Hol}(g_1) \times \mathrm{Hol}(g_2).$$

In fact, the same can occur even if (M, g) is only *reducible*, meaning that it is locally isometric to a product. If (M, g) is *irreducible* (i.e. if it is not locally isometric to a product), we have the following.

Lemma 2.11. *If (M, g) is irreducible and n-dimensional then* $\mathrm{Hol}(g)$ *acts irreducibly on* $\mathbb{R}^n$*.*

Remark. Here we should note that the holonomy group is not just given as a subgroup of $\mathrm{O}(n)$, but also its representation as a subgroup of the standard representation of $\mathrm{O}(n)$ as orthogonal transformations of $\mathbb{R}^n$ is given. This is the meaning of the action in the lemma above.

We now notice that if (M, g) is a symmetric space, which will then mean that it is of the form G/H for groups G, H with a G-invariant metric, then $\mathrm{Hol}(g) = H$. This means that classifying the holonomy of symmetric spaces is a purely algebraic problem, solved by Cartan in 1925. In fact, the same algebra problem classifies holonomy groups of *locally symmetric* spaces: those for which the Riemann curvature is parallel.

Finally, we restrict to simply connected manifolds, which we can always do by passing to the universal cover, and obtain the following classification theorem due to Berger in 1955.

Theorem 2.12 (Berger)**.** *Let (M, g) be a simply connected n-dimensional Riemannian manifold which is irreducible and not locally symmetric. Then* $\mathrm{Hol}(g)$ *can be only be one of the following groups* G *(and name of the associated type of metric):*

Dimension n	Group G	Name
any n	$\mathrm{SO}(n)$	generic
$n = 2m$	$\mathrm{U}(m)$	Kähler
$n = 2m$	$\mathrm{SU}(m)$	Calabi–Yau
$n = 4k$	$\mathrm{Sp}(k)$	hyperkähler
$n = 4k$	$\mathrm{Sp}(k) \cdot \mathrm{Sp}(1)$	quaternionic Kähler
$n = 7$	G_2	holonomy G_2
$n = 8$	$\mathrm{Spin}(7)$	holonomy $\mathrm{Spin}(7)$

Remark. Berger's theorem only tells you which groups *could* occur as holonomy groups. It does not tell you that they *do* occur. In fact, another group ($\mathrm{Spin}(9)$) was initially on the list in dimension 16, but it was shown later that this cannot occur for non-locally symmetric Riemannian manifolds.

Some comments on the list in Berger's theorem are probably in order, particularly since some of the groups are more familiar than others.

- The groups other than $\mathrm{SO}(n)$ and $\mathrm{U}(n)$ are called the *special holonomy* groups.
- The compact symplectic group $\mathrm{Sp}(k)$ is the subgroup of $\mathrm{GL}(\mathbb{H}^k)$, where $\mathbb{H}$ is the quaternions, preserving the following inner product:

$$\langle (x_1, \ldots, x_k), (y_1, \ldots, y_k) \rangle = \sum_{j=1}^{k} x_j \bar{y}_j.$$

- We then have
$$\mathrm{Sp}(k) \cdot \mathrm{Sp}(1) = \frac{\mathrm{Sp}(k) \times \mathrm{Sp}(1)}{\{(1, 1), (-1, -1)\}}.$$
Its action on $\mathbb{H}^k$ is by left multiplication by the first factor and right multiplication by the conjugate of the second factor, once one diagonally embeds $\mathrm{Sp}(1)$ in $\mathrm{Sp}(k)$.
- The last two groups (G_2 and $\mathrm{Spin}(7)$) are called the *exceptional holonomy groups*.
- The group G_2 is a 14-dimensional exceptional Lie group which is the automorphism group of the octonions $\mathbb{O}$, and so has a natural representation on $\mathbb{R}^7 = \mathrm{Im}\,\mathbb{O}$ since it must fix the unit in $\mathbb{O}$.
- The group $\mathrm{Spin}(7)$ is the double cover of $\mathrm{SO}(7)$ and has a natural action on $\mathbb{R}^8$ which is a spin representation, which also has an interpretation in terms of octonions.

Remark. The special holonomy groups all give rise to *Ricci-flat* metrics, apart from the quaternionic Kähler metrics, which are *Einstein* metrics with *non-zero* Einstein constant/scalar curvature, i.e. $\mathrm{Ric}(g) = \lambda g$ for some non-zero constant λ. This makes special holonomy metrics particularly interesting.

The reason why one should believe that holonomy restricts the curvature of a Riemannian manifold is due to the *Ambrose–Singer theorem.* This implies that, at each point $p \in (M, g)$, the Riemann curvature Rm_p when viewed as a symmetric map on $\Lambda^2 T_p^* M \cong \mathfrak{so}(n)$, is actually a symmetric map on $\mathfrak{hol}_p(g) \subset \mathfrak{so}(n)$ (the Lie algebra of $\mathrm{Hol}_p(g)$).

Now that we have candidate geometries where calibrated geometries may be interesting, we wish to explore some of these examples for the remainder of this first part of the course.

3 Complex and special Lagrangian submanifolds

In this section we study calibrated geometry associated with the first two non-trivial holonomy groups on Berger's list: $\mathrm{U}(n)$ and $\mathrm{SU}(n)$.

3.1 Kähler manifolds and complex submanifolds

To start, we would now like to address the question: where does the calibration condition come from? The answer is from *complex geometry.*

On $\mathbb{R}^{2n} = \mathbb{C}^n$ with coordinates $z_j = x_j + iy_j$, we have the complex structure J_0 and the distinguished Kähler 2-form

$$\omega_0 = \sum_{j=1}^n \mathrm{d}x_j \wedge \mathrm{d}y_j = \frac{i}{2}\sum_{j=1}^n \mathrm{d}z_j \wedge \mathrm{d}\bar{z}_j.$$

Note that both J_0 and ω_0 are constant on $\mathbb{C}^n$ and invariant under the action of $\mathrm{U}(n)$. This means that more generally we can work with a *Kähler manifold* (M, g, J, ω).

Definition 3.1. A *Kähler manifold* is a $2n$-dimensional Riemannian manifold (M, g) such that $\mathrm{Hol}(g) \subseteq \mathrm{U}(n)$. Such a manifold is endowed with a parallel 2-form ω which is identified with ω_0 at every point and is called the Kähler form. It also has a parallel complex structure J (which is identified with J_0 at every point), such that $g(Ju, v) = \omega(u, v)$ for all tangent vectors u, v. We then write (M, g, J, ω) for the choice of Kähler form and associated complex structure.

Our first key result is the following.

Theorem 3.2. *On a Kähler manifold (M, g, J, ω), $\frac{\omega^k}{k!}$ is a calibration whose calibrated submanifolds are the complex k-dimensional submanifolds: i.e. submanifolds N such that $J(T_pN) = T_pN$ for all $p \in N$.*

Since $\mathrm{d}\omega^k = k\mathrm{d}\omega \wedge \omega^{k-1} = 0$, Theorem 3.2 follows immediately from the following result.

Theorem 3.3 (Wirtinger's inequality)**.** *For any orthonormal vectors $e_1, \ldots, e_{2k} \in \mathbb{C}^n$,*

$$\frac{\omega_0^k}{k!}(e_1, \ldots, e_{2k}) \leq 1$$

with equality if and only if $\mathrm{Span}\{e_1, \ldots, e_{2k}\}$ *is a complex k-plane in $\mathbb{C}^n$.*

Before proving this we make the following observation.

Lemma 3.4. *If η is a calibration and $*\eta$ is closed then $*\eta$ is a calibration. Moreover an oriented tangent plane P is calibrated by η if and only if there is an orientation on the orthogonal complement $P^\perp$ so that it is calibrated by $*\eta$.*

Proof. Suppose that η is a calibration k-form on (M, g) with $\mathrm{d}*\eta = 0$. Let $p \in M$. Take any $n - k$ orthonormal tangent vectors $e_{k+1}, \ldots, e_n$ at p. Then there exist $e_1, \ldots, e_k \in T_pM$ so that $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for T_pM. Since $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis, we can use the definition of the Hodge star to calculate

$$*\eta(e_{k+1}, \ldots, e_n) = \eta(e_1, \ldots, e_k) \leq 1.$$

Hence $*\eta$ is a calibration by Definition 2.5. Moreover, the oriented plane $P = \mathrm{Span}\{e_{k+1}, \ldots, e_n\}$ is calibrated by $*\eta$ if and only if there is an orientation on $\mathrm{Span}\{e_1, \ldots, e_k\} = P^\perp$ so that it is calibrated by η, since $\eta(e_1, \ldots, e_k) = \pm * \eta(e_{k+1}, \ldots, e_n) = \pm 1$. $\square$

We can now prove Wirtinger's inequality.

Proof of Theorem 3.3. For ease of notation we shall write $J = J_0$ and $\omega = \omega_0$.

We see that $|\frac{\omega^k}{k!}|^2 = \frac{n!}{k!(n-k)!}$ and $\mathrm{vol}_{\mathbb{C}^n} = \frac{\omega^n}{n!}$ so $*\frac{\omega^k}{k!} = \frac{\omega^{n-k}}{(n-k)!}$. Hence, by Lemma 3.4, it is enough to study the case where $k \leq \frac{n}{2}$.

Let P be any $2k$-plane in $\mathbb{C}^n$ with $2k \leq n$. We shall find a canonical form for P. First consider $\langle Ju, v\rangle$ for orthonormal vectors $u, v \in P$. This must have a maximum, so let $\cos\theta_1 = \langle Ju, v\rangle$ be this maximum realised by some orthonormal vectors $u, v \in P$, where $0 \leq \theta_1 \leq \frac{\pi}{2}$.

Suppose that $w \in P$ is a unit vector orthogonal to $\mathrm{Span}\{u, v\}$, where $\cos\theta_1 = \langle Ju, v\rangle$. The function

$$f_w(\theta) = \langle Ju, \cos\theta v + \sin\theta w\rangle$$

has a maximum at $\theta = 0$ so $f_w'(0) = \langle Ju, w\rangle = 0$. Similarly we have that $\langle Jv, w\rangle = 0$, and thus $w \in \mathrm{Span}\{u, v, Ju, Jv\}^\perp$.

We then have two cases. If $\theta_1 = 0$ then $v = Ju$ so we can set $u = e_1, v = Je_1$ and see that $P = \mathrm{Span}\{e_1, Je_1\} \times Q$ where Q is a $2(k-1)$-plane in $\mathbb{C}^{n-1} = \mathrm{Span}\{e_1, Je_1\}^\perp$. If $\theta_1 \neq 0$ we have that $v = \cos\theta_1 Ju + \sin\theta_1 w$ where w is a unit vector orthogonal to u and Ju, so we can let $u = e_1$, $w = e_2$ and see that $P = \mathrm{Span}\{e_1, \cos\theta_1 Je_1 + \sin\theta_1 e_2\} \times Q$ where Q is a $2(k-1)$-plane in $\mathbb{C}^{n-2} = \mathrm{Span}\{e_1, Je_1, e_2, Je_2\}^\perp$.

Proceeding by induction we see that we have an oriented basis $\{e_1, Je_1, \ldots, e_n, Je_n\}$ for $\mathbb{C}^n$ so that

$$P = \mathrm{Span}\{e_1, \cos\theta_1 Je_1 + \sin\theta_1 e_2, \ldots, e_{2k-1}, \cos\theta_k Je_{2k-1} + \sin\theta_k e_{2k}\},$$

where $0 \leq \theta_1 \leq \ldots \leq \theta_{k-1} \leq \frac{\pi}{2}$ and $\theta_{k-1} \leq \theta_k \leq \pi - \theta_{k-1}$. (The reason why we cannot determine the sign of $\cos\theta_k$ in the last step is that the orientation of P is fixed.)

Since we can write $\omega = \sum_{j=1}^n e^j \wedge Je^j$ (where e^j is the covector dual to e_j) we see that $\frac{\omega^k}{k!}$ restricts to P to give a product of $\cos\theta_j$ which is certainly less than or equal to 1. Moreover, equality holds if and only if all of the $\theta_j = 0$ which means that P is complex. $\square$

Putting together Theorem 3.2 and Theorem 2.7 yields the following.

Corollary 3.5. *Compact complex submanifolds of Kähler manifolds are homologically volume-minimizing.*

We know that complex submanifolds are defined by holomorphic functions; i.e. solutions to the Cauchy–Riemann equations, which are a first-order PDE system, as one would expect for calibrated submanifolds.

Example. The 2-dimensional submanifold

$$N = \{(z_1, z_2) \in \mathbb{C}^2 \,:\, z_1 z_2 = 1\} \cong \mathbb{C}^*$$

is a complex curve in $\mathbb{C}^2$, and thus is calibrated. This example will appear again later.

Example. An important non-trivial example of a Kähler manifold is $\mathbb{CP}^n$, where the zero set of a system of polynomial equations defines a (possibly singular) complex submanifold. This is one of the first links between calibrated geometry and algebraic geometry.

Example. Consider the complex curve

$$N = \{(z_1, z_2) \in \mathbb{C}^2 \,:\, z_1^2 = z_2^3\}.$$

This is calibrated but clearly has a singularity at the origin. This shows that we must allow calibrated submanifolds to have singularities at least in *codimension* 2. In fact, a famous theorem of Almgren implies that calibrated submanifolds (or just ones that are volume-minimizing) have singularities in *at most* codimension 2. Therefore, in particular, one can speculate about the possible relationship between codimension 2 singularities of calibrated submanifolds and complex geometry.

3.2 Calabi–Yau manifolds and special Lagrangians

Complex submanifolds are very familiar, but can we find any other interesting classes of calibrated submanifolds? The answer is that indeed we can, particularly when the manifold has special holonomy as we alluded to earlier. We begin with the case of holonomy (contained in) $\mathrm{SU}(n)$ – so-called *Calabi–Yau manifolds.*

The model example for Calabi–Yau manifolds is $\mathbb{C}^n$ with complex structure J_0, Kähler form ω_0 and holomorphic volume form

$$\Upsilon_0 = \mathrm{d}z_1 \wedge \ldots \wedge \mathrm{d}z_n,$$

if $z_1, \ldots, z_n$ are complex coordinates on $\mathbb{C}^n$.

Definition 3.6. A *Calabi–Yau manifold* is a Kähler manifold (M, g, J, ω) of complex dimension n such that $\mathrm{Hol}(g) \subseteq \mathrm{SU}(n)$. In this case, there is parallel complex n-form Υ on M (called a holomorphic volume form) which is identified with Υ_0 at every point. We then write $(M, g, J, \omega, \Upsilon)$ for the choice of holomorphic volume form on the Calabi-Yau manifold.

Theorem 3.7. *Let $(M, g, J, \omega, \Upsilon)$ be a Calabi–Yau manifold. Then $\mathrm{Re}(e^{-i\theta}\Upsilon)$ is a calibration for any $\theta \in \mathbb{R}$.*

Since $\mathrm{d}\Upsilon = 0$, the result follows immediately from the following.

Theorem 3.8. *On $\mathbb{C}^n$, for all orthonormal vectors $e_1, \ldots, e_n$ we have*

$$|\Upsilon_0(e_1, \ldots, e_n)| \leq 1$$

with equality if and only if $P = \mathrm{Span}\{e_1, \ldots, e_n\}$ is a Lagrangian plane, i.e. P is an n-plane such that $\omega_0|_P \equiv 0$.

Proof. Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{R}^n$ and let P be an n-plane in $\mathbb{C}^n$. There exists $A \in \mathrm{GL}(n, \mathbb{C})$ so that $f_1 = Ae_1, \ldots, f_n = Ae_n$ is an orthonormal basis for P. Then $\Upsilon_0(Ae_1, \ldots, Ae_n) = \det_{\mathbb{C}}(A)$ (since Υ_0 is a complex volume form). Using the fact the linear algebra fact that $|\det_{\mathbb{C}}(A)|^2 = |\det_{\mathbb{R}}(A)|$, we see that

$$\begin{aligned}|\Upsilon_0(f_1, \ldots, f_n)|^2 &= |\det\nolimits_{\mathbb{C}}(A)|^2 = |\det\nolimits_{\mathbb{R}}(A)| \\ &= |f_1 \wedge Jf_1 \wedge \ldots \wedge f_n \wedge Jf_n| \leq |f_1||Jf_1| \ldots |f_n||Jf_n| = 1,\end{aligned}$$

with equality if and only if $f_1, Jf_1, \ldots, f_n, Jf_n$ are orthonormal. However, this is exactly equivalent to the Lagrangian condition, since $\omega_0(u, v) = \langle J_0 u, v\rangle$ so $\omega_0|_P \equiv 0$ if and only if $J_0 P = P^\perp$. □

Remark. We notice that by Theorem 3.8, for any oriented Lagrangian submanifold N in a Calabi–Yau manifold $(M, g, J, \omega, \Upsilon)$ we have that

$$\Upsilon|_N = e^{i\theta}\,\mathrm{vol}_N$$

for an $\mathcal{S}^1$-valued function $e^{i\theta} : N \to \mathcal{S}^1 \subseteq \mathbb{C}$, often called the *Lagrangian phase.* The *Lagrangian angle* is a choice of (possibly) multi-valued function θ on N representing the phase.

If θ can be chosen to be single-valued we call N *zero Maslov.* More generally, the class $[\mathrm{d}\theta] \in H^1(N)$ (up to a possible normalization by a multiple of π) is called the *Maslov class* of N. The Maslov class is a kind of "winding number": we can map each $p \in N$ to $T_pN \in U(n)/SO(n)$ (the oriented Lagrangian Grassmannian) and then take the determinant to give the Lagrangian phase map $e^{i\theta}$, and then pull back the fundamental class of $H^1(\mathcal{S}^1)$ to N to give $[\mathrm{d}\theta]$.

Definition 3.9. An oriented n-dimensional submanifold N of a Calabi–Yau n-fold $(M, g, J, \omega, \Upsilon)$ calibrated by $\mathrm{Re}(e^{-i\theta}\Upsilon)$, i.e. such that

$$\mathrm{Re}(e^{-i\theta}\Upsilon)|_N = \mathrm{vol}_N,$$

is called *special Lagrangian* with phase $e^{i\theta}$. If $\theta = 0$ we say that N is simply special Lagrangian.

By Theorem 3.8, we see that N is special Lagrangian if and only if $\omega|_N \equiv 0$ (i.e. N is Lagrangian) and $\mathrm{Im}\,\Upsilon|_N \equiv 0$ (up to a choice of orientation so that $\mathrm{Re}\,\Upsilon|_N > 0$).

Example. Consider $\mathbb{C} = \mathbb{R}^2$ with coordinates $z = x + iy$, complex structure J given by $Jw = iw$, Kähler form $\omega = \mathrm{d}x \wedge \mathrm{d}y = \frac{i}{2}\mathrm{d}z \wedge \mathrm{d}\bar{z}$ and holomorphic volume form $\Upsilon = \mathrm{d}z = \mathrm{d}x + i\mathrm{d}y$. We want to consider the special Lagrangians in $\mathbb{C}$, which are 1-dimensional submanifolds or curves N in $\mathbb{C} = \mathbb{R}^2$.

Since ω is a 2-form, it vanishes on any curve in $\mathbb{C}$. Hence every curve in $\mathbb{C}$ is Lagrangian. For N to be special Lagrangian with phase $e^{i\theta}$ we need that

$$\mathrm{Re}(e^{-i\theta}\Upsilon) = \cos\theta \mathrm{d}x + \sin\theta \mathrm{d}y$$

is the volume form on N, or equivalently that

$$\mathrm{Im}(e^{-i\theta}\Upsilon) = \cos\theta \mathrm{d}y - \sin\theta \mathrm{d}x$$

vanishes on N. This means that $\cos\theta\partial_x + \sin\theta\partial_y$ is everywhere a unit tangent vector to N, so N is a straight line given by $N = \{(t\cos\theta, t\sin\theta) \in \mathbb{R}^2 : t \in \mathbb{R}\}$ (up to translation), so it makes an angle θ with the x-axis, hence motivating the term "phase $e^{i\theta}$".

Notice that this result is compatible with the fact that special Lagrangians are minimal, and hence must be geodesics in $\mathbb{R}^2$; i.e. straight lines.

Example. Consider $\mathbb{C}^2 = \mathbb{R}^4$. We know that $\omega = \mathrm{d}x_1 \wedge \mathrm{d}y_1 + \mathrm{d}x_2 \wedge \mathrm{d}y_2$. Since $\Upsilon = \mathrm{d}z_1 \wedge \mathrm{d}z_2 = (\mathrm{d}x_1 + i\mathrm{d}y_1) \wedge (\mathrm{d}x_2 + i\mathrm{d}y_2)$, we also know that $\mathrm{Re}\,\Upsilon = \mathrm{d}x_1 \wedge \mathrm{d}x_2 + \mathrm{d}y_2 \wedge \mathrm{d}y_1$, which looks somewhat similar. In fact, if we let J' denote the complex structure given by $J'(\partial_{x_1}) = \partial_{x_2}$ and $J'(\partial_{y_2}) = \partial_{y_1}$, then $\mathrm{Re}\,\Upsilon = \omega'$, the Kähler form corresponding the complex structure J'. Hence special Lagrangians in $\mathbb{C}^2$ are complex curves for a different complex structure.

In fact, we have a *hyperkähler triple* of complex structures J_1, J_2, J_3, where $J_1 = J$ is the standard one and $J_3 = J_1J_2 = -J_2J_1$ so that $J_1 = J_2J_3 = -J_3J_2$ and $J_2 = J_3J_1 = -J_1J_3$, and the corresponding Kähler forms are $\omega = \omega_1$, ω_2, ω_3 which are orthogonal and the same length with $\Upsilon = \omega_2 + i\omega_3$. The transformation we just performed above is from J_1 to J_2 and is an example of a hyperkähler rotation.

This shows we should only consider complex dimension 3 and higher to find new calibrated submanifolds.

Example. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function and let $N = \mathrm{Graph}(f) \subseteq \mathbb{R}^{2n} = \mathbb{C}^n$. We want to see when N is special Lagrangian. We see that tangent vectors to N are given by

$$e_1 + i\nabla_{e_1}f, \ldots, e_n + i\nabla_{e_n}f.$$

Hence N is Lagrangian if and only if

$$\omega(e_j + i\nabla_{e_j}f, e_k + i\nabla_{e_k}f) = \nabla_{e_k}f_j - \nabla_{e_j}f_k = 0$$

for all j, k. Since $\mathbb{R}^n$ is simply connected, this occurs if and only if there exists F such that $f_j = \nabla_{e_j}F$; i.e. $f = \nabla F$.

Recall that $\Upsilon = \mathrm{d}z_1 \wedge \ldots \wedge \mathrm{d}z_n$. We know that N is special Lagrangian if and only if N is Lagrangian and $\mathrm{Im}\,\Upsilon$ vanishes on N. Now

$$\Upsilon(a_1 + ib_1, \ldots, a_n + ib_n) = \det_{\mathbb{C}}(A + iB)$$

where A, B are the matrices with columns a_i, b_j respectively. Hence

$$\Upsilon(e_1 + i\nabla_{e_1}\nabla F, \ldots, e_n + i\nabla_{e_n}\nabla F) = \det_{\mathbb{C}}(I + i\mathrm{Hess}\,F),$$

where $\mathrm{Hess}\,F = (\frac{\partial^2 F}{\partial x_i \partial x_j})$.

Therefore $N = \mathrm{Graph}(f)$ is special Lagrangian (up to a choice of orientation) if and only if $f = \nabla F$ and

$$\mathrm{Im}\det_{\mathbb{C}}(I + i\,\mathrm{Hess}\,F) = 0.$$

If $n = 2$,

$$I + i\mathrm{Hess}\,F = \begin{pmatrix} 1 + iF_{xx} & iF_{xy} \\ iF_{yx} & 1 + iF_{yy} \end{pmatrix}.$$

Therefore, the determinant gives

$$1 - F_{xx}F_{yy} + F_{xy}^2 + i(F_{xx} + F_{yy}),$$

then the imaginary part is $F_{xx} + F_{yy}$. Therefore, N is special Lagrangian if and only if $\Delta F = 0$.

As we know, a graph in $\mathbb{C}^2$ of $f = u + iv : \mathbb{C} \to \mathbb{C}$ is a complex surface if and only if $u + iv$ is holomorphic, which implies that u, v are harmonic. We know that special Lagrangians in $\mathbb{C}^2$ are complex surfaces for a different complex structure, so this is expected.

If $n = 3$,

$$I + i\mathrm{Hess}\, F = \begin{pmatrix} 1 + iF_{xx} & iF_{xy} & iF_{xz} \\ iF_{yx} & 1 + iF_{yy} & iF_{yz} \\ iF_{zx} & iF_{zy} & 1 + iF_{zz} \end{pmatrix}.$$

Hence,

$$\begin{aligned} \mathrm{Im}\det\nolimits_{\mathbb{C}}(I + i\,\mathrm{Hess}\, F) &= F_{xx} + F_{yy} + F_{zz} \\ &\quad - F_{xx}(F_{yy}F_{zz} - F_{yz}^2) - F_{xy}(F_{yz}F_{zx} - F_{xy}F_{zz}) - F_{zx}(F_{xy}F_{yz} - F_{yy}F_{zx}). \end{aligned}$$

Therefore, N is special Lagrangian if and only if

$$\begin{aligned} -\Delta F &= F_{xx} + F_{yy} + F_{zz} \\ &= F_{xx}(F_{yy}F_{zz} - F_{yz}^2) - F_{xy}(F_{xy}F_{zz} - F_{yz}F_{zx}) + F_{zx}(F_{xy}F_{yz} - F_{yy}F_{zx}) \\ &= \det \mathrm{Hess}\, F, \end{aligned}$$

which is related to the *real Monge–Ampère equation*.

We now wish to describe some very important examples of special Lagrangians, which are asymptotic to pairs of planes.

Example. It is a fact that $\mathrm{SU}(n)$ acts transitively on the space of special Lagrangian planes with isotropy $\mathrm{SO}(n)$. Hence, any special Lagrangian plane is given by $A \cdot \mathbb{R}^n$ for $A \in \mathrm{SU}(n)$ where $\mathbb{R}^n$ is the standard real $\mathbb{R}^n$ in $\mathbb{C}^n$.

Given $\theta = (\theta_1, \ldots, \theta_n)$ we can define a plane $P(\theta)$ by

$$P(\theta) = \{(e^{i\theta_1}x_1, \ldots, e^{i\theta_n}x_n) \in \mathbb{C}^n \,:\, (x_1, \ldots, x_n) \in \mathbb{R}^n\},$$

noting that we can swap orientation in this expression. We see that $P(\theta)$ is special Lagrangian if and only if $\mathrm{Re}\,\Upsilon|_P = \pm\cos(\theta_1 + \ldots + \theta_n) = 1$ so that

$$\theta_1 + \ldots + \theta_n \in \pi\mathbb{Z}.$$

Given any $\theta_1, \ldots, \theta_n \in (0, \pi)$ with $\theta_1 + \ldots + \theta_n = \pi$, there exists a special Lagrangian N (called a *Lawlor neck*) asymptotic to $P(0) \cup P(\theta)$: see, for example, [42, Example 8.3.15] or §3.3 for details. It is diffeomorphic to $\mathcal{S}^{n-1} \times \mathbb{R}$. By rotating coordinates we have a special Lagrangian with phase i asymptotic to $P(-\frac{\theta}{2}) \cup P(\frac{\theta}{2})$.

The simplest case is when $\theta_1 = \ldots = \theta_n = \frac{\pi}{n}$: here N is called the *Lagrangian catenoid*. When $n = 2$, under a coordinate change (related to the hyperkähler rotation we described earlier) the Lagrangian catenoid becomes the complex curve

$$\{(z, \frac{1}{z}) \in \mathbb{C}^2 \,:\, z \in \mathbb{C} \setminus \{0\}\}$$

that we saw before.

The case $n = 3$ is interesting for pairs of special Lagrangian planes, since the only possibilities for the angles are $\sum_i \theta_i = \pi, 2\pi$. Moreover, if $\sum_i \theta_i = 2\pi$ we can rotate coordinates and change the order of the planes so that $P(0) \cup P(\theta)$ becomes $P(0) \cup P(\theta')$ where $\sum_i \theta'_i = \pi$. Hence, given any pair of transverse special Lagrangian planes in $\mathbb{C}^3$, there exists a Lawlor neck asymptotic to their union. We shall later that this fact is important when desingularizing intersections of special Lagrangian 3-folds.

Remark. The Lawlor neck clearly comes in a 1-parameter family given just by rescaling. The Lawlor necks are also examples of *exact* Lagrangians. To define exactness, note that on $\mathbb{C}^n$ we have that the Kähler form satisfies $2\omega_0 = \mathrm{d}\lambda_0$ where

$$\lambda_0 = \sum_{j=1}^{n} x_j \mathrm{d}y_j - y_j \mathrm{d}x_j,$$

where $x_j + iy_j$, $j = 1, \ldots, n$ are complex coordinates on $\mathbb{C}^n$. Since ω_0 vanishes on any Lagrangian we have that λ_0 is a closed 1-form on any Lagrangian. We say that a Lagrangian N in $\mathbb{C}^n$ is *exact* if $\lambda_0|_L$ is exact. It is then natural to ask if these are all of the exact special Lagrangians asymptotic to a transverse pair of special Lagrangian planes are Lawlor necks.

Using complex geometry and hyperkähler rotation it is easy to classify all of the smooth special Lagrangians in $\mathbb{C}^2$ asymptotic to a pair of transverse planes, and one sees that the Lawlor necks in $\mathbb{C}^2$ are the unique *exact* special Lagrangians with this property. It is now known that the Lawlor necks are the unique smooth exact special Lagrangian asymptotic to a pair of planes in all dimensions by work of Imagi–Joyce–Oliveira dos Santos [25].

We can find special Lagrangians in Calabi–Yau manifolds using the following easy result.

Proposition 3.10. *Let $(M, g, J, \omega, \Upsilon)$ be a Calabi–Yau manifold and let $\sigma : M \to M$ be such that $\sigma^2 = \mathrm{Id}$, $\sigma^*(\omega) = -\omega$, $\sigma^*(\Upsilon) = \overline{\Upsilon}$; i.e. σ is an anti-holomorphic isometric involution on M. Then Fix(σ) is special Lagrangian, if it is non-empty.*

Proof. If $p \in N = \mathrm{Fix}(\sigma)$ then $\omega|_p = 0$ since $\sigma^*(\omega) = -\omega$ and $\mathrm{Im}\,\Upsilon|_p = 0$ since $\sigma^*\Upsilon = \overline{\Upsilon}$. □

Example. Let $M = \{[z_0, \ldots, z_4] \in \mathbb{CP}^4 : z_0^5 + \ldots + z_4^5 = 0\}$ (the *Fermat quintic*) with its Calabi–Yau structure (which exists by Yau's solution of the Calabi conjecture since the first Chern class of M vanishes). Let σ be the restriction of complex conjugation on $\mathbb{CP}^4$ to M. Then the fixed point set of σ, which is the real locus in M, is a special Lagrangian 3-fold (if it is non-empty). (There is a subtlety here: σ is certainly an anti-holomorphic isometric involution for the induced metric on M, but this is *not* the same as the Calabi–Yau metric on M. Nevertheless, it is the case that σ satisfies the conditions of Proposition 3.10. This is using the uniqueness part of Yau's resolution of the Calabi conjecture.)

Example. There exists a complete Calabi–Yau metric on $T^*\mathcal{S}^n$ (the Stenzel metric [78]) so that the base $\mathcal{S}^n$ is special Lagrangian. When $n = 2$ this is a hyperkähler metric called the Eguchi–Hanson metric [11].

It is natural to ask if the special Lagrangian condition imposes any constructs on the geometry and topology of the submanifold. The following result [3] shows otherwise in 3 dimensions.

Theorem 3.11. *Every compact oriented real analytic Riemannian 3-manifold can be isometrically embedded in a Calabi–Yau 3-fold as the fixed point set of an involution.*

Remark. It is important in Theorem 3.11 that one is allowed to *choose* the Calabi–Yau 3-fold M given the 3-manifold N that you wish to embed. In general M will be non-compact and the metric on it will be incomplete: you should imagine M as a neighbourhood of the zero section in T^*N. It is therefore still an interesting question to ask whether the geometry and topology of a special Lagrangian is constrained in a *fixed* Calabi–Yau manifold.

3.3 The angle theorem

We now discuss a very natural and elementary problem in Euclidean geometry where calibrations play a major, and perhaps unexpected, role.

We return to Question 1.1 from the introduction: when is the union of two transverse n-planes in $\mathbb{R}^{2n}$ volume-minimizing? Two such planes are determined by the n angles between them as follows.

Lemma 3.12. *Let P, Q be oriented n-planes in $\mathbb{R}^{2n}$. There exists an oriented orthonormal basis for $\mathbb{R}^{2n}$, $\{e_1, \ldots, e_{2n}\}$, such that $P = \mathrm{Span}\{e_1, \ldots, e_n\}$ and*

$$Q = \mathrm{Span}\{\cos\theta_1 e_1 + \sin\theta_1 e_{n+1}, \ldots, \cos\theta_n e_n + \sin\theta_n e_{2n}\}$$

where $0 \leq \theta_1 \leq \ldots \leq \theta_{n-1} \leq \frac{\pi}{2}$ and $\theta_{n-1} \leq \theta_n \leq \pi - \theta_{n-1}$. These angles are called the characterising angles *of P, Q.*

Proof. The proof is very similar to the argument in the proof of Wirtinger's inequality (Theorem 3.3). Choose unit $e_1 \in P$ and maximise $\langle e_1, u_1\rangle$ for $u_1 \in Q$, and let $e_{n+1} \in P^\perp$ be defined by $u_1 = \cos\theta_1 e_1 + \sin\theta_1 e_{n+1}$. Now choose $e_2 \in P \cap e_1^\perp$ and maximise $\langle e_2, u_2\rangle$ for $u_2 \in Q \cap u_1^\perp$, then proceed by induction. Notice that for the final step we cannot choose the sign of $\cos\theta_n$ because we are working with oriented planes. □

Remark. If the characterising angles of P, Q are $\theta_1, \ldots, \theta_n$, then the characterising angles of $P, -Q$ are $\psi_1, \ldots, \psi_n$ where $\psi_j = \theta_j$ for $j = 1, \ldots, n-1$ and $\psi_n = \pi - \theta_n$.

The idea of the following theorem is that the union of $P \cup Q$ is area-minimizing if $P, -Q$ are not too close together [49].

Theorem 3.13 (Angle Theorem)**.** *Let P, Q be oriented transverse n-planes in $\mathbb{R}^{2n}$ and let $\psi_1, \ldots, \psi_n$ be the characterising angles between $P, -Q$. Then $P \cup Q$ is volume-minimizing if and only if $\psi_1 + \ldots + \psi_n \geq \pi$.*

Remark. Notice that the criteria in the Angle Theorem are impossible to fulfill in 1 dimension. We also see that

$$\psi_1 + \ldots + \psi_n \geq \pi \quad \Leftrightarrow \quad \theta_n \leq \theta_1 + \ldots + \theta_{n-1}.$$

Proof. We will sketch the proof which involves calibrations in a fundamental way in both directions. For details, look at [19].

First if $P \cup Q$ does not satisfy the angle condition, we can choose coordinates by Lemma 3.12 so that

$$P = P(-\frac{\psi}{2}) \quad \text{and} \quad -Q = P(\frac{\psi}{2})$$

where $\psi = (\psi_1, \ldots, \psi_n)$ and

$$P(\psi) = \{(e^{i\psi_1}x_1, \ldots, e^{i\psi_n}x_n) \,:\, (x_1, \ldots, x_n) \in \mathbb{R}^n\}$$

as given earlier. We know that we have a special Lagrangian Lawlor neck N with phase i (so calibrated by $\mathrm{Im}\,\Upsilon$) asymptotic to $P(-\frac{\psi'}{2}) \cup P(\frac{\psi'}{2})$ for any ψ' where $\sum_{i=1}^n \psi_i' = \pi$. The claim is then that since $\sum \psi_i < \pi$ we can find ψ' so that $\sum \psi_i' = \pi$ and $N \cap P(\pm\frac{\psi'}{2})$ is a compact hypersurface (in fact, an ellipsoid). This is actually a way to characterise N.

