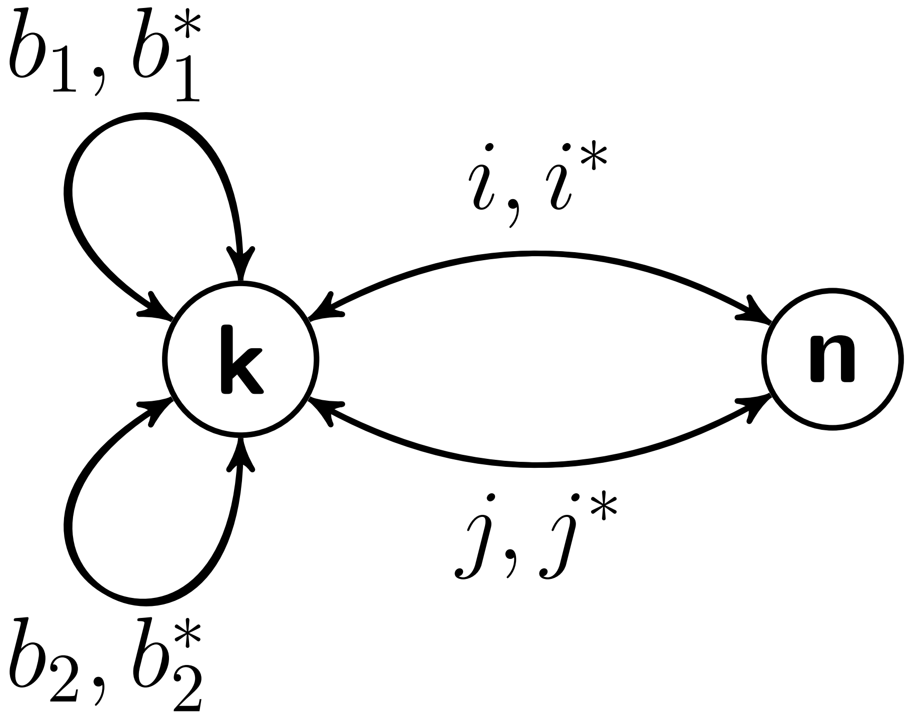
An introductory approach to instantons and the ADHM-construction

Mike Daas

July 2, 2018

Bachelor thesis Mathematics and Physics & Astronomy

Supervisors: dr. Raf Bocklandt, prof. dr. Erik Verlinde



Institute of Physics
Korteweg-de Vries Institute for Mathematics
Faculty of Sciences
University of Amsterdam

Abstract

This thesis will introduce the notion of instantons and describe the ADHM-construction in full detail. First, introductory differential geometry will be discussed, introducing concepts such as connections, curvatures, vector potentials, the exterior covariant derivative and the Hodge-star operator. These notions will be used to define G-bundles and gauge theories formally. The $U(1)$-gauge theory of electromagnetism and the magnetic monopole will be explored and some properties of the group $SU(2)$ are deduced along with the definition of the Yang-Mills action. The origin of the Pontryagin index as a result of the required convergence of the action will be outlined and some topological invariants will be derived. More general Chern forms will also be discussed. Instantons are shown to be global minima of the Yang-Mills action and the correctness of the ADHM-construction of instantons is fully proved using quaternions. Using the construction, the BPST-instanton is derived and shown to have unit topological charge. More calculations are performed and instantons of higher topological charge are constructed using block matrices. The 't Hooft instantons for an $SU(2)$-gauge theory are also presented. Lastly, representations of quivers are considered to investigate the origin of the ADHM-equations. To that end, the preprojective algebra is discussed and two moment maps are constructed, the zeroes of which correspond via the Kempf-Ness theorem to the two ADHM-equations.

Title: An introductory approach to instantons and the ADHM-construction
Author: Mike Daas, dutchmikedaas@gmail.com, 10999892
Supervisors: dr. Raf Bocklandt, prof. dr. Erik Verlinde
Second graders: dr. Diego Hofman, prof. dr. Eric Opdam
End date: July 2, 2018

Institute of Physics
University of Amsterdam
Science Park 904, 1098 XH Amsterdam
http://www.iop.uva.nl

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Science Park 904, 1098 XH Amsterdam
http://www.kdvi.uva.nl

Contents

1 Introduction

In the past century, physicists have developed an increased interest in the topic of *gauge theory*, as it provides a useful framework for classical field theories. Quantum field theory generally attempts to describe all elementary particles as emerging from the quantization of some classical field theory; hence gauge theories quickly caught the attention of many scientists. Gauge theory can be described as the study of some manifold over which particles move, along with some vector bundle which consists of the state space of some property of the particles. This property often possesses some kind of *internal* symmetry, which is most often empirically established, which gives rise to a certain *gauge group* that describes these symmetries. This in turn provides non-trivial ways to *glue together* the fibres to construct vector bundles which are of topological interest, and which induce interesting *invariants*. The main gauge group of study here will be the group $SU(2)$. The physicists Yang and Mills were the first to realize that proper physical theories were required to be invariant under so-called gauge-transformations of $SU(2)$. We will therefore study 4-dimensional spacetimes and $SU(2)$-bundles thereon.

The theory is accompanied by a specific action, called the *Yang-Mills action*. Global minima of this action are called *instantons* and are the main topic of this thesis. These objects are solutions to specific equations of motion in *Euclidean space*, and numerous applications of this fact exist. For instance, Euclidean space can be thought of Minkowski space but with *imaginary time*. Then the path integral formalism by Feynman tells us to focus on $\exp(iS)$ where S denotes the action on Minkowski space, and of which the Euclidean action is a multiple, with an additional factor i. Therefore minima of the Euclidean action correspond to maximum values of $\exp(iS)$, which will dictate the result of the path integral, see also [5]. This provides an alternative way of looking at *quantum tunneling* effects. Namely, continuing to imaginary time inverts the sign of the potential, so that we may regard the tunneling particle as if moving classically along the inverted potential. The solutions to this problem are then given by the instantons. A more in-depth analysis of the applications of instantons and a worked example of how these notions correctly describe the double well potential can be found in [10]. This thesis will not focus on these applications.

It is clear that these instantons are objects of interest, and the main topic of this thesis is to explore the famous construction which provides an explicit way to find them, called the *ADHM-construction*. In order to formally understand the notion of instantons and ultimately the ADHM-construction, we will require a number of concepts from differential geometry, which we will review first. This includes the notions of *Lie-Groups*, *vector fields*, *connections*, the *exterior covariant derivative* and the *curvature*. This introduction will be brief and most results presented will not be rigorously proved.

Subsequently we will concern ourselves with the aforementioned gauge theories and we will apply the previously introduced notions to investigate how they give rise to the topological invariants and how to exploit these to prove an equivalent statement to the definition of instantons, which will be invaluable to understanding their importance. Also, a brief side note will be made about the extensive topic of *Chern-Simons theory*, which uses the same language and concepts as those developed in the first chapter, and directly ties in with the topological invariants.

Having established that, we will in the next chapter be able to give a complete description of the ADHM-construction, which, given certain conditions, constructs a vector bundle and an associated instanton. Of course the most natural way to proceed would be to wonder what these instantons would actually look like in practice, and how convenient a way the ADHM-construction actually is to explicitly calculate instantons. This will be done, and expressions for instantons of low *topological charge* will be found. In particular, the well-known *BPST-instanton* of unit charge will be thoroughly explored. Then we will generalize to a method of finding instantons of arbitrary topological charge.

The aim of the concluding section is to motivate the origin of the conditions that appeared in the ADHM-construction, called the *ADHM-equations*. This will entail studying mathematical objects called *quivers* and their representations corresponding to the *preprojective algebra*. Lastly, two *moment maps* will be constructed whose zeroes resemble the ADHM-equations.

I would like to thank my supervisor prof. dr. Erik Verlinde for introducing me to the ADHM-construction and for his useful explanations and suggestions which shaped the lion's share of chapter 3 to chapter 5. I would also like to thank my supervisor dr. Raf Bocklandt for repeatedly finding time on short notice to discuss my progress and to answer my numerous questions about the material. This thesis would not have turned out the same if it wasn't for their indispensable contributions.

Lastly I would like to briefly thank Ruben La for telling me how to get the quiver on the front page with infinite detail. Now any reader can zoom in on it to their liking.

2 Connections and curvature

In order to understand the notion of instantons and ultimately the ADHM-construction, we will require a number of concepts from differential geometry, which we will review first. This includes the notions of *Lie-Groups*, *vector fields* on a manifold and, most importantly, *connections*, their associated *exterior covariant derivatives* and the *curvature*. This introduction will be brief and most results presented will not be rigorously proved, as they follow from direct and rather unintuitive calculations. Emphasis will be put on motivating the definitions and outlining their geometrical meanings, so as to not dwell on the details too much.

For those interested in the proofs of these perliminary results, we refer to [1].

