Notes on Seiberg-Witten Theory

Liviu I. Nicolaescu

Graduate Studies in Mathematics

Volume 28

American Mathematical Society
Providence, Rhode Island

Contents

Introduction

> *My task which I am trying to achieve is by the power of the written word, to make you hear, to make you feel - it is, before all, to make you* see. *That - and no more, and it is everything.*
>
> Joseph Conrad

Almost two decades ago, a young mathematician by the name of Simon Donaldson took the mathematical world by surprise when he discovered some "pathological" phenomena concerning *smooth* 4-manifolds. These pathologies were caused by certain behaviours of instantons, solutions of the Yang-Mills equations arising in the physical theory of gauge fields.

Shortly after, he convinced all the skeptics that these phenomena represented only the tip of the iceberg. He showed that the moduli spaces of instantons often carry nontrivial and surprising information about the background manifold. Very rapidly, many myths were shattered.

A flurry of work soon followed, devoted to extracting more and more information out of these moduli spaces. This is a highly nontrivial job, requiring ideas from many branches of mathematics. Gauge theory was born and it is here to stay.

In the fall of 1994, the physicists N. Seiberg and E. Witten introduced to the world a new set of equations which according to physical theories had to contain the same topological information as the Yang-Mills equations.

From an analytical point of view these new equations, now known as the Seiberg-Witten equations, are easier to deal with than the Yang-Mills equations. In a matter of months many of the results obtained by studying instantons were re-proved much faster using the new theory. (To be perfectly honest, the old theory made these new proofs possible since it created the right mindset to think about the new equations.) The new theory goes one step further, since it captures in a more visible fashion the interaction geometry-topology.

The goal of these notes is to help the potential reader share some of the excitement afforded by this new world of gauge theory and eventually become a player him/herself.

There are many difficulties to overcome. To set up the theory one needs a substantial volume of information. More importantly, all this volume of information is processed in a nontraditional way which may make the first steps in this new world a bit hesitant. Moreover, the large and fast-growing literature on gauge theory, relying on a nonnegligible amount of "folklore"[1], may look discouraging to a beginner.

To address these issues within a reasonable space we chose to present a few, indispensable, key techniques and as many relevant examples as possible. That is why these notes are far from exhaustive and many notable contributions were left out. We believe we have provided enough background and intuition for the interested reader to be able to continue the Seiberg-Witten journey on his/her own.

It is always difficult to resolve the conflict clarity vs. rigor and even much more so when presenting an eclectic subject such as gauge theory. The compromises one has to make are always biased and thus may not satisfy all tastes and backgrounds. We could not escape this bias, but whenever a proof would have sent us far astray we tried to present all the main concepts and ideas in as clear a light as possible and make up for the missing details by providing generous references. Many technical results were left to the reader as exercises but we made sure that all the main ingredients can be found in these notes.

Here is a description of the content. The first chapter contains preliminary material. It is clearly incomplete and cannot serve as a substitute for a more thorough background study. We have included it to present in the nontraditional light of gauge theory many classical objects which may already be familiar to the reader.

The study of the Seiberg-Witten equations begins in earnest in Chapter 2. In the first section we introduce the main characters: the monopoles, i.e. the solutions of the Seiberg-Witten equations and the group of gauge transformations, an infinite dimensional Abelian group acting on the set of monopoles. The Seiberg-Witten moduli space and its structure are described in Section 2.2 while the Seiberg-Witten invariants are presented in Section 2.3. We have painstakingly included all the details concerning orientations

[1] That is, basic facts and examples every expert knows and thus are only briefly or not at all explained in a formal setting. They are usually transmitted through personal interactions.

because this is one of the most confusing aspects of the theory. We conclude this chapter with two topological applications: the proof by P. Kronheimer and T. Mrowka of the Thom conjecture for $\mathbb{CP}^2$ and the new proof based on monopoles of Donaldson's first theorem, which started this new field of gauge theory.

In Chapter 3 we concentrate on a special, yet very rich, class of smooth 4-manifolds, namely the algebraic surfaces. It was observed from the very beginning by E. Witten that the monopoles on algebraic surfaces can be given an explicit algebraic-geometric description, thus opening the possibility of carrying out many concrete computations. The first section of this chapter is a brief and informal survey of the geometry and topology of complex surfaces together with a large list of examples. In Section 3.2 we study in great detail the Seiberg-Witten equations on Kähler surfaces and, in particular, we prove Witten's result stating the equivalence between the Seiberg-Witten moduli spaces and certain moduli spaces of divisors. The third section is devoted entirely to applications. We first prove the nontriviality of the Seiberg-Witten invariants of a Kähler surface and establish the invariance under diffeomorphisms of the canonical class of an algebraic surface of general type. We next concentrate on simply connected elliptic surfaces. We compute all their Seiberg-Witten invariants following an idea of O. Biquard based on the factorization method of E. Witten. This computation allows us to provide the complete *smooth* classification of simply connected elliptic surfaces. In §3.3.3, we use the computation of the Seiberg-Witten invariants of $K3$-surfaces to show that the smooth h-cobordism theorem fails in four dimensions. We conclude this section and the chapter with a discussion of the Seiberg-Witten invariants of symplectic 4-manifolds and we prove Taubes' theorem on the nontriviality of these invariants in the symplectic world.

The fourth and last chapter is by far the most technically demanding one. We present in great detail the cut-and-paste technique for computing Seiberg-Witten invariants. This is a very useful yet difficult technique but the existing written accounts of this method can be unbalanced as regards their details. In this chapter we propose a new approach to this technique which in our view has several conceptual advantages and can be easily adapted to other problems as well. Since the volume of technicalities can often obscure the main ideas we chose to work in a special yet sufficiently general case when the moduli spaces of monopoles on the separating 3-manifold are, roughly speaking, Bott nondegenerate.

Section 4.1 contains preliminary material mostly about elliptic equations on manifolds with cylindrical ends. Most objects on closed manifolds have cylindrical counterparts which often encode very subtle features. We discovered that a consistent use of cylindrical notions is not only æsthetically desirable, but also technically very useful. The cylindrical context highlights and coherently organizes many important and not so obvious aspects of the whole gluing problem. An important result in this section is the Cappell-Lee-Miller gluing theorem. We adapt the asymptotic language of [**110**], which is extremely convenient in gluing problems. This section ends with the long subsection §4.1.6 containing many useful and revealing examples. These are frequently used in gauge theory and we could not find any satisfactory reference for them.

In Section 4.2 we study the finite energy monopoles on cylindrical manifolds. The results are very similar to the ones in Yang-Mills equations and that is why this section was greatly inspired by [**96, 133**].

Section 4.3 is devoted to the local study of the moduli spaces of finite energy monopoles. The local structure is formally very similar to that in Yang-Mills theory with a notable exception, the computation of the virtual dimensions, which is part of the folklore. We present in detail this computation since it is often relevant. Moreover, we describe some new exact sequences relating the various intervening deformation complexes to objects covered by the Cappell-Lee-Miller gluing theorem. These exact sequences represent a departure from the mainstream point of view and play a key role in our local gluing theorem.

Section 4.4 is devoted to the study of global properties of the moduli spaces of finite energy monopoles: generic smoothness, compactness (or lack thereof) and orientability. The orientability is no longer an elementary issue in the noncompact case and we chose to present a proof of this fact only in some simpler situations we need for applications.

Section 4.5 contains the main results of this chapter dealing with the process of reconstructing the space of monopoles on a 4-manifold decomposed into several parts by a hypersurface. This manifold decomposition can be analytically simulated by a neck stretching process. During this process, the Seiberg-Witten equations are deformed and their solutions converge to a singular limit. The key issue to be resolved is whether this process can be reversed: given a singular limit can we produce monopoles converging to this singular limit?

In his dissertation [**99**], T. Mrowka proved a very general gluing theorem which provides a satisfactory answer to the above question in the related context of Yang-Mills equations. In §4.5.2, we prove a local gluing theorem, very similar in spirit to Mrowka's theorem but in an entirely new context. The main advantage of the new approach is that all the spectral estimates needed in the proof follow immediately from the Cappell-Lee-Miller gluing theorem. Moreover, the Mayer-Vietoris type local model is just a reformulation of the Cappell-Lee-Miller theorem. The asymptotic language of [**110**] has allowed us to provide intuitive, natural and explicit descriptions of the various morphisms entering into the definition of this Mayer-Vietoris model.