Now we observe that

$$\mathrm{Im}\,\Upsilon|_{P(\pm\psi/2)} = \sin(\pm\sum_j \psi_j/2)\,\mathrm{vol}_{P(\pm\psi/2)} < \mathrm{vol}_{P(\pm\psi/2)}\,.$$

Hence, since N is calibrated by $\mathrm{Im}\,\Upsilon$ and $\mathrm{Im}\,\Upsilon|_{P\cup Q} < \mathrm{vol}_{P\cup Q}$, $P \cup Q$ cannot be volume-minimizing by the usual Stokes' Theorem argument for calibrated submanifolds. Specifically, if we let $N \cap (P(-\psi/2) \cup -P(\psi/2)) = S$ and N' be the compact part of $P(-\psi/2) \cup -P(\psi/2)$ bounded by S, then

$$\mathrm{Vol}(N) = \int_N \mathrm{vol}_N = \int_N \mathrm{Im}\,\Upsilon = \int_{N'} \mathrm{Im}\,\Upsilon < \mathrm{Vol}(N'),$$

where we used that $\partial N = \partial N' = S$.

We now provide a few extra details, for which we need to describe N. For maps $z_1, \ldots, z_n : \mathbb{R} \to \mathbb{C}$ define

$$N = \{(t_1 z_1(s), \ldots, t_n z_n(s)) \in \mathbb{C}^n \,:\, s \in \mathbb{R},\, t_1, \ldots, t_n \in \mathbb{R},\, \sum_{j=1}^n t_j^2 = 1\}.$$

It is not difficult to calculate that N is special Lagrangian with phase i (so calibrated by $\operatorname{Im}\Upsilon$) if and only if

$$\overline{z_j}\frac{\mathrm{d}z_j}{\mathrm{d}s} = if_j\overline{z_1\dots z_n}$$

for positive real functions f_j.

Suppose that $f_j = 1$ for all j. Write $z_j = r_je^{i\theta_j}$, let $\theta = \sum_{j=1}^n \theta_j$ and suppose that $z_j(0) = c_j > 0$. From the differential equation, one quickly sees that $r_j^2 = c_j^2 + u$ for some function u with $u(0) = 0$ and

$$r_1 \dots r_n \cos\theta = c_1 \dots c_n.$$

If we now suppose that $u = t^2$, we see that

$$\theta_j(t) = \int_0^t \frac{a_j\mathrm{d}t}{(1 + a_jt^2)\sqrt{\frac{1}{t^2}\big((1 + a_1t^2)\dots(1 + a_nt^2) - 1\big)}}$$

where $a_j = c_j^{-2}$. We observe that $\theta \to \pm\frac{\pi}{2}$ as $t \to \pm\infty$ and hence N, which is a Lawlor neck, is asymptotic to a pair of planes where the sum of the angles is $\pm\frac{\pi}{2}$.

Now fix $t > 0$ and define

$$f : X = \{(a_1,\dots,a_n) \in \mathbb{R}^n \,:\, a_j \geq 0\} \to Y = \{(\theta_1,\dots,\theta_n) \in \mathbb{R}^n \,:\, \theta_j \geq 0, \sum_{j=1}^n \theta_j < \frac{\pi}{2}\}$$

by $f(a_1,\dots,a_n) = (\theta_1,\dots,\theta_n)$ where (recalling that t is fixed)

$$\theta_j = \int_0^t \frac{a_j\mathrm{d}t}{(1 + a_jt^2)\sqrt{\frac{1}{t^2}\big((1 + a_1t^2)\dots(1 + a_nt^2) - 1\big)}}.$$

It is clear that if $n = 1$, $f : X \to Y$ is surjective, because in this case

$$\theta_j = \int_0^t \frac{\sqrt{a_j}\mathrm{d}t}{1 + a_jt^2} = \tan^{-1}(\sqrt{a_j}t).$$

We want to show f is surjective for all n.

For $\theta \in (0, \frac{\pi}{2})$ define $H_\theta = \{(\theta_1,\dots,\theta_n) \in Y \,:\, \sum_{j=1}^n \theta_j = \theta\}$. Recall that

$$1 = \frac{r_1^2\dots r_n^2\cos^2\theta}{c_1^2\dots c_n^2} = \frac{(c_1^2 + u)\dots(c_n^2 + u)\cos^2\theta}{c_1^2\dots c_n^2} = (1 + a_1t^2)\dots(1 + a_nt^2)\cos^2\theta,$$

using the facts that $u = t^2$ and $a_j = c_j^{-2}$. We therefore see that that

$$f^{-1}(H_\theta) \subseteq S_\theta = \{(a_1,\dots,a_n) \in X \,:\, (1 + a_1t^2)\dots(1 + a_nt^2) = \cos^{-2}\theta\}.$$

Notice that if the degree of $f : \partial S_\theta \to \partial H_\theta$ is 1 then the degree of $f : S_\theta \to H_\theta$ is 1. Thus, by induction on n, we see that $f : S_\theta \to H_\theta$ is surjective.

Now, given any plane $\{(e^{i\theta_1}x_1,\dots e^{i\theta_n}x_n) \,:\, (x_1,\dots,x_n) \in \mathbb{R}^n\}$ where $(\theta_1,\dots,\theta_n) \in Y$, $\theta_j \neq 0$ for all j, we see that we can choose a Lawlor neck N which intersects the plane in a hypersurface as claimed.

We now move to the other direction in the statement of the Angle Theorem. If $P \cup Q$ does satisfy the angle condition, then (by choosing coordinates so that $P = \mathbb{R}^n$ and Q is in standard position) we claim that it is calibrated by a so-called *Nance calibration*:

$$\eta(u_1,\dots,u_n) = \operatorname{Re}\big((\mathrm{d}x_1 + u_1\mathrm{d}y_1) \wedge \dots \wedge (\mathrm{d}x_n + u_n\mathrm{d}y_n)\big)$$

where $u_1,\dots,u_n \in \mathcal{S}^2 \subseteq \operatorname{Im}\mathbb{H}$. If $u_m = i$ for all m then $\eta = \operatorname{Re}\Upsilon$, so it is believable that it is a calibration in general, but we now show that it is indeed true.

Let $x_1, y_1,\dots,x_n, y_n$ be coordinates on $\mathbb{R}^{2n}$. We call an n-form η on $\mathbb{R}^{2n}$ a torus form if η lies in the span of forms of type

$$\mathrm{d}x_{i_1} \wedge \dots \wedge \mathrm{d}x_{i_k} \wedge \mathrm{d}y_{j_1} \wedge \dots \wedge \mathrm{d}y_{j_l}$$

where $\{i_1,\ldots,i_k\}\cap\{j_1,\ldots,j_l\}=\emptyset$ and $\{i_1,\ldots,i_k\}\cup\{j_1,\ldots,j_l\}=\{1,\ldots,n\}$. We now claim that a torus form η is a calibration if and only if

$$\eta(\cos\theta_1 e_1+\sin\theta_1 e_{n+1},\ldots,\cos\theta_n e_n+\sin\theta_n e_{2n})\leq 1$$

for all $\theta_1,\ldots,\theta_n\in\mathbb{R}$.

For $n=1$, $\eta=\mathrm{d}x_1\wedge\mathrm{d}y_1$ which is a calibration. Suppose that the result holds for $n=k$. Let η be a torus form on $\mathbb{R}^{2(k+1)}$ and rescale η so that the maximum of η is 1 and is attained at some plane. The idea is to show using the argument in the proof of Wirtinger's inequality to put planes in standard position that we can write $\eta=e_1\wedge\eta_1+e_2\wedge\eta_2$ where e_1,e_2 are orthonormal and span an $\mathbb{R}^2$ and η_1,η_2 are torus forms on $\mathbb{R}^{2k}$. The claim then follows by induction on n.

Hence, the Nance calibration η above is a calibration and moreover we know $P(\theta)$ is calibrated by $\eta(u)$ if and only if

$$\prod_{j=1}^{n}(\cos\theta_j+\sin\theta_j u_j)=1.$$

We then just need to find the u_j determined by θ_j. Notice (as we did before) that the condition that $\psi_1+\ldots+\psi_n\geq\pi$ holds if and only if $\theta_n\leq\theta_1+\ldots+\theta_{n-1}$. If we write

$$\cos\theta_j+\sin\theta_j u_j=w_j\overline{w}_{j+1}$$

where $w_{n+1}=w_1$ and w_j are unit imaginary quaternions then the product condition $\prod_{j=1}^{n}(\cos\theta_j+\sin\theta_j u_j)=1$ is satisfied automatically. We then just need

$$\langle w_j,w_{j+1}\rangle=\cos\theta_j,$$

which is equivalent to finding n points on the unit 2-sphere so that $d(w_j,w_{j+1})=\theta_j$, where $\theta_n\leq\theta_1+\ldots+\theta_{n-1}$ and d is the spherical distance. This is indeed possible, by considering a suitable n-sided spherical polygon in $\mathcal{S}^2$. $\square$

4 Calibrated submanifolds and exceptional holonomy

In this section we discuss the relationship between calibrated geometry and the exceptional holonomy groups G_2 and Spin(7) in dimensions 7 and 8 respectively.

4.1 $\mathbf{G_2}$ manifolds and associative and coassociative submanifolds

We begin by introducing our calibrated geometry associated with G_2 holonomy in dimension 7. We begin with a definition.

Definition 4.1. Let $(x_1,\ldots,x_7)$ be coordinates on $\mathbb{R}^7$. We define a 3-form φ_0 on $\mathbb{R}^7$ by the following formula:

$$\varphi_0=\mathrm{d}x_{123}+\mathrm{d}x_{145}+\mathrm{d}x_{167}+\mathrm{d}x_{246}-\mathrm{d}x_{257}-\mathrm{d}x_{347}-\mathrm{d}x_{356},$$

where we use the short-hand notation $\mathrm{d}x_{ij\ldots k}=\mathrm{d}x_i\wedge\mathrm{d}x_j\wedge\ldots\wedge\mathrm{d}x_k$. If this formula looks mysterious, we see that if $u,v,w\in\mathbb{R}^7$ then

$$\varphi_0(u,v,w)=g_0(u\times v,w),$$

where g_0 is the Euclidean metric and $\times$ is the cross product on the imaginary octonions $\operatorname{Im}\mathbb{O}$ (i.e. the skew-symmetrization of the usual octonionic product). Hence φ_0 is like a "scalar triple product" on $\operatorname{Im}\mathbb{O}$.

We see that Hodge dual of φ_0 is given by

$$*\varphi_0=\mathrm{d}x_{4567}+\mathrm{d}x_{2367}+\mathrm{d}x_{2345}+\mathrm{d}x_{1357}-\mathrm{d}x_{1346}-\mathrm{d}x_{1256}-\mathrm{d}x_{1247}.$$

In fact, given the 3-form φ_0 we can recover the Euclidean metric and orientation on $\mathbb{R}^7$, and hence determine $*\varphi_0$. To see this, we observe that the following is true:

$$u\lrcorner\varphi_0\wedge v\lrcorner\varphi_0\wedge\varphi_0=6g_0(u,v)\mathrm{d}x_{1234567}.$$

This enables us to find the metric g_0 and volume form $\mathrm{vol}_{\mathbb{R}^7}=\mathrm{d}x_{1234567}$ purely algebraically from φ_0.

The fundamental fact is then the following, which is not surprising given the relationship between φ_0 and the cross product on the imaginary octonions.

Lemma 4.2. *The stabilizer of φ_0 in $\mathrm{GL}(7,\mathbb{R})$ is isomorphic to G_2, i.e.*

$$\mathrm{Stab}(\varphi_0) = \{A \in \mathrm{GL}(7,\mathbb{R}) \,:\, A^*\varphi_0 = \varphi_0\} \cong \mathrm{G}_2\,.$$

Remark. Since $\mathrm{G}_2 \subseteq \mathrm{SO}(7)$, this gives an abstract way to see that φ_0 determines g_0 and $\mathrm{vol}_{\mathbb{R}^7}$.

We can now define the setting where we want to study calibrated geometry related to G_2.

Definition 4.3. Let (M^7, g) be a Riemannian manifold so that $\mathrm{Hol}(g) \subseteq \mathrm{G}_2$. Then there exists a parallel 3-form φ on M which is identified with φ_0 at every point, which we call a G_2 *structure* on M. This 3-form φ induces g and an orientation vol_M on M (so necessarily M must be orientable) by the identity:

$$u \lrcorner \varphi \wedge v \lrcorner \varphi \wedge \varphi = 6g(u,v)\,\mathrm{vol}_M,$$

for tangent vectors u, v on M.

We then call (M, g, φ) a G_2 *manifold.* Since g is determined by φ we may also just write (M, φ). Notice that φ determines its Hodge dual $*\varphi$, which is also parallel.

The first key result is the following.

Theorem 4.4. *Let (M^7, φ) be a G_2 manifold. Then φ and $*\varphi$ are calibrations.*

Proof. Let u, v, w be oriented orthonormal vectors in $\mathbb{R}^7$. There exists an element A of G_2 so that $Au = e_1$, since G_2 acts transitively on the 6-sphere $\mathcal{S}^6$. It is then a fact that the subgroup of G_2 fixing e_1 is isomorphic to $\mathrm{SU}(3)$. We then know from the proof of Wirtinger's inequality (Theorem 3.3) there exists a (special) unitary transformation so that $v = e_2$ and $w = \cos\theta e_3 + \sin\theta v$ for some θ and v orthogonal to e_1, e_2, e_3.

Since $\varphi_0(e_1, e_2, .) = \mathrm{d}x_3$, we see that $\varphi_0(u, v, w) = \cos\theta$. Hence, since φ is closed, φ is a calibration and the calibrated planes in $\mathbb{R}^7$ are given by $A.\,\mathrm{Span}\{e_1, e_2, e_3\}$ for $A \in \mathrm{G}_2$.

By Lemma 3.4, $*\varphi$ is also a calibration. □

Let us look more closely at the calibrated planes for φ_0. If u, v, w are unit vectors in $\mathbb{R}^7 \cong \mathrm{Im}\,\mathbb{O}$ (the imaginary octonions), then $\varphi_0(u, v, w) = \langle u \times v, w\rangle = 1$ if and only if $w = u \times v$, so $P = \mathrm{Span}\{u, v, w\}$ is a copy of $\mathrm{Im}\,\mathbb{H}$ in $\mathrm{Im}\,\mathbb{O}$; in other words, $\mathrm{Span}\{1, u, v, w\}$ is an *associative* subalgebra of $\mathbb{O}$. Moreover, suppose we define a vector-valued 3-form χ_0 on $\mathbb{R}^7$ by

$$\chi_0(u, v, w) = \frac{1}{2}[u, v, w] = \frac{1}{2}\left(u(vw) - (uv)w\right),$$

where $[u, v, w]$ is known as the *associator*. Equivalently, one may define χ_0 by the formula:

$$*\varphi_0(x, u, v, w) = g_0\big(x, \chi_0(u, v, w)\big).$$

(In other words, we "raise an index" on $*\varphi_0$.) Then we observe the following.

Lemma 4.5. *A 3-plane P in $\mathbb{R}^7$ satisfies $\chi_0|_P \equiv 0$ if and only if P admits an orientation so that it is calibrated by φ_0.*

Proof. Since the associator is clearly invariant under G_2 we can put any plane P in standard position using G_2, i.e. as in the proof of Theorem 4.4, we can write $P = \mathrm{Span}\{e_1, e_2, \cos\theta e_3 + \sin\theta v\}$ for some v orthogonal to e_1, e_2, e_3. We can calculate that $[e_1, e_2, e_3] = 0$ whereas $[e_1, e_2, v] \neq 0$ for any v orthogonal to e_1, e_2, e_3. Moreover, P is calibrated by φ_0 if and only if $\theta = 0$. We thus see that P is calibrated by φ_0 (up to a choice a orientation) if and only if $\chi_0|_P \equiv 0$. □

Hence we call the φ_0-calibrated planes *associative.* In general on a G_2 manifold we can define a 3-form χ as follow.

Definition 4.6. On a G_2-manifold (M^7, φ), we define a 3-form χ with values in TM by:

$$*\varphi(x, u, v, w) = g\big(x, \chi(u, v, w)\big)$$

for tangent vectors x, u, v, w.

For $*\varphi_0$ we see that $*\varphi_0|_P = \mathrm{vol}_P$ for a plane P if and only if $\varphi_0|_{P^\perp} = \mathrm{vol}_{P^\perp}$ (for a suitable choice of orientation on $P^\perp$. Hence the planes calibrated by $*\varphi_0$ are the orthogonal complements of the associative planes, so we call them *coassociative.* We have a similar alternative characterisation for 4-planes calibrated by $*\varphi$.

Lemma 4.7. *A 4-plane P in $\mathbb{R}^7$ satisfies $\varphi_0|_P \equiv 0$ if and only if P admits an orientation so that it is calibrated by $*\varphi_0$.*

Proof. We know that given a 4-plane P we can choose coordinates such that $P^\perp = \mathrm{Span}\{e_1, e_2, \cos\theta e_3 + \sin\theta(a_4e_4 + a_5e_5 + a_6e_6 + a_7e_7)\}$ where $\sum_j a_j^2 = 1$. Then

$$\begin{aligned} P = \mathrm{Span}\{ &-\sin\theta e_3 + \cos\theta(a_je_j), a_5e_4 - a_4e_5 + a_7e_6 - a_6e_7, \\ &a_6e_4 - a_7e_5 - a_4e_6 + a_5e_7, a_7e_4 + a_6e_5 - a_5e_6 - a_4e_7\}. \end{aligned}$$

We can then see directly that $*\varphi_0|_P = \cos\theta$. We also have $\varphi_0(e_i, e_j, e_k) = 0$ for $i, j, k \in \{4, 5, 6, 7\}$ and $e_3 \lrcorner \varphi = -\mathrm{d}x_{47} - \mathrm{d}x_{56}$, so that $\varphi_0(-\sin\theta e_3 + \cos\theta(a_je_j), v, w)$ is a non-zero multiple of $\sin\theta$ for some $v, w \in P$. Hence $\varphi_0|_P = 0$ if and only if $\theta = 0$, which is if and only if P is calibrated by $*\varphi_0$ (again up to a choice of orientation). □

Remark. Using the calculations and observations above, we see that associative and coassociative Grassmannians are both isomorphic to $\mathrm{G}_2 / \mathrm{SO}(4)$. Curiously, this actually turns out to an admit a quaternionic Kähler metric.

We thus can define our calibrated submanifolds.

Definition 4.8. Let (M^7, φ) be a G_2 manifold.

The 3-dimensional submanifolds in (M^7, φ) calibrated by φ are called *associative* 3-folds. Moreover, N is associative if and only if $\chi|_N \equiv 0$ (up to a choice of orientation).

The 4-dimensional submanifolds in (M^7, φ) calibrated by $*\varphi$ are called *coassociative* 4-folds. Moreover, N is coassociative if and only if $\varphi|_N \equiv 0$ (up to a choice of orientation).

It is instructive to see the form that the associative or coassociative condition takes by studying associative or coassociative graphs in $\mathbb{R}^7$: see [20] for details.

A simple way to get associative and coassociative submanifolds is by using known geometries.

Proposition 4.9. *Let $x_1, \ldots, x_7$ be coordinates on $\mathbb{R}^7$ and let $z_j = x_{2j} + ix_{2j+1}$ be coordinates on $\mathbb{C}^3$ so that $\mathbb{R}^7 = \mathbb{R} \times \mathbb{C}^3$.*

(a) $N = \mathbb{R} \times S \subseteq \mathbb{R} \times \mathbb{C}^3$ is associative or coassociative if and only if S is a complex curve or a special Lagrangian 3-fold with phase $-i$, respectively.

(b) $N \subseteq \{0\} \times \mathbb{C}^3$ is associative or coassociative if and only if N is a special Lagrangian 3-fold or a complex surface, respectively.

Proof. Recall the Kähler form ω_0 and holomorphic volume form Υ_0 on $\mathbb{C}^3$. We can write

$$\varphi_0 = \mathrm{d}x_1 \wedge \omega_0 + \mathrm{Re}\,\Upsilon_0 \quad \text{and} \quad *\varphi_0 = \frac{1}{2}\omega_0^2 - \mathrm{d}x_1 \wedge \mathrm{Im}\,\Upsilon_0.$$

For associatives, we see that $\varphi_0|_{\mathbb{R}\times S} = \mathrm{d}x_1 \wedge \mathrm{vol}_S$ if and only if $\omega_0|_S = \mathrm{vol}_S$ and $\varphi_0|_N = \mathrm{Re}\,\Upsilon_0|_N$ for $N \subseteq \mathbb{C}^3$.

For coassociatives, we see that $*\varphi_0|_{\mathbb{R}\times S} = \mathrm{d}x_1 \wedge \mathrm{vol}_S$ if and only if $-\mathrm{Im}\,\Upsilon_0|_S = \mathrm{vol}_S$ and $*\varphi_0|_N = \frac{1}{2}\omega_0^2|_N$ for $N \subseteq \mathbb{C}^3$.

The results quickly follow. □

We can also produce examples in G_2 manifolds with an isometric involution.

Proposition 4.10. *Let (M, φ) be a G_2 manifold with an isometric involution $\sigma \neq \mathrm{id}$ such that $\sigma^*\varphi = \varphi$ or $\sigma^*\varphi = -\varphi$. Then Fix($\sigma$) is an associative or coassociative submanifold in M respectively, if it is non-empty.*

We also have explicit examples of associatives and coassociatives.

Example. The first explicit examples of associatives in $\mathbb{R}^7$ not arising from other geometries are given in [52] from symmetry and evolution equation considerations.

The first explicit non-trivial examples of coassociatives in $\mathbb{R}^7$ are given in [20]. There are two dilation families: one which has one end asymptotic to a cone C on a non-round $\mathcal{S}^3$, and one which has two ends asymptotic to $C \cup \mathbb{R}^4$. The coassociative cone C was discovered earlier by Lawson–Osserman [50] and was the first example of a volume-minimizing submanifold which is not smooth (it is Lipschitz but not C^1).

Example. In the Bryant–Salamon complete holonomy G_2 metric on the spinor bundle of $\mathcal{S}^3$ [5] (which is topologically just $\mathbb{R}^4 \times \mathcal{S}^3$), the base $\mathcal{S}^3$ is associative.

In the Bryant–Salamon complete holonomy G_2 metrics on the bundles of self-dual 2-forms over the 4-sphere and the complex projective 2-space, $\Lambda^2_+ T^*\mathcal{S}^4$ and $\Lambda^2_+ T^*(\overline{\mathbb{CP}^2})$ (where $\overline{\mathbb{CP}^2}$ means we take the opposite orientation to the usual one) [5], the bases $\mathcal{S}^4$ and $\mathbb{CP}^2$ are coassociative. (Here, a 2-form α on an oriented Riemannian 4-manifold is *self-dual* if $*\alpha = \alpha$, where $*$ is the Hodge star.)

Since we can embed any compact oriented real analytic Riemannian 3-manifold isometrically as a special Lagrangian 3-fold N in a Calabi–Yau 3-fold M, we can isometrically embed it in as an associative 3-fold $\{0\} \times N$ in the G_2 manifold $\mathbb{R} \times M$ with the product G_2 structure. We also have a similar isometric embedding result for coassociative 4-folds to the special Lagrangian 3-fold case [3].

Theorem 4.11. *Any compact oriented real analytic Riemannian 4-manifold whose bundle of self-dual 2-forms is trivial can be isometrically embedded in a* G_2 *manifold as the fixed points of an isometric involution.*

4.2 Spin(7) manifolds and Cayley submanifolds

We now discuss want to discuss our final class of calibrated submanifolds, which is associated with the group Spin(7). We begin with defining a distinguished 4-form on $\mathbb{R}^8$.

Definition 4.12. Given coordinates $(x_0, \ldots, x_7)$ on $\mathbb{R}^8$ we define a 4-form on $\mathbb{R}^8$ by:

$$\begin{aligned}\Phi_0 = {}& \mathrm{d}x_{0123} + \mathrm{d}x_{0145} + \mathrm{d}x_{0167} + \mathrm{d}x_{0246} - \mathrm{d}x_{0257} - \mathrm{d}x_{0347} - \mathrm{d}x_{0356} \\ & + \mathrm{d}x_{4567} + \mathrm{d}x_{2367} + \mathrm{d}x_{2345} + \mathrm{d}x_{1357} - \mathrm{d}x_{1346} - \mathrm{d}x_{1256} - \mathrm{d}x_{1247}.\end{aligned}$$

Equivalently, we can take coordinates $(x_0, x_1, x_2, x_3, x_0', x_1', x_2', x_3', x_4')$ on $\mathbb{R}^8$ and let

$$\begin{aligned}\omega_1 &= \mathrm{d}x_{01} + \mathrm{d}x_{23}, \quad \omega_2 = \mathrm{d}x_{02} + \mathrm{d}x_{31}, \quad \omega_3 = \mathrm{d}x_{03} + \mathrm{d}x_{12}, \\ \omega_1' &= \mathrm{d}x_{01}' + \mathrm{d}x_{23}', \quad \omega_2' = \mathrm{d}x_{02}' + \mathrm{d}x_{31}', \quad \omega_3' = \mathrm{d}x_{03}' + \mathrm{d}x_{12}',\end{aligned}$$

so that

$$\Phi_0 = \mathrm{d}x_{0123} + \mathrm{d}x_{0123}' - \sum_{j=1}^{3} \omega_j \wedge \omega_j'.$$

Notice that Φ_0 is self-dual, i.e. $*\Phi_0 = \Phi_0$. In a similar, but more complicated, way to the form φ_0 on $\mathbb{R}^7$, we can recover the Euclidean metric and volume form on $\mathbb{R}^8$ from Φ_0. We may also relate Φ_0 to the octonions $\mathbb{O}$.

We then have the following fundamental fact about the form Φ_0.

Lemma 4.13. *The stabilizer of* Φ_0 *in* $\mathrm{GL}(8, \mathbb{R})$ *is isomorphic to* $\mathrm{Spin}(7)$, *i.e.*

$$\mathrm{Stab}(\Phi_0) = \{A \in \mathrm{GL}(8, \mathbb{R}) \,:\, A^*\Phi_0 = \Phi_0\} \cong \mathrm{Spin}(7).$$

Remark. Since $\mathrm{Spin}(7) \subseteq \mathrm{SO}(8)$, this gives an alternative way to see that Φ_0 determines the Euclidean metric and volume form.

We can now define our distinguished class of 8-manifolds to study Spin(7) calibrated geometry.

Definition 4.14. Let (M^8, g) be a Riemannian manifold with $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$. Then there exists a parallel 4-form Φ on M which is identified pointwise with Φ_0, which we call a Spin(7) *structure* on M. The 4-form Φ induces the metric g and an orientation on M so that Φ is self-dual.

Since g is determined by Φ, we then call (M, Φ) a Spin(7) *manifold.*

As we would expect, we have the following important result.

Theorem 4.15. *On a* Spin(7) *manifold* (M^8, Φ)*,* Φ *is a calibration.*

Proof. Let P be a plane in $\mathbb{R}^8 \cong \mathbb{C}^4$. Since $\mathrm{SU}(4) \subseteq \mathrm{Spin}(7)$, by the proof of Wirtinger's inequality (Theorem 3.3), we can choose $A \in \mathrm{Spin}(7)$ so that $A(P)$ is spanned by

$$\{e_0, \cos\theta_1 e_1 + \sin\theta_1 e_2, e_4, \cos\theta_2 e_5 + \sin\theta_2 e_6\}.$$

Then

$$\Phi_0|_P = (\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2)\,\mathrm{vol}_P = \cos(\theta_1 - \theta_2)\,\mathrm{vol}_P\,.$$

Hence Φ_0 and thus Φ is a calibration (as it is closed). □

We can thus define our calibrated submanifolds in Spin(7) manifolds.

Definition 4.16. The oriented 4-dimensional submanifolds in a Spin(7) manifold (M^8, Φ) calibrated by Φ are called *Cayley* 4-folds.

Remark. The name Cayley submanifolds is because of the relation between the submanifolds and the octonions or Cayley numbers $\mathbb{O}$.

We can relate Cayley submanifolds to all of the other calibrated geometries we have seen.

Proposition 4.17. *(a) Complex surfaces and special Lagrangian 4-folds in* $\mathbb{C}^4$ *are Cayley in* $\mathbb{R}^8 = \mathbb{C}^4$*.*

(b) Write $\mathbb{R}^8 = \mathbb{R} \times \mathbb{R}^7$*. Then* $\mathbb{R} \times S$ *is Cayley if and only if* S *is associative in* $\mathbb{R}^7$ *and* $N \subseteq \mathbb{R}^7$ *is Cayley in* $\mathbb{R}^8$ *if and only if* N *is coassociative in* $\mathbb{R}^7$*.*

Proof. Recall the Kähler form ω_0 and holomorphic volume form Υ_0 on $\mathbb{C}^4$ and the G_2 3-form φ_0 on $\mathbb{R}^7$.

Part (a) is immediate from the formula

$$\Phi_0 = \frac{1}{2}\omega_0^2 + \mathrm{Re}\,\Upsilon_0,$$

since complex surfaces are calibrated by $\frac{1}{2}\omega_0^2$, special Lagrangians are calibrated by $\mathrm{Re}\,\Upsilon_0$, Υ_0 vanishes on complex surfaces and ω_0 vanishes on special Lagrangians.

Given the formula

$$\Phi_0 = \mathrm{d}x_0 \wedge \varphi_0 + *\varphi_0,$$

part (b) then follows. □

We can also use an isometric involution to construct Cayley submanifolds as in our previous calibrated geometries.

Proposition 4.18. *Let* (M, Φ) *be a* Spin(7) *manifold and let* $\sigma \neq \mathrm{id}$ *be an isometric involution with* $\sigma^*\Phi = \Phi$*. Then Fix*(σ) *is Cayley submanifold, if it is non-empty.*

Example. The first interesting explicit examples of Cayleys in $\mathbb{R}^8$ not arising from other geometries were given in [53] and are asymptotic to cones.

Example. The base $\mathcal{S}^4$ in the complete Bryant–Salamon holonomy Spin(7) metric on $\mathbb{S}_+(\mathcal{S}^4)$ [5] is Cayley.

Just before we conclude this section we make a couple more observations. On $\mathbb{O}$ there exists a 4-fold cross product, whose real part gives Φ_0 and whose imaginary part we call τ_0. Perhaps unsurprisingly, we have the following result, which we will leave as an exercise for the reader.

Lemma 4.19. *A 4-plane* P *in* $\mathbb{R}^8$ *satisfies* $\tau_0|_P \equiv 0$ *if and only if it admits an orientation so that it is calibrated by* Φ_0*.*

We can extend τ to a Spin(7) manifold, except that we need a rank 7 vector bundle on M in which τ takes values that plays the role of $\operatorname{Im}\mathbb{O}$ at each point. To understand this, we need to make some algebraic digression.

Since $\Lambda^2(\mathbb{R}^8)^*$ is 28-dimensional and the 21-dimensional Lie algebra of Spin(7) sits inside the space of 2-forms on $\mathbb{R}^8$, we must have a distinguished 7-dimensional subspace Λ^2_7 of 2-forms on $\mathbb{R}^8$. So what is this subspace? Let $u, v \in \mathbb{R}^8$. Then we can construct a 2-form $u \wedge v$, viewing u, v as cotangent vectors. We can also construct a 2-form from u, v by considering $\Phi_0(u, v, ., .)$. These considerations lead us to the following definition.

Definition 4.20. We define a subspace Λ^2_7 of $\Lambda^2(\mathbb{R}^8)^*$ by the following equivalent definitions:

$$\Lambda^2_7 = \{\alpha \in \Lambda^2(\mathbb{R}^8)^* \,:\, \alpha \wedge \Phi_0 = 3 * \alpha\} = \{u \wedge v + \Phi_0(u, v, ., .) \,:\, u, v \in \mathbb{R}^8\}.$$

We can then similarly define a subbundle Λ^2_7 of $\Lambda^2 T^*M$ on any Spin(7) manifold (M, Φ).

With this definition in hand, we have the following alternative characterisation of Cayley 4-folds.

Lemma 4.21. *A submanifold N in a* Spin(7) *manifold (M, Φ) is Cayley (up to a choice of orientation) if and only if $\tau \in C^\infty(\Lambda^4 T^*M; \Lambda^2_7)$ vanishes on N.*

5 Moduli problems and calibrated geometry

The study of *moduli spaces* forms a key part of geometry. Moduli spaces parametrize *deformations* of a given geometric object. In our case, we are interested in compact calibrated submanifolds N, and so the moduli spaces we wish to study describe all (nearby) compact calibrated submanifolds to N. As well as being of inherent interest, one can use the moduli space theory and its proof has a number of applications, including to constructing examples of calibrated submanifolds.

5.1 Introduction

It is easy to construct complex submanifolds in Kähler manifolds algebraically. Constructing other calibrated submanifolds is much more challenging because one needs to solve a nonlinear PDE, even in Euclidean space. There are approaches in Euclidean space and other simple spaces which have involved reducing the problem to ODEs or other problems which do not require PDE (for example, algebraic methods). For example, we have the following methods, which you can find out more about in [42] or the references provided.

- Symmetries/evolution equations [17, 20, 21, 28, 30, 31, 33, 34, 39, 40, 41, 52, 54].
- Use of integrable systems to study calibrated cones [8, 9, 22, 43, 65].
- Calibrated cones and ruled smoothings of these cones [2, 4, 13, 14, 32, 52, 53, 59].
- Vector sub-bundle constructions [27, 45, 46].
- Classification of calibrated submanifolds satisfying pointwise constraints on their second fundamental form [4, 12, 26, 59, 60].

However, an important direction which has borne fruit in calibrated geometry and special holonomy recently has been to study the nonlinear PDE head on, especially by perturbative and gluing methods. Hence, we naturally need to understand moduli problems in calibrated geometry to make such construction methods viable.

We want to solve nonlinear PDE, so how do we tackle this? The idea is to use the linear case to help. Suppose we are on a compact manifold N and recall the theory of linear *elliptic* operators L of order l on N, including:

- the definition of ellipticity of L via the *principal symbol* σ_L (which encodes the highest order derivatives in the operator) being an isomorphism;

- the use of *Hölder spaces* $C^{k,a}$ to give elliptic regularity theory (so-called *Schauder theory*), namely that if $w \in C^{k,a}$ and $Lv = w$ then $v \in C^{k+l,a}$ and there is a universal constant C so that
$$\|v\|_{C^{k+l,a}} \leq C(\|Lv\|_{C^{k,a}} + \|v\|_{C^0})$$
(and we can drop the $\|v\|_{C^0}$ term if v is orthogonal to $\operatorname{Ker} L$);
- the adjoint operator L^* and that $\sigma_{L^*} = (-1)^l \sigma_L^*$ so that L^* is elliptic if and only if L is elliptic; and
- the Fredholm theory of L, namely that $\operatorname{Ker} L$ (and hence $\operatorname{Ker} L^*$) is finite-dimensional, and we can solve $Lv = w$ if and only if $w \in (\operatorname{Ker} L^*)^\perp$.

We shall discuss this in a model example which we shall use throughout this section.