2.1 Preliminary differential geometry

We will ease into the theory by starting off with a number of fundamental definitions. These definitions will not be as precisely formulated as in any proper introduction to elementary differential geometry, but they will suffice for their purpose as a mere reminder and a reference tool. Furthermore, we will not introduce the concept of a manifold formally; roughly it is an object that locally looks like Euclidean space of a fixed dimension n, but globally may not. Thus it is a set M together with a collection of charts $\phi_\alpha : \mathbb{R}^n \supset V_\alpha \to U_\alpha \subset M$ which are bijections and such that each V_α is open and all U_α together cover all of M, subject to a whole cascade of further requirements. For more information, we refer to [4]. The following definition introduces some useful notation we will use throughout this thesis.

Definition 2.1. Given a manifold M, let $C(M)$ denote the space of all smooth functions $f : M \to \mathbb{R}$. A *tangent vector* ν at a point $p \in M$ is defined to be a map $\nu : C(M) \to \mathbb{R}$ that is linear and also satisfies the Leibniz-rule $\nu(fg) = \nu(f)g(p) + f(p)\nu(g)$, for all $f, g \in C(M)$. Write T_pM for the tangent space of a point $p \in M$, which consists of all tangent vectors at p, and $T(M)$ for the full tangent space of M, which is defined as the union of all tangent spaces of points $p \in M$. A *vector field* v on a manifold M is a map $v : M \to T(M)$ that is smooth and has the property that $v(p) \in T_pM$ for all $p \in M$. Note that we can also view v as a function $C(M) \to C(M)$ by writing $(v(f))(p) = (v(p))(f)$ for all $p \in M$ and $f \in C(M)$. Write $\mathfrak{X}(M)$ for the space of all vector fields on M, equipped with an operation, $[.,.]$, called the *Lie-bracket*, defined as $[v, w] = v \circ w - w \circ v$. Locally, we can write $\{\partial_\mu | 1 \leq \mu \leq \dim(M)\}$ for a basis of all vector fields, satisfying $[\partial_\mu, \partial_\nu] = 0$.

Note that the suggestive notation ∂_μ emphasizes the differentiation nature of the tangent vectors expressed by the Leibniz-rule, and that $[\partial_\mu, \partial_\nu] = 0$ expresses the well-known commutativity of differentiation in standard calculus. Recall that the differential $\mathrm{T}f : T(M) \to T(N)$ of any smooth map $f : M \to N$ between manifolds can be defined in a canonical way, by defining $T_p f : T_p M \to T_{f(p)}(N)$ by $T_p f(\nu)(g) = \nu(g \circ f)$ for all $\nu \in T_p M$ en $g : N \to \mathbb{R}$. For more details, we again refer to [4].

Definition 2.2. Manifolds that are equipped with a group structure are called *Lie-groups*, and are naturally equipped with left- and right-multiplication maps L_g and R_g respectively, that diffeomorphically map G onto itself. Any vector field v on a Lie-group G that satisfies $\mathrm{T}_h L_g(v(h)) = v(gh)$ is called left-invariant. We write $\mathfrak{g}$ for the space of left-invariant vector fields on G, which we will call the *Lie-algebra* of G. Note there exists an isomorphism $\mathfrak{g} \cong T_e G$ through $v \mapsto v(e)$.

We will also need a small lemma from linear algebra, which is presented below, and the proof of which will be omitted.

Lemma 2.3. *Let V be a vector space. Then $End(V) \cong V \otimes V^*$, where V^* denotes the dual space of V and $End(V)$ the space of linear endomorphisms of V. The isomorphism is given by writing $(v \otimes f)(w) = f(w)v$, where $v \in V$ and $f \in V^*$.*

It might be wise to briefly introduce a useful way of writing down equations involving many summations, called the Einstein summation convention. In short, this comes down to abbreviating

$$\sum_{\mu} v^{\mu} w_{\mu} =: v^{\mu} w_{\mu}.$$

Note that this notation is only used to shorten expressions involving a raised and a lowered index; repeated indices should be interpreted as implicit sums. This convention will make many expressions much easier to work with, and, with some experience, also much easier to read.

2.2 Connections and vector potentials

Recall that a vector bundle over a manifold M consists of a manifold E and a projection map $\pi : E \to M$ such that each fibre $\pi^{-1}(p) = E_p$ is a vector space, subject to a couple of local trivialization requirements listed in [4]. Given a manifold M and a such vector bundle E over M, one can define sections $s : M \to E$ to be functions that are smooth and also satisfy the property $\pi \circ s = \mathrm{id}$; that is, for any $p \in M$, we have that $s(p) \in E_p$, where E_p denotes the fibre of p of the vector bundle E. Now one can imagine that it is desirable to have a well defined way to compute the derivative of such sections in some given direction, but one obvious complication stands in our way; s maps to different spaces at each point. This leads to the interesting result that the derivative of a section is no longer uniquely defined, as we will see momentarily.

Definition 2.4. Let $\Gamma(E)$ denote the space of smooth sections of E. A *connection* D on M is defined to be a map $\Gamma(E) \times \mathfrak{X}(M) \to \Gamma(E)$ that is linear in both arguments and also satisfies the properties

$$D_v(fs) = v(f)s + fD_v(s) \quad \text{and} \quad D_{fv}(s) = fD_v(s),$$

for all $v \in \mathfrak{X}(M)$, $f \in C(M)$ and $s \in \Gamma(E)$.

Again a Leibniz-like rule appears in the definition of the connection, implying that this will indeed induce some sort of derivative of sections.

Example 2.5. As a rare exception, we will go through the small effort of actually calculating the value of $D_v(s)$, given arbitrary $v \in \mathfrak{X}(M)$ and $s \in \Gamma(E)$. To do this, we will work locally on some open $U \subset M$ and let $\{\partial_\mu\}$ be a local basis for $\mathfrak{X}(U)$ and $\{e_j\}$ be a basis of sections of E over U. We can then for any μ and j define functions $A^i_{\mu j}$, by

$$D_\mu(e_j) = A^i_{\mu j} e_i,$$

where we made use of the convention to write $D_{\partial_\mu} = D_\mu$. This will allow us to compute that for $v = v^\mu \partial_\mu$ and $s = s^i e_i$, we have that

$$D_v(s) = v^\mu(\partial_\mu s^i + A^i_{\mu j} s^j) e_i,$$

making repeated use of the properties of D and where we have exploited the Einstein summation convention to make the expression more manageable. △

Note the remarkable fact that, as a consequence of the Leibniz-rule, we end up with two sets of terms in the expression of $D_v(s)$; one that actually involves derivatives in the sense we're familiar with, and another one that does not; namely $v^\mu A^i_{\mu j} s^j e_i$. This observation leads us to the following definition.

Definition 2.6. One defines the *vector potential* A as the $C(M)$-linear map $\Gamma(E) \times \mathfrak{X}(U) \to \Gamma(E)$, or $\mathrm{End}(E)$-valued 1-form for short, as

$$A = A^j_{\mu i} e_j \otimes e^i \otimes dx^\mu,$$

where $\{e^i\}$ denotes the dual basis of $\{e_j\}$, and $\{dx^\mu\}$ denotes the dual basis of $\{\partial_\mu\}$. It is common to write

$$A_\mu = A^j_{\mu i} e_j \otimes e^i, \quad \text{so that} \quad A_\mu \in \Gamma(\mathrm{End}(E)) \quad \text{and} \quad A = A_\mu \otimes dx^\mu.$$

Note this definition makes use of lemma 2.3. It can be checked by direct calculation that now indeed $A(v)s = v^\mu A^i_{\mu j} s^j e_i$, so that this definition of A is in accordance with the aforementioned results. Also note that the defining difference between D and A is that A is $C(M)$-linear, that is, $A(v)(fs) = fA(v)(s)$ for all $v \in \mathfrak{X}(U)$, $s \in \Gamma(E)$ and $f \in C(M)$, whereas D need not be. We can now state a fundamental result about connections and vector potentials in general.

Theorem 2.7. *Given any connection D and any vector potential A, the operator $D + A$ is again a connection. Furthermore, given some connection D^0, locally any connection can be expressed as $D = D^0 + A$ for some vector potential A.*

The first assertion can be verified through direct computation, whereas the second one is somewhat more involved, but the proof will not be presented here and can be found in [1].

Definition 2.8. Usually D^0 is chosen to satisfy $D_v^0(s) = v(s^j)e_j$, that is, $A = 0$. Then D^0 is called the *standard flat connection*, which depends on the choice of bases and is therefore not uniquely defined. Furthermore, this definition applies only locally, and for a general vector bundle it is possible that no such a flat connection exists.

Remark 2.9. It is important to observe that for any trivial bundle $M \times V$ for some vector space $V \cong \mathbb{R}^k$, we have a canonical choice for the standard flat connection. Namely, we can consider s as a map $M \to \mathbb{R}^k$, to define $D_v^0(s) = ds(v)$, where d denotes the straightforward extension of the ordinary 1-dimensional exterior derivative of functions $f : M \to \mathbb{R}$.