The local gluing theorem we prove produces monopoles converging to a singular limit at a certain rate. If all monopoles degenerated to the singular limit set at this rate then we could conclude that the entire moduli space on a manifold with a sufficiently long neck can be reconstructed from the local gluing constructions. This issue of the surjectivity of the gluing construction is conspicuously missing in the literature and it is quite nontrivial in non-generic situations. We deal with it in §4.5.3 by relying on Łojasewicz's inequality in real algebraic geometry.

In §4.5.4 we prove two global gluing theorems, one in a generic situation and the other one in a special, obstructed setting.

Section 4.6 contains some simple topological applications of the gluing technique. We prove the connected sum theorem and the blow-up formula. Moreover, we present a new and very short proof of a vanishing theorem of Fintushel and Stern.

These notes were written with a graduate student in mind but there are many new points of view to make it interesting for experts as well (especially our new approach to the gluing theorem). The minimal background needed to go through these notes is a knowledge of basic differential geometry, algebraic topology and some familiarity with fundamental facts concerning elliptic partial differential equations. The list of contents for Chapter 1 can serve as background studying guide.

* * *

Personal note. I have spent an exciting time of my life thinking and writing these notes and I have been supported along the way by many people.

The book grew out of a year long seminar at McMaster University and a year long graduate course I taught at the University of Notre Dame. I want to thank the participants at the seminar and the course for their patience, interest, and most of all, for their many useful questions and comments.

These notes would perhaps not have seen the light of day were it not for Frank Connolly's enthusiasm and curiosity about the subject of gauge theory which have positively affected me, personally and professionally. I want to thank him for the countless hours of discussions, questions and comments which helped me crystallize many of the ideas in the book.

For the past five years, I have been inspired by Arthur Greenspoon's passion for culture in general, and mathematics in particular. His interest in these notes kept my enthusiasm high. I am greatly indebted to him for reading these notes, suggesting improvements and correcting my often liberal use of English language and punctuation.

While working on these notes I benefited from the conversations with Andrew Sommese, Stefan Stolz and Larry Taylor, who patiently answered my sometimes clumsily formulated questions and helped clear the fog.

My wife has graciously accepted my long periods of quiet meditation or constant babbling about gauge theory. She has been a constant source of support in this endeavor. I want to thank my entire family for being there for me.

Notre Dame, Indiana 1999

Hence, for large L, the trivial closed 2-form lives in the positive chamber $\mathcal{N}_{\sigma_0, g_L}$ because

$$\langle 0 - \frac{1}{\|\omega_L\|}\omega_L, 2\pi c(\sigma_0)\rangle > 0.$$

Since $\mathbf{sw}_+(\sigma_0) \neq 0$ the above conclusion implies that for all large L there exist $(g_L, 0)$-monopoles.

2.4. Applications

The theory developed so far is powerful enough to produce nontrivial topological and geometric applications. The goal of this section is to present some of them. More precisely we will present Kronheimer and Mrowka's proof of the Thom conjecture [**71**] for the projective plane and a proof of Donaldson's Theorem A on smooth, negative definite 4-manifolds [**28, 29**]. Because of its relevance in this section and later on as well, we have also included a separate technical subsection describing a few properties of the Seiberg-Witten equations on cylinders.

2.4.1. The Seiberg-Witten equations on cylinders. Suppose (N, g) is a compact, oriented, Riemannian 3-manifold. We want to describe a few particular features of the Seiberg-Witten equations on the 4-manifold $\hat{N} = [a, b] \times N$ equipped with the product metric.

Some conventions are in order for this subsection. We will denote by t the longitudinal coordinate on $\hat{N}$ and we will identify N with the slice $\{b\} \times N$ of the cylinder $\hat{N}$. To distinguish objects of similar nature on N and $\hat{N}$ we will use a hat "ˆ" to denote the objects on the 4-manifold. Thus d will denote the exterior derivative on N while

$$\hat{d} = dt \wedge \partial_t + d$$

will denote the exterior derivative on $\hat{N}$. The metric on $\hat{N}$ will be denoted by $\hat{g}$ and the corresponding Hodge operator by $\hat{*}$. Denote by $\lrcorner_t$ the contraction by the tangent vector ∂_t.

Any differential form ω on $\hat{N}$ can be uniquely written as

$$\omega = dt \wedge f + a, \quad f := \lrcorner_t \omega, \quad a := \omega - dt \wedge f.$$

Above, f and a are paths of forms on N. Observe that

$$\hat{d}(dt \wedge f^0 + a^1) = dt \wedge (\dot{a}^1 - df^0) + da^1 \tag{2.4.1}$$

and

$$\hat{*}\omega^2 := \hat{*}(dt \wedge f^1 + a^2) = dt \wedge *a^2 + *f^1 \tag{2.4.2}$$

where the dot stands for t-differentiation. Then

$$\hat{d}_+(dt\wedge f^0+a)=\frac{1}{2}(\hat{d}+\hat{*}\hat{d})(dt\wedge f^0+a^1)$$

$$=\frac{1}{2}dt\wedge(\dot{a}^1-df^0+*da^1)+\frac{1}{2}*(\dot{a}^1-df^0+da^1)$$

and

$$\hat{d}^*(dt\wedge f^0+a^1)=-\hat{*}\hat{d}\hat{*}(dt\wedge f^0+a^1)=-(\dot{f}^0-d^*a^1).$$

Fix a *spin*^{c} structure on N. It induces by pullback a *spin*^{c} structure $\hat{\sigma}$ on $\hat{N}$ with associated bundle of complex spinors

$$\hat{\mathbb{S}}_\sigma=\hat{\mathbb{S}}_\sigma^+\oplus\hat{\mathbb{S}}_\sigma^-.$$

Denote by $\hat{\mathbf{c}}$ the Clifford multiplication on $\hat{\mathbb{S}}_\sigma$. We set $J:=\hat{\mathbf{c}}(dt):\hat{\mathbb{S}}_\sigma^+\to\hat{\mathbb{S}}_\sigma^-$. Observe that J produces an isomorphism between the restrictions of $\hat{\mathbb{S}}_\sigma^\pm$ to N. We set

$$\mathbb{S}_\sigma:=\hat{\mathbb{S}}_\sigma^+\,|_N\cong\hat{\mathbb{S}}_\sigma^-\,|_N\ .$$

The bundle $\mathbb{S}_\sigma$ is equipped with a Clifford structure given by the Clifford multiplication

$$\mathbf{c}(\alpha)=J\hat{\mathbf{c}}(\alpha):\hat{\mathbb{S}}_\sigma^+\,|_N\to\hat{\mathbb{S}}_\sigma^+\,|_N\ .$$

$\mathbb{S}_\sigma$ is precisely the bundle of complex spinors associated to the *spin*^{c} structure on the *odd*-dimensional manifold N.

For any 2-form $\hat{\alpha}$ on $\hat{N}$ we have $\hat{\mathbf{c}}(\hat{\alpha}-*\hat{\alpha})=0$ on $\hat{\mathbb{S}}_{\hat{\sigma}}^+$ so that, using (2.4.2), we deduce

$$\mathbf{c}(\alpha)=\mathbf{c}(*\alpha),\ \ \forall\alpha\in\Omega^1(N)\tag{2.4.3}$$

and

$$\mathbf{c}(dv(g))=-\mathbf{1}.\tag{2.4.4}$$

Set $\det(\sigma)=\det\mathbb{S}_\sigma=\det(\hat{\sigma})\,|_N$ and fix a smooth Hermitian connection A_0 on $\det(\sigma)$. It induces by pullback a Hermitian connection on $\det(\hat{\sigma})$ which we denote by $\hat{A}_0$. A Hermitian connection $\hat{A}$ on $\det(\hat{\sigma})$ is called *temporal* if

$$\lrcorner_t(\hat{A}-\hat{A}_0)=0,$$

that is,

$$\hat{A}=\hat{A}_0+\mathbf{i}a(t)$$

where $a(t)$ is a path of 1-forms on A. We set $A(t)=A_0+\mathbf{i}a(t)$ so that $\hat{A}$ can be regarded as a path of Hermitian connections on $\det(\sigma)$. Using the identities (2.4.1) and (2.4.2) we deduce

$$F_{\hat{A}}=\mathbf{i}dt\wedge\dot{a}+F_{A(t)}\tag{2.4.5}$$

and

$$2F^+_{\hat{A}} = dt \wedge (\mathbf{i}\dot{a} + *F_{A(t)}) + *(\mathbf{i}\dot{a} + *F_{A(t)}). \tag{2.4.6}$$

Lemma 2.4.1. *If $\hat{A}$ is a smooth Hermitian connection on $\det(\hat{\sigma})$ then there exists a smooth map*

$$\hat{f} : \hat{N} \to \mathbb{R}$$

such that the connection $\exp(\mathbf{i}\hat{f}) \cdot \hat{A} := \hat{A} - 2\mathbf{i}\hat{d}\hat{f}$ is temporal.