Example. The Laplacian on functions is given by $\Delta f = \mathrm{d}^*\mathrm{d}f$ which in normal coordinates at a point is given by $f \mapsto -\sum_i \frac{\partial^2 f}{\partial x_i^2}$, so it is a linear second order differential operator. We see that its principal symbol is $\sigma_\Delta(x,\xi)f = -|\xi|^2 f$ which is an isomorphism for $\xi \in T_x^*N \setminus \{0\}$, so Δ is elliptic. We therefore have that if $h \in C^{k,a}(N)$ and $\Delta f = h$ then $f \in C^{k+2,a}(N)$, and we have an estimate
$$\|f\|_{C^{k+2,a}} \leq C(\|\Delta f\|_{C^{k,a}} + \|f\|_{C^0}).$$
We also know that $\Delta^* = \Delta$ and $\operatorname{Ker}\Delta$ is given by the constant functions (since if $f \in \operatorname{Ker}\Delta$ then
$$0 = \langle f, \Delta f\rangle_{L^2} = \langle f, \mathrm{d}^*\mathrm{d}f\rangle_{L^2} = \|\mathrm{d}f\|^2_{L^2}$$
so $\mathrm{d}f = 0$). Hence, we can solve $\Delta f = h$ if and only if h is orthogonal to the constants, i.e. $\int_N h \,\mathrm{vol}_N = 0$.

The operator defining the minimal graph equation for a hypersurface is
$$P(f) = -\operatorname{div}\left(\frac{\nabla f}{\sqrt{1+|\nabla f|^2}}\right),$$
which is a nonlinear second order operator whose linearisation $L_0 P$ at 0 is Δ. Thus P is a nonlinear elliptic operator at 0. If we linearise P at f_0 we find a more complicated expression depending on f_0, but it is still a perturbation of the Laplacian.

Suppose we are on a compact manifold N and we want to solve $P(f) = 0$ where P is the minimal graph operator on functions f. Let us consider regularity for f. We can re-arrange $P(f) = 0$ by taking all of the second derivatives to one side as:
$$R(x, \nabla f(x))\nabla^2 f(x) = E(x, \nabla f(x))$$
where $x \in N$. Since $L_0 P = \Delta$ is elliptic and ellipticity is an open condition we know that the operator L_f (depending on f) given by
$$L_f(h)(x) = R(x, \nabla f(x))\nabla^2 h(x)$$
is a *linear* elliptic operator whenever $\|\nabla f\|_{C^0}$ is small, in particular if $\|f\|_{C^{1,a}}$ is sufficiently small. The operator L_f does not have smooth coefficients, but if $f \in C^{k,a}$ then the coefficients $R \in C^{k-1,a}$.

Suppose that $f \in C^{1,a}$ and $\|f\|_{C^{1,a}}$ is small with $P(f) = 0$. Then $L_f(f) = E(f)$ and L_f is a linear *second order* elliptic operator with coefficients in $C^{0,a}$ and $E(f)$ is in $C^{0,a}$. So by elliptic regularity we can deduce that $f \in C^{2,a}$. We have gained one degree of regularity, so we can "bootstrap", i.e. proceed by induction and deduce that any $C^{1,a}$ solution to $P(f) = 0$ is smooth.

Example. $C^{1,a}$-minimal submanifolds (and thus calibrated submanifolds) are *smooth.*

Remark. More sophisticated techniques can be used to deduce that C^1-minimal submanifolds are real analytic [69]. Notice that elliptic regularity results are *not* valid for C^k spaces, so this result is not obvious.

We can also arrange our simple equation $P(f) = 0$ as $\Delta f + Q(\nabla f, \nabla^2 f) = 0$, where Q is nonlinear but linear in $\nabla^2 f$. If we know that $\int_N P(f)\,\mathrm{vol}_N = 0$, i.e. that $P(f)$ is orthogonal to the constants, then we can always solve $\Delta f_0 = -Q(\nabla f, \nabla^2 f)$. We do know that $\int_N P(f)\,\mathrm{vol}_N = 0$ since P has a divergence form. This means we are in the setting for implementing the Implicit Function Theorem for Banach spaces to conclude that we can always solve $P(f) = 0$ for some f near 0, and f will be smooth by our regularity argument above. In general, we will use the following.

Theorem 5.1 (Implicit Function Theorem)**.** *Let X, Y be Banach spaces, let $U \ni 0$ be open in X, let $P : U \to Y$ with $P(0) = 0$ and $L_0P : X \to Y$ surjective with finite-dimensional kernel K.*

Then for some U, $P^{-1}(0) = \{u \in U : P(u) = 0\}$ is a manifold of dimension $\dim K$. Moreover, if we write $X = K \oplus Z$, $P^{-1}(0) = \mathrm{Graph}\, G$ for some map G from an open set in K to Z with $G(0) = 0$.

5.2 Set-up and strategy

The basic set-up for our moduli problems is the following.

- We have a compact submanifold N of a Riemannian manifold (M, g) which is calibrated by η.
- Recall that, by the tubular neighbourhood theorem, that any nearby deformation of N is given by the graph N_v of $v \in C^\infty(\nu(N))$ (with sufficiently small C^1-norm), so that $v = 0$ corresponds to the original N.
- The *moduli space* of η-calibrated deformations of N is locally equal to

$$\mathcal{M}(N) = \{v : N_v \text{ is } \eta\text{-calibrated}\}.$$

Our aim is to describe $\mathcal{M}(N)$.

Remark. In moduli space problems, one often has to worry about the issue of redundancy in the parametrization, i.e. that one wants a one-to-one correspondence between elements of the moduli space and geometric objects, so that equivalent objects are identified. This identification typically involves taking a quotient by some infinite-dimensional group, and therefore one needs to obtain a slice for this group action to describe the moduli space. Obtaining this slice often relies on what is called *gauge-fixing* (the name is motivated by gauge theory). In our case, the group in question is the possible reparametrizations of the submanifold, and the gauge-fixing is achieved by writing the nearby submanifold as a *normal* graph.

The strategy to describe $\mathcal{M}(N)$ is the following.

- Find (sub)spaces X, Y of sections of vector bundles V, W over N, with $\nu(N) \cong V$, and a (first order) differential operator $P : X \to Y$ so that $0 \in X$ corresponds to N, $P(0) = 0$ and

$$\mathcal{M}(N) \cong \{x \in X : P(x) = 0\}.$$

- Compute the linearization L_0P of P at 0, show it is elliptic and let $\mathcal{I} = \ker L_0P$ and $\mathcal{O} = \mathrm{coker} L_0P$.
- If $P(X) \subseteq \mathrm{im}(L_0P)$, replace Y by $\mathrm{im}(L_0P)$ and apply the Implicit Function Theorem to deduce that $\mathcal{M}(N)$ is a manifold of dimension $\dim \mathcal{I}$.
- If $P(X) \nsubseteq \mathrm{im}(L_0P)$, replace X by $X \oplus \mathcal{O}$, replace P by $Q(x, y) = P(x) + y$ and apply the Implicit Function Theorem to deduce that there is an open neighbourhood of 0, $\hat{\mathcal{M}}(N) \subseteq \mathcal{I}$, and a smooth map $\pi : \hat{\mathcal{M}}(N) \to \mathcal{O}$ with $\pi(0) = 0$ so that

$$\mathcal{M}(N) \cong \pi^{-1}(0).$$

  Deduce that the *expected dimension* of $\mathcal{M}(N)$ is

$$\dim \mathcal{I} - \dim \mathcal{O} = \mathrm{ind} L_0P,$$

  the *index* of L_0P.

Remark. We call $\mathcal{I}$ the *infinitesimal deformation space* and $\mathcal{O}$ the *obstruction space*. The elements of $\mathcal{I}$ are the possible tangent vectors to paths of calibrated deformations of N. Elements of $\mathcal{O}$ represent the possible obstructions to realize the infinitesimal deformations as tangent vectors to an actual path in $\mathcal{M}(N)$.

5.3 Special Lagrangian deformations

We now follow the strategy set out above in the special Lagrangian case, in a result originally due to McLean [66].

Theorem 5.2. *Let N be a compact special Lagrangian in a Calabi–Yau manifold M. Then the moduli space of deformations of N is a smooth manifold of dimension $b^1(N)$.*

Remark. One should compare this result to the deformation theory for complex submanifolds in Kähler manifolds. There, one does not get that the moduli space is a smooth manifold: in fact, it can be singular, and one has *obstructions* to deformations. It is somewhat remarkable that special Lagrangian calibrated geometry enjoys a much better deformation theory than this classical calibrated geometry. The deformation theory of embedded compact complex submanifolds in Calabi–Yau manifolds has recently been revisited using analytic techniques [67].

Proof. The tubular neighbourhood theorem gives us a diffeomorphism $\exp : S \subseteq \nu(N) \to T \subseteq M$ which maps the zero section to N; in other words, we can write any nearby submanifold to N as the graph of a normal vector field on N. We know that N is Lagrangian, so the complex structure J gives an isomorphism between $\nu(N)$ and TN and the metric gives an isomorphism between TN and T^*N: $v \mapsto g(Jv, .) = \omega(v, .) = \alpha_v$. Therefore any deformation of N in T is given as the graph N_α of a 1-form α. In fact, using the Lagrangian neighbourhood theorem, we can arrange that any $N_\alpha \subseteq T$ which is graphical over N is the graph of a 1-form α so that, if $f_\alpha : N \to N_\alpha$ is the natural diffeomorphism, then

$$f_\alpha^*(\omega) = \mathrm{d}\alpha \quad \text{and} \quad -*f_\alpha^*(\operatorname{Im}\Upsilon) = F(\alpha, \nabla\alpha) = \mathrm{d}^*\alpha + Q(\alpha, \nabla\alpha),$$

where the second formula follows from a calculation using the special Lagrangian condition on N and the fact that the ambient structure is Calabi–Yau. Notice that the original special Lagrangian N corresponds to taking $\alpha = 0$.

Hence, N_α is special Lagrangian if and only if

$$P(\alpha) = (F(\alpha, \nabla\alpha), \mathrm{d}\alpha) = 0.$$

This means that infinitesimal special Lagrangian deformations of N, which are the elements in the kernel of linearisation L_0P of P at 0, are given by closed and coclosed 1-forms, i.e.

$$\operatorname{Ker} L_0P = \{\alpha \in C^\infty(T^*N) \,:\, \mathrm{d}\alpha = 0 = \mathrm{d}^*\alpha\}.$$

Since $\operatorname{Im}\Upsilon = 0$ on N we have that $[\operatorname{Im}\Upsilon] = 0$ on N_α, which means that $f_\alpha^*(\operatorname{Im}\Upsilon)$ is exact. Thus $F(\alpha, \nabla\alpha) = - * f_\alpha^*(\operatorname{Im}\Upsilon)$ is coexact and so

$$P : C^\infty(S) \to \mathrm{d}^*(C^\infty(T^*N)) \oplus \mathrm{d}(C^\infty(T^*N)) \subseteq C^\infty(\Lambda^0T^*N \oplus \Lambda^2T^*N).$$

If we let $X = C^{1,a}(T^*N)$, $Y = \mathrm{d}^*(C^{1,a}(T^*N)) \oplus \mathrm{d}(C^{1,a}(T^*N))$ and $U = C^{1,a}(S)$ we can apply the Implicit Function Theorem if we know that

$$L_0P : \alpha \in X \mapsto (\mathrm{d}^*\alpha, \mathrm{d}\alpha) \in Y$$

is surjective, i.e. given $\mathrm{d}\beta + \mathrm{d}^*\gamma \in Y$ does there exist α such that $\mathrm{d}\alpha = \mathrm{d}\beta$ and $\mathrm{d}^*\alpha = \mathrm{d}^*\gamma$? If we let $\alpha = \beta + \mathrm{d}f$ then we need $\Delta f = \mathrm{d}^*\mathrm{d}f = \mathrm{d}^*(\gamma - \beta)$. Since

$$\int_N \mathrm{d}^*(\gamma - \beta)\,\mathrm{vol}_N = \pm\int_N \mathrm{d} * (\gamma - \beta) = 0$$

we can solve the equation for f, and hence L_0P is surjective.

Therefore $P^{-1}(0)$ is a manifold of dimension $\dim\operatorname{Ker} L_0P = b^1(N)$ by Hodge theory. Moreover, if $P(\alpha) = 0$ then N_α is special Lagrangian, hence minimal and since $\alpha \in C^{1,a}$ we deduce that α is in fact smooth. □

Example. The special Lagrangian $\mathcal{S}^n$ in the Calabi–Yau manifold $T^*\mathcal{S}^n$ with the Stenzel metric has $b^1 = 0$ and so is rigid (i.e. it has no deformations). In fact, any compact minimal submanifold in $T^*\mathcal{S}^n$ is contained in $\mathcal{S}^n$.

Example. Observe that if we have a special Lagrangian T^n in a Calabi–Yau manifold M then $b^1(T^n) = n$. If the torus is close to flat then its deformations locally foliate M (as there will be n nowhere vanishing harmonic 1-forms), so we can hope to find special Lagrangian torus *fibrations*. This cannot happen in compact manifolds without singular fibres, but still motivates the *SYZ conjecture* in Mirror Symmetry.

Remark. Theorem 5.2 has also been extended to certain non-compact, singular and boundary settings, for example in [6, 36, 72].

Remark. The case of special Lagrangian rational homology 3-spheres in Calabi–Yau 3-folds is particularly interesting because they must be rigid (as $b^1 = 0$). Therefore, one might hope to "count" these special Lagrangians, perhaps in a fixed homology or Hamiltonian isotopy class. This could then, potentially, lead to a new *invariant* for Calabi–Yau 3-folds. There have been various attempts to pursue this programme but it currently is incomplete.

5.4 Associative and coassociative deformations

We now want to understand deformations of associatives and coassociatives, from which perturbation or gluing results will follow. We begin with associatives.

Notice that if P is an associative plane, $u \in P$ and $v \in P^\perp$ then for all $w \in P$ we have $w \times u \in P$ and hence

$$\varphi(w, u, v) = g(w, u \times v) = g(v, w \times u) = 0.$$

We deduce that $u \times v \in P^\perp$. Thus, if N is associative, cross product gives a (Clifford) multiplication

$$m : C^\infty(T^*N \otimes \nu(N)) \to C^\infty(\nu(N))$$

(viewing tangent vectors as cotangent vectors via the metric). Hence, using the normal connection

$$\nabla^\perp : C^\infty(\nu(N)) \to C^\infty(T^*N \otimes \nu(N))$$

on $\nu(N)$ we get a linear operator as follows.

Definition 5.3. Let N be an associative in a G_2-manifold. Using the notation above we define

$$\not{D} = m \circ \nabla^\perp : C^\infty(\nu(N)) \to C^\infty(\nu(N)).$$

We call $\not{D}$ the *Dirac operator.* We see that its principal symbol is given by

$$\sigma_{\not{D}}(x, \xi)v = i\xi \times v,$$

so $\not{D}$ is elliptic, and we also have that $\not{D}^* = \not{D}$.

Remark. Since an orientable 3-manifold is always spin, we have a spinor bundle $\mathbb{S}$ on N, a connection

$$\nabla : C^\infty(\mathbb{S}) \to C^\infty(T^*M \otimes \mathbb{S})$$

(a lift of the Levi-Civita connection) and we have Clifford multiplication

$$m : C^\infty(T^*M \otimes \mathbb{S}) \to C^\infty(\mathbb{S}) \qquad m(\xi, v) = \xi \cdot v.$$

Hence we have a composition

$$\not{D} = m \circ \nabla : C^\infty(\mathbb{S}) \to C^\infty(\mathbb{S}),$$

which is a first order linear differential operator called the Dirac operator. Locally it is given by

$$\not{D}v = \sum_i e_i \cdot \nabla_{e_i} v,$$

so we have that $\sigma_{\not{D}}(\xi, v) = i\xi \cdot v$. Hence $\not{D}$ is elliptic. Moreover $\not{D}$ is self-adjoint.

In fact, it is possible (see e.g. [66]) to see that the complexified normal bundle $\nu(N) \otimes \mathbb{C} = \mathbb{S} \otimes V$ for a $\mathbb{C}^2$-bundle V over N, so that the Dirac operator on $\nu(N)$ is just a "twist" of the usual Dirac operator on $\mathbb{S}$.

Consider a compact associative N. We want to describe the associative deformations of N, just as in the case of special Lagrangians above. To be consistent with that previous setting, we will now use P to denote a nonlinear deformation map: we trust that this will not cause confusion given the context.

We know that $\exp_v(N) = N_v$, which is the graph of v, is associative for a normal vector field v if and only if $*\exp_v^*(\chi) \in C^\infty(TM|_N)$ is 0. In fact, it turns out that $P(v) = *\exp_v^*(\chi) \in C^\infty(\nu(N))$ since N is associative and

$$L_0 P(v) = *\mathrm{d}(v \lrcorner \chi) = \not{D} v.$$

Here L_0P is not typically surjective so we cannot apply our Implicit Function Theorem, except when $\operatorname{Ker} \not{D} = \operatorname{Ker} \not{D}^* = \{0\}$. However, we can still say something in these circumstances, for which we make a small digression to a more general situation.

Suppose X, Y are Banach spaces. Let $U \subseteq X$ be an open set with $0 \in U$ and let $P : U \to Y$ be a smooth map with $P(0) = 0$ such that $L_0P : X \to Y$ is Fredholm.

Let $\mathcal{I} = \operatorname{Ker} L_0P$ and let $\mathcal{O}$ be such that $Y = L_0P(X) \oplus \mathcal{O}$, which exists and is finite-dimensional by the assumption that L_0P is Fredholm. We then let $Z = X \oplus \mathcal{O}$ and define $F : U \oplus \mathcal{O} \to Y$ by $F(u, y) = P(u) + y$. We see that $L_0F : X \oplus \mathcal{O} \to Y = L_0P(X) \oplus \mathcal{O}$ is given by $L_0F(x, y) = L_0P(x) + y$ which is surjective and $L_0F(x, y) = 0$ if and only if $L_0P(x) = 0$ and $y = 0$, thus $\operatorname{Ker} L_0F = \operatorname{Ker} L_0P \times \{0\}$.

There exists $W \subseteq X$ such that $\operatorname{Ker} L_0P \oplus W = X$. Applying the Implicit Function Theorem, there exist open sets $U_1 \subseteq \operatorname{Ker} L_0P$ containing 0, $U_2 \subseteq W$ containing 0 and $U_3 \subseteq \mathcal{O}$ containing 0 and smooth maps $G_2 : U_1 \to U_2$, $G_3 : U_1 \to U_3$ such that

$$F^{-1}(0) \cap U_1 \times U_2 \times U_3 = \{(u, G_2(u), G_3(u)) \,:\, u \in U_1\}.$$

We also know that $P(x) = 0$ if and only if $F(x, y) = 0$ and $y = 0$. Hence

$$P^{-1}(0) \cap U_1 \times U_2 = \{(u, G_2(u)) \,:\, u \in G_3^{-1}(0)\}.$$

Let $\mathcal{U} = U_1$ and define $\pi : \mathcal{U} \to \mathcal{O}$ by $\pi(u) = G_3(u)$. Then $P^{-1}(0) \cap U_1 \times U_2$ is a graph over $\pi^{-1}(0)$, and hence $P^{-1}(0)$ is locally homeomorphic to $\pi^{-1}(0)$.

Sard's Theorem says that generically $\pi^{-1}(y)$ is a smooth manifold of dimension

$$\dim \mathcal{I} - \dim \mathcal{O} = \dim \operatorname{Ker} L_0P - \dim \operatorname{Coker} L_0P = \operatorname{ind}(L_0P),$$

which is the index of L_0P. Hence, the expected dimension of $P^{-1}(0)$ is the index of L_0P.

In the associative setting we have that the linearisation is $\not{D}$, which is elliptic and thus Fredholm, and we know that $\operatorname{ind} \not{D} = \dim \operatorname{Ker} \not{D} - \dim \operatorname{Ker} \not{D}^* = 0$. We deduce the following [66].

Theorem 5.4. *The expected dimension of the moduli space of deformations of a compact associative 3-fold N in a G_2 manifold is 0 and infinitesimal deformations of N are given by the kernel of $\not{D}$ on $\nu(N)$. Moreover, if $\operatorname{Ker} \not{D} = \{0\}$ then N is rigid.*

Remark. The dimension of the kernel of $\not{D}$ typically depends on the metric on N rather than just the topology, so it is usually difficult to determine. However, there are some cases where one can ensure the moduli space is smooth cf. [15].

Example. For the associative $N = \mathcal{S}^3$ in the complete Bryant–Salamon holonomy G_2 manifold $\mathbb{S}(\mathcal{S}^3)$, $\nu(N) = \mathbb{S}(\mathcal{S}^3)$ so $\not{D}$ is just the usual Dirac operator. A theorem of Lichnerowicz states that $\operatorname{Ker} \not{D} = \{0\}$ as $\mathcal{S}^3$ has positive scalar curvature so N is rigid.

Example. Corti–Haskins–Nordström–Pacini construct rigid associative $\mathcal{S}^1 \times \mathcal{S}^2$s in compact holonomy G_2 manifolds, known as twisted connected sums [10].

Remark. The fact that associatives have an index zero deformation theory suggests that, perhaps in very good situations, one might be able to "count" associatives by asking simply how many there are in, say, a fixed homology class. One might then hope to define an *invariant* for G_2 manifolds from this count. This turns out to be somewhat naive, but still very motivational.

For coassociatives, the deformation theory is much better behaved, like for special Lagrangians [66].

Theorem 5.5. *Let N be a compact coassociative in a G_2 manifold (or just a 7-manifold with closed G_2 structure). The moduli space of deformations of N is a smooth manifold of dimension $b^2_+(N)$.*

Proof. Since N is coassociative the map $v \mapsto v \lrcorner \varphi = \alpha_v$ defines an isomorphism from $\nu(N)$ to a rank 3 vector bundle on N, which is $\Lambda^2_+ T^*N$, the 2-forms on N which are self-dual (so $*\alpha = \alpha$). We can therefore view nearby submanifolds to N as graphs of self-dual 2-forms.

We know that $N_v = \exp_v(N)$ is coassociative if and only if $\exp_v^*(\varphi) = 0$. We see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp_{tv}^*(\varphi)|_{t=0} = \mathcal{L}_v \varphi = \mathrm{d}(v \lrcorner \varphi) = \mathrm{d}\alpha_v.$$

Hence nearby coassociatives N' to N are given by the zeros of $P(\alpha) = \mathrm{d}\alpha + Q(\alpha, \nabla\alpha)$. Moreover, since $\varphi = 0$ on N, $[\varphi] = 0$ on N' and hence $P : C^\infty(\Lambda^2_+ T^*N) \to \mathrm{d}(C^\infty(\Lambda^2 T^*N))$.

Here P is not elliptic, but $L_0 P = \mathrm{d}$ has finite-dimensional kernel, the closed self-dual 2-forms, since $\mathrm{d}\alpha = 0$ implies that $\mathrm{d}^*\alpha = - * \mathrm{d} * \alpha = 0$ so α is harmonic. Moreover, $L_0 P$ has injective symbol so it is overdetermined elliptic, which means that elliptic regularity still holds. Another way to deal with this is to consider $F(\alpha, \beta) = P(\alpha) + \mathrm{d}^*\beta$ for β a 4-form. Now $F^{-1}(0)$ is the disjoint union of $P^{-1}(0)$ and multiples of the volume form, as exact and coexact forms are orthogonal. Moreover, $L_0 F(\alpha, \beta) = \mathrm{d}\alpha + \mathrm{d}^*\beta$ is now elliptic. Overall, we can apply our standard Implicit Function Theorem if we know that

$$\mathrm{d}(C^{k+1,a}(\Lambda^2_+ T^*N)) = \mathrm{d}(C^{k+1,a}(\Lambda^2 T^*N)).$$

This is true because by Hodge theory if α is a 2-form, we can write $\alpha = \mathrm{d}^*\beta + \gamma$ for a 3-form β and a closed form γ, so $\mathrm{d}\alpha = \mathrm{d}\mathrm{d}^*\beta = \mathrm{d}(\mathrm{d}^*\beta + *\mathrm{d}^*\beta)$ and $\mathrm{d}^*\beta + *\mathrm{d}^*\beta$ is self-dual. □

Example. The $\mathcal{S}^4$ and $\overline{\mathbb{CP}^2}$ in the Bryant–Salamon metrics on $\Lambda^2_+ T^*\mathcal{S}^4$ and $\Lambda^2_+ T^*\overline{\mathbb{CP}^2}$ have $b^2_+ = 0$ and so are rigid. (Here it is important that we write $\overline{\mathbb{CP}^2}$ since $\mathbb{CP}^2$ has $b^2_+ = 1$.)

Example. For a K3 surface and T^4 we have $b^2_+ = 3$ and Λ^2_+ is trivial, so we can hope to find coassociative K3 and T^4 *fibrations* of compact G_2 manifolds. There is a programme [47] for constructing a coassociative K3 fibration (with singular fibres). Towards completing this programme, the first examples of compact coassociative 4-folds with conical singularities in compact holonomy G_2 twisted connected sums were constructed in [61].

Remark. The deformation theory results for compact associative and coassociative submanifolds have been extended to certain non-compact, singular and boundary settings, for example in [16, 44, 48, 55, 56, 58].

5.5 Cayley deformations

To discuss deformations of a compact Cayley N, we need some further discussion of algebra related to Spin(7).

When P is a Cayley plane and $u, v \in P$ are orthogonal we see that $\Phi_0(u, v, ., .) = *_P(u \wedge v)$ so that $u \wedge v + \Phi_0(u, v, ., .)$ is self-dual on P. Since $\Lambda^2_+ P^*$ is 3-dimensional, we see that there must be a 4-dimensional space E of 2-forms on P such that $\Lambda^2_7|_P = \Lambda^2_+ P^* \oplus E$. Moreover, if $u \in P$ and $v \in P^\perp$ then $m(u, v) = u \wedge v + \Phi_0(u, v, ., .) \in E$ and the map $m : P \times P^\perp \to E$ is surjective.

Now let us move to a Cayley submanifold N in a Spin(7) manifold (M, Φ). On M we have a rank 7 bundle Λ^2_7 of 2-forms and we have that

$$\Lambda^2_7|_N = \Lambda^2_+ T^*N \oplus E$$

for some rank 4 bundle E over N. The map m above defines a (Clifford) multiplication

$$m : C^\infty(T^*N \otimes \nu(N)) \to C^\infty(E)$$

(viewing tangent vectors as cotangent vectors via the metric), and thus using the normal connection

$$\nabla^\perp : C^\infty(\nu(N)) \to C^\infty(T^*N \otimes \nu(N))$$

we get a linear first order differential operator as follows.

Definition 5.6. Let N be a Cayley 4-fold in a Spin(7) manifold (M, Φ). Using the notation above we define

$$\not{D}_+ = m \circ \nabla^\perp : C^\infty(\nu(N)) \to C^\infty(E).$$

Again this an elliptic operator called the *positive Dirac operator*, but it is not self-adjoint: its adjoint is the negative Dirac operator from E to $\nu(N)$.

Remark. If N is spin, the spinor bundle $\mathbb{S}$ splits as $\mathbb{S}_+ \oplus \mathbb{S}_-$, and the Dirac operator $\not{D}$ splits into $\not{D}_\pm$ from $\mathbb{S}_\pm$ to $\mathbb{S}_\mp$ so that $\not{D}(v_+, v_-) = (\not{D}_- v_-, \not{D}_+ v_+)$. Hence $\not{D}^* = \not{D}$ says that $\not{D}_\pm^* = \not{D}_\mp$.

It turns out (see, for example, [66]) that there exists a $\mathbb{C}^2$-bundle V on N so that $\nu(N) \otimes \mathbb{C} = \mathbb{S}_+ \otimes V$, $E \otimes \mathbb{C} = \mathbb{S}_- \otimes V$ and $\not{D}_+$ on $\nu(N)$ is a "twist" of the usual positive Dirac operator. However, not every 4-manifold is spin, so we cannot always make this identification.

Now suppose that N is a compact Cayley 4-fold. Then the zeros of the equation $F(v) = *\exp_v^*(\tau)$ for $v \in C^\infty(\nu(N))$ define Cayley deformations (as the graph of v). We know that F takes values in $\Lambda^2_7|_N = \Lambda^2_+ T^*N \oplus E$ and it turns out that

$$L_0 F(v) = *\mathrm{d}(v \lrcorner \tau) = \not{D}_+ v$$

since N is Cayley. So, we potentially have a problem because F does not necessarily take values only in E (and in general it will not just take values in E). However, the Cayley condition on N means that $F(v) = 0$ if and only $P(v) = \pi_E F(v) = 0$, where π_E is the projection onto E (again, we are using P to denote the nonlinear deformation map as in our previous discussion, and we expect it will not cause confusion given the context). Then the operator $P : C^\infty(\nu(N)) \to C^\infty(E)$ and $L_0 P = \not{D}_+$ is elliptic.

Again, we cannot say that $L_0 P$ is surjective, so we have the following using the same argument as in the lead up to Theorem 5.4, cf. [66].

Theorem 5.7. *The expected dimension of the moduli space of deformations of a compact Cayley 4-fold N in a* Spin(7) *manifold is* $\operatorname{ind} \not{D}_+ = \dim \operatorname{Ker} \not{D}_+ - \dim \operatorname{Ker} \not{D}_+^*$ *with infinitesimal deformations given by* $\operatorname{Ker} \not{D}_+$ *on* $\nu(N)$. *Moreover,*

$$\operatorname{ind} \not{D}_+ = \frac{1}{2}\sigma(N) + \frac{1}{2}\chi(N) - [N].[N],$$

where $\sigma(N) = b^2_+(N) - b^2_-(N)$ *(the signature of N),* $\chi(N) = 2b^0(N) - 2b^1(N) + b^2(N)$ *(the Euler characteristic of N) and* $[N].[N]$ *is the self-intersection of N, which is the Euler number of* $\nu(N)$.

Example. For the Cayley $N = \mathcal{S}^4$ in the holonomy Spin(7) Bryant–Salamon manifold $\mathbb{S}_+(\mathcal{S}^4)$, $\nu(N) = \mathbb{S}_+(\mathcal{S}^4)$ and $\not{D}_+$ is the usual positive Dirac operator. Again, since N has positive scalar curvature, we see that $\operatorname{Ker} \not{D}_\pm = \{0\}$ so N is rigid.

Remark. Theorem 5.7 has been extended to various other non-compact, singular and boundary settings, for example in [68, 70, 71].

5.6 Gluing problems

We now return to the issue of construction of new calibrated submanifolds, by starting with some known examples.

A well-known way to get a solution of a linear PDE from two solutions is simply to add them. However, for a nonlinear PDE $P(v) = 0$ this will not work. Intuitively, we can try to add two solutions to give us a solution v_0 for which $P(v_0)$ is small. Then we may try to perturb v_0 by v to solve $P(v + v_0) = 0$.

Geometrically, this occurs when we have two calibrated submanifolds N_1, N_2 and then glue them together to give a submanifold N which is "almost" calibrated. Then we wish deform N to become calibrated: here is where the ideas from the moduli problems comes into play.

If the two submanifolds N_1, N_2 are glued using a very long neck then one can imagine that N is almost the disjoint union of N_1, N_2 and so close to being calibrated. If instead one scales N_2 by a factor t and then glues it into a singular point of N_1, we can again imagine that as t becomes very small N resembles N_1 and so again is close to being calibrated. These two examples are in fact related, because if we rescale the shrinking N_2 to fixed size, then we get a long neck between N_1 and N_2 of length of order $-\log t$. However, although these pictures are appealing, they also reveal the difficulty in this approach: as t becomes small, N becomes more "degenerate", giving rise to analytic difficulties which are encoded in the geometry of N_1, N_2 and N.

These ideas are used extensively in geometry, and particularly successfully in calibrated geometry e.g. [7, 23, 35, 37, 38, 51, 57, 62, 73]. A particular simple case is the following, which we will describe to show the basic idea of the gluing method.

Theorem 5.8. *Let N be a compact connected 3-manifold and let $i : N \to M$ be a special Lagrangian immersion with tranverse self-intersection points in a Calabi–Yau manifold M. Then there exist embedded special Lagrangians N_t such that $N_t \to N$ as $t \to 0$.*

Remark. One might ask about the sense of convergence here: for definiteness, we can say that N_t converges to N in the sense of currents; that is, if we have any compactly supported 3-form χ on M then $\int_{N_t} \chi \to \int_N \chi$ as $t \to 0$. However, all sensible notions of convergence of submanifolds will be true in this setting.

Proof. Here we only provide a sketch of the proof: see, for example, [35, §9] for a detailed proof.

At each self-intersection point of N the tangent spaces are a pair of transverse 3-planes, which we can view as a pair of tranverse special Lagrangian 3-planes P_1, P_2 in $\mathbb{C}^3$. Since we are in dimension 3, we know that there exists a (unique up to scale) special Lagrangian Lawlor neck L asymptotic to $P_1 \cup P_2$. We can then glue tL into N near each intersection point to get a compact embedded submanifold $S_t = N \# tL$ (if we glue in a Lawlor neck for every self-intersection point). We can also arrange that S_t is Lagrangian, i.e. that it is a Lagrangian connect sum.

Now we want to perturb S_t to be special Lagrangian. Since S_t is Lagrangian, by the deformation theory we can write any nearby submanifold as the graph of a 1-form α, and this graph will be special Lagrangian if and only if (using the same notation as in our deformation theory discussion)

$$P_t(\alpha) = (-*f_\alpha^*(\operatorname{Im}\Upsilon), f_\alpha^*(\omega)) = 0.$$

Since S_t is Lagrangian but not special Lagrangian we have that

$$f_\alpha^*(\omega) = \mathrm{d}\alpha \quad \text{and} \quad -*f_\alpha^*(\operatorname{Im}\Upsilon) = P_t(0) + \mathrm{d}_t^*\alpha + Q_t(\alpha, \nabla\alpha)$$

where $P_t(0) = -*\operatorname{Im}\Upsilon|_{S_t}$ and $\mathrm{d}_t^* = L_0 P_t$, which is a perturbation of the usual d^* since we are no longer linearising at a point where $P_t(0) = 0$. By choosing $\alpha = \mathrm{d}f$, we then have to solve

$$\Delta_t f = -P_t(0) - Q_t(\nabla f, \nabla^2 f)$$

where Δ_t is a perturbation of the Laplacian.

For simplicity, let us suppose that Δ_t is the Laplacian on S_t. The idea is to view our equation as a fixed point problem. We know that if we let $X^k = \{f \in C^{k,a}(N) : \int_N f \operatorname{vol}_N = 0\}$ then $\Delta_t : X^{k+2} \to X^k$ is an isomorphism so it has an inverse G_t. We know by elliptic regularity that there exists a constant $C(\Delta_t)$ such that

$$\|f\|_{C^{k+2,a}} \leq C(\Delta_t)\|\Delta_t f\|_{C^{k,a}} \Leftrightarrow \|G_t h\|_{C^{k+2,a}} \leq C(\Delta_t)\|h\|_{C^{k,a}}$$

for any $f \in X^{k+2}$, $h \in X^k$.

We thus see that $P_t(f) = 0$ for $f \in X^{k+2}$ if and only if

$$f = G_t(-P_t(0) - Q_t(f)) = F_t(f).$$

The idea is now to show that F_t is a contraction sufficiently near 0 for all t small enough. Then it will have a (unique) fixed point near 0, which will also be smooth because it satisfies $P_t(f) = 0$ and hence defines a special Lagrangian as the graph of $\mathrm{d}f$ over S_t.