We will introduce one final operation that follows naturally from the notion of a connection.

Definition 2.10. Given a section s of some vector bundle E over a manifold M, we can turn s into an *E-valued 1-form* $d_D s$ by letting $d_D s(v) = D_v(s)$ for all $v \in \mathfrak{X}(M)$; we will refer to this operation as the *E-valued derivative* of a section s.

2.3 Curvature and the exterior covariant derivative

Armed with this basic understanding of connections and vector potentials, we will now be able to define a key concept in the theory of differential geometry. It may come across quite arbitrary at first sight, but nonetheless it is an extremely powerful tool with a wide range of applications.

Definition 2.11. The *curvature* $F : \Gamma(E) \times \mathfrak{X}(M)^2 \to \Gamma(E)$ is an $\mathrm{End}(E)$-valued 2-form, defined in terms of a connection D, to satisfy

$$F(v, w) = D_v D_w - D_w D_v - D_{[v,w]} = [D_v, D_w] - D_{[v,w]}.$$

Given some basis $\{\partial_\mu\}$ of $\mathfrak{X}(M)$, we denote

$$F_{\mu\nu} := F(\partial_\mu, \partial_\nu) \in \mathrm{End}(E), \quad \text{so that} \quad F(v, w) = v^\mu w^\nu F_{\mu\nu}.$$

Note the antisymmetry; for any μ, ν we have that $F_{\mu\nu} = -F_{\nu\mu}$. Recall that $[\partial_\mu, \partial_\nu] = 0$, so that $F_{\mu\nu} = [D_\mu, D_\nu]$. Lastly, given a dual basis $\{dx^\mu\}$ of $\{\partial_\mu\}$, we can express

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu,$$

where the factor $1/2$ accounts for the double-counting that arises from the antisymmetry of both $F_{\mu\nu}$ and $dx^\mu \wedge dx^\nu$.

The observant reader will have raised an eyebrow whilst reading the above definition, and eagerly awaits the following lemma, which can be checked by direct computation.

Lemma 2.12. *The addition* End(E)-valued form *in the definition of F is justified, i.e. F is indeed linear in all arguments and also $C(M)$-linear. That is, for any $f \in C(M)$ and $s \in \Gamma(E)$, we have*

$$F(fv, w)(s) = F(v, fw)(s) = F(v, w)(fs) = fF(v, w)(s).$$

Example 2.13. In order to achieve a more concrete view of the behaviour of the curvature, we will once again introduce $\{e_i\}$ as a local basis of sections, along with its dual $\{e^j\}$. Direct computation then shows us that, should one introduce the functions $F^j_{\mu\nu i}$, we may write

$$F_{\mu\nu} = F^j_{\mu\nu i} e_j \otimes e^i, \quad \text{where} \quad F^j_{\mu\nu i} = \partial_\mu A^j_{\nu i} - \partial_\nu A^j_{\mu i} + A^j_{\mu k} A^k_{\nu i} - A^j_{\nu k} A^k_{\mu i},$$

which is often abbreviated to

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu].$$

In the last expression all indices regarding the *local* basis of sections have been repressed, and therefore care should be taken in its use. △

Lest the reader think that this is all there is to say about the curvature, we hasten to introduce one final, novel concept, which will prove to be imperative in the description of Yang-Mills theory.

It is not hard to imagine that every End(E)-valued p-form can be written as a linear combination of $T \otimes \omega$ for some $T \in \Gamma(\mathrm{End}(E))$ and p-form ω. We will use this fact without giving a proof, to help us define an important operation on End(E)-valued forms.

Definition 2.14. We will constructively define the *exterior covariant derivative* $d_\mathcal{D}$ of End(E)-valued differential forms, by

$$d_{\mathcal{D}^0} T = \partial_\mu T \otimes dx^\mu,$$

for all $T \in \Gamma(\mathrm{End}(E))$ and $s \in \Gamma(E)$, where ∂_μ acts on T in an analogous way to how $\partial_\mu A_\nu$ was defined in the above expression for the curvature. Again, $\{dx^\mu\}$ denotes the dual basis of the basis vector fields $\{\partial_\mu\}$. We then proceed to define for any differential form ω,

$$d_{\mathcal{D}^0}(T \otimes \omega) = d_{\mathcal{D}^0} T \wedge \omega + T \otimes d\omega,$$

where d denotes the usual exterior derivative on differential forms, and where the wedge-product is defined as

$$(T \otimes \omega) \wedge \eta = T \otimes (\omega \wedge \eta),$$

where η is also an ordinary differential form. For any other connection D, define for any End(E)-valued p-form η

$$d_\mathcal{D} \eta = d_{\mathcal{D}^0} \eta + A \wedge \eta - (-1)^p \eta \wedge A.$$

Definition 2.15. Consider two End(E)-valued forms; $S \otimes \omega$ and $T \otimes \eta$. Then define the wedge product

$$(S \otimes \omega) \wedge (T \otimes \eta) = (S \circ T) \otimes (\omega \wedge \mu)$$

and for general End(E)-valued forms this definition can be extended linearly.

Lemma 2.16. *For any End(E)-valued p-form ω and End(E)-valued form η, it holds that*

$$d_{\mathcal{D}}(\omega \wedge \eta) = d_{\mathcal{D}}\omega \wedge \eta + (-1)^p \omega \wedge d_{\mathcal{D}}\eta.$$

Furthermore, for any flat connection D^0, it holds that $d^2_{\mathcal{D}^0} = 0$.

This will allow us to formulate a non-trivial, yet very elegant property of the curvature that holds very generally, but can be applied in numerous ways.

Theorem 2.17. *We always have the* Bianchi identity*;*

$$d_{\mathcal{D}}F = 0.$$

This equation is equivalent to its local form

$$D_\mu F_{\nu\lambda} + D_\nu F_{\lambda\mu} + D_\lambda F_{\mu\nu} = 0.$$

Example 2.18. Consider the 2-form

$$d_{\mathcal{D}^0}A + A \wedge A,$$

where D^0 is some flat connection. We investigate which section of End(E) is coupled to $dx^\mu \wedge dx^\nu$. Write $A = A_\mu dx^\mu$. Then $d_{\mathcal{D}^0}A$ contributes $\partial_\mu A_\nu - \partial_\nu A_\mu$ to the $dx^\mu \wedge dx^\nu$-term, since the $A_\mu ddx^\mu$ terms in the definition of $d_{\mathcal{D}^0}$ vanish. Somewhat more directly, it follows that the $A \wedge A$ term contributes $A_\mu A_\nu - A_\nu A_\mu = [A_\mu, A_\nu]$. Combining these results with example 2.13, we conclude that we may locally write

$$F = d_{\mathcal{D}^0}A + A \wedge A.$$

This pretty and concise formula can be applied globally if the vector bundle admits a global flat connection. $\triangle$

In general, the following result can be obtained.

Proposition 2.19. *Let D^0 be a connection on a vector bundle over a manifold, with associated curvature F^0. If $D = D^0 + A$ is another connection, it holds that*

$$F = F^0 + d_{\mathcal{D}^0}A + A \wedge A.$$

2.4 The Hodge-star operator

At the center of Yang Mills Theory will be a special operator: $\star$. Just like the connection on a vector bundle, this operator will not be unique in the sense that there is a canonical way to define it, given just a manifold. We will require to impose some additional structure on M, which will allow us to define the inner-product of two ordinary differential p-forms.

Definition 2.20. Let M be a semi-Riemannian manifold; that is, there exists a non-degenerate inner-product $g_p : T_pM \times T_pM \to \mathbb{R}$ that is not necessarily positive-definite. Let $\{\partial_\mu\}$ be a basis of vector fields, so that for $v, w \in \mathfrak{X}(M)$ we may write $v = v^\mu \partial_\mu$ and $w = w^\mu \partial_\mu$. Define the functions

$$g_{\mu\nu}(p) = g_p(\partial_\mu(p), \partial_\nu(p)).$$

We may then view g as the matrix with entries $g_{\mu\nu} : M \to \mathbb{R}$, so that

$$g(v, w) = g_{\mu\nu} v^\mu w^\nu.$$

Denote the entries of the matrix g^{-1} by $g^{\mu\nu}$. Then we define the inner-product of two 1-forms $\omega = w_\mu dx^\mu$ and $\eta = \eta_\nu dx^\nu$ by

$$\langle \omega, \eta \rangle = g^{\mu\nu} \omega_\mu \eta_\nu.$$

Then the *inner-product of two p-forms* ω and η can be computed by linearly continuing the map $\langle ., . \rangle$, defined by

$$\langle dx_1^i \wedge \ldots \wedge dx^{i_p}, dx^{j_1} \wedge \ldots \wedge dx^{j_p} \rangle = \det[(\langle dx^{i_n}, dx^{j_m} \rangle)_{n,m}].$$

We say that $\{dx^\mu\}$ is an orthonormal basis of 1-forms if $\langle dx^\mu, dx^\nu \rangle = \pm \delta_{\mu\nu}$, where the sign is usually denoted by $\epsilon(\mu)$.