Proof We write

$$\hat{A} = \hat{A}_0 + \mathbf{i}dt \wedge g(t) + \mathbf{i}a(t)$$

where $g(t) \oplus a(t)$ is a path of sections of $(\Lambda^0 \oplus \Lambda^1)T^*N$. Any function $\hat{f} : \hat{N} \to \mathbb{R}$ can be viewed as a path $f(t)$ of 0-forms on N. The condition

$$\lrcorner_t(\, \exp(\mathbf{i}\hat{f})(\hat{A} - \hat{A}_0)\,) = 0$$

is equivalent to

$$\mathbf{i}(g(t) - 2\dot{f}(t)) = 0.$$

We can define

$$\hat{f}(t,x) = \frac{1}{2}\int_a^t g(s,x)ds, \quad \forall t \in [a,b],\ x \in N. \ \blacksquare$$

Suppose now that $\hat{\mathsf{C}} = (\hat{\psi}, \hat{A})$ is a $\hat{g}$-monopole on $\hat{N}$. Modulo a $\widehat{\mathfrak{G}}_{\hat{\sigma}}$-change we can assume $\hat{A}$ is temporal so we can identify it with a path $A(t)$ of connections on $\det(\sigma)$. The spinor $\hat{\psi}$ can be viewed as a path $\psi(t)$ of sections in $\mathbb{S}_\sigma$. The connection $\hat{\nabla}^{\hat{A}}$ induced by $\hat{A}$ on $\hat{\mathbb{S}}_{\hat{\sigma}}$ has the form

$$\hat{\nabla}^{\hat{A}} = dt \otimes \partial_t + \nabla^{A(t)}$$

where $\nabla^{A(t)}$ is the connection induced by $A(t)$ on $\hat{\mathbb{S}}_{\hat{\sigma}}\,|_N \cong \mathbb{S}_\sigma \oplus \mathbb{S}_\sigma$. If (e_i) is a local orthonormal frame on N and (e^i) denotes is dual coframe then we have

$$\hat{\not{\mathfrak{D}}}_{\hat{A}} = \hat{\mathbf{c}} \circ \hat{\nabla}^{\hat{A}} = \hat{\mathbf{c}}(dt)\partial_t + \sum_i \hat{\mathbf{c}}(e^i)\nabla^{A(t)}_{e_i} = J\left(\partial_t - \sum_i \mathbf{c}(e^i)\nabla^{A(t)}_{e_i}\right)$$

$$= J\left(\partial_t - \mathfrak{D}_{A(t)}\right)$$

where $\mathfrak{D}_{A(t)}$ denotes the geometric Dirac operator induced by the connection $A(t)$. Using the above identity, (2.4.3) and (2.4.6) we deduce that $\hat{\mathsf{C}} = (\psi(t), A(t) = \hat{A}_0 + \mathbf{i}a(t))$ satisfies the "evolution" equations

$$\left\{\begin{array}{rcl} \frac{d\psi}{dt} & = & \mathfrak{D}_{A(t)}\psi(t) \\ \mathbf{i}\dot{a} & = & \frac{1}{2}\mathbf{c}^{-1}(\, q(\psi(t))\,) - *F_{A(t)} \end{array}\right. . \tag{2.4.7}$$

To proceed further we imitate the four-dimensional situation and consider

$$\mathcal{C}_\sigma = \Gamma(\mathbb{S}_\sigma) \times \mathfrak{A}_\sigma$$

where $\mathfrak{A}_\sigma$ denotes the affine space of Hermitian connections on $\det(\sigma)$. Now define

$$\mathcal{E}_\sigma : \mathcal{C}_\sigma \to \mathbb{R},$$

by

$$\mathcal{E}_\sigma(\psi, A) = \frac{1}{2}\int_N (A - A_0) \wedge (F_A + F_{A_0}) + \frac{1}{2}\int_N \mathfrak{Re}\langle \mathfrak{D}_A\psi, \psi\rangle dv_g \tag{2.4.8}$$

We claim that the gradient of this functional (with respect to the L^2-metric on $\mathcal{C}_\sigma$) is given by precisely the right-hand side of (2.4.7).

The proof of this claim relies on the following technical result.

Exercise 2.4.1. Prove that for any real 1-form α on N we have

$$2|\alpha(x)|^2 = 2|*\alpha(x)|^2 = |\mathbf{c}(\alpha(x))|^2 := -\operatorname{tr}(\mathbf{c}(\alpha(x))^2), \ \ \forall x \in N.$$

(**Note** the factor of 2 and **compare** to the analogous identity in Lemma 2.1.5 in §2.1.1 concerning self-dual forms.)

To verify this claim set $\mathbf{i}a := A - A_0 \in \mathbf{i}\Omega^1(N)$ (so that $\mathfrak{D}_A = \mathfrak{D}_{A_0} + \frac{1}{2}\mathbf{c}(\mathbf{i}a)$) and write $\mathcal{E}_\sigma(\psi, a)$ instead of $\mathcal{E}_\sigma(\psi, A)$. We have

$$\frac{d}{dt}|_{t=0}\, \mathcal{E}_\sigma(\psi + t\dot{\psi}, a + t\dot{a}) = \frac{1}{2}\int_N \mathbf{i}\dot{a} \wedge (\mathbf{i}da + 2F_{A_0}) + \frac{1}{2}\int_N \mathbf{i}a \wedge d\mathbf{i}\dot{a}$$

$$+\frac{1}{2}\int_N \Big(\frac{1}{2}\langle \mathbf{c}(\mathbf{i}\dot{a})\psi, \psi\rangle + 2\mathfrak{Re}\,\langle \mathfrak{D}_A\psi, \dot{\psi}\rangle\Big) dv_g$$

(use Stokes' theorem in the second integral)

$$= \frac{1}{2}\int_N \mathbf{i}\dot{a} \wedge (\mathbf{i}da + 2F_{A_0}) + \frac{1}{2}\int_N \mathbf{i}\dot{a} \wedge \mathbf{i}da$$

$$+ \int_N \mathfrak{Re}\langle \mathfrak{D}_A\psi, \dot{\psi}\rangle dv_g + \frac{1}{4}\int_N \langle \mathbf{c}(\mathbf{i}\dot{a})\psi, \psi\rangle dv_g$$

(use $\langle \mathbf{c}(\mathbf{i}\dot{a})\psi, \psi\rangle = \mathfrak{Re}\operatorname{tr}(\mathbf{c}(\mathbf{i}\dot{a})q(\psi)) := \langle q(\psi), \mathbf{c}(\mathbf{i}\dot{a})\rangle$)

$$= \int_N \mathbf{i}\dot{a} \wedge F_A + \int_N \mathfrak{Re}\langle \mathfrak{D}_A\psi, \dot{\psi}\rangle dv_g + \frac{1}{4}\int_N \langle \mathbf{c}(\mathbf{i}\dot{a}), q(\psi)\rangle dv_g$$

$$= -\int_N \langle \mathbf{i}\dot{a}, *F_A\rangle dv_g + \int_N \mathfrak{Re}\langle \mathfrak{D}_A\psi, \dot{\psi}\rangle dv_g + \frac{1}{4}\int_N \langle \mathbf{c}(\mathbf{i}\dot{a}), q(\psi)\rangle dv_g$$