We know that $F_t : X^{k+2} \to X^{k+2}$ with

$$\|F_t(f_1) - F_t(f_2)\|_{C^{k+2,a}} = \|G_t(Q_t(f_1) - Q_t(f_2))\|_{C^{k+2,a}} \le C(\Delta_t)\|Q_t(f_1) - Q_t(f_2)\|_{C^{k,a}}.$$

Since Q_t and its first derivatives vanish at 0 we know that

$$\|Q_t(f_1) - Q_t(f_2)\|_{C^{k,a}} \le C(Q_t)\|f_1 - f_2\|_{C^{k+2,a}}(\|f_1\|_{C^{k+2,a}} + \|f_2\|_{C^{k+2,a}}).$$

We deduce that

$$\|F_t(f_1) - F_t(f_2)\|_{C^{k+2,a}} \le C(\Delta_t)C(Q_t)\|f_1 - f_2\|_{C^{k+2,a}}(\|f_1\|_{C^{k+2,a}} + \|f_2\|_{C^{k+2,a}})$$

and

$$\|F_t(0)\|_{C^{k+2,a}} = \|G_t(P_t(0))\|_{C^{k+2,a}} \le C(\Delta_t)\|P_t(0)\|_{C^{k,a}}.$$

Hence, F_t is a contraction on $\overline{B_{\epsilon_t}(0)} \subseteq X^{k+2}$ if we can choose ϵ_t so that

$$2C(\Delta_t)\|P_t(0)\|_{C^{k,a}} \le \epsilon_t \le \frac{1}{4C(\Delta_t)C(Q_t)}.$$

(This also proves Theorem 5.2, where we used the Implicit Function Theorem, by hand since there $P_t(0) = P(0) = 0$ so we just need to take ϵ_t small enough.) In other words, we need that

- $P_t(0)$ is small, so S_t is "close" to being calibrated and is a good approximation to $P_t(f) = 0$;
- $C(\Delta_t), C(Q_t)$, which are determined by the linear PDE and geometry of N, L and S_t, are well-controlled as $t \to 0$.

The statement of the theorem is then that there exists t sufficiently small and ϵ_t so that the contraction mapping argument works.

This is a delicate balancing act since as $t \to 0$ parts of the manifold are collapsing, so the constants $C(\Delta_t), C(Q_t)$ above (which depend on t) can and typically do blow-up as $t \to 0$. To control this, we need to understand the Laplacian on N, L and S_t and introduce "weighted" Banach spaces so that tL gets rescaled to constant size (independent of t), and S_t resembles the union of two manifolds with a cylindrical neck (as we described earlier). It is also crucial to understand the relationship between the kernels and cokernels of the Laplacian on the *non-compact* N (with the intersection points removed), L and compact S_t: here is where connectedness is important so that the kernel and cokernel of the Laplacian is 1-dimensional. □

Remark. In more challenging gluing problems it is not possible to show that the relevant map is a contraction, but rather one can instead appeal to an alternative theorem (e.g. Schauder fixed point theorem) to show that it still has a fixed point.

Example. Suppose we have a 1-parameter family of G_2-manifolds (M^7, φ_s) for $s \in (-\epsilon, \epsilon)$ we have a pair of associatives $N_{1,s}$ and $N_{2,s}$ in each (M^7, φ_s) so that $N_{j,s}$ vary smoothly with s.

Since associatives are 3-dimensional, we would expect them not to intersect. However, since $3 + 3 = 7 - 1$, we would expect that in a generic 1-parameter family (M^7, φ_s) we would have that associatives would intersect at a point for some s. Therefore, suppose that $N_{1,s} \cap N_{2,s} = \emptyset$ for $s \neq 0$, but that $N_{1,0} \cap N_{2,0} = \{x\}$, where they intersect transversely. Under some additional hypothesis, it was shown by Nordström that for $s > 0$ (say) there exists a new associative $N_s \cong N_{1,s} \# N_{2,s}$. The idea is to use the moduli space theory of associatives and use Lawlor necks to resolve the intersection point in a similar manner to what we saw above in the special Lagrangian setting.

An important consequence of this result is that new associatives can be "born" or "appear" as we vary in the family (M^7, φ_s) and cross $s = 0$. (By reversing the process, we can also have that associatives "die" or "disappear".) This shows one of the many inherent difficulties in trying to define an invariant by counting associatives in G_2-manifolds.

6 Introduction to gauge theory

We now want to move onto the second topic of the course: gauge theory. As we said at the very start of the course, gauge theory has proved to be a powerful tool in geometry and topology, most notably the study of topology of 3 and 4-dimensional manifolds. We are going to focus on gauge theory in higher dimensions, which are motivated by the low-dimensional gauge theories, but which have the additional feature that our manifolds require more structure: in fact, they will have special holonomy. Later, we will see that gauge theory in higher dimensions is intimately related to the calibrated geometry which have studied in the earlier part of the course.

Gauge theory is the study of connections on manifolds. It has often proved useful to study connections on Riemannian manifolds, even if ultimately one show that the invariants that one gets from the theory is in some sense independent of the choice of Riemannian metric (for example, as in Donaldson theory on 4-manifolds). In this section we will briefly review the basics of gauge theory, then discuss some of the most important features of gauge theory in low dimensions that will resurface when we discuss higher-dimensional gauge theory.

We shall therefore consider a Riemannian manifold (M, g) and a vector bundle E of rank k over M. We could also take a principal G-bundle P over M, where the group G is taken to be a compact Lie group. We would often restrict to the cases where G is U(1), SO(3) or SU(2), with the last case typically of primary interest. Later, it will be necessary for our vector bundle to be equipped with Euclidean metrics on its fibres: one could think of the tangent bundle or any other bundle of forms over M.

6.1 Connections and curvature

Definition 6.1. We shall let $\mathcal{A}$ denote the space of *connections* on E, which we can view as covariant derivatives $\nabla_A : C^\infty(E) \to C^\infty(T^*M \otimes E)$. This is an affine space modelled on $C^\infty(T^*M \otimes \mathrm{End}(E))$, which is the 1-forms on M with values in $\mathrm{End}(E)$, the endomorphism bundle of E. Note that, this means that the difference of two connections is a 1-form with values in $\mathrm{End}(E)$.

Remark. If E has more structure, say it has a metric or a complex structure on each fibre, then we would restrict to connections preserving or compatible with that structure, which would then restrict the space of endomorphisms of E we consider.

Remark. Once we pick a reference connection A_0 every other connection A can be written $A = A_0 + a$ where $a \in C^\infty(T^*M \otimes \mathrm{End}(E))$. In terms of covariant derivatives,

$$\nabla_A = \nabla_{A_0} + a.$$

We now give an example that we will keep returning to.

Example. In a local trivialization of E, we can pick the trivial connection A_0 as our reference and identify all the fibres of E in this trivialization with $\mathbb{R}^k$. Since endomorphisms of $\mathbb{R}^k$ are just given by multiplication by $k \times k$ matrices, we can write any connection A as a matrix-valued 1-form on this trivialization. It is worth noting that the $k \times k$ matrices are the Lie algebra $\mathfrak{gl}(k, \mathbb{R})$ of the Lie group $\mathrm{GL}(k, \mathbb{R})$.

In terms of covariant derivatives, in this local trivialization we can write a section s of E as

$$s = \sum_{j=1}^{k} s_j \otimes e_j$$

with respect to some constant basis $e_1, \ldots, e_k$ on E, for some functions s_j, and then

$$\nabla_A s = \sum_{j=1}^{k} (\mathrm{d}s_j + A s_j) \otimes e_j.$$

We often say that, locally, $\nabla_A = \mathrm{d} + A$.

Remark. The above example is helpful, but is dependent on the choice of local trivialization. What happens if we change trivialization? A straightforward calculation shows that if we change the trivialization on an open set U in M using $g : U \to \mathrm{GL}(k, \mathbb{R})$ then the matrix A changes by

$$g : A \mapsto g^{-1}Ag + g^{-1}\mathrm{d}g.$$

This shows that A does *not* define a tensor on M: this should be familiar from the study of Christoffel symbols in Riemannian geometry.

Example. In Riemannian geometry, $E = TM$ and since we have an inner product on each fibre of E (given by the Riemannian metric), we can talk about orthogonal transformations of E. If A (or ∇_A) is the Levi-Civita connection then in local trivializations it is a 1-form taking values in the Lie algebra of the orthogonal group, i.e. the skew-symmetric matrices.

Definition 6.2. We let $\mathcal{G}$ denote the space of *gauge transformations* of E: the sections g of $\mathrm{End}(E)$ which are invertible on each fibre of E. These act on connections by

$$g : \nabla_A \mapsto g^{-1}\nabla_A g$$

for $g \in \mathcal{G}$.

We now introduce the key space which is studied in gauge theory.

Definition 6.3. The *moduli space* of connections is $\mathcal{A}/\mathcal{G}$.

We now want to discuss the *curvature* of a connection. To do that, we need to first discuss the *exterior covariant derivative* d_A associated to a connection A, which is just the extension of ∇_A to E-valued forms. Explicitly, we have the following.

Definition 6.4. If we have an E-valued l-form of the form $\sigma \otimes s$, where σ is a usual l-form and s is a section of E, then the *exterior covariant derivative* d_A is just defined by the "Leibniz rule"

$$\mathrm{d}_A(\sigma \otimes s) = \mathrm{d}\sigma \otimes s + (-1)^l \sigma \otimes \nabla_A s.$$

We can then extend d_A to all E-valued forms by linearity.

Given this definition, we can then introduce the notion of *curvature*.

Definition 6.5. Given a connection A on E, the *curvature* F_A of A, is the $\mathrm{End}(E)$-valued 2-form on M satisfying

$$\mathrm{d}_A \circ \mathrm{d}_A = F_A$$

on all E-valued forms. We see that if we choose a reference connection A_0 and write $A = A_0 + a$ for an $\mathrm{End}(E)$-valued 1-form a, then

$$F_A = F_{A_0} + \mathrm{d}_{A_0}a + a \wedge a.$$

Note that even though A cannot (in general) be viewed globally as an $\mathrm{End}(E)$-valued form on M, its curvature can be.

Recall that if we had the trivial connection, then $\mathrm{d}_A = \mathrm{d}$ and so $\mathrm{d}_A^2 = \mathrm{d}^2 = 0$ so $F_A = 0$. We can therefore think of F_A as a measure of the failure of d_A^2 to be zero. This is still quite abstract, so let's return to our local trivialization.

Example. In a local trivialization as before, where we view the connection A as a $k \times k$ matrix of 1-forms, we can compute that the curvature is

$$F_A = \mathrm{d}A + A \wedge A,$$

where the wedge product includes matrix multiplication and wedge product on the 1-forms.

If we choose local coordinates $(x_1, \ldots, x_n)$ on M in this local trivialization and let ∂_i and ∇_i denote the tangent vector and covariant derivative (given by ∇_A) in the x_i direction, then

$$F_A(\partial_i, \partial_j) = \nabla_i \circ \nabla_j - \nabla_j \circ \nabla_i.$$

This is reminiscent of the usual definition of curvature in terms of second derivatives and the failure for the commutator of first derivatives to be zero.

Moreover, under a gauge transformation g, we see that the action is

$$g : F_A \mapsto g^{-1} F_A g,$$

which confirms that F_A is indeed a tensor, since it transforms in the correct manner.

Example. In the Riemannian geometry case, where $E = TM$ and A is the Levi-Civita connection, F_A takes values in the skew-symmetric transformations of the fibres, so F_A actually takes values in the 2-forms on M. This reflects two of the symmetries of the Riemann curvature tensor R_{abcd}: that it is skew in the first and last pair of indices, i.e. $R_{bacd} = R_{abdc} = -R_{abcd}$. Another famous symmetry, that it is invariant when we swap the first and last pairs of indices ($R_{cdab} = R_{abcd}$) is the statement that F_A is actually a symmetric map from 2-forms to 2-forms.

We also want to make one more observation about the curvature of a connection, which we leave as an exercise.

Lemma 6.6. *The curvature F_A of a connection A satisfies the Bianchi identity:*

$$\mathrm{d}_A F_A = 0.$$

Remark. Before we conclude this section, we wish to ask the question: how can we describe the points in the moduli space $\mathcal{A}/\mathcal{G}$? As we discussed before, this is the notion of *gauge-fixing*. If we fix a reference connection A_0, then all other connections are given by $A = A_0 + a$ for $\mathrm{End}(E)$-valued 1-forms a. It turns out that the linearisation at A_0 of the action of the gauge group at the identity is given by:

$$A_0 \mapsto A_0 + \mathrm{d}_{A_0} u$$

for a section u of $\mathrm{End}(E)$. Hence, if we let $\mathrm{d}_{A_0}^*$ denote the formal adjoint of d_{A_0}, then we can impose the *Coulomb gauge* condition

$$\mathrm{d}_{A_0}^* a = 0$$

to take care of the gauge freedom, at least near A_0. The reason why is that if we impose $\mathrm{d}_{A_0}^* \mathrm{d}_{A_0} u = 0$ then

$$0 = \langle \mathrm{d}_{A_0}^* \mathrm{d}_{A_0} u, u \rangle_{L^2} = \|\mathrm{d}_{A_0} u\|_{L^2}^2 \quad \Rightarrow \quad \mathrm{d}_{A_0} u = 0,$$

as long as integration by parts is valid (say if M is compact or u is compactly supported. This is what we mean by taking a slice for the action of the gauge group $\mathcal{G}$. We then can describe points in $\mathcal{A}/\mathcal{G}$ near A_0 as $\mathrm{End}(E)$-valued 1-forms a satisfying $\mathrm{d}_{A_0}^* a = 0$.

6.2 Yang–Mills functional

We now wish to discuss one of the important objects in the study of gauge theory, which originates from the study of particle physics. From this point on we need a Euclidean metric on the fibres of our vector bundle E, so we shall assume this going forward. In particular, this means that the connections $\mathcal{A}$ will be compatible with this metric and elements of the gauge group $\mathcal{G}$ will define fibrewise orthogonal transformations on E.

Definition 6.7. For a connection A with curvature F_A on an oriented Riemannian manifold (M, g), we define the *Yang–Mills energy* of A by

$$\mathcal{E}(A) = \int_M |F_A|^2 = \|F_A\|_{L^2}^2,$$

when this is well-defined (for example, if M is compact). Since $|g^{-1} F_A g|^2 = |F_A|^2$ for $g \in \mathcal{G}$, the Yang–Mills energy defines the *Yang–Mills functional* $\mathcal{E} : \mathcal{A}/\mathcal{G} \to \mathbb{R}$ on the moduli space $\mathcal{A}/\mathcal{G}$.

Remark. The Yang–Mills functional on the moduli space of connections will play a similar role to the volume functional on the space of submanifolds. We did not stress the point at the time, but the volume functional is also defined on a quotient space: the space of immersions modulo reparametrization by diffeomorphisms of the domain, since the volume of submanifold is independent of its parametrization.

A natural question to ask, just as we did for the volume functional, is: what are the critical points for the Yang–Mills functional? To answer this, suppose we have a connection A which is critical for the Yang–Mills functional, and let a be any $\mathrm{End}(E)$-valued 1-form. We may then compute for a real number t (recalling the formula for F_{A+ta}):

$$\begin{aligned}\mathcal{E}(A+ta) &= \int_M |F_{A+ta}|^2 \\ &= \int_M |F_A + t\mathrm{d}_A a + t^2 a\wedge a|^2 \\ &= \mathcal{E}(A) + 2t\langle F_A, \mathrm{d}_A a\rangle_{L^2} + O(t^2).\end{aligned}$$

Differentiating this formula and setting $t=0$ gives:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(A+ta)|_{t=0} &= 2\langle F_A, \mathrm{d}_A a\rangle_{L^2} \\ &= 2\langle \mathrm{d}_A^* F_A, a\rangle_{L^2},\end{aligned}$$

where d_A^* is the formal adjoint of d_A that we saw before. We deduce the following result.

Lemma 6.8. *A connection A is a critical point for the Yang–Mills functional if and only if* $\mathrm{d}_A^* F_A = 0$.

Hence, we make a corresponding definition.

Definition 6.9. A connection A is *Yang–Mills* if $\mathrm{d}_A^* F_A = 0$.

Remark. Notice the similarities with the minimal submanifold condition: again, the Yang–Mills condition is a critical point condition for a functional and it is a *second order* partial differential equation on the connection (since the curvature is first order in the connection).

The Yang–Mills condition is also elliptic (modulo gauge): this is not immediately clear, since we only have the equation $\mathrm{d}_A^* F_A = 0$, but if we recall the Bianchi identity $\mathrm{d}_A F_A = 0$, then we see that Yang–Mills connection can equivalently be defined by the elliptic system

$$\mathrm{d}_A F_A = 0 \quad \text{and} \quad \mathrm{d}_A^* F_A = 0.$$

Before we go any further, we should try to find some examples of Yang–Mills connections.

Example. A flat connection, i.e. with $F_A = 0$, is a Yang–Mills connection.

Remark. The previous example of flat connections may seem like a trivial example, but it really isn't! Flat connections are related to representations of the fundamental group of the manifold, via their holonomy: if you take a loop γ based at a point p and compute the holonomy of the connection around γ then this will be independent of the choice of representative of the homotopy class of γ in $\pi_1(M)$ by the flatness condition. Hence, flat connections are important from a topological viewpoint.

More than that, it was shown by Taubes that one can use this relation to interpret the *Casson invariant* for rational homology 3-spheres in terms of gauge theory: specifically, as the Euler characteristic of the moduli space $\mathcal{A}/\mathcal{G}$, where one considers E to have structure group $\mathrm{SU}(2)$ (so E is a complex rank 2 vector bundle).

We now want to give a non-trivial example of a Yang–Mills connection, which we do on 4-dimensional Euclidean space.

Example. Let (x_0, x_1, x_2, x_3) be coordinates on $\mathbb{R}^4$. We can identify $\mathbb{R}^4$ with the quaternions $\mathbb{H}$ by

$$(x_0, x_1, x_2, x_3) \mapsto x = x_0 + ix_1 + jx_2 + kx_3,$$

where i, j, k are the imaginary units in $\mathbb{H}$ satisfying $ij = k$ etc. We may then define a family of connections A_c for $c > 0$ on $\mathbb{R}^4 = \mathbb{H}$ on a trivial vector bundle of rank 4 (with fibres identified with $\mathbb{H}$ on which $\mathrm{SU}(2) = \mathrm{Sp}(1)$ acts, viewed as multiplication by unit quaternions) by the formula:

$$A_c = \frac{c^2}{1 + c^2|x|^2}\,\mathrm{Im}(\bar{x}\mathrm{d}x),$$

where $\bar{x}$ denotes the quaternionic conjugate of x (i.e. $\bar{x} = x_0 - ix_1 - jx_2 - kx_3$ for x as above).

We can compute the curvature F_c of A_c as follows:

$$\begin{aligned} F_c &= \mathrm{d}A_c + A_c \wedge A_c \\ &= \frac{c^2}{(1+c^2|x|^2)^2}\mathrm{d}\bar{x} \wedge \mathrm{d}x \\ &= \frac{2}{(1+c^2|x|^2)^2}(i\omega_1 + j\omega_2 + k\omega_3) \end{aligned}$$

where

$$\omega_1 = \mathrm{d}x_0 \wedge \mathrm{d}x_1 - \mathrm{d}x_2 \wedge \mathrm{d}x_3, \quad \omega_2 = \mathrm{d}x_0 \wedge \mathrm{d}x_2 - \mathrm{d}x_3 \wedge \mathrm{d}x_1, \quad \omega_3 = \mathrm{d}x_0 \wedge \mathrm{d}x_3 - \mathrm{d}x_1 \wedge \mathrm{d}x_2,$$

which are the standard *anti-self-dual* 2-forms on $\mathbb{R}^4$ we have seen before.

We therefore see that, since $F_c = - * F_c$ that the Bianchi identity implies that

$$\mathrm{d} * F_c = -\mathrm{d}F_c = 0 \quad \text{and hence} \quad \mathrm{d}^* F_c = 0.$$

(We could also explicitly check this, of course.) This means that A_c is a family of Yang–Mills connections on $\mathbb{R}^4$.

We may also compute their Yang–Mills energy, using the fact that $|\omega_1|^2 = |\omega_2|^2 = |\omega_3|^2 = 2$:

$$\begin{aligned} \mathcal{E}(A_c) &= \int_{\mathbb{R}^4} |F_c|^2 \\ &= \int_{\mathbb{R}^4} \frac{48c^4}{(1+c^2|x|^2)^4} \\ &= 48\,\mathrm{Vol}(\mathcal{S}^3) \int_0^\infty \frac{r^3}{(1+r^2)^4}\mathrm{d}r \\ &= 4\,\mathrm{Vol}(\mathcal{S}^3) = 8\pi^2, \end{aligned}$$

which is *independent* of c! Notice that if we let $\sigma_c : x \mapsto cx$ then $A_c = \sigma_c^* A_1$, so that we can think of the 1-parameter family as really defined by A_1 up to rescaling.

Remark. We notice that as $c \to 0$ in A_c we obtain the trivial flat connection A_0. However, as $c \to \infty$ we see that away from $x = 0$ we again obtain a flat connection (since $|F_c|^2 \to 0$ as $c \to \infty$ for $x \neq 0$), but there is a *singularity* at $x = 0$. This singularity is *removable* in the sense that, we can define a new connection away from $x = 0$ which is just the flat connection. This is the model example of how one and take a limit of Yang–Mills connections and, although a singularity forms, it can be removed to still give you a Yang–Mills connection. We also see that, in the notation above, $\sigma_{c^{-1}}^* A_c = A_1$, so if $c \to \infty$ we can perform rescalings (with scale c^{-1} going to zero) to just see the fixed BPST instanton A_1 in the limit.

Example. Suppose we take a trivial complex line bundle over $\mathbb{R}^4$ with structure group U(1). Then a connection is just an imaginary 1-form iA and its curvature is an imaginary 2-form iF. (The fact that they are imaginary is that the Lie algebra of U(1) is $\mathfrak{u}(1) = i\mathbb{R}$.) If we split $\mathbb{R}^4 = \mathbb{R} \oplus \mathbb{R}^3$ with $t \in \mathbb{R}$ we can suggestively write

$$F = -\mathrm{d}t \wedge \mathbf{E} + *_{\mathbb{R}^3}\mathbf{B}$$

for 1-forms (on equivalently vector fields) $\mathbf{E}$ and $\mathbf{B}$ on $\mathbb{R}^3$ (which depend on t). We see that

$$\mathrm{d}F = \mathrm{d}t \wedge *_{\mathbb{R}^3}\left(\frac{\partial \mathbf{B}}{\partial t} + *\mathrm{d}\mathbf{E}\right) + \mathrm{d} *_{\mathbb{R}^3} \mathbf{B}.$$

We may also see that if we use the Minkowski metric on $\mathbb{R} \oplus \mathbb{R}^3$ then $*(iF)$ is identified with

$$\mathrm{d}t \wedge \mathbf{B} + *\mathbf{E}$$

and hence its exterior derivative is

$$\mathrm{d}t \wedge *(\frac{\partial \mathbf{E}}{\partial t} - *\mathrm{d}\mathbf{B}) + \mathrm{d} * \mathbf{E}.$$

We therefore see that the Bianchi identity and the Yang–Mills conditions on the connection are equivalent to

$$\mathrm{div}(\mathbf{B}) = 0, \qquad \frac{\partial \mathbf{B}}{\partial t} = -\operatorname{curl}(\mathbf{E})$$
$$\mathrm{div}(\mathbf{E}) = 0, \qquad \frac{\partial \mathbf{B}}{\partial t} = \operatorname{curl}(\mathbf{E}),$$

which are, of course, *Maxwell's equations* for the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ in the vacuum.

If we write

$$A = -\phi \mathrm{d}t + \mathbf{A}$$

for a 1-form (or vector field) $\mathbf{A}$ and function ϕ on $\mathbb{R}^3$ depending on t, then

$$F = \mathrm{d}t \wedge \left(\frac{\partial \mathbf{A}}{\partial t} + \mathrm{d}\phi\right) + \mathrm{d}\mathbf{A},$$

which says that

$$\mathbf{E} = -\operatorname{grad}\phi - \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \operatorname{curl}\mathbf{A},$$

which is the way to view the electromagnetic field in terms of potentials. We see that the Bianchi identity is automatically satisfied, and the gauge fixing condition $\mathrm{d}^*(iA) = 0$ becomes

$$\mathrm{div}(\mathbf{A}) = -\frac{\partial \phi}{\partial t},$$

so that Maxwell's equations (or the Yang–Mills condition) just becomes the *wave equation* on $\mathbf{A}$ and ϕ in this gauge. (If we had used the Euclidean metric, we would have instead arrived at *Laplace's equation*, as we might expect.)

6.3 Instantons in 4 dimensions

The examples of Yang–Mills connections we have seen are rather special because they satisfy *first order* PDEs (e.g. $F_A = 0$) on the connection that imply the second order PDE $\mathrm{d}_A^* F_A = 0$. So, how can we understand these special connections? Well, clearly, flat connections are absolute minimizers for the Yang–Mills functional. Can we find other first-order PDE that imply that the connections which satisfy it are absolute minimizers for $\mathcal{E}$? (Clearly such connections would be somewhat analogous to calibrated submanifolds.)

We are therefore led to make the following definition.

Definition 6.10. Let (M^4, g) be an oriented Riemannian 4-manifold. A connection A over M is an *ASD instanton* if its curvature satisfies:

$$F_A = -*F_A \quad \Leftrightarrow F_A^+ = 0,$$

where $F_A^{\pm} = \frac{1}{2}(F_A \pm *F_A)$ is the projection onto the anti-self-dual or self-dual 2-forms. Here, ASD stands for anti-self-dual, as one would expect. We can similarly define SD (or self-dual) instantons, but these can just be exchanged with ASD instantons by changing the orientation on M.

Remark. If one works in a suitable gauge, one can view the ASD instanton condition as an elliptic equation. In fact, if we consider the connection $A + ta$ where A is an ASD instanton and a is an $\mathrm{End}(E)$-valued 1-forms then

$$F_A^+ = (F_A + t\mathrm{d}_A a + t^2 a \wedge a)^+ = t\mathrm{d}_A^+ a + O(t^2).$$

Hence, combining this observation with the Coulomb gauge condition, we see that the linearization of the ASD instanton condition modulo gauge is

$$(\mathrm{d}_A^+ + \mathrm{d}_A^*)a = 0,$$

which is an elliptic system on $\mathrm{End}(E)$-valued 1-forms. We also see that the expected dimension of the moduli space of ASD instantons can be computed from the index of $\mathrm{d}_A^+ + \mathrm{d}_A^*$, which is a Dirac-type operator.

Example. The examples A_c that we saw before on $\mathbb{R}^4$ are ASD instantons, and A_1 is called the *BPST* instanton or *standard* instanton on $\mathbb{R}^4$.

Remark. We define the Yang–Mills connection A_c on $\mathbb{R}^4$, but since the energy is finite and the anti-self-duality condition $F_A = -*F_A$ is *conformally invariant*, it also defines a Yang–Mills connection on $\mathcal{S}^4$. In fact, it extends to a bundle E with structure group SU(2) and second Chern class $c_2(E) = 1$ over $\mathcal{S}^4$.

Example. If we took the trivial complex line bundle over $\mathbb{R}^4$ with structure group U(1), then we can view an ASD instanton A as simply a 1-form a on $\mathbb{R}^4$. The ASD condition $\mathrm{d}^+a = 0$ means that $\mathrm{d}a = -*\mathrm{d}a$, so differentiating both sides gives $\mathrm{d}^*\mathrm{d}a = 0$. Whenever integration by parts is valid (which is when F_A is L^2-integrable), one deduces that $\mathrm{d}a = 0$, which means that $F_A = 0$, i.e. A is flat.

We can now make some elementary but important observations.

Lemma 6.11. *ASD instantons are Yang–Mills.*

Proof. Since $F_A = -*F_A$, applying d_A and using the Bianchi identity gives

$$\mathrm{d}_A * F_A = -\mathrm{d}_A F_A = 0,$$

and thus $\mathrm{d}_A^* F_A = 0$. □

Our next result shows another similarity between ASD instantons and calibrated submanifolds.

Lemma 6.12. *If M^4 is compact, ASD instantons are absolute minimizers of the Yang–Mills energy.*

Proof. For any connection A, since anti-self-dual and self-dual 2-forms are orthogonal on M, we have that

$$\mathcal{E}(A) = \|F_A^+\|_{L^2}^2 + \|F_A^-\|_{L^2}^2,$$

where $F_A^{\pm}$ denotes the projection of F_A to the anti-self-dual or self-dual 2-forms depending on the sign.

We now consider

$$\kappa(E) = -\int_M \mathrm{tr}(F_A \wedge F_A),$$

where tr denotes the trace of the endomorphism part of $F_A \wedge F_A$, so that $\kappa(E) \in \mathbb{R}$. (It is important to note that the trace does not depend on any choice of basis for the fibres of E.) Chern–Weil theory implies that integrals of invariant polynomials in the curvature F_A will lead to topological quantities, i.e. only depending on the topology of the bundle E. In our setting, $\mathrm{tr}(F_A \wedge F_A)$ is invariant because of the invariance of the trace. Hence, $\kappa(E)$ depends only on the topology of E: it can be expressed in terms of the first and second Chern classes of E when it is a complex vector bundle.

However, using the fact that $F_A^+ \wedge F_A^- = 0$, we compute that

$$\begin{aligned}\kappa(E) = -\int_M \mathrm{tr}(F_A \wedge F_A) &= -\int_M \mathrm{tr}(F_A^+ \wedge F_A^+) - 2\int_M \mathrm{tr}(F_A^+ \wedge F_A^-) - \int_M \mathrm{tr}(F_A^- \wedge F_A^-) \\ &= -\int_M \mathrm{tr}(F_A^+ \wedge *F_A^+) + \int_M \mathrm{tr}(F_A^- \wedge *F_A^-) \\ &= -\|F_A^+\|_{L^2}^2 + \|F_A^-\|_{L^2}^2.\end{aligned}$$

Hence, we see that

$$\mathcal{E}(A) = 2\|F_A^+\|^2 + \kappa(E),$$

and thus ASD instantons (for which $F_A^+ = 0$) are the absolute minimizers of $\mathcal{E}(A)$. □

Remark. We see that the same proof shows that SD instantons are also absolute minimizers of the Yang–Mills energy, as we would expect. It also shows that for ASD instantons to exist on E we need $\kappa(E) \geq 0$, and that $\kappa(E) = 0$ implies that ASD instantons on E must be flat.

Remark. It is possible to find Yang–Mills connections on $\mathbb{R}^4$ which are *not* ASD instantons, but since they are not minimizers and one needs to solve a *second order* PDE, rather than a first order one, these are not so easy to find or describe.

ASD instantons are the backbone of *Donaldson theory* which defines invariants for smooth 4-manifolds. Though one needs the Riemannian metric to define ASD instantons, in good cases one can show that the invariant one defines are independent of the choice of generic Riemannian metric. Surprisingly, the invariants are able to detect to different smooth structures on the same underlying topological manifolds.

We are not going to get into the details of Donaldson theory, but a crucial point is that one needs to understand the *moduli space* of ASD instantons.

Example. Consider the moduli space $\mathcal{M}$ of ASD instantons on $\mathbb{R}^4$ modulo gauge, with L^2-integrable curvature, which are asymptotic to the trivial flat connection (modulo gauge) and with a given trivialization at infinity (known as a framing). Suppose that the bundle E also has structure group $\mathrm{SU}(l)$. Then one finds that the Yang–Mills energy of any ASD instanton A in $\mathcal{M}$ is

$$\mathcal{E}(A) = 8\pi^2 k$$

where $k \in \mathbb{N}$ is called the *charge* (or instanton number) of the ASD instanton. Notice that the Yang–Mills energy of the ASD instantons on $\mathbb{R}^4$ is "quantized" to be integer multiples of $8\pi^2$. Notice that the BPST instanton has charge $k = 1$.

One finds that the moduli space $\mathcal{M}$ has dimension $4kl$: this is a multiple of 4, so maybe there's some kind of quaternionic structure around? In fact, there is a natural *hyperkähler* metric on $\mathcal{M}$, which is just the L^2 metric.

One particular issue we need to understand the *non-compactness* of the moduli space of ASD instantons. We have already seen that non-compactness can occur, even in the simple case of $\mathbb{R}^4$ (or $\mathcal{S}^4$), so what can we do to deal with this? The answer is the following important result, which is due to the work primarily of Uhlenbeck.

Theorem 6.13. *Let (M^4, g) be a compact, oriented, Riemannian 4-manifold and let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ be a sequence of ASD instantons on the bundle E. After passing to a subsequence (which we still denote by (A_n)), there is*

- *a finite set of points $N = \{p_1, \dots, p_m\} \subseteq M$,*
- *an ASD instanton A_∞ on a bundle E_∞ on M and*
- *a sequence of isomorphisms $\rho_n : E_\infty|_{M\setminus N} \to E|_{M\setminus N}$*

such that

- *$\rho_n^* A_n \to A_\infty$ as $n \to \infty$,*
- *for each $p_j \in N$, if we take a sequence of suitable rescalings around p_j then A_n converges to an ASD instanton on $\mathbb{R}^4 = T_{p_j}M$, and*
- *$|F_{A_n}|^2 \to |F_{A_\infty}|^2 + 8\pi^2 \sum_{j=1}^m \delta_{p_j}$ as currents, so $\kappa(E) = \kappa(E_\infty) + 8\pi^2 m$.*

Remark. It is no coincidence that the $8\pi^2$ appears, which is the quantization of the energy of ASD instantons on $\mathbb{R}^4$ that we saw earlier. This theorem gives an example of the "bubbling phenomenon" that one sees ASD instantons at each point $p_j \in N$ "bubbling off" in the limit, but this information is retained in the drop in energy of A_∞ relative to A_n (which can also be measured purely in terms of the topology of the bundles E and E_∞). We also see the "removable singularity" phenomenon, that says that we get a limit A_∞ away from the points in N, but then A_∞ extends as an ASD instanton across N but on a possibly different bundle.

The theorem we have just described is crucial to developing the relevant compactness theory to develop Donaldson theory, and any other application of ASD instantons to geometry and topology.

6.4 Chern–Simons in 3 dimensions

We have so far seen two types of Yang–Mills connections: flat connections and ASD connections. In fact, the two are related as follows.

Suppose our Riemannian manifold (M^3, g) over which we are studying gauge theory is compact and 3-dimensional. (We would also typically take E to be a complex rank 2 vector bundle with structure group SU(2).) Pick a reference connection $A_0 \in \mathcal{A}$ and write $A \in \mathcal{A}$ as $A = A_0 + a$ as usual. On $[0,1] \times M$ we can pullback the bundle E and define a connection $\mathbf{A}$ on this pullback bundle F by

$$\mathbf{A} = A_0 + sa$$

where $s \in [0,1]$. We let the curvature of $\mathbf{F}$ and then make the following definition.

Definition 6.14. We define the *Chern–Simons functional* $\mathcal{F}$ on $\mathcal{A}$ (with respect to the reference A_0) by

$$\mathcal{F}(A) = \int_{[0,1]\times M} \operatorname{tr}(\mathbf{F} \wedge \mathbf{F}).$$

Equivalently,

$$\mathcal{F}(A) = \int_M \operatorname{tr}\left(a \wedge \left(2F_{A_0} + \mathrm{d}_{A_0} a + \frac{2}{3} a \wedge a\right)\right).$$

when $A = A_0 + a$.

To see the equivalence of the definitions we see that the curvature of $\mathbf{A}$ is given by

$$\begin{aligned}\mathbf{F} &= \mathrm{d}s \wedge \frac{\partial \mathbf{A}}{\partial s} + F_{A_0 + sa} \\ &= \mathrm{d}s \wedge a + F_{A_0} + s\mathrm{d}_{A_0} a + s^2 a \wedge a.\end{aligned}$$

Given this we just derived, we see that

$$\begin{aligned}\mathcal{F}(A) &= 2\int_0^1 \int_M \mathrm{d}s \wedge \operatorname{tr}\left(a \wedge \left(F_{A_0} + s\mathrm{d}_{A_0} a + s^2 a \wedge a\right)\right) \\ &= \int_M \operatorname{tr}\left(a \wedge \left(2F_{A_0} + \mathrm{d}_{A_0} a + \frac{2}{3} a \wedge a\right)\right)\end{aligned}$$

as claimed.