Inverting g in the above definition arises from the underlying concept of the metric allowing us to freely transform vector fields into 1-forms and vice versa. Namely,

$$v \mapsto g(v, \cdot) = g_{\mu\nu} v^\mu dx^\nu$$

transforms a vector field into a 1-form, and

$$\omega \mapsto g^{\mu\nu} w_\mu \partial_\nu$$

transforms this 1-form back into its vector field.

Also observe that the inner product of two differential forms is *not* a number, but instead a *function.* Therefore it is fundamentally different from an ordinary inner product on vector spaces. With these notions in hand, it is just a small step to defining the long-awaited operator. Recall that an orientation on a manifold M is a smooth choice of orientations of its tangent spaces T_pM for all $p \in M$.

In the remainder of this section we will assume that the manifold of interest, M, is *flat* in the sense that its Riemannian curvature, not to be confused with the curvature 2-form, vanishes. We opt not to go into the details of this broad field of study any further, but all that matters for now is that it ensures the existence of a positively oriented *orthonormal* basis of 1-forms; something that we will need for the following definition.

Definition 2.21. We define the *volume form* vol on a flat and oriented n-dimensional manifold M by

$$\mathrm{vol} = dx^1 \wedge \ldots \wedge dx^n,$$

where $\{dx^1, \ldots, dx^n\}$ is a positively oriented orthonormal basis of 1-forms. Usually the definition of the volume form would also comprise a factor of $\sqrt{|\det g|}$, but this vanishes by our assumptions about our basis. Write $\Omega^p(M)$ for the space of differential forms on M. Then there exists an operator $\star : \Omega(M) \to \Omega(M)$ such that for all $p \in \{1, \ldots, n\}$ and $\omega, \eta \in \Omega^p(M)$, we have

$$\omega \wedge \star\eta = \langle \omega, \eta \rangle \mathrm{vol},$$

called the *Hodge-star operator*. We will henceforth refer to $\star\omega$ as the *dual* of ω. Explicitly, if $\{dx^1, \ldots, dx^n\}$ is a positively oriented orthonormal basis of 1-forms, we have that

$$\star(dx^{i_1} \wedge \ldots \wedge dx^{i_p}) = \mathrm{sgn}(\tau)\epsilon(i_1) \cdots \epsilon(i_p) dx^{i_{p+1}} \wedge \ldots \wedge dx^{i_n},$$

where $\{i_{p+1}, \ldots i_n\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_p\}$ and τ denotes the permutation $(1, \ldots n) \mapsto (i_1, \ldots, i_n)$.

Example 2.22. As an easy example, consider the general form of the curvature F on a flat and oriented 4-dimensional manifold, given by

$$\begin{aligned} F = {} & F_{12} dx^1 \wedge dx^2 + F_{13} dx^1 \wedge dx^3 + F_{14} dx^1 \wedge dx^4 \\ & + F_{23} dx^2 \wedge dx^3 + F_{24} dx^2 \wedge dx^4 + F_{34} dx^3 \wedge dx^4. \end{aligned}$$

We define the Hodge star operator on $\mathrm{End}(E)$-valued forms by $\star(T \otimes \omega) = T \otimes \star\omega$ for all $T \in \Gamma(\mathrm{End}(E))$ en $\omega \in \Omega(M)$. Choosing $\{dx^1, dx^2, dx^3, dx^4\}$ as an orthonormal basis with all $\epsilon(i) = 1$, keeping track of the proper signs, we obtain

$$\begin{aligned} \star F = {} & (+1) F_{12} dx^3 \wedge dx^4 + (-1) F_{13} dx^2 \wedge dx^4 + (+1) F_{14} dx^2 \wedge dx^3 \\ & + (+1) F_{23} dx^1 \wedge dx^4 + (-1) F_{24} dx^1 \wedge dx^3 + (+1) F_{34} dx^1 \wedge dx^2. \end{aligned}$$

Note that we could rid some minus signs exploiting the antisymmetry $F_{\mu\nu} = -F_{\nu\mu}$. $\triangle$

Remark 2.23. We remark that for any Riemannian manifold M, we have for any flat orthonormal basis of 1-forms $\{dx^\mu\}$ that $\epsilon(\mu) = 1$, so that we may write that

$$\star(dx^{i_1} \wedge \ldots \wedge dx^{i_p}) = \mathrm{sgn}(\tau) dx^{i_{p+1}} \wedge \ldots \wedge dx^{i_n},$$

where τ still denotes the permutation $\{1, \ldots n\} \mapsto \{i_1, \ldots, i_n\}$. This will be useful later.

3 Chern-Simons and Yang-Mills Theory

The two main sources used for this chapter were chapter 1 and 2 from [1] and chapter 3 from [2]; additional information on the topics discussed can be found there.

3.1 Electromagnetism on a trivial bundle

First we will try to motivate the mathematical concepts introduced in the previous chapter, by applying them to the well-known theory of classical electromagnetism. To start off this section, we will introduce a concept regarding the extensive field of study involving *gauge theory*, which turns out to be the physical framework within which the applications are most abundant.

Definition 3.1. Let G be a Lie-group with Lie-algebra $\mathfrak{g}$ and with some representation ρ on some vector space V, where a *representation* of G is defined to be a homomorphism $\rho : G \to \mathrm{End}(V)$. For a manifold M, define $E = M \times V$ to be a *trivial G-bundle*; in this setting, G is called the *gauge-group*. We say that some $T \in \Gamma(\mathrm{End}(E))$ *lives in* $\mathfrak{g}$ if for all $p \in M$, we have that $T(p, v) = (p, \mathrm{T}\rho(X_p)v)$ for some $X_p \in \mathfrak{g}$ and for all $v \in V$. Here $\mathrm{T}\rho$ denotes the differential of the map between manifolds $\rho : G \to \mathrm{End}(V)$. Note that here we identified $\mathrm{End}(V)$ with its tangent space at each point. Futhermore, we say that some connection D is a *G-connection* if the corresponding A_μ live in $\mathfrak{g}$ for all μ. Note that G acts on E by $g \cdot (p, v) = (p, \rho(g)v)$.

Example 3.2. Consider the simple case of $G = U(1) = \{e^{i\phi} \mid \phi \in [0, 2\pi)\}$, with fundamental representation $\rho : G \to \mathbb{C}$ that is just the embedding. Then $\mathrm{End}(\mathbb{C}) = \mathbb{C}$; namely, all endomorphisms of $\mathbb{C}$ are given by $z \mapsto \alpha z$ for some $\alpha \in \mathbb{C}$. Futhermore, since ρ is simply an embedding, we have that $\mathrm{T}\rho$ is also just the identity map. Suppose we are dealing with a G-connection D. This means that the components of the vector potential must live in $\mathfrak{g} = \mathfrak{u}(1) \cong \mathbb{R}$, where the last isomorphism holds since the tangent space to a point in the unit circle is nothing but a straight line. In other words, if A_μ lives in $\mathfrak{u}(1)$, we must have that $A_\mu(p)$ acts on $\mathbb{C}$ by nothing more than scalar multiplication by some real number; that is, A_μ maps to $\mathbb{R}$, making in this case A nothing more than an ordinary 1-form on M. △

Remark 3.3. It should be noted that in the above example we technically have that $\mathfrak{u}(1) = i\mathbb{R}$, so that A_μ is actually a purely complex valued 1-form on M. However, we opt to omit this additional factor of i in the forthcoming example, since it will only clutter the computations.

Example 3.4. Recall the Hodge-star operator $\star$ introduced in chapter 2. To see why this is an interesting operator to study, we will consider an important example that arises from the physics of electromagnetism. This is based on the observation that, when one considers the right manifold and the right forms, it is possible to reformulate Maxwell's equations more concisely using the language of differential geometry. Essentially, the *electric field* E is represented by

$$E = E_x dx + E_y dy + E_z dz$$

and the *magnetic field* B by

$$B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy,$$

which will both live on $\mathbb{R}^4$, representing Minkowski spacetime, on which we will use the convention that $x^0 = t$. The first two Maxwell equations,

$$\nabla \cdot \vec{B} = 0 \quad \text{and} \quad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0,$$

are then simply equivalent to writing

$$d_S B = 0 \quad \text{and} \quad \partial_t B + d_S E = 0,$$

where d_S denotes the spacelike part of the exterior derivative; it disregards any time-dependence of the functions making up the differential form.