($*$ denotes the *complex linear* Hodge operator, and we use Exercise 2.4.1 in the last integral above)

$$= \int_N \langle \mathbf{i}\dot{a}, \frac{1}{2}\mathbf{c}^{-1}(q(\psi)) - *F_A\rangle dv_g + \int_N \mathfrak{Re}\langle \mathfrak{D}_A\psi, \dot{\psi}\rangle dv_g.$$

The functional $\mathcal{E}_\sigma$ is not $\mathfrak{G}_\sigma = \mathrm{Map}\,(N, S^1)$-invariant. In fact $\forall \gamma \in \mathfrak{G}_\sigma$ and $\mathsf{C} \in \mathcal{C}_\sigma$ we have

$$\mathcal{E}_\sigma(\gamma \cdot \mathsf{C}) = \mathcal{E}_\sigma(\mathsf{C}) - \int_N \frac{d\gamma}{\gamma} \wedge (F_A + F_{A_0})$$

$$= \mathcal{E}_\sigma(\mathsf{C}) - 4\pi^2 \int_M \frac{1}{2\pi\mathbf{i}} \frac{d\gamma}{\gamma} \wedge \frac{\mathbf{i}}{2\pi}(F_A + F_{A_0})$$

$$= \mathcal{E}_\sigma(\mathsf{C}) - 8\pi^2 \int_N \deg\gamma \wedge c_1(\det(\sigma)) \tag{2.4.9}$$

where $\deg\gamma \in H^1(N, \mathbb{Z})$ is the cohomology class $\gamma^*(\frac{1}{2\pi}d\theta)$. In particular, we deduce that $\mathcal{E}_\sigma$ is $\mathfrak{G}_\sigma$-invariant if and only if $c_1(\det\sigma)$ is a torsion class.

Definition 2.4.2. The critical points of the functional $\mathcal{E}_\sigma$ are called *g-monopoles* on N corresponding to the *spin*^{c} structure σ.

Remark 2.4.3. We want to point out a curious and somewhat confusing fact. More precisely, observe that the energy functional $\mathcal{E}_\sigma$ is *orientation sensitive*. By changing the orientation of N respecting the normalization (2.4.4) the energy function changes to $-\mathcal{E}_\sigma$.

Inspired by the results in §2.1.1 we define the *energy* of a configuration $\hat{\mathsf{C}} = (\hat\psi, \hat A)$ on $\hat N$ by

$$E(\hat{\mathsf{C}}) := \int_{\hat N} (|\hat\nabla^{\hat A}\hat\psi|^2 + \frac{\hat s}{4}|\hat\psi|^2 + \frac{1}{8}|q(\hat\psi)|^2 + |F_{\hat A}|^2) dv(\hat g)$$

where $\hat s$ denotes the scalar curvature of $\hat g$. If $\hat A$ is temporal, $\hat A = A(t) = \hat A_0 + \mathbf{i}a(t)$ then using (2.4.5) and the identity $|q(\psi)|^2 = \frac{1}{2}|\psi|^4$ we deduce

$$E(\hat\psi, \hat A) = \int_a^b dt \int_N (|\dot\psi|^2 + |\dot a|^2) dv(g)$$

$$+ \int_a^b \int_N (|\nabla^{A(t)}\psi(t)|^2 + \frac{s}{4}|\psi(t)|^2 + \frac{1}{16}|\psi(t)|^4 + |F_{A(t)}|^2) dv(g)$$

where s denotes the scalar curvature of g. (Observe that on the cylinder $\hat N$ we have $s = \hat s$.)

Lemma 2.4.4. (Main energy identity) *Suppose $\hat{\mathsf{C}} = (\hat\psi, \hat A)$ is a monopole on $\hat N$ such that $\hat A$ is temporal, $\hat A = A(t) = A_0 + \mathbf{i}a(t)$. Then*

$$\int_a^b dt \int_N \Big(|\dot\psi(t)|^2 + |\dot a(t)|^2\Big) dv(g)$$

$$= \int_a^b dt \int_N \Big(|\nabla^{A(t)}\psi|^2 + \frac{s}{4}|\psi|^2 + \frac{1}{16}|\psi|^4 + |F_{A(t)}|^2\Big) dv(g)$$

$$= \frac{1}{2} E(\hat{\psi}, \hat{A}).$$

Proof For brevity, we will write A instead of $A(t)$ and ψ instead of $\psi(t)$. Using the first equation in (2.4.7) we deduce

$$\int_N |\dot{\psi}|^2 dv(g) = \int_N |\mathfrak{D}_A \psi|^2 dv(g)$$

(use the Weitzenböck formula for $\mathfrak{D}_A$ and integration by parts)

$$= \int_N \Big(|\nabla^A \psi|^2 + \frac{s}{4}|\psi|^2 + \frac{1}{2}\mathfrak{Re}\langle \mathbf{c}(F_A)\psi, \psi \rangle \Big) dv(g).$$

Using the second equation in (2.4.7) and Exercise 2.4.1 we deduce

$$2\int_N |\dot{a}|^2 dv(g) = \int_N |\mathbf{c}(\dot{a})|^2 dv(g) = \int_N |\frac{1}{2} q(\psi) - \mathbf{c}(F_A)|^2 dv(g)$$

$$= \int_N \Big(|\frac{1}{2} q(\psi)|^2 + |\mathbf{c}(F_A)|^2 - \mathfrak{Re}\langle q(\psi), \mathbf{c}(F_A) \rangle \Big) dv(g)$$

(use Exercise 2.4.1 again)

$$= \int_N \Big(\frac{1}{8}|\psi|^4 + 2|F_A|^2 - \langle \mathbf{c}(F_A)\psi, \psi \rangle \Big) dv(g).$$

The energy identity is now obvious. ■

Remark 2.4.5. We want to point out a nice feature of the main energy identity. Its right-hand side is manifesly gauge independent while the left-hand side is apparently gauge dependent since the configuration $(\hat{\psi}, \hat{A})$ was chosen so that $\hat{A}$ is temporal.

The functional $\mathcal{E}_\sigma$ has nice variational properties, reminiscent of the Palais-Smale condition.

Proposition 2.4.6. *Suppose $\mathsf{C}_n = (\psi_n, A_n)$ is a sequence of smooth configurations such that*

$$\|\psi_n\|_\infty = O(1), \quad as \;\; n \to \infty \tag{2.4.10}$$

and

$$\|\nabla \mathcal{E}_\sigma(\mathsf{C}_n)\|_{L^2} = o(1), \quad as \;\; n \to \infty. \tag{2.4.11}$$

Then there exists a sequence $\gamma_n \in \mathfrak{G}_\sigma$ such that $\gamma_n \cdot \mathsf{C}_n$ converges in any Sobolev norm to a critical point C_∞ of $\mathcal{E}_\sigma$

$$\nabla \mathcal{E}_\sigma(\mathsf{C}_\infty) = 0.$$

Proof The condition (2.4.11) implies

$$\|\mathfrak{D}_{A_n}\psi_n\|_2 = o(1) \tag{2.4.12}$$

and

$$\|F_{A_n}\|_2 = \|\frac{1}{2}q(\psi_n)\|_2 + o(1). \tag{2.4.13}$$

Using the sup-bound on ψ_n in the last inequality we deduce

$$\|F_{A_n}\|_2 = O(1).$$

Modulo changes of gauge, which can be used to reduce the size of the harmonic part of F_{A_n} below a fixed, geometrically determined constant, the last inequality leads to $L^{1,2}$-bounds for $\mathbf{i}a_n := A_n - A_0$.

Throw this information back in (2.4.12) to obtain

$$\mathfrak{D}_{A_0}\psi_n = -\mathbf{c}(\mathbf{i}a_n)\psi_n + o(1).$$

The elliptic estimates coupled with the sup-bound on ψ_n and the $L^{1,2}$-bound on a_n lead to $L^{1,2}$-bounds on ψ_n. Bootstrap to obtain bounds on (a_n, ψ_n) in arbitrary norms. These coupled with compact Sobolev embeddings allows us now to conclude that a subsequence of C_n converges in any Sobolev norm to some smooth $\mathsf{C}_\infty \in \mathcal{C}_\sigma$. The conclusion in the proposition now follows using (2.4.11) once again. ■

The last proposition has an important consequence.