Example. If we work in a local trivialization and choose A_0 to be the trivial flat connection, then in that trivialization we have that the Chern–Simons integrand (also sometimes called the transgression form) is

$$\operatorname{tr}\left(a \wedge \mathrm{d}a + \frac{2}{3} a \wedge a \wedge a\right),$$

which may be a more familiar formula. Sometimes we can even choose A_0 globally to be trivial, which means this formula becomes global too.

We now consider the derivative of $\mathcal{F}$:

$$\begin{aligned}\mathcal{F}(A + t\dot{a}) &= \mathcal{F}(A) + t\int_M \operatorname{tr}(\dot{a} \wedge 2F_{A_0}) + t\int_M \operatorname{tr}(\dot{a} \wedge \mathrm{d}_{A_0} a) + t\int_M \operatorname{tr}(a \wedge \mathrm{d}_{A_0}\dot{a}) + 2t\int_M \operatorname{tr}(\dot{a} \wedge a \wedge a) \\ &\quad + O(t^2) \\ &= \mathcal{F}(A) + 2t\int_M \operatorname{tr}(\dot{a} \wedge F_A) + O(t^2),\end{aligned}$$

using the fact that

$$\begin{aligned}\int_M \operatorname{tr}(a \wedge \mathrm{d}_{A_0}\dot{a}) &= -\int_M \operatorname{tr}(\mathrm{d}_{A_0}(a \wedge \dot{a})) + \int_M \operatorname{tr}(\mathrm{d}_{A_0} a \wedge \dot{a}) \\ &= \int_M \operatorname{tr}(\dot{a} \wedge \mathrm{d}_{A_0} a)\end{aligned}$$

by Stokes Theorem. We deduce the following.

Lemma 6.15. *The differential of $\mathcal{F}$ at A is*

$$\mathrm{d}\mathcal{F}_A(\dot{a}) = 2\langle \dot{a}, *F_A\rangle_{L^2}.$$

Hence, $\mathrm{d}\mathcal{F}$ *is a well-defined closed 1-form on* $\mathcal{A}/\mathcal{G}$ *and the critical points of* $\mathcal{F}$ *(or zeros of* $\mathrm{d}\mathcal{F}$*) are flat connections.*

Proof. The fact that the critical points are flat connections is clear. The fact that $\mathrm{d}\mathcal{F}$ is closed is because it is (at least locally) the derivative of a functional: one can also check explicitly that the formula defines a closed 1-form. Finally, one sees that

$$\begin{aligned}\mathrm{d}\mathcal{F}_A(\mathrm{d}_A u) &= 2\int_M \mathrm{tr}(\mathrm{d}_A u \wedge F_A)\\ &= \int_M \mathrm{tr}(\mathrm{d}_A(u \wedge F_A)) = 0,\end{aligned}$$

using Stokes Theorem and the Bianchi identity $\mathrm{d}_A F_A = 0$. Thus, $\mathrm{d}\mathcal{F}_A$ vanishes in the directions tangent to the gauge orbits at A, and so descends to the quotient $\mathcal{A}/\mathcal{G}$. □

This lemma shows that $\mathcal{F}$ well-defined locally on $\mathcal{A}/\mathcal{G}$, but what about globally? The answer is that it is not well-defined globally, but only up to the *periods* of $\mathrm{d}\mathcal{F}$ which is related to $H_1(\mathcal{A}/\mathcal{G})$. In the case when E is an SU(2) bundle, one finds that $\mathcal{F}$ is defined up to integers (or, more accurately, integer multiples of $8\pi^2$), and so $\mathcal{F}$ can be viewed as an $\mathcal{S}^1$-valued function. One is then tempted to consider $\mathcal{F}$ as a Morse function, so one wants to think about its gradient flow lines. One quickly sees from the formula above that $A(t)$ is a negative gradient flow line for $\mathcal{F}$ if and only if

$$\frac{\partial}{\partial t}A(t) = - * F_{A(t)},$$

which is equivalent to saying that $\mathbf{A} = A(t)$ on $\mathbb{R} \times M$ has curvature $\mathbf{F}$ satisfying

$$\begin{aligned}\mathbf{F} &= \mathrm{d}t \wedge \frac{\partial}{\partial t}A(t) + F_{A(t)}\\ &= -\mathrm{d}t \wedge *F_{A(t)} + F_{A(t)}\\ &= - * \mathbf{F},\end{aligned}$$

i.e. $\mathbf{F}$ is an ASD instanton on $\mathbb{R}\times M$. It is this relation that is at the heart of Floer theory for 3-manifolds and Taubes's work on the Casson invariant.

Remark. It is interesting to notice that the negative gradient flow equation for $A(t)$ above is not even well-posed as an evolution equation on $A(t)$. Therefore, the fact that it can be interpreted as the elliptic equation (modulo gauge) given by the ASD instanton equation is crucial for the development of Floer theory in this setting.

6.5 Monopoles in 3 dimensions

We have seen that by studying flat connections on 3-manifolds using Chern–Simons theory, we naturally arrive at the ASD instanton equations in 4 dimensions. The question I want to now ask is: what happens if we instead start with the ASD instanton equations in 4 dimensions and try to go down to 3 dimensions? Do we just get flat connections or something else?

To this end, let us suppose that our 4-manifold is of the form $\mathbb{R} \times M^3$ for an oriented Riemannian 3-manifold (M, g) and we have the product metric on $\mathbb{R} \times M$. Suppose that our bundle E is the pullback of a bundle from M. Then, if $t \in \mathbb{R}$, we can write any connection on $\mathbb{R} \times M$ as

$$\mathbf{A} = \phi(t)\mathrm{d}t + A(t)$$

where $A(t)$ is a family of connections on M and $\phi(t)$ is a family of sections of $\mathrm{End}(E)$. Let us suppose that (A, ϕ) is *independent* of t and that $\mathbf{A}$ is an ASD instanton: this is the process of *dimensional reduction* of the ASD condition.

We see that the curvature $\mathbf{F}$ of $\mathbf{A}$ is:

$$\mathbf{F} = -\mathrm{d}t \wedge \mathrm{d}_A\phi + F_A,$$

so we see that $\mathbf{A}$ is an ASD instanton if and only if (A, ϕ) is a *monopole* in the following sense.

Definition 6.16. A *monopole* on an oriented Riemannian 3-manifold (M^3, g) is a pair (A, ϕ) of a connection A on a bundle E over M and a section ϕ of $\mathrm{End}(E)$ satisfying

$$F_A = *\mathrm{d}_A\phi.$$

These are sometimes called the *Bogomolny equations*. The section ϕ is often called a *Higgs field*.

So it seems that we might get more than just flat connections when we reduce the ASD instantons from 4 dimensions to 3. However, the next lemma shows that this is really not the case, at least for compact manifolds.

Lemma 6.17. *If M^3 is compact, then a monopole (A, ϕ) necessarily has $F_A = 0$, i.e. A is flat, and $\mathrm{d}_A\phi = 0$.*

Proof. If we differentiate both sides of the Bogomolny equations, we see that

$$\mathrm{d}_A * \mathrm{d}_A\phi = \mathrm{d}_A F_A = 0$$

by the Bianchi identity. Hence, our usual integration by parts trick shows that $\mathrm{d}_A\phi = 0$ and thus $F_A = 0$. □

Hence, for monopoles to be interesting, we have to go to non-compact manifolds or allow our Higgs fields to be singular. We will take the first approach and consider the case where $M = \mathbb{R}^3$ with the Euclidean metric.

Example. If we take E to be a complex line bundle over $\mathbb{R}^3 \setminus \{0\}$ with structure group $\mathrm{U}(1)$, then we can identify the connection A with a vector field $\mathbf{A}$ and ϕ with a function V satisfying

$$\mathrm{curl}\,\mathbf{A} = \mathrm{grad}\,V,$$

which is possibly familiar from the study of electromagnetism. Here, $\mathbf{B} = \mathrm{curl}\,\mathbf{A}$ is the magnetic field, and the fact that $\mathrm{curl} \circ \mathrm{grad} = 0$ means that $\mathrm{curl}\,\mathbf{B} = 0$ and so $\mathbf{B}$ is an isolated, radial magnetic field (and there is no electric field). This is the reason why monopoles are sometimes called *magnetic monopoles*.

Example. If we use spherical polar coordinates (r, θ, ψ) on $\mathbb{R}^3 \setminus \{0\}$, so $(r\sin\theta\cos\psi, r\sin\theta\sin\psi, r\cos\theta)$ are Euclidean coordinates, then an explicit monopole with $\mathrm{U}(1)$ gauge group is given for $k \in \mathbb{Z}$ by

$$A = \begin{cases} \frac{ik}{2}(1-\cos\theta)\mathrm{d}\psi, & \cos\theta \neq 1, \\ -\frac{ik}{2}(1+\cos\theta)\mathrm{d}\psi, & \cos\theta \neq -1, \end{cases} \quad \text{and} \quad \phi = i\left(m - \frac{k}{2r}\right),$$

using that we have a transition function of $e^{ik\psi}$ on our complex line bundle on the overlap region where $\cos\theta \notin \{\pm 1\}$, which is consistent with the fact that on the overlap region the difference in the definitions of A is $ik\mathrm{d}\psi$. Note that

$$F_A = \frac{ik}{2}\sin\theta\mathrm{d}\theta \wedge \mathrm{d}\psi = \frac{ik}{2r^2}\,\mathrm{vol}_{\mathbb{R}^3}$$

and that ϕ is indeed a harmonic function.

This is the *Dirac monopole* with (signed) charge k, which we note is *singular* as we approach 0, as we would expect.

Example. We now want to build a monopole with gauge group $\mathrm{SU}(2)$ on a trivial rank 4 vector bundle on $\mathbb{R}^3$. We identify $\mathbb{R}^3 = \mathrm{Im}\,\mathbb{H}$, the imaginary quaternions as follows:

$$(x_1, x_2, x_3) \mapsto x = x_1 i + x_2 j + x_3 k.$$

We then define a connection and Higgs field by

$$A = a(r)\,\mathrm{Im}(\bar{x}\mathrm{d}x) \quad \text{and} \quad \phi = b(r)x$$

for functions a, b depending only on $r = |x|$. The monopole equation then becomes ODEs on the functions a, b, which one can solve explicitly. Up to possible sign mistakes and constant factors, the smooth solutions are (A_m, ϕ_m) for $m > 0$ given by:

$$A_m = \left(\frac{m}{\sinh(mr)} - \frac{1}{r}\right)\frac{\mathrm{Im}(\bar{x}\mathrm{d}x)}{r} \quad \text{and} \quad \phi_m = \left(\frac{m}{\tanh(mr)} - \frac{1}{r}\right)\frac{x}{r}.$$

This called the *BPS monopole* on $\mathbb{R}^3$ with *mass* $m > 0$. Notice that, although it looks singular at $r = 0$, one sees that it is in fact well-defined as $r \to 0$, and has that $\phi_m(0) = 0$.

Example. If we consider monopoles on $\mathbb{R}^3$ with gauge group SU(2) so that $|\phi|$ has a finite limit at infinity, then (under a few more technical assumptions, such as a framing at infinity) one has that

$$|\phi| = m - \frac{k}{2r} + o(\frac{1}{r}),$$

where $k \in \mathbb{N}$ is called the *charge* of the monopole and $m > 0$ is called the *mass*.

The moduli space of such framed monopoles with fixed mass has dimension $4k$: this is again a multiple of 4! So, we can ask: is there a quaternionic structure? The answer is yes: the L^2 metric on the moduli space is *hyperkähler*.

Example. Since translations of $\mathbb{R}^3$ act on monopoles, we consider taking the quotient by this action and removing the framing condition to obtain the moduli space of *centred* monopoles on $\mathbb{R}^3$ of fixed mass $m > 0$. If we again choose gauge group SU(2), this moduli space now has dimension $4(k-1)$ for charge k. In the case $k = 1$ we obtain a point, which is the BPS monopole we saw above. For $k = 2$ we obtain a hyperkähler 4-manifold, which turns out to be SU(2)-invariant and is called the *Atiyah–Hitchin* manifold.

Remark. We see that monopoles are critical points for the *Yang–Mills–Higgs functional*

$$\mathcal{E}(A, \phi) = \int_M |F_A|^2 + |\mathrm{d}_A\phi|^2,$$

whenever this is well-defined.

7 Gauge theory and Kähler geometry

We now would like to start to discuss gauge theory in higher dimensions, and the easiest setting to do so, just as in the case of calibrated geometry, is when the ambient manifold is Kähler. We recall that means that we have a $2n$-dimensional manifold M with a metric g with holonomy contained in $\mathrm{U}(n)$, so there is a complex structure J on M and a Kähler form ω, which is a parallel non-degenerate 2-form on M.

7.1 Hermitian connections and holomorphic bundles

We will now restrict the bundles and connections we consider as follows.

Definition 7.1. Let E be a complex vector bundle of rank m with a Hermitian metric h on the fibres of E. We say that a connection A on E is *Hermitian* (or *unitary*) if it is compatible with h, in the sense that

$$\mathrm{d}(h(u, v)) = h(\mathrm{d}_A u, v) + h(u, \mathrm{d}_A v)$$

for all sections u, v of E. The curvature F_A of a Hermitian connection is skew-Hermitian, since $\mathrm{d}^2 = 0$ means that

$$0 = \mathrm{d}^2(h(u, v)) = h(F_A u, v) + h(u, F_A v),$$

for any sections u, v. In each fibre, F_A takes values in $\mathfrak{u}(m)$, the Lie algebra of $\mathrm{U}(m)$.

We now have some basic algebraic facts that will be useful when we look at the curvature of our connections.

We recall that on $\mathbb{C}^n$ with complex coordinates $(z_1,\ldots,z_n)$ we can decompose the complex k-forms $\Lambda^k(\mathbb{C}^n)^*$ on $\mathbb{C}^n$ into *types*:

$$\Lambda^k(\mathbb{C}^n)^* = \oplus_{p+q=k}\Lambda^{(p,q)}(\mathbb{C}^n)^*,$$

where $\Lambda^{(p,q)}(\mathbb{C}^n)^*$ has basis given by

$$\{\mathrm{d}z_{i_1}\wedge\ldots\wedge\mathrm{d}z_{i_p}\wedge\mathrm{d}\bar{z}_{j_1}\ldots\mathrm{d}\bar{z}_{j_q} : 1\leq i_1,\ldots,i_p,j_1,\ldots,j_q\leq n\}.$$

We are only concerned with $k=1$ and $k=2$ which give

$$\Lambda^1(\mathbb{C}^n)^* = \Lambda^{(1,0)}(\mathbb{C}^n)^*\oplus\Lambda^{(0,1)}(\mathbb{C}^n)^* \quad\text{and}\quad \Lambda^2(\mathbb{C}^n)^* = \Lambda^{(2,0)}(\mathbb{C}^n)^*\oplus\Lambda^{(1,1)}(\mathbb{C}^n)^*\oplus\Lambda^{(0,2)}(\mathbb{C}^n)^*.$$

This discussion clearly generalizes to any complex manifold, and hence to our Kähler manifold M.

Example. The Kähler form ω is a real $(1,1)$-form.

Example. Consider the curvature F_A of a Hermitian connection A, which we know takes values in skew-Hermitian matrices at each point. It follows from this that the $(0,2)$ part $F_A^{(0,2)}$ of F_A determines the $(2,0)$ part of F_A.

The type decomposition allows us to decompose the exterior derivative as

$$\mathrm{d} = \partial + \bar{\partial}$$

where

$$\partial : \Omega^{(p,q)}(M)\to\Omega^{(p+1,q)}(M) \quad\text{and}\quad \bar{\partial} : \Omega^{(p,q)}(M)\to\Omega^{(p,q+1)}(M).$$

Example. On functions on $\mathbb{C}$, $\bar{\partial}$ is just the Cauchy–Riemann operator, i.e. $\bar{\partial}f=0$ if and only if f is holomorphic.

These considerations allow us to discuss holomorphic vector bundles and Cauchy–Riemann operators.

Definition 7.2. Let E be a complex vector bundle over a complex manifold M.

We say that E is *holomorphic* if it admits a complex structure so that projection from E to M is holomorphic and is locally holomorphically trivial.

We can consider differential operators

$$\bar{\partial} : \Gamma(E)\to\Omega^{(0,1)}(M)\otimes\Gamma(E)$$

on which are $\mathbb{C}$-linear and satisfy the Leibniz rule: these are *Dolbeault operators.* We say that $\bar{\partial}$ is a *Cauchy–Riemann operator* on E if $\bar{\partial}^2=0$.

It turns out that E is holomorphic if and only if it admits a Cauchy–Riemann operator $\bar{\partial}$, which is why $\bar{\partial}$ is often called a *holomorphic structure* on E. We often write a holomorphic vector bundle with the choice of holomorphic structure as $(E,\bar{\partial})$.

Remark. If we are given any Hermitian connection A we can take the projection of d_A to the $(0,1)$ part to obtain a Dolbeault operator $\mathrm{d}_A^{(0,1)} : \Gamma(E)\to\Omega^{(0,1)}(M)\otimes\Gamma(E)$. However, we would not necessarily have that $\mathrm{d}_A^{(0,1)}$ squares to zero, so it would not necessarily be a Cauchy–Riemann operator.

The previous remark suggests that there should be a condition on Hermitian connections which guarantees that they induce a Cauchy–Riemann operator, or equivalently a holomorphic structure, on E. This is indeed the case as we now see.

Proposition 7.3. *Let A be a Hermitian connection on E satisfying $F_A^{(0,2)}=0$. Then there is a unique holomorphic structure $\bar{\partial}$ on E such that $\mathrm{d}_A^{(0,1)}=\bar{\partial}$.*

Conversely, if E admits a holomorphic structure $\bar{\partial}$, then there is a unique Hermitian connection A on E such that $\mathrm{d}_A^{(0,1)}=\bar{\partial}$ and $F_A^{(0,2)}=0$. This connection is called the Chern connection of $(E,\bar{\partial})$.

This result suggests that, since we are primarily interested in holomorphic bundles, we should restrict attention to Hermitian connections satisfying $F_A^{(0,2)}=0$. Recall that such connections would automatically solve $F_A^{(2,0)}=0$ as well.

7.2 Hermitian Yang–Mills connections

We are now in the position to start looking at the gauge theory we are interested in, but to motivate it I would like to do a little bit more algebra of complex forms.

Consider the following map on complex 2-forms α on $\mathbb{C}^n$:

$$L(\alpha) = *\left(\alpha \wedge \frac{\omega_0^{n-2}}{(n-2)!}\right),$$

where we have to take care about how we extend the Hodge star to complex forms. We see that L takes 2-forms to 2-forms and, if we choose the right extension of the Hodge star, it preserves type decomposition, and so it is naturally to ask what the eigenvalues and eigenspaces of L are. We see that L is self-adjoint, so it has real eigenvalues and must decompose the complex 2-forms on $\mathbb{C}^n$ into eigenspaces.

Proposition 7.4. *The complex 2-forms on $\mathbb{C}^n$ decompose into the following eigenspaces of L:*

$$\begin{aligned}\{\alpha : L(\alpha) = (n-1)\alpha\} &= \mathrm{Span}\{\omega_0\},\\ \{\alpha : L(\alpha) = \alpha\} &= \Lambda^{(2,0)}(\mathbb{C}^n)^* \oplus \Lambda^{(0,2)}(\mathbb{C}^n)^*,\\ \{\alpha : L(\alpha) = -\alpha\} &= \Lambda_0^{(1,1)}(\mathbb{C}^n)^* \cong \mathfrak{su}(n)\otimes\mathbb{C}\\ &= \{\alpha \in \Lambda^{(1,1)}(\mathbb{C}^n)^* \,:\, \alpha\wedge\omega_0^{n-1} = 0\} = \{\alpha \in \Lambda^{(1,1)}(\mathbb{C}^n)^* \,:\, \Lambda_{\omega_0}\alpha = 0\},\end{aligned}$$

where Λ_{ω_0} is the adjoint of wedge product with $\frac{\omega_0^{n-1}}{(n-1)!}$.

Proof. This calculation is easy once one observes that L is invariant under the action of $\mathrm{U}(n)$, so by Schur's lemma the eigenspaces will correspond to irreducible representations of $\mathrm{U}(n)$.

We see that $\Lambda^{(2,0)}(\mathbb{C}^n)^*$ is an irreducible representation of $\mathrm{U}(n)$ so to find the eigenvalue of L on this space we need only pick one element. It is straightforward to compute that

$$L(\mathrm{d}z_1\wedge \mathrm{d}z_2) = \mathrm{d}z_1\wedge\mathrm{d}z_2,$$

from which it follows that L has eigenvalue 1 on this space. The same also holds on $\Lambda^{(0,2)}(\mathbb{C}^n)^*$.

We clearly have that ω_0 lies in $\Lambda^{(1,1)}(\mathbb{C}^n)^*$ and $\mathrm{U}(n)$ acts trivially on its span, so its span corresponds to the trivial representation. We easily compute that

$$L(\omega_0) = *\frac{\omega_0^{n-1}}{(n-2)!} = (n-1)*\frac{\omega_0^{n-1}}{(n-1)!} = (n-1)\omega_0$$

as we wanted.

We see that the complex dimensions of the following spaces are

$$\dim\Lambda^2(\mathbb{C}^n)^* = \binom{2n}{2} = n(2n-1) \quad\text{and}\quad \dim\Lambda^{(2,0)}(\mathbb{C}^n)^* = \dim\Lambda^{(0,2)}(\mathbb{C}^n)^* = \binom{n}{2} = \frac{n(n-1)}{2}.$$

This means that

$$\dim\Lambda_0^{(1,1)}(\mathbb{C}^n)^* = n^2-1.$$

Since the complexification of $\mathfrak{su}(n)$ must lie in $\Lambda^2(\mathbb{C}^n)^*$, by dimension counting and matching, we see that this must be isomorphic to $\Lambda_0^{(1,1)}(\mathbb{C}^n)^*$. We see that $\mathrm{U}(n)$ acts irreducibly on $\mathfrak{su}(n)$, which means that again we can restrict to a single element. For example, one may easily see that

$$\alpha = \frac{i}{2}(\mathrm{d}z_1\wedge\mathrm{d}\bar{z}_1 - \mathrm{d}z_2\wedge\mathrm{d}\bar{z}_2) \in \Lambda_0^{(1,1)}(\mathbb{C}^n)^*$$

and that $L(\alpha) = -\alpha$ as claimed. ☐

Remark. Notice that the trace of the map L is 0.

Example. Note that if $\alpha = f\omega_0$ then

$$\alpha\wedge\frac{\omega_0^{n-1}}{(n-1)!} = nf\frac{\omega_0^n}{n!} = nf\,\mathrm{vol}_{\mathbb{C}^n},$$

so we see that

$$\Lambda_{\omega_0}\alpha = nf.$$

Thus, if $\alpha \in \Lambda^{(1,1)}(\mathbb{C}^n)^*$, we can decompose it as

$$\alpha = \frac{\Lambda_{\omega_0}\alpha}{n}\omega_0 + \alpha_0$$

where $\alpha_0 \in \Lambda_0^{(1,1)}(\mathbb{C}^n)^*$. We can view $\Lambda_{\omega_0}\alpha$ as the "trace" of α with respect to ω_0.

Hence, if $\alpha \in \Lambda^{(1,1)}(\mathbb{C}^n)^*$ we see that

$$\begin{aligned} L(\alpha) &= \frac{(n-1)\Lambda_{\omega_0}\alpha}{n}\omega_0 - \alpha_0 \\ &= (\Lambda_{\omega_0}\alpha)\omega_0 - \alpha. \end{aligned}$$

We will see the utility of this formula shortly.

We now want to consider the interaction between the Hermitian condition and the Yang–Mills condition on a connection A on a Kähler manifold M. As we indicated earlier, we want to consider holomorphic bundles over M, and so we impose the condition $F_A^{(0,2)} = 0$. In this case, we see from the example above that

$$F_A \wedge \frac{\omega^{n-2}}{(n-2)!} = (\Lambda_\omega F_A)\frac{\omega^{n-1}}{(n-1)!} - *F_A.$$

We can differentiate the left-hand side and see that since $\mathrm{d}\omega = 0$ and $\mathrm{d}_A F_A = 0$ by the Bianchi identity, we have that

$$\mathrm{d}_A\left(F_A \wedge \frac{\omega^{n-2}}{(n-2)!}\right) = 0.$$

Therefore, again using $\mathrm{d}\omega = 0$, we know that

$$\mathrm{d}_A * F_A = \mathrm{d}_A(\Lambda_\omega F_A)\frac{\omega^{n-1}}{(n-1)!}.$$

We deduce the following, where id_E denotes the identity endomorphism on E, and recalling that F_A is skew-Hermitian. For the statement, we recall that a connection A is *irreducible* if it does not preserve any proper subbundle of E.

Proposition 7.5. *An irreducible Hermitian connection A with $F_A^{(0,2)} = 0$ is Yang–Mills if and only if $\Lambda_\omega F_A = i\lambda\,\mathrm{id}_E$ for a real constant λ.*

This motivates the following definition.

Definition 7.6. A Hermitian connection A on E is *Hermitian Yang–Mills* (HYM) if

$$F_A^{(0,2)} = 0 \quad \text{and} \quad \Lambda_\omega F_A = i\lambda\,\mathrm{id}_E$$

for a real constant λ. Note that the argument leading up to Proposition 7.5 shows that HYM connections are Yang–Mills, so the name is not too bad!

Example. Let M be a Kähler–Einstein manifold, i.e. the Ricci curvature of the Kähler metric g is a constant multiple of g:

$$\mathrm{Ric}(g) = \kappa g.$$

Then one sees that if we take A to be the Levi–Civita connection on TM then A is Hermitian and

$$F_A = i\kappa\omega \otimes \mathrm{id}_{TM} + (F_A)_0^{(1,1)}.$$

Hence, A is Hermitian Yang–Mills.

We now observe that we are actually not free to choose the constant λ in the definition of Hermitian Yang–Mills connections, if M is compact. In this case we see that, by Chern–Weil theory, that

$$\frac{i}{2\pi}\int_M \mathrm{tr}(F_A \wedge \omega^{n-1}) = \int_M c_1(E) \wedge \omega^{n-1} := \deg(E) \in \mathbb{Z},$$

the degree of E, which only depends on the first Chern class $c_1(E)$ and the cohomology class $[\omega]$ of ω. On the other hand, if A is Hermitian Yang–Mills then

$$\begin{aligned}\frac{i}{2\pi}\operatorname{tr}(F_A\wedge\omega^{n-1}) &= -\frac{1}{2\pi n}\operatorname{tr}(\lambda\omega^n\otimes\mathrm{id}_E)\\ &= -\frac{\lambda\operatorname{rank}(E)}{2\pi n}\omega^n.\end{aligned}$$

Integrating over M we deduce that

$$\deg(E) = -\frac{(n-1)!\lambda\operatorname{rank}(E)\operatorname{Vol}(M)}{2\pi},$$

since $\omega^n/n!$ is the volume form on M.

This leads us to make the following definition.

Definition 7.7. We define the *slope* $\mu(E)$ of the bundle E over M as

$$\mu(E) = \frac{\deg(E)}{\operatorname{rank}(E)}.$$

We see that if M is compact, then a Hermitian Yang–Mills connection with constant λ can only exist on E if

$$\lambda = \lambda(E) := -\frac{2\pi\mu(E)}{(n-1)!\operatorname{Vol}(M)},$$

which is determined by the topology of E as the volume of M is fixed.

Remark. Complex line bundles are classified by their degree. In particular, a complex line bundle has degree zero if and only if it is trivial.

For a complex vector bundle of rank at least 2, the degree can be zero and the bundle can still be topologically interesting. For example, an important subclass of bundles E are those whose structure group is a special unitary group $\mathrm{SU}(m)$, who then have $c_1(E) = 0$ automatically and thus always have $\deg(E) = 0$.

Our remark shows that $\deg(E) = 0$ can be interesting and in this setting we can rephrase the Hermitian Yang–Mills condition in a neat way.

Proposition 7.8. *Suppose that E over the compact Kähler manifold M of complex dimension n has $\deg(E) = 0$. Then a Hermitian connection A on E is Hermitian Yang–Mills if and only if*

$$F_A \wedge \frac{\omega^{n-2}}{(n-2)!} = -*F_A,$$

which is equivalent to saying that F_A is of type $\Lambda_0^{(1,1)}$.

Proof. This results follows from two observations. The first is that F_A lies in $\Lambda_0^{(1,1)}$ if and only $L(F_A) = -F_A$ by Proposition 7.4. The second is that since $\deg(E) = 0$ the Hermitian Yang–Mills equations reduce to $F_A^{(0,2)} = 0$ and $\Lambda_\omega F_A = 0$. Since we observed that F_A is skew-Hermitian, we know that $F_A^{(0,2)} = 0$ if and only if $F^{(2,0)} = 0$ as well, the result follows. □

This looks suspiciously similar to the ASD condition, and it can be seen as a generalization to higher dimensions in the following sense.

Example. Let M be a compact Kähler surface (i.e. with complex dimension $n = 2$) and let E have $\deg(E) = 0$ with a Hermitian connection A. Then A is an HYM connection if and only if it is an ASD instanton.

Given this observation we can also ask whether HYM connections minimize the Yang–Mills energy in the $\deg(E) = 0$ setting. The answer is as expected.

Proposition 7.9. *Let A be a Hermitian Yang–Mills connection on E with $\deg(E) = 0$ on a compact Kähler manifold M. Then A is an absolute minimizer of the Yang–Mills energy on E.*

Proof. We first see that, since the decomposition of the 2-forms into eigenspaces of L is orthogonal and $(2,0)$ and $(0,2)$ forms are orthogonal, we have for any Hermitian connection A:

$$\mathcal{E}(A) = \frac{1}{n}\|\Lambda_\omega F_A\|^2_{L^2} + \|F_A^{(2,0)}\|^2_{L^2} + \|F_A^{(0,2)}\|^2_{L^2} + \|(F_A)_0^{(1,1)}\|^2_{L^2},$$

using the obvious notation for $(F_A)_0^{(1,1)}$ as the part of F_A of type $\Lambda_0^{(1,1)}$. We can also see that

$$\kappa(E) = -\int_M \operatorname{tr}\left(F_A \wedge F_A \wedge \frac{\omega^{n-2}}{(n-2)!}\right)$$

is purely topological by Chern–Weil theory and we can compute:

$$= -\frac{n-1}{n}\|\Lambda_\omega F_A\|^2_{L^2} - \|F_A^{(2,0)}\|^2_{L^2} - \|F_A^{(0,2)}\|^2_{L^2} + \|(F_A)_0^{(1,1)}\|^2_{L^2},$$

using Proposition 7.4. We deduce that

$$\mathcal{E}(A) = \|\Lambda_\omega F_A\|^2_{L^2} + 2\|F_A^{(2,0)}\|^2_{L^2} + 2\|F_A^{(0,2)}\|^2_{L^2} + \kappa(E).$$

Since we have assumed $\deg(E) = 0$, we deduce that HYM connections are indeed absolute minimizers of the Yang–Mills energy $\mathcal{E}$. □

Remark. Notice that HYM connections on degree 0 bundles minimize the Yang–Mills energy amongst *all* connections on E, not just the Hermitian ones.

7.3 Stability

We now come to one of the cornerstones in the study of Hermitian Yang–Mills connection, which is that we can determined when such a connection exists purely in terms of *complex algebraic geometry.* This is surprising because the Hermitian Yang–Mills condition is a nonlinear partial differential equation, and so seems to be very much in the realm *differential geometry.* The theory of HYM connections is now one of the paradigms showing how we can have an equivalence between algebro-geometric and differential geometric problems.

Remark. A similar, but more complicated, version of this relation between algebraic and differential geometry comes in the study of Kähler–Einstein metrics.

The key notion we need to introduce is *stability* for a holomorphic vector bundle.

Definition 7.10. A holomorphic bundle E on compact M is *stable* if and only if for all proper, non-zero, coherent subsheafs E' of E we have that the slopes satisfy

$$\mu(E') < \mu(E).$$

(A coherent sheaf is a weaker notion than holomorphic vector bundle for which things like first Chern class and rank, and thus μ, make sense.) We say that E is *polystable* if it is a direct sum of stable bundles, possibly with the same slope.

Informally, a bundle is stable if it cannot be broken up into smaller pieces for which the slope μ is larger.

The fundamental result in the field is the following theorem, often called the Donaldson–Uhlenbeck–Yau Theorem because it was first proved by Donaldson for Kähler surfaces and then proved in general by Uhlenbeck–Yau. To make the statement cleaner we restrict attention to *irreducible* connections on E, meaning ones which do not preserve any proper subbundles of E.

Theorem 7.11. *A holomorphic bundle over a compact Kähler manifold admits an irreducible Hermitian Yang–Mills connection if and only if it is stable.*

Remark. It we do not restrict to irreducible connections, then we have to allow the bundle to be polystable, as should be clear.

Remark. You may be wondering: where does the stability condition come from? Unfortunately this will take me too long to explain, but it is from geometric invariant theory (GIT). The idea is that we can complexify the action of the gauge transformations and look at the orbits of this complexified group action. We can view the Hermitian Yang–Mills connections as critical points for the norm squared of the moment map for the gauge group action, so we are motivated by the classical Kempf–Ness theorem: that the norm will have a minimum on the complexified orbit if and only if the orbit is closed. This is where stability comes from: it enables us to determine when the orbit will be closed (and this also shows the links to coherent sheaves, because we need to look at "limits" of holomorphic bundles). Thus the Donaldson–Uhlenbeck–Yau theorem can be seen as an infinite-dimensional analogue of the Kempf–Ness theorem.

Example. To give some feeling for the stability condition, consider a proper holomorphic subbundle $(E', \bar\partial')$ of a holomorphic bundle $(E, \bar\partial)$. Suppose that $(E, \bar\partial)$ admits a Hermitian metric so that the Chern connection A is Hermitian Yang–Mills. (This can always be done if E admits an HYM connection.) We can then restrict this metric to E' and get a Chern connection A' on E'.

Since E' is a subbundle, we have an orthogonal decomposition $E = E' \oplus F$ and with respect to this splitting we can then write

$$\bar\partial = \begin{pmatrix} \bar\partial' & \beta \\ 0 & \bar\partial_F \end{pmatrix},$$

where β is called the *second fundamental form* of E'.

If we let π' denote the endomorphism of E which projects orthogonally to E', we have the following equation, which we can think of like the Gauss equation from the study of Riemannian submanifolds:

$$F_{A'} = \pi' \circ F_A \circ \pi' + \beta \wedge \beta^*.$$

We then note that, since A is HYM, we have that

$$\Lambda_\omega(\pi' \circ F_A \circ \pi') = i\lambda(E)\,\mathrm{id}_{E'}$$

and thus

$$\begin{aligned} \mathrm{tr}(\pi' \circ F_A \circ \pi' \wedge \omega^{n-1}) &= \frac{i\lambda(E)}{n}\,\mathrm{tr}(\mathrm{id}_{E'})\omega^n \\ &= -\frac{2\pi i\mu(E)\,\mathrm{rank}(E')}{\mathrm{Vol}(M)}\frac{\omega^n}{n!}. \end{aligned}$$

We also see that

$$i\,\mathrm{tr}\left(\Lambda_\omega(\beta \wedge \beta^*)\right) \le 0$$

with equality only if the connection A preserves E'.

Hence, we see that

$$\begin{aligned} \mu(E') &= \frac{\deg(E')}{\mathrm{rank}(E')} \\ &= \frac{i}{2\pi\,\mathrm{rank}(E')}\int_M \mathrm{tr}(F_{A'} \wedge \omega^{n-1}) \\ &= \frac{i}{2\pi\,\mathrm{rank}(E')}\int_M \mathrm{tr}(\pi' \circ F_A \circ \pi' \wedge \omega^{n-1}) + \frac{i}{2\pi n\,\mathrm{rank}(E')}\int_M \mathrm{tr}\,\Lambda_\omega(\beta \wedge \beta^*)\omega^n \\ &\le \frac{\mu(E)}{\mathrm{Vol}(M)}\int_M \frac{\omega^n}{n!} = \mu(E), \end{aligned}$$

and the equality must be strict if A is irreducible.