In this setting, consider the trivial $U(1)$-bundle $E = \mathbb{R}^4 \times \mathbb{C}$ as in example 3.2. Since we may choose our $U(1)$-connection freely, we will opt for the vector potential A the differential 1-form

$$A = (\varphi/c)dt + \mathcal{A}_1 dx + \mathcal{A}_2 dy + \mathcal{A}_3 dz,$$

where $\mathcal{A}$ denotes the magnetic potential, defined by $\nabla \times \mathcal{A} = \vec{B}$, and φ the electric potential, defined to satisfy $\vec{E} = -\nabla\varphi - \frac{\partial \mathcal{A}}{\partial t}$. Note that it is a priori not at all clear that such A and φ should exist, but for now, we assume that they do. We can then calculate the curvature F, in this setting better known as the *electromagnetic tensor*, using

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

where we have used the commutivity of the components of A to rid the commutator $[A_\mu, A_\nu]$. Direct computation gives us then that

$$F = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy + (E_x/c)dx \wedge dt + (E_y/c)dy \wedge dt + (E_z/c)dz \wedge dt.$$

Note that F can also be described by the much simpler equation

$$F = B + (E/c) \wedge dt.$$

Crucial is to observe that, in hindsight, we did not really need the vectorpotential to define the curvature at all. It is then easy to see that the first two Maxwell equations are even more concisely described by

$$dF = 0;$$

that is, they are nothing but the Bianchi identity for the curvature, where we use that $d_{\mathcal{D}} = d$ in this case. Using the previous example for the B-field, but remembering that $\langle dt, dt\rangle = -1$ in the Minkowski case for the E-field, we are able to calculate

$$\star F = -B_x dx\wedge dt - B_y dy\wedge dt - B_z dz\wedge dt + (E_x/c)dy\wedge dz + (E_y/c)dz\wedge dx + (E_z/c)dx\wedge dy.$$

In vacuum, the last two Maxwell equations are given by

$$\nabla\cdot\vec{E} = 0 \quad \text{and} \quad \nabla\times\vec{B} - \frac{\partial\vec{E}}{\partial t} = 0,$$

which can also by direct computation be verified to be described by

$$d\star F = 0.$$

It follows that the Hodge star operator takes on a real physical meaning as that of transforming the curvature F that describes the geometry of electromagnetism, into a different 2-form that instead describes the dynamics. The equation above is often referred to as the *Yang-Mills equation.* △

Note that any curvature that would happen to possess the property $F = \pm\star F$, would automatically satisfy the Yang Mills equation, since the Bianchi identity holds in general. This observation may look trivial, but it will be of crucial importance towards the end of this chapter.

We conclude by describing what a general G-bundle is for some Lie-group G. The idea is essentially that one glues together a couple of the aforementioned trivial G-bundles, with transition maps that satisfy certain properties.

Definition 3.5. Let M be a manifold and consider a cover $\{U_\alpha\}_{\alpha\in I}$ consisting of open neighbourhoods to which a certain atlas of charts for M maps. Also consider a Lie-group G with some representation ρ onto some vector space V. To each U_α, stick a trivial vector bundle $U_\alpha\times V$, and consider the disjoint union $\bigsqcup_{\alpha\in I} U_\alpha\times V$. In order to turn this into a single vector bundle over M, to each $p\in M$ we should associate precisely one vector space. For $p\in U_\alpha\cap U_\beta$, this is not yet the case. Hence we must identify the two copies of V that have been coupled to p, and the idea is that given our group G, there are non-trivial ways to do this. Namely, we define *transition functions* $g_{\alpha\beta}: U_\alpha\cap U_\beta\to G$, and we identify

$$U_\alpha\times V\ni(p,v)\sim(p,v')\in U_\beta\times V \iff v = \rho(g_{\alpha\beta}(p))v'.$$

Formally, we also require $g_{\alpha\alpha} = 1$ for points inside a single neighbourhood, and in order to not run into trouble for points in $U_\alpha\cap U_\beta\cap U_\gamma$, we also require that $g_{\alpha\beta}g_{\beta\gamma}g_{\gamma\alpha} = 1$; which is called the *cocycle condition.*

Definition 3.6. A *gauge transformation* g is defined to be a map $M\to G$ that acts on a vector bundle by $g(p,v) = (p,\rho(g(p))v)$. Observe that we may regard $\rho\circ g$ as a section of $\mathrm{End}(E)$, since $\rho(g(p))\in\mathrm{End}(E)$ for all $p\in M$.

In electromagnetism the gauge group considered is $U(1)$, but in Yang Mills theory, what will be discussed next, a more convoluted gauge group is often the center of attention; the group $SU(2)$.

3.2 The Yang-Mills action

The idea of studying gauge theories arose from the observation that some sets of elementary particles shared a great many properties when considered interacting with the strong nuclear force. For example, protons and neutrons are both nucleons, spin-1/2 fermions to be precise, with virtually indistinguishable masses and similar behaviour when subjected to the strong nuclear force. It therefore led to the idea of *isospin*; that maybe protons and neutrons were just two different flavours of the same kind. Generally the *isospin-doublets* are defined to be

$$p \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad n \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

It turns out that the group that formally described many of the observed symmetries was $SU(2)$, a group of 2×2-matrices which will be introduced momentarily. This idea caught the interest of many physicists when it turned out that the pions $\pi^{\pm}$ and π^0, when described by an isospin triplet, had symmetries that could be described by not a trival representation of $SU(2)$, but rather a 3-dimensional respresentation of the same group. Therefore the group $SU(2)$ became of particular interest to physicists, and the search for theories that were invariant under the action of $SU(2)$ began. Yang and Mills were the first to realize that requiring only invariance of this kind was insufficient, and they proposed to study theories that were invariant not only under fixed elements of $SU(2)$, but even under *any* gauge-transformation of $SU(2)$. We will therefore study 4-dimensional spacetimes and $SU(2)$-bundles thereon.

First we will deduce some properties of $SU(2)$ that will be of use to us later.

Definition 3.7. Define $SU(2) \subset \mathbb{C}^{2\times 2}$ as the Lie-group of unitary 2×2 matrices having unit determinant. Explicitly, we can write

$$SU(2) = \left\{ \begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} : |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

As a result, we have a diffeomorphism $SU(2) \to S^3$. Namely, for $\alpha = x_1 + ix_2$ and $\beta = x_3 + ix_4$, we have a smooth and bijective map

$$\begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} \mapsto (x_1, x_2, x_3, x_4) \in S^3.$$

In particular we find that $SU(2)$ is a manifold of real dimension 3.

Proposition 3.8. *The Lie-algebra of $SU(2)$, denoted by $\mathfrak{su}(2)$, is given by all skew-hermitian matrices with vanishing trace.*

Proof. Consider the following three paths in $SU(2)$:

$$\gamma_1(t) = \begin{pmatrix} 0 & -e^{-it} \\ e^{it} & 0 \end{pmatrix}, \quad \gamma_2(t) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \quad \text{and} \quad \gamma_3(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}.$$

All these curves have the property that $\gamma_i(0) = \mathrm{id}_2$, so that their derivatives at zero are part of the Lie-algebra. We compute that

$$\gamma_1'(0) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \gamma_2'(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \gamma_3'(0) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

These matrices are most clearly linearly independent, and since the tangent space to $SU(2)$ has real dimension 3, it follows that they span the full tangent space. These matrices are trivially skew-hermitian and they also have vanishing trace. Also, for a general 2×2 skew-hermitian matrix with vanishing trace, we find that

$$\begin{pmatrix} ix & -y + iz \\ y + iz & -ix \end{pmatrix} = z \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} + y \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + x \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

for arbitrary $x, y, z \in \mathbb{R}$. It follows that the span of the three matrices equals the space of skew-hermitian matrices with vanishing trace. This completes the proof. $\square$

Most principles in physics can be distilled to one very fundamental principle; the action principle. Naively, it states that a (most often dynamical) process will happen if and only if the process extremizes something called the *action*,

$$S = \int_t L \, dt,$$

where L is called the Lagrangian and is dependent on the *path* taken by the process in the space of possibilities. In classical mechanics for instance, L is defined to be the difference between the kinectic and the potential energy, the time-dependence of which is determined by the path taken by the object in question. Extremizing the action defined in this way leads to the famous Euler-Lagrange equations, which in a sense fully describe classical mechanics.

Of course we are not prohibited from defining different actions, in hope of extending this idea to a more abstract and general setting. In order to define the action we will more closely inspect momentarily, we will first need one quick definition to help us get started.

Definition 3.9. Let E be a vector bundle over some manifold M and consider $T \in \Gamma(\mathrm{End}(E))$. Then define the *trace* of $T = T^i_j e_i \otimes e^j$ to be

$$\mathrm{tr}(T) = \sum_i T^i_i$$

as a function $M \to \mathbb{R}$. For any $\mathrm{End}(E)$-valued form $T \, dx_{i_1} \otimes \wedge \ldots \wedge dx_{i_k}$, we set

$$\mathrm{tr}(T \otimes dx_{i_1} \wedge \ldots \wedge dx_{i_k}) = \mathrm{tr}(T) \; dx_{i_1} \wedge \ldots \wedge dx_{i_k},$$

which is an ordinary differential form.