Corollary 2.4.7. *Suppose $\hat{\mathsf{C}} = (\hat{\psi}, \hat{A})$ is a smooth finite energy monopole on $\hat{N}_\infty := \mathbb{R} \times N$ such that $\hat{A}$ is temporal and*

$$\|\hat{\psi}\|_\infty < \infty.$$

Then there exists a sequence $t_n \to \infty$ such that, modulo $\mathfrak{G}_\sigma$, the configurations $(\psi(t_n), A(t_n))$ converge in any Sobolev norm to a critical point of $\mathcal{E}_\sigma$.

Proof Using the main energy identity we deduce

$$\int_{-\infty}^{\infty} dt \int_N \Big(|\dot{\psi}(t)|^2 + |\dot{a}(t)|^2 \Big) dv(g) < \infty$$

so that there exists a sequence $t_n \to \infty$ such that

$$\Big\| \nabla\mathcal{E}_\sigma\big(\psi(t_n), A(t_n)\big) \Big\|_{L^2}^2 = \int_N \Big(|\dot{\psi}(t_n)|^2 + |\dot{a}(t_n)|^2 \Big) dv(g) = o(1).$$

The desired conclusion now follows from Proposition 2.4.6. ■

2.4.2. The Thom conjecture. To put the Thom conjecture in the proper context we begin by recalling a classical algebraic-geometry result. We will denote the tensor multiplication of line bundles additively, by $+$.

Proposition 2.4.8. (Adjunction formula) *Suppose (X, J) is an almost complex manifold of dimension $2n$ and $Y \subset X$ is a submanifold of dimension $2(n-1)$ such that the natural inclusion*

$$TY \hookrightarrow TX\,|_Y$$

is a morphism of complex bundles. Then

$$K_Y = K_X\,|_Y + N_Y$$

where N_Y denotes the complex normal line bundle, $N_Y := TX\,|_Y\,/TY$ determined by the embedding $Y \hookrightarrow X$, and K denotes the canonical line bundle, $K_M = \det(T^*M)^{1,0} = \det(T^{0,1}M)$.

Proof Along $Y \hookrightarrow X$ we have the isomorphism of *complex* vector bundles

$$TX^{1,0}\,|_Y \cong TY^{1,0} \oplus N_Y.$$

By passing to determinants we deduce

$$-K_X\,|_Y = -K_Y + N_Y. \quad \blacksquare$$

Suppose now that (X, ω) is a Kähler manifold of complex dimension two and $\Sigma \hookrightarrow X$ is a *smooth complex curve on X*, i.e. a compact, connected, complex submanifold of X. Using the adjunction formula we deduce

$$K_\Sigma = K_X\,|_\Sigma + N_\Sigma.$$

Again we identify the complex line bundles with their first Chern class c_1^{top}. Integrating (=Kronecker pairing) the above equality over Σ we deduce

$$\langle K_\Sigma, \Sigma\rangle = \langle K_X, \Sigma\rangle + \Sigma \cdot \Sigma$$

since, according to the Gauss-Bonnet theorem, the pairing $\langle N_\Sigma, \Sigma\rangle$ is the self-intersection of $\Sigma \hookrightarrow X$. Using Gauss-Bonnet again we deduce

$$\langle K_\Sigma, \Sigma\rangle = 2g(\Sigma) - 2$$

where $g(\Sigma)$ is the genus of the Riemannian surface Σ. This yields the *genus formula*

$$g(\Sigma) = 1 + \frac{1}{2}(K_X \cdot \Sigma + \Sigma \cdot \Sigma). \tag{2.4.14}$$

We specialize further and we assume $X = \mathbb{CP}^2$ and $\Sigma \to \mathbb{CP}^2$ is a smooth complex curve of degree d, i.e.

$$[\Sigma] = dH, \text{ in } H_2(\mathbb{CP}^2, \mathbb{Z}).$$

Using the equality $K_{\mathbb{CP}^2} = -3H$ established in §2.3.4 we deduce

$$g(\Sigma) = 1 + \frac{d(d-3)}{2}. \tag{2.4.15}$$

Kervaire and Milnor (see [**56, 62**]) have shown that if the homology class $dH \in H_2(\mathbb{CP}^2, \mathbb{Z})$ is characteristic for the intersection form (i.e. d is odd) and can be represented by an embedded sphere then

$$1 = \tau(\mathbb{CP}^2) \equiv d^2 \mod 16.$$

In particular this shows that the class $3H$ cannot be represented by an embedded sphere.

To connect this fact with the genus formula (2.4.15) we introduce

$$g_{min} : H_2(\mathbb{CP}^2, \mathbb{Z}) \to \mathbb{Z}_+$$

where $g_{min}(dH)$ denotes the minimum of the genera of smoothly embedded Riemann surfaces $\Sigma \hookrightarrow \mathbb{CP}^2$ carrying the homology class dH. The above result of Kervaire and Milnor implies

$$g_{min}(dH) \geq 1, \quad d = 3.$$

The equality is optimal for $d = 3$ since according to (2.4.15) the curves of degree 3 on $\mathbb{CP}^2$ have genus 1. In particular this shows that

$$g_{min}(dH) = 1 + \frac{d(d-3)}{2}, \quad d = 1, 2, 3.$$

A famous conjecture, usually attributed to R. Thom, states that the above equality holds for all $d \geq 0$. Using the genus formula we can rephrase this by saying that the complex curves are genus minimizing amongst the smoothly embedded surfaces within a given homology class. The methods developed so far are powerful enough to offer a solution to this conjecture.

Theorem 2.4.9. *For every $d \geq 0$ we have the equality*

$$g_{min}(dH) = 1 + \frac{d(d-3)}{2}.$$

Proof We follow closely the ideas of Kronheimer and Mrowka [**71**]. The above observations show that it suffices to consider only the case $d > 3$.

Suppose $\Sigma \hookrightarrow \mathbb{CP}^2$ is a smoothly embedded surface such that $[\Sigma] = dH$, $d > 3$. Then

$$\Sigma \cdot \Sigma = k := d^2.$$

We blow up $\mathbb{CP}^2$ k times $\mathbb{CP}^2 \dashrightarrow \mathbb{CP}^2 \# k\overline{\mathbb{CP}}^2$ and denote by π the natural projection

$$M := \mathbb{CP}^2 \# k\overline{\mathbb{CP}}^2 \to \mathbb{CP}^2.$$

As in Example 2.3.36 denote by E_i, $i = 1, \cdots, k$ the homology classes carried by the exceptional divisors. Consider the proper transform $\tilde{\Sigma}$ in the blow-up in the sense of algebraic geometry. Topologically this means $\tilde{\Sigma}$ is the connected sum with the k spheres representing the classes $-E_i$. Thus

$$\tilde{\Sigma} \cdot \tilde{\Sigma} = 0.$$

We now follow closely the geometric situation in Example 2.3.14. Denote by U a small tubular neighborhood of $\tilde{\Sigma} \hookrightarrow M$ diffeomorphic to $D^2 \times \tilde{\Sigma}$ and set $N = \partial U \cong S^1 \times \tilde{\Sigma}$. Equip $\tilde{\Sigma}$ with a metric g_0 of constant scalar curvature s_0. The Gauss-Bonnet theorem implies

$$\frac{1}{4\pi} \int_{\tilde{\Sigma}} s_0 dv(g_0) = 2 - 2g(\tilde{\Sigma}) = 2 - 2g(\Sigma)$$

so that

$$s_0 = \frac{8\pi}{\mathrm{vol}_{g_0}(\tilde{\Sigma})} (\, 1 - g(\Sigma)\,). \tag{2.4.16}$$

When no confusion is possible we will continue to denote by g_0 the product metric on $N = S^1 \times \tilde{\Sigma}$.