This does not quite show holomorphic bundles which admit HYM connections are polystable because we need to worry about the possibility that E' is only a proper coherent subsheaf of E, but it is a good step in that direction.

The previous example gives the key argument at the heart of showing the "easy" direction in the Donaldson–Uhlenbeck–Yau theorem, namely that holomorphic bundles admitting an irreducible HYM connections are stable. To prove the other direction is "hard" and involves some sophisticated analysis.

The analytic approach that works in generality is due to Uhlenbeck–Yau and is an example of the *continuity method*. This method works schematically as follows.

We let $(*_t)$ denote a family of equations for $t \in [0,1]$ so that $(*_1)$ is the equation we actually want to solve (in our case, this is the HYM condition) and so that $(*_0)$ is easy to solve. The strategy then is to introduce the set

$$\mathcal{S} = \{t \in [0,1] \,:\, (*_t) \text{ has a solution}\}$$

and show that $\mathcal{S}$ is:

- non-empty,
- open and
- closed.

If we show this, then $\mathcal{S} = [0,1]$ as $[0,1]$ is connected so, in particular, $(*_1)$ has a solution, which is what we wanted.

Let me comment on each of the steps.

- $\mathcal{S} \neq \emptyset$: this is trivial or easy because of our choice that $(*_0)$ is easy to solve. Thus, $0 \in \mathcal{S}$.
- $\mathcal{S}$ is open: this typically amounts to an Implicit Function Theorem type argument. We suppose that $t_0 \in \mathcal{S}$, so $(*_{t_0})$ has a solution. We then linearize $(*_{t_0})$ and aruge that that the linearization has an appropriate surjectivity property, which then shows that we can solve $(*_t)$ for all t sufficiently near t_0.
- $\mathcal{S}$ is closed: this is usually the hardest part. The challenge is that one needs to worry about solutions to a sequence of equations $(*_{t_j})$ where $t_j \to t_0$. The difficulty is that we have to argue that we can take an appropriate limit of our solutions to get a solution to $(*_{t_0})$. It is at this step that one can encounter *singular* objects, since we are taking limits.

In our setting, the final step gives further credence that only may need to consider more singular objects than subbundles.

One can say a lot more about the theory of Hermitian Yang–Mills connections but we shall leave it there for now in these notes. Suffice it to say, they remain an active area of study and the links between algebraic and differential geometry in this context continue to be explored.

We shall see the HYM connections re-appear in the context of special holonomy in the next section.

8 Gauge theory and special holonomy

Whilst the theory of Hermitian Yang–Mills connections is very interesting, it has a rather different character to the low-dimensional gauge theory we saw earlier. We therefore now turn to the proposal for *gauge theory in higher dimensions* due to Donaldson–Thomas, and subsequently expanded by Donaldson–Segal. The idea here is to try to emulate the theory we saw when studying ASD instantons, Chern–Simons theory and monopoles. To do so, we shall need to use extra structure on our ambient manifold, namely *special holonomy*.

We continue to assume that E is a vector bundle of rank m over our manifold M with Euclidean metrics on the fibres as usual.

8.1 Gauge theory on Calabi–Yau manifolds

We begin our discussion by returning to the Kähler setting, but we now additionally assume that M is Calabi–Yau and of complex dimension at least 3. Recall that, if M has real dimension $2n$ (which we are assuming is at least 6) then we have a metric g on M with $\mathrm{Hol}(g) \subseteq \mathrm{SU}(n)$, a Kähler form ω and a holomorphic volume form Υ.

We start with a simple but important observation.

Lemma 8.1. *Let α be a 2-form on $\mathbb{C}^n$ for $n \geq 3$, and let ω_0 and Υ_0 be the standard Kähler form and holomorphic volume form on $\mathbb{C}^n$.*

Then,

$$\alpha \wedge \operatorname{Re}\Upsilon_0 = 0 \quad \Leftrightarrow \quad \alpha \wedge \operatorname{Im}\Upsilon_0 = 0 \quad \Leftrightarrow \quad \alpha^{(2,0)+(0,2)} = 0,$$

where $\alpha^{(2,0)+(0,2)}$ denotes the projections to the forms of type $\Lambda^{(2,0)+(0,2)}$.

Proof. Note that $\operatorname{Re}\Upsilon_0$, $\operatorname{Im}\Upsilon_0$ are of type $\Lambda^{(3,0)+(0,3)}$. We therefore see that the wedge product of α with either of these forms is of type $\Lambda^{(3,2)+(2,3)}$ and is zero if and only if $\alpha^{(2,0)+(0,2)}$ vanishes, using the fact that Υ_0 is nowhere vanishing. □

Proposition 8.2. *Let A be a connection on E over the Calabi–Yau M with Kähler form ω and holomorphic volume form Υ. Then the following are equivalent:*

- $F_A \wedge \dfrac{\omega^{(n-2)}}{(n-2)!} = - * F_A$;
- $F_A \wedge \operatorname{Im}\Upsilon = 0$ *and* $F_A \wedge \omega^{n-1} = 0$;
- $F_A \wedge \operatorname{Re}\Upsilon = 0$ *and* $F_A \wedge \omega^{n-1} = 0$;
- $F_A^{(2,0)+(0,2)} = 0$ *and* $\Lambda_\omega F_A = 0$;
- F_A *is of type* $\Lambda^{(1,1)}_0 \cong \mathfrak{su}(n)$.

Proof. The equivalence of the middle lines follows from the previous lemma and the definition of Λ_ω as the adjoint of wedge product with ω^{n-1}. The equivalence of the first and last two lines follows from Proposition 7.4. □

This leads us to the following definition.

Definition 8.3. We say that a connection A on a Calabi–Yau n-fold M is an $\mathrm{SU}(n)$ *instanton* if

$$F_A \wedge \frac{\omega^{(n-2)}}{(n-2)!} = - * F_A,$$

which is equivalent to saying that F_A is pointwise of type $\Lambda^{(1,1)}_0 \cong \mathfrak{su}(n)$.

Example. Suppose that E is a Hermitian vector bundle and A is Hermitian. Then A is an $\mathrm{SU}(n)$ instanton if and only if it is Hermitian Yang–Mills with $\lambda(E) = 0$.

Remark. The name $\mathrm{SU}(n)$ instanton is not standard. Motivated by the previous example they are often simply called Hermitian Yang–Mills. However, it is worth noting that $\mathrm{SU}(n)$ instantons make sense even when the bundle E (and thus the connection) is not Hermitian (or even complex), so it would seem better to use a different name.

We quickly see the following result, which is proved in an identically manner to the Hermitian Yang–Mills case with $\lambda(E) = 0$.

Proposition 8.4. *An $\mathrm{SU}(n)$ instanton A on a Calabi–Yau n-fold M is Yang–Mills. Moreover, if M is compact, then A is an absolute minimizer for the Yang–Mills energy.*

Example. If $n = 2$, so we are talking about Calabi–Yau 2-folds, then the $\mathrm{SU}(2)$ instanton condition is just that A is an ASD instanton.

Example. Consider the Levi-Civita connection A on the tangent bundle of a Calabi–Yau n-fold M. Since the metric g on M has holonomy $\mathrm{Hol}(g) \subseteq \mathrm{SU}(n)$, by the Ambrose–Singer Theorem we know that F_A takes values in $\mathfrak{su}(n)$ at each point, viewed as a subspace of the 2-forms at that point; i.e.

$$F_A(u, v) \in \Lambda^{(1,1)}_0 \cong \mathfrak{su}(n)$$

for all u, v.

However, $F_A : \Lambda^2 T^*M \to \Lambda^2 T^*m$ is symmetric by the symmetries of the Riemann curvature tensor, so that means F_A must pointwise be of type $\Lambda^{(1,1)}_0$, i.e. A is an $\mathrm{SU}(n)$ instanton.

We will leave the discussion of gauge theory on Calabi–Yau manifolds here for now, given the overlap with the study of Hermitian Yang–Mills connections. However, they shall reappear later in this section.

8.2 Instantons in 8 dimensions

We now move on to the Donaldson–Thomas picture for gauge theory in higher dimensions, for which one needs not just special holonomy, but *exceptional holonomy.*

We recall that we have parallel 4-form Φ_0 on $\mathbb{R}^8$ which has stabilizer Spin(7). We may also recall that there was distinguished subspace Λ^2_7 of the 2-forms on $\mathbb{R}^8$. If we consider the map on 2-forms given

$$L(\alpha) = *(\alpha \wedge \Phi_0),$$

then Λ^2_7 was the eigenspace with eigenvalue 3. Since L is self-adjoint, it must decompose the 2-forms on $\mathbb{R}^8$ into orthogonal eigenspaces with real eigenvalues, and as L is Spin(7)-invariant these eigenspaces must correspond to representations of Spin(7).

The orthogonal complement of Λ^2_7 in $\Lambda^2 T^*M$ has rank 21 and so we write it as Λ^2_{21}. Moreover, Λ^2_{21} must be isomorphic to the Lie algebra $\mathfrak{spin}(7)$ of Spin(7), since $\mathfrak{spin}(7) \subseteq \mathfrak{so}(8) \cong \Lambda^2(\mathbb{R}^8)^*$ and Λ^2_7 the representation of Spin(7) as the double cover of SO(7) on $\mathbb{R}^7$. A short calculation, which we leave as an exercise, yields the following result, which computes the eigenvalue of L on Λ^2_{21}.

Lemma 8.5. *We have an orthogonal decomposition* $\Lambda^2(\mathbb{R}^8)^* = \Lambda^2_7 \oplus \Lambda^2_{21}$ *where*

$$\Lambda^2_7 = \{\alpha \in \Lambda^2(\mathbb{R}^8)^* \,:\, \alpha \wedge \Phi_0 = 3 * \alpha\} = \{u \wedge v + \Phi_0(u, v, .,.) \,:\, u, v \in \mathbb{R}^8\},$$
$$\Lambda^2_{21} = \{\alpha \in \Lambda^2(\mathbb{R}^8)^*, : \alpha \wedge \Phi_0 = - * \alpha\} \cong \mathfrak{spin}(7).$$

Remark. Knowing the eigenvalue of L on Λ^2_7 determines what it must be on Λ^2_{21} since the trace of L must be zero.

These observations lead us to define the following type of connections on Spin(7) manifolds (M^8, Φ), noting that the decomposition of 2-forms on $\mathbb{R}^8$ extends to M since Φ is pointwise identified with Φ_0.

Definition 8.6. Let (M^8, Φ) be a Spin(7) manifold and let π_7 be the projection to the 2-forms of type Λ^2_7. A connection A is called a Spin(7) *instanton* if its curvature 2-form satisfies

$$F_A \wedge \Phi = - * F_A \quad \Leftrightarrow \quad \pi_7(F_A) = 0.$$

Note that the 2-form components of the curvature of a Spin(7) instanton take values in $\mathfrak{spin}(7)$ at each point.

Remark. The fact that these connections are called Spin(7) instantons further motivates the name of SU(n) instantons we used earlier.

The definition of Spin(7) instanton should remind us somewhat of the ASD instanton condition in 4 dimensions. With that in mind, the next result should not be a surprise.

Proposition 8.7. *A* Spin(7) *instanton is Yang–Mills. Moreover, if the* Spin(7) *manifold* (M, Φ) *is compact, then* Spin(7) *instantons are absolute minimizers of the Yang–Mills energy.*

Proof. By the now familiar trick, we differentiate both sides of the Spin(7) instanton condition and see that

$$\begin{aligned}\mathrm{d}_A * F_A &= -\mathrm{d}_A(F_A \wedge \Phi)\\ &= -\mathrm{d}_A F_A \wedge \Phi - F_A \wedge \mathrm{d}\Phi = 0,\end{aligned}$$

by the Bianchi identity and the fact that Φ is parallel and thus closed.

If M is compact, we can do better and if we let π_7, π_{21} denote the obvious projections on the 2-forms on M then the Yang–Mills energy of any connection A on a bundle E is

$$\mathcal{E}(A) = \|\pi_7(F_A)\|^2_{L^2} + \|\pi_{21}(F_A)\|^2_{L^2}.$$

On the other hand, Chern–Weil theory states that

$$\kappa(E) = -\int_M \mathrm{tr}(F_A \wedge F_A \wedge \Phi)$$

is a topological quantity: this uses the fact that Φ is closed. Using the description of the decomposition of the 2-forms into eigenspaces of the operator L we introduced above, we see that

$$\kappa(E) = \|\pi_{21}(F_A)\|^2_{L^2} - 3\|\pi_7(F_A)\|^2_{L^2},$$

and hence that

$$\mathcal{E}(A) = 4\|\pi_7(F_A)\|^2_{L^2} + \kappa(E).$$

This gives the claimed result. □

We now claim that the Spin(7) instanton equation is elliptic modulo augue. To see this, if we let

$$P(A) = F_A \wedge \Phi + *F_A,$$

whose zeros are Spin(7) instantons, and we linearize this equation at a Spin(7) instanton A, then we obtain the operator $L_A P$ which acts on $\mathrm{End}(E)$-valued 1-forms a by:

$$L_A P(a) = \mathrm{d}_A a \wedge \Phi + *\mathrm{d}_A a = \pi_7 \mathrm{d}_A a.$$

This equation is not elliptic, because it maps between bundles of the different ranks: $T^*M \otimes \mathrm{End}(E)$ and $\Lambda^2_7 \otimes \mathrm{End}(E)$. However, they differ in rank by 1, so if we include the Coulomb gauge condition $\mathrm{d}_A^* a = 0$, we obtain the equation:

$$a \mapsto \mathrm{d}_A^* a + \pi_7 \mathrm{d}_A a,$$

which is now elliptic modulo gauge.

In fact, this operator may be identified with a negative Dirac operator, since there is a natural spin structure on M so that the positive spinors are sections of $\Lambda^0 \oplus \Lambda^2_7$ and negative spinors are sections of T^*M. Thus the index of the operator can be calculated, and so give us the expected dimension of the moduli space of Spin(7) instantons as follows.

Theorem 8.8. *Let A be a* Spin(7) *instanton on a bundle E with compact semi-simple structure group* G *over a compact* Spin(7) *manifold M. Then the expected dimension of the moduli space of* Spin(7) *instantons near A is*

$$\dim \mathfrak{g}(b^1(M) - b^0(M) - b^2_7(M)) + \frac{1}{24}\int_M p_1(X)p_1(F) - \frac{1}{12}\int_M p_1(F)^2 - 2p_2(F),$$

where $\mathfrak{g}$ is the Lie algebra of G, *$b^2_7(M)$ is the space of harmonic forms of type Λ^2_7 on M, p_1, p_2 are the first and second Pontryagin classes, and F is the endomorphism bundle of E or, equivalently, we can take F to be the adjoint bundle $\mathfrak{g}_E$.*

The next obvious question is whether there are any Spin(7) instantons. We now tackle this question.

Example. A flat connection on a Spin(7) manifold is a Spin(7) instanton.

This example may seem trivial, but we have already seen the importance of flat connections in low dimensions.

Example. Let A be the Levi-Civita connection on a Spin(7) manifold with metric g. Then since $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, we must have by the Ambrose–Singer theorem that

$$F_A(u,v) \in \Lambda^2_{21} \cong \mathfrak{spin}(7),$$

for all u, v. However, since $F_A : \Lambda^2 T^*M \to \Lambda^2 T^*M$ is symmetric by the symmetries of the Riemann curvature tensor, we deduce that

$$\pi_7(F_A) = 0$$

as required.

Example. Recall that we can write the Spin(7) 4-form on $\mathbb{R}^8 = \mathbb{R}^4 \oplus \mathbb{R}^4$ as

$$\Phi_0 = \mathrm{d}x_{0123} + \mathrm{d}x'_{0123} - \sum_{j=1}^{3} \omega_j \wedge \omega'_j,$$

where we have coordinates $(x_0, x_1, x_2, x_3, x_0', x_1', x_2', x_3', x_4')$ on $\mathbb{R}^4 \oplus \mathbb{R}^4$ and

$$\omega_1 = \mathrm{d}x_{01} + \mathrm{d}x_{23}, \quad \omega_2 = \mathrm{d}x_{02} + \mathrm{d}x_{31}, \quad \omega_3 = \mathrm{d}x_{03} + \mathrm{d}x_{12},$$
$$\omega_1' = \mathrm{d}x_{01}' + \mathrm{d}x_{23}', \quad \omega_2' = \mathrm{d}x_{02}' + \mathrm{d}x_{31}', \quad \omega_3' = \mathrm{d}x_{03}' + \mathrm{d}x_{12}'.$$

Hence, if we let B be an ASD instanton on $\mathbb{R}^4$ and let $\pi : \mathbb{R}^8 \to \mathbb{R}^4$ be the projection onto the first $\mathbb{R}^4$, then we see that $A = \pi^* B$ is a Spin(7) instanton:

$$F_A \wedge \Phi_0 = F_A \wedge \mathrm{d}x_{0123}' = - *_{\mathbb{R}^4} F_A \wedge \mathrm{d}x_{0123}' = - * F_A,$$

where we used that $F_A \wedge \omega_j = 0$ for all j.

Example. Suppose that M is a Calabi–Yau 4-fold with Kähler form ω and holomorphic volume form Υ. Then M is a Spin(7) manifold with

$$\Phi = \frac{1}{2}\omega^2 + \operatorname{Re}\Upsilon.$$

Suppose that A is an SU(4) instanton. Then we know that

$$F_A \wedge \operatorname{Re}\Upsilon = 0 \quad \text{and} \quad F_A \wedge \frac{\omega^2}{2} = - * F_A.$$

Hence, A is a Spin(7) instanton.

Example. Consider the trivial bundle E with structure group Spin(7) on $\mathbb{R}^8$. There exists a 1-parameter family $(A_c)_{c>0}$ of Spin(7) instantons on E which are invariant under Spin(7) acting in the usual way on $\mathbb{R}^8$, discovered independently by Fairlie–Nuyts and Fubini–Nicolai.

We may compute that, if r denotes the distance to the origin in $\mathbb{R}^8$, then

$$|F_{A_c}|^2 \sim \frac{c^2}{(1+cr^2)^4},$$

just like for the standard instanton we saw on $\mathbb{R}^4$. We see that as $c \to 0$ we obtain the trivial flat connection and as $c \to \infty$ we obtain a connection which is singular at the origin but flat everywhere else, just as in the $\mathbb{R}^4$ case.

However, we notice that the Yang–Mills energy of A_c is always infinite, which is different from the $\mathbb{R}^4$ setting. In fact, it is known that there are no non-flat finite energy Yang–Mills connections on $\mathbb{R}^n$ except when $n = 4$.

Again, the instantons are all equivalent after rescaling and A_1 is often called the *standard* Spin(7) *instanton* on $\mathbb{R}^8$.

We have some non-trivial Spin(7) instantons but they all have large gauge group, so we can we get anything smaller?

Example. Let E be the trivial bundle with structure group SU(2) over the Bryant–Salamon Spin(7) manifold $(\mathbb{S}_-(\mathcal{S}^4), \Phi)$. There is a non-trivial 1-parameter family $(A_c)_{c>0}$ of Spin(7) instantons on E, due to Clarke–Oliveira. As $c \to 0$ one obtains the trivial flat connection, but as $c \to \infty$ something more interesting occurs, which we shall discuss later.

The idea of the construction is to use the high degree of symmetry of the problem to reduce the Spin(7) instanton condition to ODEs, which one can then solve explicitly.

Given the previous example, it is natural to ask if one can find Spin(7) instantons with gauge group SU(2) on *compact* Spin(7) manifolds. This can be done and was first achieved by Lewis. We shall discuss how the construction goes later on.

The idea behind studying Spin(7) instantons is that one would want to "count" how many there are on a given bundle E on a compact Spin(7) manifold (M, Φ). The hope would be that this count would be an *invariant* of (M, Φ), meaning that if we deform the Spin(7) form Φ, then the count should stay the same. This turns out to be too naive, but is still motivational.

8.3 Chern–Simons in 7 dimensions

Given that we are trying to draw an analogy between the theory we have seen in low dimensions and the higher dimensional gauge theories, and that we have a notion of instantons in 8 dimensions, it is not unreasonable to study a Chern–Simons theory in 7 dimensions.

Suppose now that we have a compact 7-dimensional Riemannian manifold (M^7, g) where $\mathrm{Hol}(g) \subseteq \mathrm{G}_2$. As we have seen, on such a G_2 manifold there is parallel 3-form φ which in fact induces the metric g and an orientation on M (and thus a Hodge star operator).

We pick a reference connection $A_0 \in \mathcal{A}$, the space of connections on our vector bundle E, and write $A \in \mathcal{A}$ as $A = A_0 + a$ for an $\mathrm{End}(E)$-valued 1-form a. On the 8-manifold $[0,1] \times M$ we can take the pullback of E and a connection $\mathbf{A}$ on it with curvature $\mathbf{F}$ by

$$\mathbf{A} = A_0 + sa \quad \text{and} \quad \mathbf{F} = \mathrm{d}s \wedge a + F_{A_0} + s\mathrm{d}_{A_0}a + s^2 a \wedge a,$$

where $s \in [0,1]$. Using this we have the following definition.

Definition 8.9. The *Chern–Simons functional* $\mathcal{F}$ on $\mathcal{A}$ (with respect to the reference A_0) on a compact G_2 manifold (M, φ) is given by

$$\mathcal{F}(A) = \int_{[0,1]\times M} \mathrm{tr}(\mathbf{F} \wedge \mathbf{F} \wedge *\varphi),$$

where we pullback $*\varphi$ from M to $[0,1] \times M$. Equivalently,

$$\mathcal{F}(A) = \int_M \mathrm{tr}\left(\left(2a \wedge F_{A_0} + a \wedge \mathrm{d}_{A_0}a + \frac{2}{3}a \wedge a \wedge a\right) \wedge *\varphi\right)$$

for $A = A_0 + a$.

Of course, the dependence on A (or a) in this Chern–Simons functional is really the same as in the 3-dimensional case, so the same argument as there leads us to compute the differential of $\mathcal{F}$.

Lemma 8.10. *The differential of* $\mathcal{F}$ *at* A *is*

$$\mathrm{d}\mathcal{F}_A(\dot{a}) = 2\langle \dot{a}, *(F_A \wedge *\varphi)\rangle_{L^2}.$$

Hence $\mathrm{d}\mathcal{F}$ *is a well-defined closed 1-form on* $\mathcal{A}/\mathcal{G}$.

From this we are led to the following definition.

Definition 8.11. A connection A on a bundle over a G_2 manifold (M, φ) is a G_2 *instanton* if and only if

$$F_A \wedge *\varphi = 0.$$

These connections are the critical points of the Chern–Simons functional $\mathcal{F}$.

Before we return to the Chern–Simons functional, we want to make some algebraic observations about 2-forms on $\mathbb{R}^7$. Consider the map

$$L(\alpha) = *(\alpha \wedge \varphi_0)$$

acting from 2-forms to 2-forms on $\mathbb{R}^7$. This map is G_2-invariant and so its eigenspaces must correspond to representations of G_2 on the 2-forms. In fact, we have the following.

Lemma 8.12. *We can decompose* $\Lambda^2(\mathbb{R}^7)^* = \Lambda^2_7 \oplus \Lambda^2_{14}$ *where*

$$\Lambda^2_7 = \{\alpha \in \Lambda^2(\mathbb{R}^7)^* : \alpha \wedge \varphi_0 = 2 * \alpha\} = \{v \lrcorner \varphi_0 : v \in \mathbb{R}^7\}$$
$$\Lambda^2_{14} = \{\alpha \in \Lambda^2(\mathbb{R}^7)^* : \alpha \wedge \varphi_0 = - * \alpha\} = \{\alpha \in \Lambda^2(\mathbb{R}^7)^* : \alpha \wedge *\varphi_0 = 0\} \cong \mathfrak{g}_2,$$

the Lie algebra of G_2.

Proof. We can prove this as follows. Write $\mathbb{R}^7 = \mathbb{R}_x \oplus \mathbb{C}^3$ and then we can write

$$\varphi_0 = \mathrm{d}x \wedge \omega_0 + \mathrm{Re}\,\Upsilon_0,$$

where ω_0 is the Kähler form on $\mathbb{C}^3$ and Υ_0 is the holomorphic volume form on $\mathbb{C}^3$. Then, we see that if we let $v = \partial_x$ then

$$\begin{aligned}
*(v \lrcorner \varphi_0 \wedge \varphi_0) &= *\Big(\omega_0 \wedge (\mathrm{d}x \wedge \omega_0 + \operatorname{Re} \Upsilon_0)\Big) \\
&= 2 * \mathrm{d}x \wedge \frac{\omega_0^2}{2!} \\
&= 2\omega_0 \\
&= 2(v \lrcorner \varphi_0).
\end{aligned}$$

Since G_2 acts transitively on the 6-sphere $\mathcal{S}^6$, we deduce that $v \lrcorner \varphi_0$ is in the 2-eigenspace of L for all $v \in \mathbb{R}^7$. Moreover, the map $v \mapsto v \lrcorner \varphi_0$ is injective, so this means the 2-eigenspace of L is at least 7-dimensional.

We now observe that G_2 is the subgroup of Spin(7) fixing any non-zero element, so $\mathfrak{g}_2 \subseteq \mathfrak{spin}(7)$ in particular. Therefore, if we write $\mathbb{R}^8 = \mathbb{R}_t \oplus \mathbb{R}^7$ we see that the constant elements $\alpha \in \mathfrak{spin}(7) \subseteq \Lambda^2(\mathbb{R}^8)^*$ which have no $\mathrm{d}t$ component are the same as the elements which span $\mathfrak{g}_2 \subseteq \Lambda^2(\mathbb{R}^7)^*$. We then recall that

$$\alpha \wedge \Phi_0 = - *_{\mathbb{R}^8} \alpha$$

on $\mathbb{R}^8$ and that

$$\Phi_0 = \mathrm{d}t \wedge \varphi_0 + *\varphi_0.$$

Since

$$*_{\mathbb{R}^8}\alpha = \mathrm{d}t \wedge *_{\mathbb{R}^7}\alpha$$

and

$$\begin{aligned}
\alpha \wedge \Phi_0 &= \alpha \wedge (\mathrm{d}t \wedge \varphi_0 + *\varphi_0) \\
&= \mathrm{d}t \wedge \alpha \wedge \varphi_0 + \alpha \wedge *\varphi_0,
\end{aligned}$$

we see that L acts as -1 on $\mathfrak{g}_2$ then the whole lemma is proved. □

A corollary of this is the following.

Proposition 8.13. *A connection A on a G_2 manifold (M^7, φ) is a G_2 instanton if and only if*

$$F_A \wedge \varphi = - * F_A \quad \Leftrightarrow \quad \pi_7(F_A) = 0,$$

where π_7 is the orthogonal projection to Λ^2_7. Hence, G_2 instantons are Yang-Mills.

Proof. The first part is immediate from the previous lemma. Differentiating both sides of the equation as usual yields

$$\mathrm{d}_A * F_A = - * \mathrm{d}_A(F_A \wedge \varphi) = 0,$$

using the Bianchi identity $\mathrm{d}_A F_A = 0$ and $\mathrm{d}\varphi = 0$. □

Unsurprisingly, we have the following minimization property for G_2 instantons, whose proof follows exactly the same lines as for Spin(7) instantons.

Proposition 8.14. *If the G_2 manifold (M^7, φ) is compact, then G_2 instantons are absolute minimizers for the Yang–Mills energy.*

Proof. If we let

$$\kappa(E) = -\int_M \operatorname{tr}(F_A \wedge F_A \wedge \varphi)$$

for any connection A, then we see that the Yang–Mills energy of A is

$$\mathcal{E}(A) = 3\|\pi_7(F_A)\|^2_{L^2} + \kappa(E),$$

which gives the result as $\kappa(E)$ only depends on the topology of the bundle and $[\varphi]$ by Chern–Weil theory. □

We see that for the Chern–Simons functional on (M^7, φ), the G_2 instantons seem to play an analogous role to the flat connections in Chern–Simons theory on 3-manifolds and we might be inclined to think about Floer theory for $\mathcal{F}$ to build an *invariant* for G_2 manifolds. For this, we would need to understand gradient flow lines of $\mathcal{F}$.

Lemma 8.15. *Negative gradient flows lines for $\mathcal{F}$ on (M^7, φ) are* Spin(7) *instantons on $\mathbb{R}_t \times M$ with* Spin(7) *form $\Phi = \mathrm{d}t \wedge \varphi + *\varphi$.*

Proof. A negative gradient flow line for $\mathcal{F}$ is a family of connections $A(t)$ on M satisfying

$$\frac{\partial}{\partial t}A(t) = - * (F_{A(t)} \wedge *\varphi)$$

by Lemma 8.10. We see that if we let $\mathbf{A} = A(t)$ be the connection we can define on $\mathbb{R}_t \times M$, then its curvature $\mathbf{F}$ satisfies

$$\begin{aligned} - * \mathbf{F} &= - * \left(\mathrm{d}t \wedge \frac{\partial}{\partial t}A(t) + F_{A(t)}\right) \\ &= * \left(\mathrm{d}t \wedge *_M(F_{A(t)} \wedge *_M\varphi) - F_{A(t)}\right) \\ &= F_{A(t)} \wedge *\varphi - \mathrm{d}t \wedge *_M F_{A(t)} \\ &= \mathbf{F} \wedge \Phi \end{aligned}$$

using some G_2 algebra. □

This shows that we are, at least formally, well set up for studying Floer theory for the Chern–Simons functional on compact G_2 manifolds. There are, of course, some key analytic issues to overcome to make this picture anything like rigorous!

8.4 Monopoles in 7 dimensions

We have seen that G_2 instantons and Spin(7) instantons are related, but what happens if we just dimensionally reduce the Spin(7) instantons to 7 dimensions? Do we just get G_2 instantons or something else?

To answer this question, we suppose that we have $\mathbb{R} \times M^7$ where (M^7, φ) is a G_2 manifold and we put the product Spin(7) structure on $\mathbb{R} \times M^7$, i.e.

$$\Phi = \mathrm{d}t \wedge \varphi + *\varphi,$$

where t is the coordinate on $\mathbb{R}$. Suppose that E is a bundle on M which we pull back to $\mathbb{R} \times M$. We suppose further that we have a connection $\mathbf{A}$ on E over $\mathbb{R} \times M$ in the form

$$\mathbf{A} = \phi \mathrm{d}t + A$$

where (A, ϕ) are a t-independent connection on M and section of $\mathrm{End}(E)$ respectively. This is the familiar dimension reduction technique, now applied to Spin(7) instantons.

As we saw before, the curvature $\mathbf{F}$ of $\mathbf{A}$ is:

$$\mathbf{F} = -\mathrm{d}t \wedge \mathrm{d}_A\phi + F_A.$$

Hence, we see on the one hand that

$$\begin{aligned} \mathbf{F} \wedge \Phi &= (-\mathrm{d}t \wedge \mathrm{d}_A\phi + F_A) \wedge (\mathrm{d}t \wedge \varphi + *\varphi) \\ &= \mathrm{d}t \wedge (F_A \wedge \varphi - \mathrm{d}_A\phi \wedge *\varphi) + F_A \wedge *\varphi. \end{aligned}$$

On the other hand, we have that

$$- * \mathbf{F} = -\mathrm{d}t \wedge *F_A + *\mathrm{d}_A\phi.$$

Using the G_2 algebra we have seen previously, we deduce that $\mathbf{A}$ is a Spin(7) instanton if and only if (A, ϕ) is a G_2 *monopole* in the following sense.

Definition 8.16. A pair (A, ϕ) of a connection A on E and a section ϕ of $\mathrm{End}(E)$ (known as a *Higgs field*) on a G_2 manifold (M^7, φ) is a G_2 *monopole* if

$$F_A \wedge *\varphi = *\mathrm{d}_A\phi.$$

Equivalently, G_2 monopoles satisfy

$$F_A \wedge \varphi + *F_A = \mathrm{d}_A\phi \wedge *\varphi.$$

Notice that G_2 instantons are G_2 monopoles with $\phi = 0$.

We therefore see that the dimensional reduction of the Spin(7) instanton equation to 7 dimensions is the G_2 monopole equation. In particular, we note the following.

Example. Let (M^7, φ) be a G_2 manifold and let A be a G_2 instanton on a bundle E over M. Then $(\mathbb{R} \times M, \Phi)$ is a Spin(7) manifold if we choose

$$\Phi = \mathrm{d}t \wedge \varphi + *\varphi,$$

where t is the $\mathbb{R}$ coordinate. By the discussion above, the pullback of A to $(\mathbb{R} \times M, \Phi)$ is a Spin(7) instanton on the pullback of E.

However, just as in lower dimensions, we see that G_2 monopoles do not yield anything beyond what we have already seen on compact manifolds.

Lemma 8.17. *Let (M^7, φ) be a compact G_2 manifold and let (A, ϕ) be a G_2 monopole. Then A is a G_2 instanton and $\mathrm{d}_A\phi = 0$.*

Proof. We differentiate both sides of the G_2 monopole equation to see that

$$\mathrm{d}_A * \mathrm{d}_A\phi = \mathrm{d}_A(F_A \wedge *\varphi) = 0$$

since $\mathrm{d}_A F_A = 0$ (the Bianchi identity) and $\mathrm{d} * \varphi = 0$. Integration by parts then yields $\mathrm{d}_A\phi = 0$ as required, as then A is a G_2 instanton by definition. □

Remark. Even though G_2 monopoles only give G_2 instantons on compact G_2 manifolds we shall see that this is not the case on non-compact G_2 manifolds. In fact, part of the Donaldson–Segal programme involves the study of G_2 monopoles on non-compact manifolds.

This observation that G_2 monopoles and G_2 instantons essentially coincide on a compact G_2 manifold (M^7, φ) is actually rather useful. The reason is that we know that A is a G_2 instanton if and only if there is Higgs field ϕ such that (A, ϕ) is a G_2 monopole. Moreover, if A is irreducible then we know that $\mathrm{d}_A\phi = 0$ forces ϕ to be constant.

If we then linearize the G_2 monopole equation at (A, ϕ) in the direction (a, ψ) we obtain the equation

$$\mathrm{d}_A a \wedge *\varphi = *\mathrm{d}_A\psi \quad \Leftrightarrow -\mathrm{d}_A\psi + *(\mathrm{d}_A a \wedge *\varphi) = 0.$$

If we include the Coulomb gauge fixing condition $\mathrm{d}_A^* a = 0$, then we see that the linearized G_2 instanton condition can be expressed as the zeros of the map L from $\Omega^0(\mathrm{End}(E)) \oplus \Omega^1(\mathrm{End}(E))$ to itself given by

$$L \begin{pmatrix} \psi \\ a \end{pmatrix} = \begin{pmatrix} 0 & -\mathrm{d}_A^* \\ -\mathrm{d}_A & *(\mathrm{d}_A \cdot \wedge * \varphi) \end{pmatrix} \begin{pmatrix} \psi \\ a \end{pmatrix}.$$

We see that the map L is *elliptic* and *self-adjoint*, which means that the G_2 instanton condition is elliptic and index 0. We deduce the following.

Theorem 8.18. *Let A be a G_2 instanton on a compact G_2 manifold. The expected dimension of the moduli space of G_2 instantons near A is zero.*

Remark. A direct consequence of this result is that one might hope, under favourable circumstances, that the moduli space of G_2 instantons on a compact G_2 manifold is really 0 and, in fact, only consists of finitely many points. Then one could simply count these points (possibly with signs) to get a possible *invariant* for G_2 manifolds. This turns out to be too naive, but nonetheless forms part of the Donaldson–Thomas/Donaldson–Segal programme for gauge theory in higher dimensions.