Lemma 3.10. *For an End(E)-valued p-form ω and an End(E)-valued q-form η, we have that*

$$\mathrm{tr}(\omega \wedge \eta) = (-1)^{pq}\mathrm{tr}(\eta \wedge \omega).$$

As a result, $\mathrm{tr}(d_{\mathcal{D}}\eta) = \mathrm{tr}(d_{\mathcal{D}^0}\eta)$*, from which it follows that*

$$d\,\mathrm{tr}(\eta) = \mathrm{tr}(d_{\mathcal{D}}\eta).$$

Proof. The proof can be found in [1]. □

It is most common to consider the *Yang-Mills Lagrangian*, which is given by

$$L = \mathrm{tr}(F \wedge \star F), \quad \text{so that} \quad S = \int_M \mathrm{tr}(F \wedge \star F).$$

It can be shown that in Euclidean space, this action is extremized precisely when

$$d_{\mathcal{D}} \star F = 0;$$

that is, when the Yang-Mills equation holds. Solving this equation in general is extremely difficult, so from now on we will assume that M is a 4-dimensional manifold; that is, M can be thought of as some general kind of space-time.

Example 3.11. Recall example 3.4; we remark that the quantity

$$F \wedge \star F = \left[-B_x^2 - B_y^2 - B_z^2 + (E_x/c)^2 + (E_y/c)^2 + (E_z/c)^2\right] dx \wedge dy \wedge dz \wedge dt,$$

should remind the reader of the Lagrangian of classical electromagnetism. This indicates that $F \wedge \star F$ is not such a foreign object to study, as it may appear at first sight. △

3.3 The Pontryagin index

We will consider ordinary Euclidean space and an $SU(2)$-bundle thereon, and we will limit ourselves to studying curvatures for which the Yang-Mills action is finite. The reason for this is *not* that when considering the path integral formalism, solutions with infinite action do not contribute to $\exp(iS)$, but rather that they are useful in doing semi-classical approximations; for more details, see [10]. In order to ensure this finiteness, we should at the very least have that $F \to 0$ sufficiently fast as $x \to \infty$. What this means for the vectorpotential will be clear from the following lemma.

Lemma 3.12. *Let g be a gauge-transformation. Then there exists a connection D' that satisfies $D'_v(\rho(g)s) = \rho(g)D_v(s)$ for all $v \in \mathfrak{X}(M)$ and $s \in \Gamma(E)$, which is explicitly given by $D'_v(s) = \rho(g)D_v(\rho(g)^{-1}s)$. Let A' be the vector potential associated by D', and F' its curvature. Then we have that*

$$A'_\mu = \rho(g)A_\mu\rho(g)^{-1} + \rho(g)\partial_\mu\rho(g)^{-1} \quad \textit{and} \quad F'_{\mu\nu} = \rho(g)F_{\mu\nu}\rho(g)^{-1}.$$

Proof. The proof can be found in [1]. □

As a brief side-note, we remark that the above lemma is sufficient to prove that the invariance under gauge-transformations, that what Yang and Mills so adamantly sought, is satisfied by their action.

Proposition 3.13. *The Yang Mills action is gauge invariant.*

Proof. Observe that $T\rho(g)^{-1} \wedge \rho(g)S = T \wedge S$, since the definition of the wedge product involves composition of the associated sections of $\mathrm{End}(E)$. Therefore we find that

$$\begin{aligned}\operatorname{tr}\left(F' \wedge \star F'\right) &= \operatorname{tr}\left(\rho(g)F\rho(g)^{-1} \wedge \rho(g)(\star F)\rho(g)^{-1}\right) = \operatorname{tr}\left(\rho(g)F \wedge (\star F)\rho(g)^{-1}\right) \\ &= \operatorname{tr}\left((\star F)\rho(g)^{-1} \wedge \rho(g)F\right) = \operatorname{tr}(\star F \wedge F) = \operatorname{tr}(F \wedge \star F),\end{aligned}$$

where we used proposition 3.10 twice. This proves the claim. □

It follows that if $F = 0$ we need not necessarily have $A = 0$; namely, any gauge-transformation g will change $A = 0$ to a vector potential $A' = \rho(g)\partial_\mu\rho(g)^{-1}$. It turns out that these are the only vector potentials that produce vanishing curvature; such vector potentials are called *pure gauge*. Therefore we opt to only consider vector potentials A for which $A \to \rho(g)\partial_\mu\rho(g)^{-1}$ for some gauge transformation g when $|x| \to \infty$. Important to observe is that this gauge transformation need not be defined on the full space. Indeed, to ensure the correct asymptotic behaviour we only need a gauge transformation defined on $\{|x| > R\}$ for some large R. This transformation need not necessarily extend to the full space. Namely, consider the transformation restriced to $\{|x| = R\} \cong S^3$. Since also $SU(2) \cong S^3$, we find that the transformation g can be viewed as a map from the 3-sphere to itself. It is a well-known fact from topology that such maps are classified by the third homotopy group of the 3-sphere; $\pi_3(S^3) \cong \mathbb{Z}$. The associated integers are called the *winding numbers*, or in this particular context, the *Pontryagin index*. Such a map can only be continuously extended to $\{|x| \leq R\}$ if the associated Pontryagin index is zero. One can naturally use this gauge transformation to define a G-bundle using the gauge transformation in a neighbourhood of $\{|x| = R\}$, which is only trivial if the winding number is zero.

Another way of thinking about the Pontryagin index, is to think of the different ways there are to construct an $SU(2)$-bundle over S^4. Namely, we may consider S^4 to be the 1-point compactification of $\mathbb{R}^4$, which can be seen by applying the stereographic projection. Therefore we may regard a curvature F that vanishes at infinity as a curvature defined on S^4 instead of only $\mathbb{R}^4$. Recall the definition of a G-bundle. Since we may cover S^4 by precisely two charts, one that covers the northern hemisphere and one the covers the southern with some overlap in the middle, it follows that in order to define an $SU(2)$-bundle over S^4, it suffices to give a map from this overlap region, which is homotopy equivalent to S^3, to $SU(2) \cong S^3$. We again find that vector bundles over S^4 can be classified by some integer; the winding number.

The main result of this section is that the curvature F can be used to calculate the Pontryagin index by evaluating the right integral. As a result, we find that this integral defines a topological invariant for a given vector bundle, a fact we will exploit at the end of this chapter. We can prove this invariance directly, without depending on its connection to the Pontryagin index.

Theorem 3.14. *Given a vector bundle E over a 4-dimensional Riemannian manifold M and a curvature F, the* topological charge n *is defined to be*

$$8\pi^2 \mathrm{n} = \int_M \mathrm{tr}(F \wedge F).$$

If F' is another curvature on E with topological charge n', then $\mathrm{n}' = \mathrm{n}$.

Proof. Let A and A' be two different vector potentials that live on M, and F and F' their respective curvatures. We will show that $\mathrm{tr}(F \wedge F) - \mathrm{tr}(F' \wedge F')$ is an exact 2-form. It will then follow by Stokes's theorem that

$$\int_M \mathrm{tr}(F \wedge F) - \int_M \mathrm{tr}(F' \wedge F') = \int_M d\omega = 0,$$

for some 2-form ω, since M has no boundary.

Write $\delta A = A' - A$, and let F_t be the curvature associated with $A_t = A + t\delta A$, so that $F_0 = F$ and $F_1 = F'$. Recall lemma 2.19, so that we can write for some fixed F^0;

$$\begin{aligned} F_t &= F^0 + d_{\mathcal{D}^0} A_t + A_t \wedge A_t \\ &= F^0 + d_{\mathcal{D}^0} A + t d_{\mathcal{D}^0} \delta A + \big(A \wedge A + tA \wedge \delta A + t\delta A \wedge A + t^2 \delta A \wedge \delta A\big) \\ &= F + t(d_{\mathcal{D}^0} \delta A + A \wedge \delta A + \delta A \wedge A) + t^2 \delta A \wedge \delta A \\ &= F + t d_{\mathcal{D}_A} \delta A + t^2 \delta A \wedge \delta A, \end{aligned}$$

where we used the definition of the exterior covariant derivative. It then follows that

$$\begin{aligned} \frac{d}{dt} F_t &= \frac{d}{dt} \big(F + t d_{\mathcal{D}_A} \delta A + t^2 \delta A \wedge \delta A\big) \\ &= d_{\mathcal{D}_A} \delta A + 2t \delta A \wedge \delta A \\ &= d_{\mathcal{D}^0} \delta A + A \wedge \delta A + t\delta A \wedge \delta A + \delta A \wedge A + \delta A \wedge t\delta A \\ &= d_{\mathcal{D}^0} \delta A + A_t \wedge \delta A + \delta A \wedge A_t \\ &= d_{\mathcal{D}_{A_t}} \delta A, \end{aligned}$$

where we made repeated use the definition of the exterior covariant derivative and the definition of A_t. Observe that from lemma 2.16

$$d_{\mathcal{D}_{A_t}}(\delta A \wedge F_t) = d_{\mathcal{D}_{A_t}} \delta A \wedge F_t + \delta A \wedge d_{\mathcal{D}_{A_t}} F_t = d_{\mathcal{D}_{A_t}} \delta A \wedge F_t,$$