Now consider again the metric g_n, $n \gg 1$, of Example 2.3.14 so that a tubular neighborhood of $N \hookrightarrow M$ is isometric to the metric $dt^2 + d\theta^2 + g_0$ on $[-n, n] \times S^1 \times \tilde{\Sigma}$. Set $\hat{N}_n := [-n, n] \times N$. Again denote by $\hat{\sigma}_0$ the *spin*^{c} structure induced by the natural complex structure on M so that $\det(\hat{\sigma}_0) = -K_M = 3H - \sum_i E_i$. Denote by σ_0 the restriction of $\hat{\sigma}_0$ to N. We saw in that example that there exist (smooth) $(\hat{\sigma}_0, g_n, 0)$-monopoles $\hat{\mathsf{C}}_n = (\hat{\psi}_n, \hat{A}_n)$ for all $n \gg 1$.

Lemma 2.4.10. *There exists a constant $C > 0$, such that $\forall n \gg 1$ we have*

$$\|\hat{\psi}_n\|_{L^\infty(M)} < C \tag{2.4.17}$$

and

$$E(\hat{\mathsf{C}}_n\,|_{\hat{N}_n}) < C. \tag{2.4.18}$$

Proof Denote by $s_n(x)$ the scalar curvature of the metric g_n. Along the long neck $s_n(x)$ is comparable to s_0 while away from the neck it is bounded above by a constant independent of n since the metric g_n varies very little in that region. The inequality (2.4.17) is thus a consequence of the **Key Estimate** in §2.2.1.

To prove the second inequality denote by R the complement of the neck in M and let E_n denote the energy of $\hat{\mathsf{C}}_n$ on M. Since $\hat{\mathsf{C}}_n$ is a $(\sigma_0, g_n, 0)$-monopole we deduce from Proposition 2.1.4 that

$$E_n = -2\pi^2 \int_M c_{\hat{\sigma}_0}^2 = -2\pi^2 K_M^2 = 2\pi^2 (k - 9).$$

We deduce

$$E(\hat{\mathsf{C}}_n\,|_{\hat{N}_n}) = E_n - E(\hat{\mathsf{C}}_n\,|_R) \leq E_n - \int_R \frac{s_n(x)}{4} |\hat{\psi}_n(x)|^2 dv(g_n).$$

Since $s_n(x)$ and $|\hat{\psi}_n(x)|$ are bounded independent of n and R has finite volume, independent of n, we deduce that the right-hand side of the above inequality is bounded from above by a constant independent of n. This concludes the proof of the lemma. ■

Modulo a gauge transformation we can assume $\hat{\mathsf{C}}_n = (\hat{\psi}_n, \hat{A}_n)$ is temporal so that we can write

$$\hat{\psi}_n\,|_{\hat{N}_n} = \psi_n(t) \text{ and } \hat{A}_n = \hat{A}_0 + \mathbf{i}a_n(t).$$

Since

$$E(\hat{\mathsf{C}}_n\,|_{\hat{N}_n}) < C$$

there exists $|k_n| < n$ such that

$$E(\hat{\mathsf{C}}_n\,|_{[k_n,k_n+1]\times N}) < C/2n.$$

Using the main energy identity we deduce

$$\int_{k_n}^{k_n+1} dt \int_N |\dot{\psi}_n(t)|^2 + |\dot{a}_n(t)|^2 dv(g_0) < C/n.$$

Thus there exists $t_n \in [k_n, k_n+1]$ such that

$$\int_N |\dot{\psi}_n(t_n)|^2 + |\dot{a}_n(t_n)|^2 dv(g_0) < C/n. \tag{2.4.19}$$

Set

$$\mathsf{C}_n = \hat{\mathsf{C}}_n(t_n) = (\psi_n(t_n), A_0 + \mathbf{i}a_n(t_n)).$$

Lemma 2.4.10 and (2.4.19) show that the sequence C_n satisfies all the assumptions in Proposition 2.4.6. This leads to the conclusion that

◇ *there exist g_0-monopoles on $N = S^1 \times \tilde{\Sigma}$ corresponding to the spin^c structure $\sigma_0 = \hat{\sigma}_0\,|_N$.*

To conclude the proof of Theorem 2.4.9 we will show that the existence of monopoles on N imposes restrictions on $g(\tilde{\Sigma})$.

Observe first that any *spin*^{c} structure σ on $\tilde{\Sigma}$ induces by pullback via $p : N \to \tilde{\Sigma}$ a *spin*^{c} structure $p^*\sigma$ on N. Next observe that

$$\sigma_0 = \hat{\sigma}_0\,|_N = p^*\hat{\sigma}_0\,|_{\tilde{\Sigma}}$$

so that

$$\det(\sigma_0) = p^*\big(\det(\hat{\sigma}_0\,|_{\tilde{\Sigma}})\big) = p^*(-K_M\,|_{\tilde{\Sigma}}).$$

The surface $\tilde{\Sigma}$ can be naturally viewed as a submanifold in N which is the total space of a *trivial* S^1-bundle over $\tilde{\Sigma}$. The above equality implies

(2.4.20)
$$\int_{\tilde{\Sigma}} c_{\sigma_0} = -K_M \cdot \tilde{\Sigma} = (3H - \sum_{i=1}^k E_i) \cdot (dH - \sum_{i=1}^k E_i) = 3d - k = d(3-d).$$

If $\mathsf{C} = (\psi, A)$ is a g_0-monopole on N

$$\left\{ \begin{array}{rcl} \mathfrak{D}_A\psi & = & 0 \\ \mathbf{c}(*F_A) & = & \frac{1}{2}q(\psi) \end{array} \right. \tag{2.4.21}$$

then arguing **exactly** as in the proof of the **Key Estimate** in §2.2.1 we deduce

$$\|\psi\|_\infty^2 \le -2\min_{x\in N} \bar{s}_0(x)$$

where $\bar{s}_0(x)$ denotes the scalar curvature of the metric g_0 on N. Now observe that since $N = S^1 \times \tilde{\Sigma}$ is equipped with the product metric the scalar curvature $\bar{s}_0$ at $(\theta, z) \in S^1 \times \tilde{\Sigma}$ is equal to $s_0(z)$ and using (2.4.16) we conclude

$$\|\psi\|_\infty^2 \le \frac{16\pi}{\mathrm{vol}_{g_0}(\tilde{\Sigma})}(g(\Sigma) - 1). \tag{2.4.22}$$

Using Exercise 2.4.1 and (2.4.22) in the second equation of (2.4.21) we deduce

$$\sqrt{2}|F_A| = |\mathbf{c}(*F_A)| = \frac{1}{2}|q(\psi)| = \frac{1}{2\sqrt{2}}|\psi|^2 \le \frac{4\sqrt{2}\pi}{\mathrm{vol}_{g_0}(\tilde{\Sigma})}(g(\Sigma) - 1)$$

so that

$$|F_A| \le \frac{4\pi}{\mathrm{vol}_{g_0}(\tilde{\Sigma})}(g(\Sigma) - 1). \tag{2.4.23}$$

Using (2.4.20) and the assumption $d > 3$ we deduce

$$d(d-3) = \left| \int_{\tilde{\Sigma}} c_{\sigma_0} \right| \le \frac{1}{2\pi}\int_{\tilde{\Sigma}} |F_A| dv(g_0) \stackrel{(2.4.23)}{\le} 2(g(\Sigma) - 1).$$

This is exactly the content of Theorem 2.4.9. ■

Remark 2.4.11. (a) Presently the validity of the genus minimizing conjecture of Thom has been established in its full generality in the more general context of symplectic manifolds; see [**97, 114**] or the discussion at the end of §4.6.2. In this case the genus minimizing surfaces in a given homology class are precisely the symplectically embedded ones.

(b) In [**97, 101**] one can find a detailed and explicit description of the monopoles on $S^1 \times \Sigma$. For the more general case of circle bundles over a Riemann surface we refer to [**106**].

2.4.3. Negative definite smooth 4-manifolds. To help the reader better enjoy the beauty and the depth of the main result of this subsection we begin by surveying some topological facts. For more details we refer to [**29**, Chap. 1], [**51, 87**].

The world of topological 4-manifolds is very unruly and currently there is no best way to organize it, and not for lack of trying.

The fundamental group, which does wonders in dimension two and is sufficiently powerful in dimension three, is less effective in dimension four for a simple reason: every finitely presented group is the fundamental group of a smooth manifold (even symplectic, according to [**51**]). This shows that the algorithmic classification of 4-manifolds is more complicated than that of finitely presented groups, which is impossible. It is thus reasonable to try to understand first the simply connected 4-manifolds and in this dimension we have to be very specific whether we talk about topological or smooth ones.