A key issue in obtaining the expected dimension of 0 is one needs a suitable space to perturb the ambient G_2 3-form φ in on the compact 7-manifold M^7. If one restricts to G_2 manifolds, then one needs φ to stay closed and coclosed (for the metric and orientation it defines), i.e. $\mathrm{d}\varphi = 0$ and $\mathrm{d}^*_\varphi\varphi = 0$. It turns out, perhaps surprisingly, that locally (up to diffeomorphisms), such deformations of φ are parametrized by elements of $H^3(M)$: this is a result of Bryant and Joyce (independently).

Much of our discussion so far concerning G_2 instantons only required *algebraic* properties of φ, rather than differential ones, so many things pass through to the setting of general φ (not necessarily closed and coclosed). However, one sees that if we define G_2 instantons by

$$F_A \wedge *\varphi = 0,$$

then differentiating both sides leads to

$$F_A \wedge \mathrm{d} * \varphi = 0$$

by the Bianchi identity. Therefore, generically, this will give us an extra constraint on a G_2 instanton A and make the equation overdetermined, meaning that G_2 instantons will not exist (even locally). Therefore, it is natural to impose the constraint

$$\mathrm{d}^*_\varphi\varphi = 0$$

when studying G_2 instantons. It turns out that weakening to this setting, where φ is only coclosed, gives an infinite-dimensional space in which to deform φ and allows us to guarantee that irreducible G_2 instantons on compact M^7 do indeed have 0-dimensional moduli spaces under generic choices of φ.

It should however be noted that the fact that φ need not be closed means not only that G_2 instantons are no longer necessarily Yang–Mills, but there is no topological energy identity for such G_2 instantons. This is a problem if one wants to show that the moduli space of G_2 instantons is compact. If time permits, we shall return to this issue later in this course and discuss possible remedies to this key problem.

Before we go any further with this discussion of moduli spaces and so on we should ask: are there any G_2 instantons or G_2 monopoles? We now indicate that indeed there are many examples.

Example. Decompose $\mathbb{R}^7 = \mathbb{R}^3 \oplus \mathbb{R}^4$ with the flat Euclidean metric and let $\pi_{\mathbb{R}^3}$, $\pi_{\mathbb{R}^4}$ be the natural projections.

- Let B be an ASD instanton on $\mathbb{R}^4$. Consider $A = \pi^*_{\mathbb{R}^4}B$. Then we can write

$$*\varphi = \mathrm{vol}_{\mathbb{R}^4} - \mathrm{d}x_2 \wedge \mathrm{d}x_3 \wedge \omega_1 - \mathrm{d}x_3 \wedge \mathrm{d}x_1 \wedge \omega_2 - \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \omega_3,$$

  where (x_1, x_2, x_3) are Euclidean coordinates on $\mathbb{R}^3$ and $\omega_1, \omega_2, \omega_3$ is the standard triple of self-dual 2-forms on $\mathbb{R}^4$ we have seen before. We therefore see immediately that, since B is ASD,

$$F_B \wedge \omega_i = 0 \text{ for all } i \quad \Rightarrow \quad F_A \wedge *\varphi = 0.$$

  Hence, A is a G_2 instanton.

- Let (B, ψ) be a monopole on $\mathbb{R}^3$ and let (A, ϕ) be the pullback of (B, ψ) under $\pi_{\mathbb{R}^3}$. Then

$$F_A \wedge *\varphi = F_A \wedge \mathrm{vol}_{\mathbb{R}^4} = *_{\mathbb{R}^3}\mathrm{d}_B\psi \wedge \mathrm{vol}_{\mathbb{R}^4} = *\mathrm{d}_A\phi,$$

  and so (A, ϕ) is a G_2 monopole.

These constructions give us many non-trivial examples of G_2 instantons and monopoles on $\mathbb{R}^7$.

We now give a particularly important application of this construction.

Example. Use the notation of the previous example.

- Let B_c be the BPST or standard instanton on $\mathbb{R}^4$ with parameter $c > 0$. Then $A_c = \pi^*_{\mathbb{R}^4}B_c$ is a G_2 instanton for all c, all of which have infinite Yang–Mills energy. We also see that as $c \to 0$ A_c tends to the trivial flat connection, but as $c \to \infty$ $|F_{A_c}|^2$ concentrates along the *associative* 3-plane $\mathbb{R}^3 \times \{0\}$.

- Let (B_m, ϕ_m) be the BPS monopole on $\mathbb{R}^3$ with mass $m > 0$. Then the pullback (A_m, ϕ_m) to $\mathbb{R}^7$ by $\pi_{\mathbb{R}^3}$ is a G_2 monopole. We also see that as $m \to \infty$, $|F_{A_m}|^2$ concentrates along the *coassociative* 4-plane $\{0\} \times \mathbb{R}^4$ and that ϕ_m vanishes along this same 4-plane for all m.

Example. On any G_2 manifold, the Levi-Civita connection is a G_2 instanton by the same Ambrose–Singer argument we have seen in other settings.

Example. Let E be the trivial $\mathbb{R}^7$ bundle over Euclidean $\mathbb{R}^7$ with structure group G_2. There is a 1-parameter family $(A_c)_{c>0}$ of G_2 instantons on E, constructed by Günaydin–Nicolai, which all have infinite Yang–Mills energy since

$$|F_{A_c}|^2 \sim \frac{c^2}{(1+cr^2)^4}$$

where r is the radial distance from 0.

They are all equivalent up to rescaling and are referred to as the *standard* G_2 instanton on $\mathbb{R}^7$ by analogy with the standard ASD instanton on $\mathbb{R}^4$. Again we notice that as $c \to 0$ we obtain the trivial flat connection, and as $c \to \infty$ the curvature becomes zero everywhere except at the origin where we have a singularity.

Example. The first three complete holonomy G_2 manifolds (which were constructed by Bryant–Salamon) are $\Lambda^2_+ T^*\mathcal{S}^4$, $\Lambda^2_+ T^*\overline{\mathbb{CP}^2}$ and $\mathbb{S}(\mathcal{S}^3)$.

- On $\Lambda^2_+ T^*\mathcal{S}^4$ and $\Lambda^2_+ T^*\overline{\mathbb{CP}^2}$, Oliveira constructed 1-parameter families $(A_m, \phi_m)_{m>0}$ of irreducible G_2 monopoles with gauge group SU(2) and SO(3) respectively. The parameter m is the *mass* of the monopole, by analogy with what we saw for monopoles in $\mathbb{R}^3$. One way to understand the mass is it is the limiting value of $|\phi_m|$ as we tend to infinity along the fibres in the bundles of self-dual 2-forms.

  We again see that $|F_{A_m}|^2$ concentrates along the coassociatives $\mathcal{S}^4$ or $\overline{\mathbb{CP}^2}$ as $m \to \infty$ and that ϕ_m vanishes there for all m. Moreover, at each point x in $\mathcal{S}^4$ or $\overline{\mathbb{CP}^2}$, on the normal space, which is a copy of $\mathbb{R}^3$ and the fibre of Λ^2_+ at x, one sees that (after rescaling) the pair (A_m, ϕ_m) restricted to the normal space converges to the same BPS monopole on $\mathbb{R}^3$ (for every x) as $m \to \infty$.

- On $\mathbb{S}(\mathcal{S}^3)$, work by Clarke, L.–Oliveira and (very recently) Stein–Turner shows that there are two 1-parameter families of irreducible G_2 instantons with gauge group SU(2).

  In the Clarke family one sees that there is a parameter $c > 0$ so that as $c \to \infty$ the curvature of the G_2 instantons concentrates along the associative $\mathcal{S}^3$. Moreover, at each point x in $\mathcal{S}^3$, on the normal space (which is now a copy of $\mathbb{R}^4$ and is the fibre of $\mathbb{S}$ at x) as $c \to \infty$ we see the same BPST instanton for every x as the limit of the G_2 instantons as $c \to \infty$.

We now have G_2 instantons and monopoles in many settings, but what we are missing is G_2 instantons with small gauge group (i.e. SU(2) or SO(3)) on compact holonomy G_2 manifolds. These have been shown to exist by Sá Earp–Walpuski, Walpuski, Menet–Sá Earp–Walpuski and Platt. We will discuss one of these construction later, which involves *calibrated geometry.*

8.5 Instantons, Chern–Simons and monopoles in 6 dimensions

To finish this section I want to return to Calabi–Yau 3-folds to complete the Donaldson–Thomas picture. To begin with I want to make an elementary observation.

Example. Let (M^6, ω, Υ) be a Calabi–Yau 3-fold and let A be an SU(3) instanton on a bundle E over M. We can define a G_2 manifold $(\mathbb{R} \times M, \varphi)$ by choosing

$$\varphi = \mathrm{d}t \wedge \omega + \operatorname{Re}\Upsilon,$$

where t is the coordinate on $\mathbb{R}$. Then the pullback of A to $(\mathbb{R} \times M, \varphi)$ is a G_2 instanton. This is most easily seen by noting that

$$*\varphi = \frac{1}{2}\omega^2 - \mathrm{d}t \wedge \operatorname{Im}\Upsilon$$

and recalling that

$$F_A \wedge \omega^2 = 0 \quad \text{and} \quad F_A \wedge \operatorname{Im} \Upsilon = 0$$

by the various equivalent ways of viewing SU(3) instantons.

In this way we can construct G_2 instantons from SU(3) instantons, but what about the other way round, i.e. what is the dimensional reduction of the G_2 instanton equation to 6 dimensions?

We see that if we write a connection $\mathbb{A}$ on $(\mathbb{R} \times M, \varphi)$ as in the example above as

$$\mathbb{A} = \phi \mathrm{d}t + A$$

in the now familiar way, we can then ask what conditions we obtain on (A, ϕ) which are equivalent to $\mathbb{A}$ being a G_2 instanton. If we let $\mathbb{F}$ be the curvature of $\mathbb{A}$ then

$$\begin{aligned}\mathbb{F} \wedge *\varphi &= (-\mathrm{d}t \wedge \mathrm{d}_A\phi + F_A) \wedge \left(\frac{1}{2}\omega^2 - \mathrm{d}t \wedge \operatorname{Im}\Upsilon\right) \\ &= -\mathrm{d}t \wedge \left(F_A \wedge \operatorname{Im}\Upsilon + \mathrm{d}_A\phi \wedge \frac{\omega^2}{2}\right) + F_A \wedge \frac{\omega^2}{2}.\end{aligned}$$

We also compute that

$$\begin{aligned}\mathbf{F} \wedge \varphi + *\mathbf{F} &= (-\mathrm{d}t \wedge \mathrm{d}_A\phi + F_A) \wedge (\mathrm{d}t \wedge \omega + \operatorname{Re}\Upsilon) + (\mathrm{d}t \wedge *F_A - *\mathrm{d}_A\phi) \\ &= \mathrm{d}t \wedge (F_A \wedge \omega + *F_A - \mathrm{d}_A\phi \wedge \operatorname{Re}\Upsilon) + F_A \wedge \operatorname{Re}\Upsilon - *\mathrm{d}_A\phi.\end{aligned}$$

We deduce that $\mathbb{A}$ is a G_2 instanton if and only if (A, ϕ) is a *Calabi–Yau monopole* in the following sense.

Definition 8.19. A pair (A, ϕ) of a connection A on a bundle E and a section ϕ of $\operatorname{End}(E)$ (called a *Higgs field*) on a Calabi–Yau 3-fold (M^6, ω, Υ) is a *Calabi–Yau monopole* if

$$F_A \wedge \omega^2 = 0 \quad \text{and} \quad F_A \wedge \operatorname{Re}\Upsilon = *\mathrm{d}_A\phi.$$

One can check that this is precisely the dimensional reduction of the G_2 instanton equation to Calabi–Yau 3-folds given the calculations above. The second equation can also be equivalently be written

$$F_A \wedge \operatorname{Im}\Upsilon + d_A\phi \wedge \frac{\omega^2}{2} = 0.$$

In fact, the equation

$$F_A \wedge \omega + *F_A = \mathrm{d}_A\phi \wedge \operatorname{Re}\Upsilon$$

encodes the whole G_2 monopole condition, but this formulation is actually rarely used.

As should now be familiar, we see that Calabi–Yau monopoles reduce to familiar objects when M^6 is compact.

Lemma 8.20. *If (A, ϕ) is a Calabi–Yau monopole on a compact Calabi–Yau 3-fold, then $\mathrm{d}_A\phi = 0$ and A is an SU(3) instanton.*

Proof. This follows from the standard integration by parts argument. □

Therefore again if we want Calabi–Yau monopoles which do not simply define SU(3) instantons, we need to work on a noncompact Calabi–Yau 3-fold. Luckily, we have such an example.

Example. Recall that $T^*\mathcal{S}^3$ admits a complete Calabi–Yau metric due to Stenzel. Oliveira showed that there is a 1-parameter family (A_m, ϕ_m) for $m > 0$ of Calabi–Yau monopoles on $T^*\mathcal{S}^3$. The parameter, which is again the *mass*, is the limiting value of $|\phi_m|$ at infinity. As perhaps one would expect, ϕ_m vanishes on $\mathcal{S}^3$ for all m and, as $m \to \infty$, $|F_{A_m}|^2$ concentrates along the special Lagrangian $\mathcal{S}^3$. Moreover, one sees the same BPS monopole on $T^*_x\mathcal{S}^3 \cong \mathbb{R}^3$ for each $x \in \mathcal{S}^3$ as the limit of (A_m, ϕ_m) after rescaling as $m \to \infty$.

I now want to conclude this section by completing the formal Donaldson–Thomas picture for gauge theory in higher dimensions, which I will supplement with some ideas of Haydys. The key idea is to introduce a *complex* Chern–Simons functional as follows.

We have a vector bundle E over a compact Calabi–Yau 3-fold (M^6, ω, Υ) and we fix a reference connection A_0 on E. We then consider $[0,1] \times M$ with parameter $s \in [0,1]$ and for any connection $A = A_0 + a$ on E we define a connection $\mathbb{A}$ with curvature $\mathbb{F}$ on the pullback of E to $[0,1] \times M$ by

$$\mathbf{A} = A_0 + sa \quad \text{and} \quad \mathbf{F} = \mathrm{d}s \wedge a + F_{A_0} + s\mathrm{d}_{A_0}a + s^2 a \wedge a,$$

An important change now is that we consider a restricted space of connections

$$\mathcal{A}_0 = \{A \in \mathcal{A} \,:\, F_A \wedge \omega^2 = 0\}.$$

Notice that the gauge group $\mathcal{G}$ still acts on $\mathcal{A}_0$ by the gauge invariance of the curvature.

Remark. For those familiar with related theories, we are viewing $F_A \wedge \omega^2$, or equivalently $\Lambda_\omega F_A$, as a (real) moment map and so we are asking for connections whose curvature lies in the zero level set of the moment map.

Definition 8.21. We define the *complex Chern–Simons functional* $\mathcal{F}^{\mathbb{C}} : \mathcal{A}_0 \to \mathbb{C}$ (with reference A_0) on the compact Calabi–Yau 3-fold (M^6, ω, Υ) by

$$\mathcal{F}^{\mathbb{C}}(A) = \frac{1}{2}\int_{[0,1]\times M^6} \mathrm{tr}(\mathbf{F}^2) \wedge \Upsilon.$$

Here, Υ is pulled back to $[0,1] \times M$ and I have finally introduced the factor of $1/2$ for convenience, which I could have done all along! Equivalently, we may write

$$\mathcal{F}^{\mathbb{C}}(A) = \int_M \mathrm{tr}\left(\left(a \wedge F_{A_0} + \frac{1}{2}a \wedge \mathrm{d}_{A_0}a + \frac{1}{3}a \wedge a \wedge a\right) \wedge \Upsilon\right)$$

for $A = A_0 + a$.

Now, when we think about gradient flow lines we have to be more careful because they are real objects and $\mathcal{F}^{\mathbb{C}}$ is complex. To remedy this, we choose a phase and consider gradient flow lines with the given phase. In other words, we can look at the differential of $\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$. By our previous calculations, we can readily compute the differential of $\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$ as follows.

Proposition 8.22. *The differential of* $\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$ *at* $A \in \mathcal{A}_0$ *is given by*

$$\mathrm{d}\big(\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})\big)_A(\dot{a}) = -\langle \dot{a}, *(F_A \wedge \mathrm{Re}(e^{-i\theta}\Upsilon))\rangle_{L^2}.$$

Hence, critical points of $\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$ *on* $\mathcal{A}_0$ *are* $\mathrm{SU}(3)$ *instantons. Moreover,* $\mathrm{d}(\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}}))$ *is a well-defined closed 1-form on* $\mathcal{A}_0/\mathcal{G}$*.*

Proof. The proof is almost identical to the real case in 3 or 7 dimensions. The only main new point is that, since we restrict to $\mathcal{A}_0$, we already have that $F_A \wedge \omega^2 = 0$ and then $F_A \wedge \mathrm{Re}(e^{-i\theta}\Upsilon) = 0$ for any θ forces A to be an $\mathrm{SU}(3)$ instanton. The other point is the change of sign (i.e. the additional minus sign) which arises since $*^2 = -1$ on 5-forms in 6 dimensions. □

Remark. One can work with complex vector bundles and look at varying only the $(0,1)$ part of A, which is the same as looking at Dolbeault operators $\bar{\partial}_A$ on E, with critical points then being Cauchy–Riemann operators satisfying the extra constraint given by $\Lambda_\omega F_A = 0$. This is the point of view taken in the original Donaldson–Thomas paper. What we do here is very similar but less tied to complex geometry.

I now want to ask: what are the gradient flow lines for $\mathrm{Re}\,\mathcal{F}^{\mathbb{C}}$? The answer is, perhaps, unsurprising.

Lemma 8.23. *Negative gradient flow lines for* $\mathrm{Re}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$ *on* $\mathbb{R}_t \times M^6$ *are* G_2 *instantons with respect to the product* G_2 *structure*

$$\varphi = \mathrm{d}t \wedge \omega + \mathrm{Re}(e^{-i\theta}\Upsilon) \quad \textit{and} \quad *\varphi = \frac{1}{2}\omega^2 - \mathrm{d}t \wedge \mathrm{Im}(e^{-i\theta}\Upsilon).$$

Proof. Consider a negative gradient flow line $\mathbf{A} = A(t)$, which is the statement that

$$\frac{\partial A(t)}{\partial t} = *(F_{A(t)} \wedge \mathrm{Re}(e^{-i\theta}\Upsilon)).$$

We see that if we let $\mathbf{F}$ be the curvature of $\mathbf{A}$ then

$$- * \mathbf{F} = -\mathrm{d}t \wedge *F_A - *\frac{\partial A}{\partial t}$$

and

$$\begin{aligned}\mathbf{F} \wedge \varphi &= \left(\mathrm{d}t \wedge \frac{\partial A}{\partial t} + F_A\right) \wedge (\mathrm{d}t \wedge \omega + \mathrm{Re}(e^{-i\theta}\Upsilon)) \\ &= \mathrm{d}t \wedge \left(\frac{\partial A}{\partial t} \wedge \mathrm{Re}(e^{-i\theta}\Upsilon) + F_A \wedge \omega\right) + F_A \wedge \mathrm{Re}(e^{-i\theta}\Upsilon).\end{aligned}$$

Using the fact that $*^2 = -1$ on 1-forms on a 6-manifold and that G_2 instantons are defined by the equation $\mathbf{F} \wedge \varphi = - * \mathbf{F}$, we obtain the desired result. □

The big difference between our previous discussion is now that we have different flow lines depending on the phase, so we can ask now whether we can connect two flow lines, i.e. two G_2 instantons. This would then be a connection on an 8-manifold $\mathbb{R}_s \times \mathbb{R}_t \times M^6$

$$\mathbb{A} = A(s,t)$$

for connections $A(s,t)$ on M^6. These connections will have curvature

$$\mathbb{F} = \mathrm{d}s \wedge \frac{\partial A}{\partial s} + \mathrm{d}t \wedge \frac{\partial A}{\partial t} + F_A.$$

Since we are on an 8-manifold, it is perhaps natural to ask: when is $\mathbb{A}$ a Spin(7) instanton?

We first observe that

$$- * \mathbb{F} = -\mathrm{d}s \wedge *\frac{\partial A}{\partial t} + \mathrm{d}t \wedge *\frac{\partial A}{\partial s} - \mathrm{d}s \wedge \mathrm{d}t \wedge *F_A.$$

We then compute

$$\begin{aligned}\Phi &= \mathrm{d}s \wedge \varphi + *\varphi \\ &= \mathrm{d}s \wedge (\mathrm{d}t \wedge \omega + \mathrm{Re}(e^{-i\theta}\Upsilon)) + \frac{1}{2}\omega^2 - \mathrm{d}t \wedge \mathrm{Im}(e^{-i\theta}\Upsilon)\end{aligned}$$

is our natural Spin(7) form on $\mathbb{R}_s \times \mathbb{R}_t \times M^6$. We then see that

$$\begin{aligned}\mathbb{F} \wedge \Phi &= \left(\mathrm{d}s \wedge \frac{\partial A}{\partial s} + \mathrm{d}t \wedge \frac{\partial A}{\partial t} + F_A\right) \wedge \left(\mathrm{d}s \wedge \mathrm{d}t \wedge \omega + \mathrm{d}s \wedge \mathrm{Re}(e^{-i\theta}\Upsilon) + \frac{1}{2}\omega^2 - \mathrm{d}t \wedge \mathrm{Im}(e^{-i\theta}\Upsilon)\right) \\ &= \mathrm{d}s \wedge \left(\frac{\partial A}{\partial s} \wedge \frac{1}{2}\omega^2 + F_A \wedge + \mathrm{Re}(e^{-i\theta}\Upsilon)\right) + \mathrm{d}t \wedge \left(\frac{\partial A}{\partial t} \wedge \frac{1}{2}\omega^2 - F_A \wedge \mathrm{Im}(e^{-i\theta}\Upsilon)\right) \\ &\quad + \mathrm{d}s \wedge \mathrm{d}t \wedge \left(\frac{\partial A}{\partial s} \wedge \mathrm{Im}(e^{-i\theta}\Upsilon) + \frac{\partial A}{\partial t} \wedge (e^{-i\theta}\Upsilon) + F_A \wedge \omega\right).\end{aligned}$$

As usual, what looks like many equations is the same as just one, so we can look just at the dt term:

$$*\frac{\partial A}{\partial s} = \frac{\partial A}{\partial t} \wedge \frac{1}{2}\omega^2 - F_A \wedge \mathrm{Im}(e^{-i\theta}\Upsilon).$$

Re-arranging and noting that $*^2 = -1$ on 1-forms and that the complex structure J on M acts on 1-forms by

$$J(a) = *\left(a \wedge \frac{1}{2}\omega^2\right),$$

we deduce that

$$\frac{\partial A}{\partial s} + J\frac{\partial A}{\partial t} = *(F_A \wedge \mathrm{Im}(e^{-i\theta}\Upsilon)).$$

Of course, $\frac{\partial}{\partial s} + J\frac{\partial}{\partial t}$ is nothing other than the *Cauchy–Riemann operator* on $\mathbb{R}_s \times \mathbb{R}_t = \mathbb{C}_{s+it}$ and the right-hand side is the gradient of $\mathrm{Im}(e^{-i\theta}\mathcal{F}^{\mathbb{C}})$. We can therefore view our equation as a *perturbed J-holomorphic curve equation.*

Proposition 8.24. *There is a correspondence between* $\mathrm{Spin}(7)$*-instantons on* $\mathbb{R}_s \times \mathbb{R}_t \times M^6$ *for*

$$\Phi = \mathrm{d}s \wedge (\mathrm{d}t \wedge \omega + \mathrm{Re}(e^{-i\theta}\Upsilon)) + \frac{1}{2}\omega^2 - \mathrm{d}t \wedge \mathrm{Im}(e^{-i\theta}\Upsilon)$$

and solutions $A(s,t)$ *to the perturbed* J*-holomorphic curve equation*

$$\frac{\partial A}{\partial s} + J\frac{\partial A}{\partial t} = *(F_A \wedge \mathrm{Im}(e^{-i\theta}\Upsilon))$$

on $\mathbb{C}_{s+it} \times M^6$*.*

Remark. In the original Donaldson–Thomas paper, one obtains SU(4) instantons rather than just Spin(7) instantons because of the specializations they make there as discussed earlier.

Given all of this discussion, what is the upshot? The point is that one now can define a homology theory associated to $\mathcal{F}^{\mathbb{C}}$. The generators are gradient flow lines $A(t)$ and the differential is given by counting solutions $A(s,t)$ to the perturbed J-holomorphic curve equation with boundary conditions given by the gradient flows lines. In our setting, that means counting Spin(7) instantons $\mathbb{A}$ on $\mathbb{R}_s \times \mathbb{R}_t \times M$ which interpolate between two G_2 instantons $\mathbf{A}$ on $\mathbb{R}_t \times M$.

The input for this homology theory is therefore a pair of critical points A_-, A_+ for $\mathcal{F}^{\mathbb{C}}$, which are SU(3) instantons on M or, equivalently, stable holomorphic vector bundles (or coherent sheaves). The G_2 instantons $\mathbf{A}$ then interpolate between A_-, A_+, so we have our homology $H_*(A_-, A_+)$.

Altogether we obtain a *category*, now often called the *homology* or *Donaldson* category. The objects in the category are the critical points A of $\mathcal{F}^{\mathbb{C}}$, so the SU(3) instantons or stable bundles/sheaves on M, and the morphisms between objects A_-, A_+ in the category are the homology groups $H_*(A_-, A_+)$ given by the pair of critical points, whose generators are G_2 instantons and differential is defined by Spin(7) instantons.

In fact, this category has less structure than one would like, and so one defines a new category, often called a *Fukaya* or *Fukaya–Seidel* category, where one has the same critical points (so SU(3) instantons) but now the morphisms are not the homology groups $H_*(A_-, A_+)$ but the full chain complex $C_*(A_-, A_+)$, equipped with the differential ∂ given by counting Spin(7) instantons as we have discussed.

The hope then is that this category $\mathcal{C}(M^6)$ is then an *invariant* of the Calabi–Yau 3-fold M, which is independent of the choice of Calabi–Yau structure (ω, Υ) up to continuous deformation (or, perhaps, changes in some controlled way). One can then hope to extract other invariants, such as numbers, from this category. This is the main part of the overall picture suggested by Donaldson–Thomas for gauge theory in higher dimensions.

Remark. Th Fukaya–Seidel category is an example of an A_∞*-category*, which I won't explain, but means that the category has a lot more structure on it than I have explained.

9 Links between calibrated submanifolds and gauge theory

We have already seen started to see some hints about how calibrated submanifolds and gauge theory are related. In this final section of material in the course, we will discuss these relations in more detail, and see how this is related to key questions in the field.

9.1 Bubbling

When discussing the deformation theory problems for calibrated submanifolds or for the various notions of instantons we have introduced, we have focused on the issues of the dimension and smoothness (or otherwise) of the moduli space. These are *local* properties of the moduli space. However, when seeking to define invariants, as we hope to do, one has to be concerned with *global* properties of the moduli space. In particular, there is the key issue of *compactness*. More specifically, it is will typically be the case that the moduli space is noncompact, and so one has to understand what the appropriate notion of *compactification* should be.

The compactification problem is not one that is solved in general. However, we do know a lot about the potential issues that will arise to build the compactification, which turn out to be quite general. To begin with, I will start with the setting that we shall have throughout this section.

- Let (M^n, g) be a compact, oriented, n-dimensional Riemannian manifold.
- Let η be an $(n-4)$-form on M which is a calibration.
- Let E be a vector bundle over M with Euclidean metrics on its fibres and let $\mathcal{A}$ be the space of connections on E which are compatible with these metrics.

Remark. Not all of the assumptions on M and E are strictly necessary. One can allow M to be noncompact and for E to have noncompact gauge group, but then not all of the statements which I will make will hold, or will only hold under appropriate technical conditions.

We now want to define the class of connections for which we can examine the compactness problem.

Definition 9.1. A connection $A \in \mathcal{A}$ on (M, g, η) is an *η-instanton* if and only if

$$F_A \wedge \eta = - * F_A,$$

where F_A is the curvature of A. (Notice that the η-instanton condition makes sense precisely because η has degree $n-4$.)

We see that all of the instantons we have seen are η-instantons.

Example. If $n = 4$, then we can just take $\eta = 1$ and see that η-instantons are ASD instantons.

Example. If $n = 2m$ and (M, g) is Calabi–Yau with Kähler form ω then $\mathrm{SU}(m)$ instantons are η-instantons for $\eta = \omega^{m-2}/(m-2)!$.

Example. If $n = 7$ and (M, g) is a G_2 manifold with 3-form φ, then taking $\eta = \varphi$ gives that η-instantons are G_2 instantons.

Example. If $n = 8$ and (M, g) is a $\mathrm{Spin}(7)$ manifold with 4-form Φ, then taking $\eta = \Phi$ gives that η-instantons are $\mathrm{Spin}(7)$ instantons.

The usual arguments we have seen give the following.

Proposition 9.2. *An η-instanton is Yang–Mills and the Yang–Mills energy of any η-instanton is determined topologically by*

$$-\int_M \mathrm{tr}(F_A \wedge F_A \wedge \eta).$$

We then want to discuss of how compactness can fail. The results we describe are based on the combined work of a large number of people, including Uhlenbeck, Price, Nakajima, Tian, Tao–Tian and Smith–Uhlenbeck.

We begin with the following key result, which bears many similarities to the ASD case.

Theorem 9.3. *Let $(A_k)_{k\in\mathbb{N}} \subseteq \mathcal{A}$ be a sequence of η-instantons on E over M. After passing to a subsequence (which we still denoted by (A_k)), there is*

- *a closed subset $N \subseteq M$ of (Hausdorff) dimension at most $n-4$ with finite $(n-4)$-Hausdorff measure (so $\mathcal{H}^{n-4}(N) < \infty$),*
- *an η-instanton A_∞ on a bundle E_∞ on $M \setminus N$ and*
- *for any neighbourhood T of $N \subseteq M$ a sequence of isomorphisms $\rho_k : E_\infty|_{M\setminus T} \to E|_{M\setminus T}$*

such that

- *$\rho_k^* A_k \to A_\infty$ as $k \to \infty$,*
- *for $\mathcal{H}^{n-4}$-almost every point $p \in N$, the normal space $\nu_p(N) \cong \mathbb{R}^4$ exists and if we take a sequence of suitable rescalings around p, then $A_k|_{\nu_p(N)}$ converges to a nontrivial ASD instanton $B(p)$ on $\mathbb{R}^4 = \nu_p(N)$, and*
- *$|F_{A_k}|^2 \,\mathrm{vol}_M \to |F_{A_\infty}|^2 \,\mathrm{vol}_M + \delta_N$ as currents, where the support of the current δ_N is N.*

Remark. Here we notice a subtle point: it is not currently known in general that the bundles E_∞ and E are in fact isomorphic outside of N. This is why we introduce the neighbourhood T in the statement above. We can replace with T with N above in the case of $\mathrm{SU}(m)$ instantons.

This is what we mean by *bubbling*: along the subset N, we see ASD instantons concentrating on the normal spaces to N. We say that the ASD instanton $B(p)$ has "bubbled off" at p. We might also call N the *bubbling locus*.

Example. Let B_c be the standard ASD instanton on $\mathbb{R}^4$ with scale $c > 0$. Taking $c = k$ for $k \in \mathbb{N}$ and gives a sequence B_k of ASD instantons on $\mathbb{R}^4$ concentrating at 0. Letting $\pi : \mathbb{R}^7 \to \mathbb{R}^4$ be the projection onto a coassociative $\mathbb{R}^4$ and letting $A_k = \pi^* B_k$ gives a sequence of G_2 instantons on $\mathbb{R}^7$. We see that everything holds as in the theorem above where $M = \mathbb{R}^7$, A_∞ is the trivial flat connection and $N = \mathbb{R}^3$ is associative.

Given the example above, we would expect that the bubbling locus N has more structure, and indeed it does. We recall that for possibly singular objects like N we do not expect to obtain a tangent space when we zoom in at a point. However, we should expect to see a *tangent cone* since "zooming in" means acting by dilations, and so the object we see should be invariant under dilations. However, a fundamental result of Tian says that N is actually quite close to being smooth.

Theorem 9.4. *Recall the bubbling locus N above. For $\mathcal{H}^{n-4}$-almost all $p \in N$ we have that the tangent space T_pN exists and*

$$\eta|_{T_pN} = \mathrm{vol}_{T_pN},$$

i.e. the tangent spaces to N are η-calibrated.

Remark. In fact, N is rectifiable, so is essentially the union of a C^1-submanifold and a null set.

Example. For G_2 instantons, we have that the smooth part of the bubbling locus N is associative.

We should pause for a moment and notice a major difference with the ASD case: the η-instanton A_∞ we find is currently only defined on $M \setminus N$. In fact, we can do better as follows.

Theorem 9.5. *There is a closed subset $S \subseteq N$ with zero $(n-4)$-Hausdorff measure such that (E_∞, A_∞) extends to $M \setminus S$ such that A_∞ is an η-instanton.*

Remark. The set S is called the *singular* set of A_∞. In the $n = 4$ case, S is necessarily the empty set, which is a consequence of the removable singularities theorem of Uhlenbeck.

Finally, we should talk about the analogue of the energy conservation result we had for ASD instantons. Here, the answer isn't quite as nice, but is the following, where we need to recall that $\mathcal{E}$ is the Yang–Mills energy and $B(p)$ is the ASD instanton bubbling off at $p \in N$.

Theorem 9.6. *There is an upper semi-continuous function $\Theta : N \to (0, \infty)$ such that*

$$\Theta(p) \geq \mathcal{E}(B(p))$$

for almost all $p \in N$ and

$$\int_M |F_{A_k}|^2 \,\mathrm{vol}_M \to \int_M |F_{A_\infty}|^2 \,\mathrm{vol}_M + \int_N \Theta\eta|_N.$$

The reason why the function Θ appears is that more than just the ASD instanton $B(p)$ could bubble off. At different scales one could see more bubbles than just the biggest one: imagine a big sphere with a bunch of smaller spheres attached. The technical term is that one has a "*bubble tree*". The expectation is that $\Theta(p)$ should be the energy of the whole bubble tree which appears at p, and not just $\mathcal{E}(B(p))$.

Remark. Recall that if we have an $\mathrm{SU}(3)$ or G_2 monopole (A, ϕ) on a noncompact Calabi–Yau or G_2 manifold M, then we have a notion of mass $m > 0$ of the monopole, which is determined by the limit of $|\phi|$ at infinity. If we have a sequence of such monopoles $(A_k, \phi_k)_{k\in\mathbb{N}}$ with *fixed* mass $m_k = m$ then, under suitable technical assumptions due to the noncompactness of M, one expects to obtain the same bubbling results as above in the instanton setting.

However, there is an additional phenomenon which we can expect to see if the mass m_k is allowed to tend to infinity. In this case, we now should have an additional bubbling locus which is a rectifiable codimension 3 subset N along which we see monopoles on $\mathbb{R}^3$ bubbling off (at almost every point). Moreover, the smooth part of N should be special Lagrangian or coassociative depending on whether we are in the Calabi–Yau or G_2 setting, and the limiting monopole (A_∞, ϕ_∞) that we find has a Higgs field which should vanish along N.

These predictions are proved under some additional technical hypotheses by Fadel–Nagy–Oliveira, but are expected to hold in more generality.

9.2 Reverse bubbling for G_2 instantons

We now return to the problem that we have left open until now, which is the question of existence of G_2 instantons with "small", but nonabelian, gauge group G (which will be SO(3) for examples, but is in any case always compact and semisimple) on compact 7-manifolds with G_2 holonomy. The key idea is to try to "reverse" the bubbling process which we have described above.

Remark. Similar results will also hold for Spin(7) instantons on compact manifolds with Spin(7) holonomy, but we will restrict to the G_2 case just to streamline the discussion and because it is slightly simpler and further developed.

We start with the following data.