where we used the Bianchi-identity. Now we find that

$$\begin{aligned} \frac{d}{dt} \mathrm{tr}\,(F_t \wedge F_t) &= \mathrm{tr}\left(\frac{d}{dt} F_t \wedge F_t + F_t \wedge \frac{d}{dt} F_t\right) \\ &= \mathrm{tr}\left(d_{\mathcal{D}_{A_t}} \delta A \wedge F_t + F_t \wedge d_{\mathcal{D}_{A_t}} \delta A\right) \\ &= 2\mathrm{tr}\left(d_{\mathcal{D}_{A_t}} \delta A \wedge F_t\right) \\ &= 2\mathrm{tr}\left(d_{\mathcal{D}_{A_t}} (\delta A \wedge F_t)\right) \\ &= 2d\mathrm{tr}\,(\delta A \wedge F_t)\,, \end{aligned}$$

where we used proposition 3.10. Finally, we find by the fundamental theorem of calculus that

$$\begin{aligned}\mathrm{tr}(F' \wedge F') - \mathrm{tr}(F \wedge F) &= \int_0^1 \frac{d}{dt}\mathrm{tr}(F_t \wedge F_t)\ dt = \\ &= \int_0^1 2d\mathrm{tr}\left(\delta A \wedge F_t\right)\ dt \\ &= d\left(\int_0^1 2\mathrm{tr}\left(\delta A \wedge F_t\right)\ dt\right),\end{aligned}$$

which proves the claim. □

We now present the connection to the aforementioned topological invariant and the Pontryagin index introduced above.

Theorem 3.15. *For any $SU(2)$-bundle E and a curvature F thereon, we have that the Pontryagin index equals minus the topological charge. As a result, the topological charge defined above is an integer.*

Proof. The proof can be found in [2]. □

3.4 The magnetic monopole

Recall electromagnetism formalized using differential forms on a $U(1)$-bundle. In this section we will only consider situations in which $E = 0$; that is, the electric field vanishes. Furthermore, we will forget about time for a second, and just consider statics that takes place in 3-dimensional space. Therefore we have that $F = B + (E/c) \wedge dt = B$. The Hamiltonian for a particle with charge q in this situation is

$$H = \frac{1}{2m}\left(\frac{\hbar}{i}\nabla - q\mathcal{A}\right) + V.$$

Interesting is to observe that $\frac{\hbar}{i}\nabla$ most closely resembles the standard flat connection, so that $q\mathcal{A}$ must take on the role of the vector potential. Of course, this is to be expected, as $\mathcal{A}$ is often referred to as the (magnetic) vector potential. It can be shown that the solution to the above equation can be written as

$$\psi = e^{ig}\psi_0, \quad \text{where} \quad \nabla g = (q/\hbar)\mathcal{A},$$

and where ψ_0 is defined to be the solution to the Schrödinger equation for $\mathcal{A} = 0$. As a result, any charged particle that moves around a closed loop γ in some vector potential A, picks up a phase factor equal to

$$\theta = e^{\frac{i}{\hbar}q\oint_\gamma \mathcal{A}} = e^{\frac{i}{\hbar}q\int_D B},$$

where we used Stokes's theorem and the assumption that γ ran around the disk D with $\partial D = \gamma$ in the right direction. This allows us to talk about a phase-difference without explicitly demanding the existence of a magnetic potential. This is also closely related to the Berry-phase; for a more detailed discussion on the topic, we refer to [6]. Now consider the space $\mathbb{R}^3 \setminus \{0\}$; that is, just ordinary space with the origin taken out. Consider the magnetic field

$$B = \frac{m}{4\pi} \sin(\phi) d\theta \wedge d\phi,$$

for some $m \in \mathbb{R}$. Observe that it is not properly defined at the origin, but that doesn't matter, since we took that point out anyway. It can be shown that B satisfies all the properties that a magnetic field is expected to have, but the important difference is that there exists no magnetic potential A such that $B = dA$. This is a powerful property, for it allows

$$\int_{S^2} B = m$$

to be non-zero, whereas Stokes's theorem would have given us that for any $B = dA$, we would have had that

$$\int_{S^2} B = \int_{S^2} dA = \int_{\partial S^2} A = 0, \quad \text{since} \quad \partial S^2 = \varnothing.$$

This is the familiar result of classical electrodynamics and it is a direct consequence of the assumption that *magnetic monopoles* do not exist. The fact that it does not hold in our case, is understood by simply observing that the magnetic field defined above is just that; a magnetic monopole. The interesting physics arises when one considers a charged partical moving around the origin in a circle, say, in the counter-clockwise direction. In the presence of a magnetic potential the resulting phase factor was most clearly uniquely defined, but now some uncertainty arises, for the disk D over which we will integrate, may be chosen in many different ways. For example, it can be checked that setting D as the northern hemisphere, will yield

$$\theta = e^{\frac{i}{\hbar} qm/2},$$

whereas the southern hemisphere will have the opposite orientation, so that

$$\theta = e^{-\frac{i}{\hbar} qm/2}.$$

Now should we force these two answers to be the same, it immediately follows that $qm/\hbar$ must be an integer multiple of 2π. In other words,

$$\int_{S^2} F = \int_{S^2} B = m = \frac{kh}{q},$$

for some integer k. Thus any monopole F and any charge q should satisfy this relationship. Should we work in natural units, setting $q = 1$ and $\hbar = 1$, we find that

$$\int_{S_2} F = 2\pi k$$

for some integer k. This shows that $\int F$ is quantized in a sense; it can only take on particular values. Therefore curvatures could be classified by their associated integers k. In order to extend F to a curvature on the full of $\mathbb{R}^3$, one can only simply continue continuously to the origin when the associated vector bundle is trivial. For non-trivial vector bundles however, only some clever pasting can in fact extend F to $\mathbb{R}^3$. It turns out that this yields a classification of vector bundles, each group with their own value of k. This is a very powerful tool, since it implies the existence of an invariant for all curvatures of a given bundle, which can be used to prove some very fundamental results.

We will make the above *physical proof* of the quantization of $\int_{S^2} F$ more precise, and we will proceed in a similiar fashion to the proof of theorem 3.14. To that end, fix some $U(1)$-bundle over S^2, let A and A' be two different vector potentials that live on S^2, and F and F' their respective curvatures. We will show that $F - F'$ is an exact 2-form, so that again the claim will follow by Stokes's theorem.

As we saw before, we are working with ordinary differential forms now. We also see that since $\mathrm{End}(E_x) = \mathbb{C}$ for all $x \in S^2$, their definitions tell us that two exterior covariant derivatives differ by a term $A \wedge \eta - (-1)^p \eta \wedge A$. Standard calculus tells us that this is identically zero, thus we have that $d_{\mathcal{D}} = d_{\mathcal{D}^0} = d$, where now d is the ordinary exterior derivative of differential forms.

Again write $\delta A = A' - A$, and let F_t be the curvature associated with $A_t = A + t\delta A$, so that $F_0 = F$ and $F_1 = F'$. From the proof of theorem 3.14 we have that

$$\frac{d}{dt} F_t = d_{\mathcal{D}_{A_t}} \delta A = d\delta A.$$

Therefore we find by the fundamental theorem of calculus that

$$F' - F = \int_0^1 \frac{d}{dt} F_t \; dt = \int_0^1 d\delta A \; dt = d\left(\int_0^1 \delta A \; dt\right) = d\delta A,$$

which proves the claim.

We have therefore another topological invariant, but this time one defined on S^2. Entirely analogously to before one can relate the winding number, now just the familiar winding number of the fundamental group of the circle, to the integral over F. It would appear that in this case the prefactor in front of the integral is not $1/8\pi^2$, but instead $1/2\pi$. Indeed this turns out to be the case.

3.5 Chern-Simons forms

We have seen two different instances of particular expressions involving the curvature F, which, integrated over the whole manifold M, provided a topological invariant of the vector bundle E; that is, a quantity that does not depend on the connection chosen. For a $U(1)$-bundle over a 2-dimensional manifold we found that F was such an expression, but more generally it turns out that $\mathrm{tr}(F)$ is the desired invariant. This is in accordance with our result for the 4-dimensional manifolds, where we found $\mathrm{tr}(F \wedge F)$. One might suspect this to extend to higher dimensional theories; perhaps $\mathrm{tr}(F \wedge F \wedge \ldots \wedge F) = \mathrm{tr}(F^k)$

also describes an invariant. This turns out to be the case, and the proof is nothing but a mild generalization of the proof of the invariance in the 4-dimensional case. Now it follows that

$$\frac{d}{dt}\mathrm{tr}(F_t^k) = k \cdot d\,\mathrm{tr}(\delta A \wedge F_t^{k-1}),$$

so that

$$\mathrm{tr}(F'^k) - \mathrm{tr}(F^k) = d\left(\int_0^1 k\mathrm{tr}(\delta A \wedge F_t^{k-1})\ dt\right),$$

which is again exact. For their special properties, these forms have received a special name; the *Chern forms*. As it turns out, these forms are also closed. To see this, observe that from the Bianchi-identity we have that $d_{\mathcal{D}}F = 0$, so that by induction it follows that

$$d_{\mathcal{D}}F^k = d_{\mathcal{D}}F \wedge F^{k-1} + F \wedge d_{\mathcal{D}}F^{k-1} = 0,$$

for all natural numbers k. Therefore

$$d\mathrm{tr}\left(F^k\right) = \mathrm{tr}\left(d_{\mathcal{D}}F^k\right) = 0,$$

which proves the claim. We conclude that any Chern form defines a cohomology class, called the *Chern class* of that dimension. On trivial bundles, we can prove that these Chern-classes are in fact trivial themselves. To that end, assume that the vector bundle admits a flat connection D^0, so that we may drop the F^0, and write $F = d_{\mathcal{D}^0}A + A \wedge A$.