The intersection form of simply connected topological 4-manifolds is a powerful invariant: it classifies them up to homotopy equivalence (according to J.H.C. Whitehead [**147**]) and almost up to a *homeomorphism* according to the award winning results of M. Freedman [**38**]. Recall that the intersection form of a closed 4-manifold is a symmetric, unimodular, bilinear map

$$q : \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{Z}.$$

Unimodularity in this case means that the matrix describing q with respect to some integral basis of $\mathbb{Z}^n$ has determinant 1.

To each intersection form one can associate three invariants: its rank, n in this case, its signature and its type. The signature, $\tau(q)$, is defined as the difference between the number of positive eigenvalues and the number of negative eigenvalues of the symmetric matrix representing q with respect to some basis of $\mathbb{Z}^n$. The intersection forms are of two types: even, if

$$q(x,x) \equiv 0 \mod 2, \ \ \forall x \in \mathbb{Z}$$

and odd, if it's not even. Observe that q is even if and only if the matrix representing q with respect to an arbitrary basis of $\mathbb{Z}^n$ has even diagonal entries. A quadratic form q is called *positive/negative* if $\tau(q) = \pm \operatorname{rank} q$ and indefinite otherwise.

Two integral quadratic forms q_1, q_2 of the same rank n are *isomorphic* if there exists $T \in GL(n, \mathbb{Z})$ such that

$$q_1(Tx, Tx) = q_2(x,x), \ \ \forall x \in \mathbb{Z}^n.$$

The quadratic forms over $\mathbb{Q}$ or $\mathbb{R}$ are completely classified up to isomorphism by their rank and signature. The situation is considerably more complicated in the integral case.

Example 2.4.12. The *diagonal definite form* of rank n is the quadratic form $q = \langle \mathbf{1} \rangle_n$ whose matrix with respect to the canonical basis of $\mathbb{Z}^n$ is the identity matrix. More generally, a quadratic form is said to be *diagonal*(izable) if it is isomorphic to the direct sum $\langle \mathbf{1} \rangle_n \oplus \langle -\mathbf{1} \rangle_m$. The form E_8 is the positive definite quadratic form of rank 8 given by the symmetric matrix

$$E_8 = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 \end{pmatrix}. \tag{2.4.24}$$

A more efficient and very much used way of describing this matrix is through its *Dynkin diagram* (see Figure 2.9). The $\bullet$'s describe a basis $v_1, \cdots, v_8$ of $\mathbb{Z}^8$.

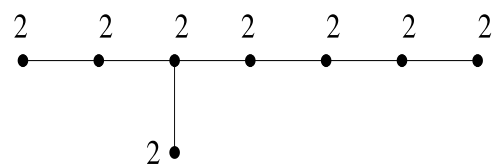

Figure 2.9. *The Dynkin diagram of E_8*

The 2's indicate that $q(v_i, v_i) = 2$ and the edges indicate that $q(v_i, v_j) = 1$ if and only if v_i and v_j are connected by an edge. E_8 is even and positive definite. E_8 is not diagonalizable over $\mathbb{Z}$. We also want to point out that often E_8 is described by a matrix very similar to the one in (2.4.24) where the 1's are replaced by -1's. The two descriptions are equivalent and correspond to the change of basis $v_i \to (-1)^i v_i$.

Another important example of quadratic form is the *hyperbolic* form H given by the matrix

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

It is even, indefinite, has zero signature and it is not diagonalizable.

The examples presented above generate a large chunk of the set of isomorphism classes of integral, unimodular, quadratic forms. More precisely, we have the following result, whose proof can be found in [**121**].

Theorem 2.4.13. *(a) Any odd,* indefinite *quadratic form is diagonalizable.*

(b) Suppose q is an even *form. Then*

$$\tau(q) \equiv 0 \mod 8.$$

(c) If q is even, indefinite *and $\tau(q) \geq 0$ then*

$$q \cong aE_8 \oplus bH := (\underbrace{E_8 \oplus \cdots \oplus E_8}_{a}) \oplus (\underbrace{H \oplus \cdots \oplus H}_{b})$$

where $\tau(q) = 8a$ and $8a + 2b = \operatorname{rank}(q)$. (When $\tau(q) < 0$ use $-q$ instead.)

The classification of definite forms is a very complicated problem. It is known that the number of nonisomorphic definite quadratic, unimodular forms of rank n goes very rapidly to ∞ as $n \to \infty$ (see [**121**]). The diagonal one however plays a special role. To describe one of its special features we need to introduce a new concept.

Suppose q is a quadratic unimodular form of rank n. A vector $\mathbf{x}_0 \in \mathbb{Z}^n$ is called a *characteristic vector* of q if

$$q(\mathbf{x}_0, y) \equiv q(y, y) \mod 2, \ \ \forall y \in \mathbb{Z}^n.$$

If we represent q by a symmetric matrix S using a basis of $\mathbb{Z}^n$ then a vector $\mathbf{x}$ is characteristic if its coordinates (x_i) with respect to the chosen basis have the same parity as the diagonal elements of S, i.e.

$$x_i \equiv s_{ii} \mod 2, \ \ \forall i = 1, \cdots, n.$$

We see that q is even if and only if 0 is a characteristic vector.

Example 2.4.14. (Wu's formula) Suppose M is a closed, compact oriented smooth 4-manifold with intersection form q_M. A special case of Wu's formula (see [**93**]) shows that the mod 2 reduction of any characteristic vector $\mathbf{x}$ of q_M is precisely the second Stiefel-Whitney class $w_2(M)$. In particular, this implies that any smooth 4-manifold admits *spin*^{c} structures (since any such structure corresponds to an integral lift of $w_2(M)$) and moreover,

$$\langle w_2(M), \alpha\rangle \equiv q_M(\alpha, \alpha) \mod 2, \ \ \forall \alpha \in H_2(M, \mathbb{Z}).$$

As explained in [**51**, Sec. 1.4], the last identity should be regarded as a mod 2 version of the adjunction formula.

The congruence (b) in Theorem 2.4.13 admits the following generalization (see [**121**]).

Proposition 2.4.15. *If q is an integral, unimodular, quadratic form and* $\mathbf{x}$ *is a characteristic vector of q then*

$$q(\mathbf{x}, \mathbf{x}) \equiv \tau(q) \mod 8.$$

Following [**32**] we introduce the *Elkies invariant* $\Theta(q)$ of a *negative definite* quadratic form q as

$$\Theta(q) := \operatorname{rank}(q) + \max\{q(\mathbf{x},\mathbf{x});\ \ \mathbf{x}\ \text{a characteristic vector}\}.$$

Observe that since q is negative definite $\Theta(q) \leq \operatorname{rank}(q) = -\tau(q)$ with equality if and only if q is even. Moreover, by Proposition 2.4.15 we have $\Theta(q) \in 8\mathbb{Z}$. We have the following nontrivial result.

Theorem 2.4.16. (Elkies, [32]) *For any negative definite quadratic form q we have*

$$\Theta(q) \geq 0$$

with equality if and only if q is diagonal.

Roughly speaking, this theorem says that if q is not diagonal then the positive form $-q$ has short characteristic vectors.

We now return to topology. Michael Freedman's classification theorem states that given any *even* quadratic form there exists a unique, up to homeomorphism, simply connected (s.c.) topological 4-manifold with this intersection form. Moreover he showed that given any *odd* quadratic form there exist exactly *two* nonhomeomorphic topological s.c. 4-manifolds with this intersection form and at most one of them is smoothable (that is it admits smooth structures). We deduce the following remarkable consequence.

Corollary 2.4.17. *Two simply connected* smooth 4*-manifolds are* homeomorphic *if and only if they have isomorphic intersection forms.*

In the early 50's, Vladimir Rohlin ([**118**]) has showed that if the even form q is the intersection form of a *smooth* s.c. 4-manifold then

$$\tau(q) \equiv 0 \mod 16.$$

According to Michael Freedman's classification, there exists a unique s.c. topological 4-manifold with intersection form E_8. The signature of E_8 is $8 = \operatorname{rank}(E_8)$. This topological 4-manifold cannot support smooth structures!!!