- Let (M^7, φ) be a compact G_2 manifold
- Let N be a compact associative 3-fold M.
- Let G be a compact semisimple Lie group and let E be a principal G-bundle over M.
- Let A be a connection on E which is G_2 instanton on (M, φ).

Remark. It may seem strange to assume the existence of a G_2 instanton when we want to construct one, but the point is that A could be a trivial example of a G_2 instanton. In fact, as we shall see, for the concrete examples we will take A to be a flat connection, which is easy to find.

Our goal then is to realize A as the limit of a 1-parameter family of G_2 instantons on M, where we bubble off ASD instantons $B(p)$ on $\mathbb{R}^4 = \nu_p(N)$ along the bubbling locus given by our initial choice of associative N.

As it stands, this is not a reasonable hope since, as we know, G_2 instantons are expected to be appear in 0-dimensional moduli spaces, not continuous families. To remedy this, we suppose the following.

- Let $(\varphi_t)_{t\in(-\epsilon,\epsilon)}$ for $\epsilon > 0$ be a 1-parameter family of closed 3-forms with $\varphi_0 = \varphi$ all inducing metrics with holonomy contained in G_2 (so they are parallel for the metrics they define).

Remark. Such a family is easy to find, given φ, since it can be defined by any choice of cohomology class in $H^3(M)$, by work of Joyce. More specifically, if we take a harmonic representative ζ of a class in $H^3(M)$ (using the metric defined by φ) then φ_t will be a small perturbation of $\varphi + t\zeta$ for t sufficiently small.

Given this family $(\varphi_t)_{t\in(-\epsilon,\epsilon)}$, our aim then is to try to find G_2 instantons on (M, φ_t) converging to A, at least for some values of t arbitrarily close to 0.

Remark. As we shall see, we will only find G_2 instantons t on one side of 0, i.e. for either $t > 0$ or $t < 0$ but not both.

We now want to study the "bubbles". To that end, let us suppose that we fix a moduli space $\mathcal{M}$ of framed ASD instantons on $\mathbb{R}^4$ with finite Yang–Mills energy on a principal G-bundle E_0. We then want to choose our ASD instanton "bubbles" $B(p) \in \mathcal{M}$ for $p \in N$.

The issue that we now have to face is: how do the ASD instantons $B(p)$ *vary* with $p \in N$? To understand this, we first make a construction.

Definition 9.7. We have a natural action of SO(4) on $\mathcal{M}$ and on the frame bundle $\mathrm{Fr}(\nu(N))$ of the normal bundle of N. We also have actions of the gauge group G on $\mathcal{M}$ and on $E|_N$. We can then define the *moduli bundle* $\mathcal{B}$ over N as an associated bundle as follows:

$$\mathcal{B} = (\mathrm{Fr}(\nu(N)) \times E|_N) \times_{\mathrm{SO}(4)\times\mathrm{G}} \mathcal{M}.$$

The fibres of $\mathcal{B}$ are each a copy of the moduli space $\mathcal{M}$ of ASD instantons on $\mathbb{R}^4$.

Remark. The point of the construction of $\mathcal{B}$ is that we coherently identify ASD instantons on $\mathbb{R}^4$ with connections on $\nu_p(N)$ and the bundles E_0 and $E|_N$ at each point $p \in N$.

Using this construction, our question becomes: how do we choose a distinguished section B of $\mathcal{B}$? As we might expect, it should be an elliptic first order PDE on B. Therefore, we need to understand the derivative of B. Since we have connections on $\nu(N)$ and $E|_N$, we have a natural connection and thus a covariant derivative $\nabla^\perp$ on sections of $\mathcal{B}$. If we have a vector field X on N and B is a section of $\mathcal{B}$ then we can ask, where does $\nabla^\perp_X B$ live? The answer is the following.

Definition 9.8. We define the *vertical moduli bundle* $\mathcal{V}$ as

$$\mathcal{V} = (\mathrm{Fr}(\nu(N)) \times E|_N) \times_{\mathrm{SO}(4)\times\mathrm{G}} T\mathcal{M},$$

which we can view naturally as a vector bundle over $\mathcal{B}$, where the fibres are tangent spaces to the moduli space $\mathcal{M}$ of ASD instantons.

Then, for a section B of $\mathcal{B}$ and a vector field X on N, we have that $\nabla^\perp_X B$ is a section of $B^*\mathcal{V}$.

The fibres of $\mathcal{V}$ are copies of $T\mathcal{M}$, on which we have a quaternionic structure (as we noted earlier) given by three orthogonal complex structures (J_1, J_2, J_3) satisfying the quaternionic relation $J_1J_2J_3 = -1$. Recall that, given an oriented orthonormal basis $\{e_1, e_2, e_3\}$ for N (which actually exists globally), we had a quaternionic structure on each normal space of N, so $\nu_p(N) \cong \mathbb{H}$, identifying cross product by e_1, e_2, e_3 with the action of J_1, J_2, J_3 on $\mathbb{H}$ (or i, j, k if you prefer).

Example. If we take the Dirac operator $\not{D}$ on $\nu(N)$, then it can be written as

$$\not{D}v = \sum_{i=1}^{3} e_i \times \nabla^\perp_{e_i} v = \sum_{i=1}^{3} J_i \nabla^\perp_{e_i} v$$

for sections v of $\nu(N)$.

With this in mind, we can clearly build the following operator on N.

Definition 9.9. The *Fueter operator* $\mathcal{F}$ is defined on sections B of $\mathcal{B}$ by

$$\mathcal{F}(B) = \sum_{i=1}^{3} J_i \nabla^\perp_{e_i} B \in \Gamma(B^*\mathcal{V}),$$

where $\{e_1, e_2, e_3\}$ is an oriented orthonormal frame on N. A solution B to

$$\mathcal{F}(B) = 0$$

is called a *Fueter section*.

Remark. You should think of the Fueter operator as a nonlinear version of the Dirac operator. It is nonlinear because the map B takes values in the manifold $\mathcal{M}$ at every point, rather than a vector space (as in the case of normal vector fields, for example).

An obvious question is: why do Fueter sections have anything to do with G_2 instantons? We can understand this as follows. At every point $p \in N$, we can identify $\nu_p(N) = \mathbb{R}^4$ and write $*\varphi$ at p as

$$*\varphi = \mathrm{vol}_{\mathbb{R}^4} - e^2 \wedge e^3 \wedge \omega_1 - e^3 \wedge e^1 \wedge \omega_2 - e^1 \wedge e^2 \wedge \omega_3,$$

where $\{e^1, e^2, e^3\}$ is an oriented orthonormal coframe at $p \in N$ and $\omega_1, \omega_2, \omega_3$ are the usual triple of orthogonal self-dual 2-forms on $\mathbb{R}^4 = \nu_p(N)$. Given a section B of $\mathcal{B}$ we can build a connection A_0 on $E|_N$ which so that $A_0|_{\nu_p(N)}$ is an ASD instanton for all p. Hence, we see that impose the G_2 instanton condition

$$F_{A_0} \wedge *\varphi = 0$$

will lead to the following.

- If we take components α of F_{A_0} at p which are 2-forms on $\nu_p(N) = \mathbb{R}^4$ then $\alpha \wedge *\varphi = 0$ is the statement that α is anti-self-dual. This is exactly guaranteed by the choice that B is a section of $\mathcal{B}$ and so $A_0|_{\nu_p(N)}$ is an ASD instanton.
- If we take components α of F_{A_0} at p which are combinations of 1-forms on $\nu_p(N)$ and T_pN, then one can check that $\alpha \wedge *\varphi = 0$ precisely gives that B is a Fueter section, i.e. $\mathcal{F}(B) = 0$.

Remark. One might ask about the third possibility that we take components of F_{A_0} which are 2-forms on N. These components do not play a role because when we rescale as the bubbles form these terms in $F_{A_0} \wedge *\varphi$ will automatically go to zero.

We are now have everything we need to reverse the bubbling construction. The idea is to construct a family of connections A_t which are approximately G_2 instantons using the Fueter section $\mathcal{B}$ for t small, then perturb them to genuine G_2 instantons using the Implicit Function Theorem. However, to make this strategy work, we need to know that all the operators we need to use are surjective. To do that, we make additional assumptions about N, A and B.

Recall the Dirac operator $\not{D}$ on N that we introduced when discussing the deformations of N and which we saw again above.

Definition 9.10. A compact associative N in (M, φ) is *unobstructed* if $\mathrm{coker}\not{D} = \{0\}$, i.e. $\not{D}$ is surjective. Since $\not{D}$ is self-adjoint, this is the same as saying that $\ker \not{D} = \{0\}$, i.e. N is rigid.

We suppose that N is unobstructed so that the linearized deformation operator $\not{D}$ is surjective.

We also need the following condition.

Definition 9.11. Let A be a G_2 instanton on E. Consider the elliptic self-dual complex

$$\Omega^0(M; \mathfrak{g}_E) \xrightarrow{\mathrm{d}_A} \Omega^1(M; \mathfrak{g}_E) \xrightarrow{*\varphi \wedge \mathrm{d}_A} \Omega^6(M; \mathfrak{g}_E) \xrightarrow{\mathrm{d}_A} \Omega^7(M; \mathfrak{g}_E),$$

where $\mathfrak{g}_E$ is the adjoint bundle. We say that A is *acyclic* if the cohomology of the above complex vanishes. In particular, this means that A is rigid and unobstructed.

We therefore suppose that A is acyclic, which in particular means that the linearized deformation operator $*\varphi \wedge \mathrm{d}_A$ surjects onto the kernel of d_A acting on $\mathfrak{g}_E$-valued 6-forms.

We are therefore only left with worrying about the surjectivity of the linearized Fueter operator $\mathcal{L} = \mathrm{d}\mathcal{F}_B$ at the Fueter section B. We observe that $\mathcal{L}$ is a self-adjoint elliptic operator on sections of $B^*\mathcal{V}$ (since it is a Dirac operator). As we now see, this means that $\mathcal{L}$ *cannot* be surjective.

Example. Suppose that B is a Fueter section. Since there is an action of rescaling on the fibres of $\nu(P)$ and on $\mathbb{R}^4$ preserving the ASD instanton condition, we have an action of dilation σ_λ on $\mathcal{B}$ for $\lambda > 0$. Then $\sigma_\lambda^* B$ is a Fueter section for all $\lambda > 0$, so Fueter sections always come in at least 1-parameter families.

The example above shows that $\mathcal{L}$ always has at least 1-dimensional kernel since it contains the element

$$v_B = \frac{\mathrm{d}}{\mathrm{d}\lambda} \sigma_\lambda^* B|_{\lambda=1},$$

which will be non-zero. Hence, $\mathcal{L}$ can not be surjective as it is self-adjoint. Therefore the best case scenario is that we suppose that

$$\dim \ker \mathcal{L} = 1$$

and thus $\ker \mathcal{L} = \mathrm{coker}\mathcal{L}$ is spanned by v_B.

We now just need one last piece of the puzzle to finish our discussion. Since we have assumed that A is acyclic (and thus rigid) and that N is rigid, we can apply the Implicit Function Theorem to deduce the following.

Lemma 9.12. *Making ϵ smaller if necessary, for all $t \in (-\epsilon, \epsilon)$ there exists a G_2 instanton A_t on E over (M, φ_t) and a compact associative N_t in (M, φ_t), which are deformations of $A_0 = A$ and $N_0 = N$.*

As a result, we can build moduli bundles $\mathcal{B}_t$ and Fueter operators $\mathcal{F}_t$. Since we have assumed that the linearization of $\mathcal{F} = \mathcal{F}_0$ has 1-dimensional kernel, which is spanned by v_B, we may deduce the following.

Lemma 9.13. *Let B be a Fueter section such that $\mathcal{L} = \mathrm{d}\mathcal{F}_B$ has 1-dimensional kernel. Then, making ϵ smaller if necessary, for all $t \in (-\epsilon, \epsilon)$ there are sections B_t with $B_0 = B$ of the moduli bundles $\mathcal{B}_t$ satisfying*

$$\mathcal{F}_t(B_t) = -\mu(t)\frac{\mathrm{d}}{\mathrm{d}\lambda}\sigma_\lambda^* B_t|_{\lambda=1}$$

for a smooth function $\mu : (-\epsilon, \epsilon) \to \mathbb{R}$ with $\mu(0) = 0$.

We now have our final key definition.

Definition 9.14. We say that a Fueter section B is *unobstructed* if

$$\dim\ker \mathrm{d}\mathcal{F}_B = 1 \quad \text{and} \quad \mu'(0) \neq 0.$$

In other words, that the function μ crosses through 0 transversely.

We therefore suppose we have an unobstructed Fueter section B.

The idea now is to glue the G_2 instantons A_t to the family of ASD instantons along N given by the unobstructed Fueter section B to give an approximate G_2 instanton. The sign of μ is then crucial: when μ is positive is when one can choose ASD instantons of non-zero size. These connections we get will be approximate G_2 instantons which are then perturbed to genuine G_2 instantons $\tilde{A}_t$ via the Implicit Function Theorem.

If we assume that $\mu'(0) > 0$, we obtain the following result, due to Walpuski.

Theorem 9.15. *Let $\epsilon > 0$ and suppose that*

- *(M^7, φ_t) is a smooth family of compact G_2 manifolds for $t \in (-\epsilon, \epsilon)$;*
- *N^3 is a compact unobstructed associative in (M, φ_0);*
- *A is an acyclic G_2 instanton on a principal G-bundle E over (M, φ_0);*
- *B is an unobstructed Fueter section with $\mu'(0) > 0$.*

Then, making ϵ smaller if necessary, there exist G_2 instantons $\tilde{A}_t$ on a bundle $\tilde{E}$ over (M, φ_t) for $t \in (0, \epsilon)$ such that $\tilde{A}_t \to A$ away from N as $t \to 0$ and, after rescaling, the ASD instanton $B(p)$ bubbles off at each $p \in N$ as $t \to 0$.

Remark. This theorem has the following disturbing consequence. We see that on the bundle $\tilde{E}$ a G_2 instanton on (M, φ_t) spontaneously appears as we cross from $t < 0$ to $t > 0$. This strongly suggests that a naive count of G_2 instantons on a compact G_2 manifold cannot produce an invariant, since the count changes by 1 as we cross $t = 0$ in the family of G_2 manifolds (M, φ_t). It also suggests that to build an invariant from G_2 instantons one also needs to take into account how many unobstructed associatives there are. This shows the complexity of the problem but also how inextricably linked calibrated geometry and gauge theory are in higher dimensions.

9.3 $\mathrm{G_2}$ instantons on compact $\mathrm{G_2}$ manifolds

We now want to use the ideas from the previous subsection to construct examples of G_2 instantons with structure group SO(3) on compact G_2 manifolds. It turns out that we have to modify the construction very slightly, but the idea is really the same.

To start with, we need to know which compact G_2 manifold we are looking at. This is provided by work of Joyce.

Consider the flat 7-torus T^7 with the standard 3-form φ_0 (which is defined on $\mathbb{R}^7$ but we can equally define it on flat T^7). There is a finite subgroup $\Gamma \cong \mathbb{Z}_2^3$ of G_2 acting on T^7 by isometries preserving φ_0 so that $M_0 = T^7/\Gamma$ is a flat G_2 orbifold and so that the orbifold locus (which are the fixed points of Γ) is a compact associative 3-fold N.

The associative 3-fold N is a disjoint union of a finite number of flat totally geodesic T^3's and $T^3/\mathbb{Z}_2$'s. At each point $p \in N$, the normal space $\nu_p(N) = \mathbb{R}^4/\mathbb{Z}_2$ (which is not surprising as N is the orbifold locus).

Now, we know that $T^*\mathcal{S}^2$ is a hyperkähler 4-manifold which is asymptotic to $\mathbb{R}^4/\mathbb{Z}_2$ and that $N \times T^*\mathcal{S}^2$ has a natural 3-form which defines a metric with holonomy contained in G_2. In fact, there is a 1-parameter family of such 3-forms for $t > 0$ given by:

$$\varphi_t^N = \mathrm{d}\theta_1 \wedge \mathrm{d}\theta_2 \wedge \mathrm{d}\theta_3 - t^2 \sum_{i=1}^{3} \mathrm{d}\theta_i \wedge \omega_i,$$

where $(\theta_1, \theta_2, \theta_3)$ are standard coordinates on N and $\{\omega_1, \omega_2, \omega_3\}$ is the triple of orthogonal closed self-dual 2-forms on $T^*\mathcal{S}^2$ defining the hyperkähler structure. Here t^2 corresponds to the size of the $\mathcal{S}^2$ in $T^*\mathcal{S}^2$.

Given this, we can glue $T^*\mathcal{S}^2$ along N to give a smooth 7-manifold M and glue φ_0 to φ_t^N for each $t > 0$ to get a closed 3-form φ_t which is approximately coclosed. Fundamental work of Joyce, using a very delicate perturbation argument, gives the following.

Theorem 9.16. *For all $t \in (0, \epsilon)$ for $\epsilon > 0$ sufficiently small, there is a small perturbation $\tilde{\varphi}_t$ of φ_t on M, which lies in the same cohomology class in $H^3(M)$, such that $(M, \tilde{\varphi}_t)$ is a compact G_2 manifold with holonomy equal to G_2.*

We now see how to modify the reverse bubbling construction we had above. We have the compact unobstructed associative N in M_0 because of its description as the orbifold locus. We can find flat SO(3) connections A on M_0 which are acyclic using appropriate elements of the orbifold fundamental group. We then just need to replace $\mathbb{R}^4$ by $T^*\mathcal{S}^2$ and consider moduli spaces $\mathcal{M}$ of ASD instantons on $T^*\mathcal{S}^2$ instead. In fact, there is a moduli space $\mathcal{M}$ of ASD instantons on $T^*\mathcal{S}^2$ with structure group SO(3) which consists of a single point B. Hence, the Fueter section that we pick is just the constant section B, which is then automatically unobstructed for trivial reasons.

Altogether we can apply the same ideas as in the reverse bubbling to get the following result, also due to Walpuski.

Theorem 9.17. *Making $\epsilon > 0$ smaller if necessary, for all $t \in (0, \epsilon)$ there is an irreducible G_2 instanton $\tilde{A}_t$ on a principal SO(3)-bundle $\tilde{E}$ over $(M, \tilde{\varphi}_t)$.*

Remark. We also see that the G_2 instantons bubble off the ASD instanton B along N as $t \to 0$ and $\tilde{A}_t$ converges to the original flat connection A away from N.

9.4 Mirror Symmetry

In this final subsection, I want to discuss one more link between calibrated geometry and gauge theory which is motivated by *Mirror Symmetry.*

I will discuss the simplest version of Mirror Symmetry which says that compact Calabi–Yau 3-folds can come in mirror pairs: $(M_{\mathrm{A}}, \omega_{\mathrm{A}}, \Upsilon_{\mathrm{A}})$ and $(M_{\mathrm{B}}, \omega_{\mathrm{B}}, \Upsilon_{\mathrm{B}})$. The idea is that the symplectic geometry of $(M_{\mathrm{A}}, \omega_{\mathrm{A}})$ (with the choice of holomorphic volume Υ_{A}) is "exchanged" or is related to the complex geometry of $(M_{\mathrm{B}}, \Upsilon_{\mathrm{B}})$ (with choice of Kähler form ω_{B}). The two sides are called the A-side or A-model and the B-side or B-model, respectively.

Under Mirror Symmetry, certain calibrated geometry on the A-side is supposed to correspond to certain gauge theory on the B-side. More specifically, on the A-side one is supposed to think about special Lagrangian submanifolds N, which are defined using the symplectic structure ω_{A}, since this defines the Lagrangians, plus the choice of Υ_{A}. On the B-side, instead one is supposed to think about Hermitian Yang–Mills connections A, particularly on degree 0 bundles which are then SU(3) instantons, but in any case are defined primarily using the complex geometry of Υ_{B}, as well as with the Kähler form ω_{B}.

The easiest case is where one takes an SU(3) instanton A on the B-side on a complex line bundle with gauge group U(1). In this case, one is supposed to look at compact special Lagrangians N on the A-side. However, it is easy to see that this correspondence cannot make sense because the SU(3) instanton A has a complex moduli space, whereas the moduli space of N is definitely not complex.

However, there is a natural way to complexify the space of special Lagrangians, and that is to think of pairs (N, B) where N is a compact special Lagrangian and B is a flat U(1)-connection on N. Remember that the moduli space of deformations of N has dimension $b^1(N)$ and deformations of flat U(1)-connections are also locally described by $b^1(N)$, so at least the numbers make sense. One can argue more convincingly why the space of pairs (N, B) is a complexification of the special Lagrangian moduli space, but I will not discuss this.

The proposal then is that the space of pairs (N, B) of a compact special Lagrangian and a flat U(1)-connection is related to the space of SU(3) instantons on the mirror Calabi–Yau 3-fold in some way.

Let us consider the simplest possible case of all which is when $M_A = M_B = \mathbb{C}^3$. Of course, this is not compact, but you can just replace $\mathbb{C}^3$ by flat torus if you like. Let us suppose that N is a special Lagrangian which is a graph of $u : \mathbb{R}^3 \to \mathbb{R}^3$, i.e.

$$N = \{x + iu(x) \;\; x \in \mathbb{R}^3\}.$$

Suppose that B is a flat U(1) connection on N. Then we can pullback B via u to be a flat connection on $\mathbb{R}^3$. Since the bundle over $\mathbb{R}^3$ is trivial, we can write this pullback of B, which we still call B, as

$$B = ib$$

for a real closed 1-form b on $\mathbb{R}^3$.

Now, what is Mirror Symmetry suppose to do to the pair (N, B) or, equivalently, the pair (u, b)? Well, if we choose coordinates (x_1, x_2, x_3) on $\mathbb{R}^3$ and we write

$$u = (u_1, u_2, u_3) \quad \text{and} \quad b = b_1 \mathrm{d}x_1 + b_2 \mathrm{d}x_2 + b_3 \mathrm{d}x_3$$

for functions $u_i : \mathbb{R}^3 \to \mathbb{R}$, then the "mirror" U(1)-connection A on $\mathbb{C}^3$ should be given by

$$A = ib_1 \mathrm{d}x_1 + ib_2 \mathrm{d}x_2 + ib_3 \mathrm{d}x_3 + iu_1 \mathrm{d}y_1 + iu_2 \mathrm{d}y_2 + iu_3 \mathrm{d}y_3$$

where $(x_1 + iy_1, x_2 + iy_2, x_3 + iy_3)$ are the complex coordinates on $\mathbb{C}^3$.

So, is A an SU(3) instanton? Well, one can can check that the pair (u, b) defines and pair (N, B) of a special Lagrangian and a flat U(1) connection on N if and only if A as defined satisfies

$$F_A \wedge \Upsilon = 0 \quad \text{and} \quad \mathrm{Im}(\omega + F_A)^3 = 0.$$

The second equation is equivalent to

$$F_A \wedge \frac{\omega^2}{2} + \frac{1}{6} F_A^3 = 0.$$

We see that this is *not* the same as the Hermitian Yang–Mills or SU(3) instanton conditions where the $\frac{1}{6} F_A^3$ term is removed. These connections have a name and are called *deformed Hermitian Yang–Mills connections*. They can be defined more generally on Kähler manifolds and have received quite a lot of study recently amid hopes that one can obtain a similar theory to Hermitian Yang–Mills connections, for example a Donaldson–Uhlenbeck–Yau-type theorem or Hitchin–Kobayashi correspondence.

References

[1] S. Brendle, *Embedded minimal tori in S^3 and the Lawson conjecture*, Acta Math. **211** (2013), no. 2, 177–190.

[2] R. L. Bryant, *Submanifolds and special structures on the octonians*, J. Differential Geom. **17** (1982), 185–232.

[3] R. L. Bryant, *Calibrated embeddings in the special Lagrangian and coassociative cases*, Ann. Global Anal. Geom. **18** (2000), no. 3-4, 405–435.

[4] R. L. Bryant, *Second order families of special Lagrangian 3-folds*, Perspectives in Riemannian geometry, pp. 63–98, CRM Proc. Lecture Notes **40**, Amer. Math. Soc., Providence, RI, 2006.

[5] R. L. Bryant and S. Salamon, *On the construction of some complete metrics with exceptional holonomy*, Duke Math. J. **58** (1989), 829–850.

[6] A. Butscher, *Deformations of minimal Lagrangian submanifolds with boundary*, Proc. Amer. Math. Soc. **131** (2002) 1953–1964.

[7] A. Butscher, *Regularizing a singular special Lagrangian variety*, Comm. Anal. Geom. **12** (2004), 733–791.

[8] E. Carberry, *Associative cones in the imaginary octonions*, Riemann surfaces, harmonic maps and visualization, pp. 249–263, OCAMI Stud. **3**, Osaka Munic. Univ. Press, Osaka, 2010.

[9] E. Carberry and I. McIntosh, *Minimal Lagrangian 2-tori in $\mathbb{CP}^2$ come in real families of every dimension*, J. London Math. Soc. (2) **69** (2004), no. 2, 531–544.

[10] A. Corti, M. Haskins, J. Nordström and T. Pacini, G_2*-manifolds and associative submanifolds via semi-Fano 3-folds*, Duke Math. J. **164** (2015), no. 10, 1971–2092.

[11] T. Eguchi and A. J. Hanson, *Asymptotically flat self-dual solutions to Euclidean gravity*, Phys. Lett. **74B** (3) (1978), 249–251.

[12] D. Fox, *Second order families of coassociative 4-folds*, Thesis (Ph.D.) Duke University, ProQuest LLC, Ann Arbor, MI, 2005.

[13] D. Fox, *Coassociative cones ruled by 2-planes*, Asian J. Math. **11** (2007), no. 4, 535–553.

[14] D. Fox, *Cayley cones ruled by 2-planes: desingularization and implications of the twistor fibration*, Comm. Anal. Geom. **16** (2008), no. 5, 937–968.

[15] D. Gayet, *Smooth moduli spaces of associative submanifolds*, Q. J. Math. **65** (2014), no. 4, 1213–1240.

[16] D. Gayet and F. Witt, *Deformations of associative submanifolds with boundary*, Adv. Math. **226** (2011), no. 3, 2351–2370.

[17] E. Goldstein, *Calibrated fibrations on noncompact manifolds via group actions*, Duke Math. J. **110** (2001), no. 2, 309–343.

[18] A. Gray, E. Abbena and S. Salamon, *Modern differential geometry of curves and surfaces with Mathematica®*, Third edition, Studies in Advanced Mathematics, Chapman & Hall/CRC, Boca Raton, FL, 2006.

[19] R. Harvey, *Spinors and calibrations*, Perspectives in Mathematics **9**, Academic Press, Inc., Boston, MA, 1990.

[20] R. Harvey and H. B. Lawson, *Calibrated geometries*, Acta Math. **148** (1982), 47–157.

[21] M. Haskins, *Special Lagrangian cones*, Amer. J. Math. **126** (2004), no. 4, 845–871.

[22] M. Haskins, *The geometric complexity of special Lagrangian T^2-cones*, Invent. Math. **157** (2004), no. 1, 11–70.

[23] M. Haskins and N. Kapouleas, *Special Lagrangian cones with higher genus links*, Invent. Math. **167** (2007), 223–294.

[24] G. Huisken and T. Ilmanen, *The inverse mean curvature flow and the Riemannian Penrose inequality*, J. Differential Geom. **59** (2001), no. 3, 353–437.

[25] Y. Imagi, D. D. Joyce and J. Oliveira dos Santos, *Uniqueness results for special Lagrangians and Lagrangian mean curvature flow expanders in* $\mathbb{C}^m$, Duke Math. J. **165** (2016), no. 5, 847–933.

[26] M. Ionel, *Second order families of special Lagrangian submanifolds in* $\mathbb{C}^4$, J. Differential Geom. **65** (2003), no. 2, 211–272.

[27] M. Ionel, S. Karigiannis and M. Min-Oo, *Bundle constructions of calibrated submanifolds in* $\mathbb{R}^7$ *and* $\mathbb{R}^8$, Math. Res. Lett. **12** (2005), no. 4, 493–512.

[28] M. Ionel and M. Min-Oo, *Cohomogeneity one special Lagrangian 3-folds in the deformed and the resolved conifolds*, Illinois J. Math. **52** (2008), no. 3, 839–865.

[29] K. Irie, F. Marques and A. Neves, *Density of minimal hypersurfaces for generic metrics*, Ann. of Math. (2) **187** (2018), no. 3, 963–972.

[30] D. D. Joyce, *Evolution equations for special Lagrangian 3-folds in* $\mathbb{C}^3$, Ann. Global Anal. Geom. **20** (2001), no. 4, 345–403.

[31] D. D. Joyce, *Constructing special Lagrangian m-folds in* $\mathbb{C}^m$ *by evolving quadrics*, Math. Ann. **320** (2001), no. 4, 757–797.

[32] D. D. Joyce, *Ruled special Lagrangian 3-folds in* $\mathbb{C}^3$, Proc. London Math. Soc. (3) **85** (2002), no. 1, 233–256.

[33] D. D. Joyce, *Special Lagrangian m-folds in* $\mathbb{C}^m$ *with symmetries*, Duke Math. J. **115** (2002), no. 1, 1–51.

[34] D. D. Joyce, $U(1)$*-invariant special Lagrangian 3-folds in* $\mathbb{C}^3$ *and special Lagrangian fibrations*, Turkish J. Math. **27** (2003), no. 1, 99–114.

[35] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. V. Survey and applications*, J. Differential Geom. **63** (2003), 279–347.

[36] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. II. Moduli spaces*, Ann. Global Anal. Geom. **25** (2004), 301–352.

[37] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. III. Desingularization, the unobstructed case*, Ann. Global Ann. Geom. **26** (2004), 1–58.

[38] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. IV. Desingularization, obstructions and families*, Ann. Global Ann. Geom. **26** (2004), 117–174.

[39] D. D. Joyce, $U(1)$*-invariant special Lagrangian 3-folds. I. Nonsingular solutions*, Adv. Math. **192** (2005), no. 1, 35–71.

[40] D. D. Joyce, $U(1)$*-invariant special Lagrangian 3-folds. II. Existence of singular solutions*, Adv. Math. **192** (2005), no. 1, 72–134.

[41] D. D. Joyce, $U(1)$*-invariant special Lagrangian 3-folds. III. Properties of singular solutions*, Adv. Math. **192** (2005), no. 1, 135–182.

[42] D. D. Joyce, *Riemannian holonomy groups and calibrated geometry*, Oxford Graduate Texts in Mathematics **12**, Oxford University Press, Oxford, 2007.

[43] D. D. Joyce, *Special Lagrangian 3-folds and integrable systems*, Surveys on geometry and integrable systems, pp. 189–233, Adv. Stud. Pure Math. **51**, Math. Soc. Japan, Tokyo, 2008.

[44] D. D. Joyce and S. Salur, *Deformations of asymptotically cylindrical coassociative submanifolds with fixed boundary*, Geom. Topol. **9** (2005), 1115–1146.

[45] S. Karigiannis and N. C.-H. Leung, *Deformations of calibrated subbundles of Euclidean spaces via twisting by special sections*, Ann. Global Anal. Geom. **42** (2012), no. 3, 371–389.

[46] S. Karigiannis and M. Min-Oo, *Calibrated subbundles in noncompact manifolds of special holonomy*, Ann. Global Anal. Geom. **28** (2005), no. 4, 371–394.

[47] A. G. Kovalev, *Coassociative* $K3$ *fibrations of compact* G_2*-manifolds*, arXiv:math/0511150.

[48] A. Kovalev and J. D. Lotay, *Deformations of compact coassociative 4-folds with boundary*, J. Geom. Phys. **59** (2009), 63–73.

[49] G. Lawlor, *The angle criterion*, Invent. Math. **95** (1989), 437–446.

[50] H. B. Lawson and R. Osserman, *Non-existence, non-uniqueness and irregularity of solutions to the minimal surface system*, Acta Math. **139** (1977), 1–17.

[51] Y.-I. Lee, *Embedded special Lagrangian submanifolds in Calabi–Yau manifolds*, Comm. Anal. Geom. **11** (2003), 391–423.

[52] J. D. Lotay, *Constructing associative 3-folds by evolution equations*, Comm. Anal. Geom. **13** (2005), 999–1037.

[53] J. D. Lotay, 2*-ruled calibrated* 4*-folds in* $\mathbb{R}^7$ *and* $\mathbb{R}^8$, J. London Math. Soc. **74** (2006), 219–243.

[54] J. D. Lotay, *Calibrated submanifolds of* $\mathbb{R}^7$ *and* $\mathbb{R}^8$ *with symmetries*, Q. J. Math. **58** (2007), 53–70.

[55] J. D. Lotay, *Coassociative 4-folds with conical singularities*, Comm. Anal. Geom. **15** (2007), 891–946.

[56] J. D. Lotay, *Deformation theory of asymptotically conical coassociative 4-folds*, Proc. London Math. Soc. **99** (2009), 386–424.

[57] J. D. Lotay, *Desingularization of coassociative 4-folds with conical singularities*, Geom. Funct. Anal. **18** (2009), 2055–2100.

[58] J. D. Lotay, *Asymptotically conical associative 3-folds*, Q. J. Math. **62** (2011), 131–156.

[59] J. D. Lotay, *Ruled Lagrangian submanifolds of the 6-sphere*, Trans. Amer. Math. Soc. **363** (2011), 2305–2339.

[60] J. D. Lotay, *Associative submanifolds of the 7-sphere*, Proc. London Math. Soc. **105** (2012), 1183–1214.

[61] J. D. Lotay, *Stability of coassociative conical singularities*, Comm. Anal. Geom. **20** (2012), 803–867.

[62] J. D. Lotay, *Desingularization of coassociative 4-folds with conical singularities: obstructions and applications*, Trans. Amer. Math. Soc. **366** (2014), 6051–6092.

[63] F. C. Marques and A. Neves, *Min-max theory and the Willmore conjecture*, Ann. of Math. (2) **179** (2014), no. 2, 683–782.

[64] F. C. Marques and A. Neves, *Existence of infinitely many minimal hypersurfaces in positive Ricci curvature*, Invent. Math. **209** (2017), no. 2, 577–616.

[65] I. McIntosh, *Special Lagrangian cones in* $\mathbb{C}^3$ *and primitive harmonic maps*, J. London Math. Soc. (2) **67** (2003), no. 3, 769–789.

[66] R. C. McLean, *Deformations of calibrated submanifolds*, Comm. Anal. Geom. **6** (1998), 705–747.

[67] K. Moore, *Cayley deformations of compact complex surfaces*, arXiv:1710.08799.

[68] K. Moore, *Deformations of conically singular Cayley submanifolds*, J. Geom. Anal. (2018).

[69] C. B. Morrey, *Multiple Integrals in the Calculus of Variations*, Grundlehren Series Volume 130, Springer–Verlag, Berlin, 1966.

[70] M. Ohst, *Deformations of compact Cayley submanifolds with boundary*, arXiv:1405.7886.

[71] M. Ohst, *Deformations of asymptotically cylindrical Cayley submanifolds*, arXiv:1506.00110.

[72] T. Pacini, *Special Lagrangian conifolds, I: moduli spaces*, Proc. London Math. Soc. (3) **107** (2013), 198–224.

[73] T. Pacini, *Special Lagrangian conifolds, II: gluing constructions in* $\mathbb{C}^m$, Proc. London Math. Soc. (3) **107** (2013), no. 2, 225–266.

[74] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, arXiv:math/0307245.

[75] R. Schoen and S. T. Yau, *On the proof of the positive mass conjecture in general relativity*, Comm. Math. Phys. **65** (1979), no. 1, 45–76.

[76] R. Schoen and S. T. Yau, *Positive scalar curvature and minimal hypersurface singularities*, arXiv:1704.05490.

[77] A. Song, *Existence of infinitely many minimal hypersurfaces in closed manifolds*, arXiv:1806.08816.

[78] M. B. Stenzel, *Ricci-flat metrics on the complexification of a compact rank one symmetric space*, Manuscripta Math. **80** (1993), no. 2, 151–163.