In the early 80's, Simon Donaldson ([**28**]) showed that this surprising fact is not singular.

Theorem 2.4.18. (Donaldson, [28, 29]) *If M is a smooth, compact, oriented* 4*-manifold with negative definite intersection form q_M then q_M is diagonal.*

This theorem shows that of the infinitely many negative definite quadratic forms only the diagonal ones can be the intersection forms of some *smooth* 4-manifold. Thus any negative definite topological 4-manifold with nondiagonalizable intersection form does not admit smooth structures !!!

Proof of Theorem 2.4.18 We will argue by contradiction. Assume q_M is not diagonal. We distinguish two cases.

• Assume first that $b_1(M) = \dim H^1(M, \mathbb{R}) = 0$. Then $\chi(M) = 2 + b_2$, $\tau(M) = -b_2$ so that for all $\sigma \in Spin^c(M)$ we have

$$d(\sigma) = \frac{1}{4}(c_\sigma^2 + b_2 - 4) = \frac{1}{4}(q_M(c_\sigma, c_\sigma) + \operatorname{rank}(q_M)) - 1.$$

By Wu's formula c_σ is a characteristic vector. Since q_M is not diagonal we deduce from Elkies' theorem that $\Theta(q_M) > 0$ and we can find $\sigma \in Spin^c(M)$ such that $d(\sigma) = \frac{1}{4}\Theta(q_M) - 1 > 0$. Since $\Theta(q_M) \in 8\mathbb{Z}$ we deduce $d(\sigma) \in 2\mathbb{Z} + 1$.

For any closed 2-form η on M and any metric g there exist reducible (g, η)-monopoles corresponding to the σ. They are determined by the condition

$$F_A^+ + \mathbf{i}\eta^+ = 0. \tag{2.4.25}$$

As in §2.2.3 we write $\eta = [\eta] + d\alpha$ and fix a connection A_0 such that

$$[F_{A_0}] = -2\pi\mathbf{i}[c_\sigma].$$

Any solution of (2.4.25) can be written as $A = A_0 - \mathbf{i}\alpha + \mathbf{i}\beta$ where β is a closed 1-form. (Observe that such an A satisfies $F_A = F_{A_0} - \mathbf{i}d\alpha$. Since M is negative definite it automatically satisfies (2.4.25) because there are no self-dual harmonic 2-forms.) On the other hand, since $b_1(M) = 0$ any closed 1-form is exact so that $\beta = -2df$. This shows that all the solutions of (2.4.25) are $\mathfrak{G}_\sigma$ equivalent.

Using the Sard-Smale theorem as in §2.2.3 we can pick η so that any (g, η)-monopole C is regular, i.e. the second cohomology group $\mathcal{H}^2_{\mathsf{C}}$ of the deformation complex at C is trivial. Denote by $\mathsf{C}_0 = (0, A_0)$ the unique (mod $\mathfrak{G}_\sigma$) reducible (g, η)-monopole. In this case, using the Kuranishi picture we deduce that away from C_0 the moduli space is a smooth manifold while a neighborhood of C_0 in the moduli space $\mathfrak{M}_\sigma(g, \eta)$ is homeomorphic to

$$\mathcal{H}^1_{\mathsf{C}_0}/S^1.$$

In this case $\mathcal{H}^1_{\mathsf{C}_0} \cong \ker \mathfrak{D}_{A_0}$. Since $\operatorname{coker} \mathfrak{D}_{A_0} = \mathcal{H}^2_{\mathsf{C}_0} = 0$ we deduce

$$\dim_{\mathbb{C}} \mathfrak{D}_{A_0} = \operatorname{ind}_{\mathbb{C}} \mathfrak{D}_{A_0} = \frac{1}{8}(c_\sigma^2 - \tau(M)) = \frac{1}{8}\Theta(q_M) = \frac{d(\sigma) + 1}{2}.$$

Thus, if $d(\sigma) = 1$ near C_0 the moduli space is homeomorphic to the segment $[0, 1)$ while if $d(\sigma) > 1$ it looks like a cone over $\pm\mathbb{CP}^{\frac{d(\sigma)-1}{2}}$.

If we chop out a small neighborhood of C_0 in $\mathfrak{M}_\sigma(g, \eta)$ we obtain a *smooth, compact, orientable* manifold X with boundary $\pm\mathbb{CP}^{\frac{d(\sigma)-1}{2}}$.

If $d(\sigma) = 1$ then X is a smooth, compact oriented one-dimensional manifold with boundary consisting of only one component. This is plainly impossible.

If $d(\sigma) > 1$ observe that the restriction of the universal line bundle $\mathbb{U}_\sigma$ to ∂X is $\pm$ the tautological line bundle over $\pm\mathbb{CP}^{\frac{d(\sigma)-1}{2}}$ and thus is nontrivial. More precisely ($\Omega_\sigma = c_1(\mathbb{U}_\sigma)$)

$$\int_{\partial X} \Omega_\sigma^{\frac{d(\sigma)-1}{2}} = \pm 1.$$

The last equality is impossible since $\mathbb{U}_\sigma$ extends over X and by Stokes' theorem we have

$$\int_{\partial X} \Omega_\sigma^{\frac{d(\sigma)-1}{2}} = \int_X d\Omega_\sigma^{\frac{d(\sigma)-1}{2}} = 0.$$

This contradiction completes the proof of Theorem 2.4.18 in the case $b_1(M) = 0$.

• $b_1(M) > 0$. We will reduce this case to the previous situation by a simple topological trick.

Choose a basis $c_1, \cdots, c_{b_1}$ of $H_1(M, \mathbb{Z})/\text{Tors}$ and represent each of these cycles by smoothly embedded S^1's. We can "kill" the homology class carried by each of these cycles by surgery (see [**51**]). This operation can be briefly described as follows.

Observe first that a tubular neighborhood N of a smoothly embedded $S^1 \hookrightarrow M$ is diffeomorphic to $D^3 \times S^1$ where D^k denotes the unit ball in $\mathbb{R}^k$. Fix such a diffeomorphism so that $\partial N \cong S^2 \times S^1$. Now remove N to obtain a manifold with boundary $S^2 \times S^1$ to which we attach the handlebody $H = S^2 \times D^2$ (which has $\partial H = S^2 \times S^1$). This operation will kill each of the classes c_i but will not affect H_2/Tors and the intersection form of M since the classes c_i are not torsion classes (use the Poincaré duals of c_i to see this). In the end we obtain a smooth manifold with the same intersection form but with $b_1 = 0$. This places us in the previous situation. The proof of Theorem 2.4.18 is now complete. ■

Exercise 2.4.2. Prove that the above sequence of surgeries does not affect the intersection form, as claimed.

Remark 2.4.19. Donaldson's theorem states that a smooth, simply connected, negative definite 4-manifold X cannot be too complicated arithmetically: its intersection form is the simplest possible.

If we remove the negativity assumption, so that the intersection form q_X is indefinite, then q_X has a much simpler from. If X is not spin then q_X

is odd and thus diagonal.[1] If X is spin then q_X is even and thus it has the form

$$q_X = aE_8 + bH, \quad a = \frac{1}{8}\tau(q), \quad 8|a| + 2b = \operatorname{rank}(q).$$

In this case the integers (a, b), $b > 0$, represent a measure of the complexity of q_X. Rohlin's theorem states there are restrictions on (a, b). More precisely, a must be an even integer. The celebrated 11/8-th conjecture states that there are even more drastic restrictions in this case, more precisely

$$11|a| = \frac{11}{8}|\tau(q)| \leq \operatorname{rank}(q_X) = 8|a| + 2b.$$

This inequality is optimal because equality is achieved when X is the $K3$ surface (see the next chapter). Using Seiberg-Witten theory M. Furuta has proved a 10/8-th theorem (see [**45**], or the simpler approach in [**19**]). More precisely, he showed that

$$10|a| + 1 \leq \operatorname{rank}(q) = 8|a| + 2b.$$

[1] The example $m\mathbb{CP}^2 \# n\overline{\mathbb{CP}}^2$ shows that any odd form is the intersection form of a smooth, s.c. 4-manifold.