Example 3.16. In the simplest of cases, we find that

$$d\mathrm{tr}(A) = \mathrm{tr}(d_{\mathcal{D}^0}A) = \mathrm{tr}(d_{\mathcal{D}^0}A + A \wedge A) = \mathrm{tr}(F),$$

where we use that lemma 3.10 tells us that $\mathrm{tr}(A \wedge A) = 0$. Thus indeed, $\mathrm{tr}(F)$ is exact now. △

Example 3.17. A less trivial case, is

$$\begin{aligned}
&d\mathrm{tr}\left(A \wedge d_{\mathcal{D}^0}A + \frac{2}{3}A \wedge A \wedge A\right) \\
&\quad = \mathrm{tr}\left(d_{\mathcal{D}^0}\left(A \wedge d_{\mathcal{D}^0}A + \frac{2}{3}A \wedge A \wedge A\right)\right) \\
&\quad = \mathrm{tr}\Big(d_{\mathcal{D}^0}A \wedge d_{\mathcal{D}^0}A + A \wedge d_{\mathcal{D}^0}^2 A + \frac{2}{3}\big(d_{\mathcal{D}^0}(A \wedge A) \wedge A + (A \wedge A) \wedge d_{\mathcal{D}^0}A\big)\Big) \\
&\quad = \mathrm{tr}\Big(d_{\mathcal{D}^0}A \wedge d_{\mathcal{D}^0}A + \frac{2}{3}\big((d_{\mathcal{D}^0}A \wedge A - A \wedge d_{\mathcal{D}^0}A) \wedge A + A \wedge A \wedge d_{\mathcal{D}^0}A\big)\Big) \\
&\quad = \mathrm{tr}\big(d_{\mathcal{D}^0}A \wedge d_{\mathcal{D}^0}A + 2A \wedge A \wedge d_{\mathcal{D}^0}A\big),
\end{aligned}$$

where we made repeated use of lemma 2.16 and of

$$\mathrm{tr}(d_{\mathcal{D}^0}A \wedge A \wedge A) = -\mathrm{tr}(A \wedge d_{\mathcal{D}^0}A \wedge A) = \mathrm{tr}(A \wedge A \wedge d_{\mathcal{D}^0}A),$$

as can be seen from proposition 3.10. Now the same proposition gives us that $\mathrm{tr}(A \wedge A \wedge A \wedge A) = 0$. Therefore

$$\begin{aligned} d\mathrm{tr}\left(A \wedge d_{\mathcal{D}^0}A + \frac{2}{3}A \wedge A \wedge A\right) &= \mathrm{tr}\big(d_{\mathcal{D}^0}A \wedge d_{\mathcal{D}^0}A + 2A \wedge A \wedge d_{\mathcal{D}^0}A + A \wedge A \wedge A \wedge A\big) \\ &= \mathrm{tr}\big((d_{\mathcal{D}^0}A + A \wedge A) \wedge (d_{\mathcal{D}^0}A + A \wedge A)\big) \\ &= \mathrm{tr}(F \wedge F), \end{aligned}$$

as desired. △

The fact that these special forms exist, called *Chern-Simons* forms, might come across as some magic coincidence, but it is far from so. Namely, the generalization of the proof of theorem 3.14 actually gives us a general way to compute these forms. Namely, since $A = 0$ describes a connection, we may take $A_t = tA$ for some vector potential A, from which it follows that $F_t = td_{\mathcal{D}^0}A + t^2 A \wedge A$.

Example 3.18. We may compute that

$$\begin{aligned} F_t \wedge F_t &= (td_{\mathcal{D}^0}A + t^2 A \wedge A) \wedge (td_{\mathcal{D}^0}A + t^2 A \wedge A) \\ &= t^2 d_{\mathcal{D}^0}A \wedge d_{\mathcal{D}^0}A + t^3(d_{\mathcal{D}^0}A \wedge A \wedge A + A \wedge A \wedge d_{\mathcal{D}^0}A) + t^4 A \wedge A \wedge A \wedge A. \end{aligned}$$

Then applying proposition 3.10 we find that

$$\begin{aligned} \mathrm{tr}(A \wedge F_t \wedge F_t) &= \mathrm{tr}\left(t^2 A \wedge (d_{\mathcal{D}^0}A)^2 + t^3(A \wedge d_{\mathcal{D}^0}A \wedge A^2 + A^3 \wedge d_{\mathcal{D}^0}A) + t^4 A^5\right) \\ &= \mathrm{tr}\left(t^2 A \wedge (d_{\mathcal{D}^0}A)^2 + 2t^3 A^3 \wedge d_{\mathcal{D}^0}A + t^4 A^5\right). \end{aligned}$$

Lastly, evaluating the integral yields

$$\mathrm{tr}\left(F \wedge F \wedge F\right) = d\left(\int_0^1 3\mathrm{tr}(A \wedge F_t \wedge F_t)\ dt\right) = d\left(A \wedge (d_{\mathcal{D}^0}A)^2 + \frac{3}{2}A^3 \wedge d_{\mathcal{D}^0}A + \frac{3}{5}A^5\right).$$

This is the third Chern-Simons form. △

Now one could naively think that in general it holds that

$$\mathrm{tr}\left(A \wedge F_t^{k-1}\right) = \sum_{n=0}^{k-1} \binom{k-1}{n} t^{2k-2-n}\mathrm{tr}\left((d_{\mathcal{D}^0}A)^n \wedge A^{2k-1-2n}\right),$$

so that after evaluating the integral

$$\mathrm{tr}\left(F^k\right) = d\left(\sum_{n=0}^{k-1} \binom{k-1}{n} \frac{k}{2k-1-n} \mathrm{tr}\left((d_{\mathcal{D}^0}A)^n \wedge A^{2k-1-2n}\right)\right).$$

This is **incorrect**. The reason for this will become apparent from writing out

$$\begin{aligned} \mathrm{tr}\left(A \wedge F_t^3\right) = \mathrm{tr}\big(&t^3 A \wedge (d_{\mathcal{D}^0}A)^3 + t^4[2A^3 \wedge (d_{\mathcal{D}^0}A)^2 + A \wedge d_{\mathcal{D}^0}A \wedge A^2 \wedge d_{\mathcal{D}^0}A] \\ &+ 3t^5 A^5 \wedge d_{\mathcal{D}^0}A + t^6 A^7\big). \end{aligned}$$

By no means proposition 3.10 ensures us that $\mathrm{tr}(A^3 \wedge (d_{\mathcal{D}^0} A)^2) = \mathrm{tr}(A \wedge d_{\mathcal{D}^0} A \wedge A^2 \wedge d_{\mathcal{D}^0} A)$, and this is generally not true at all. A last attempt could be to check whether or not these terms differ by some closed form, so that in calculating their exterior derivatives, their distinction would vanish, but as a direct computation shows, this is unfortunately not the case. After evaluating the integral we conclude that the fourth Chern-Simons form is given by

$$A \wedge (d_{\mathcal{D}^0} A)^3 + \frac{8}{5} A^3 \wedge (d_{\mathcal{D}^0} A)^2 + \frac{4}{5} A \wedge d_{\mathcal{D}^0} A \wedge A^2 \wedge d_{\mathcal{D}^0} A + 2A^5 \wedge d_{\mathcal{D}^0} A + \frac{4}{7} A^7.$$

It is possible to calculate the k-th Chern-Simons form for any k, but as the above example shows, in general their expressions will become increasingly messy. We therefore opt to conclude our short discussion on the Chern-Simons forms here, now that the expressions still fit in a single line, and to continue to the main topic of this thesis.

3.6 Instantons

It will be shown that the instantons are solutions to the equations of motion that globally extremize the Yang Mills action, and which are therefore of particular interest to study. Having established that, we will in the next chapter be able to give a complete description of the ADHM construction, which, given certain conditions, constructs a vector bundle and an associated instanton. This will involve many of the concepts previously discussed and shows their use in quite a spectacular fashion.

Definition 3.19. A curvature 2-form F that satisfies the equation $F = \star F$ is called *self-dual*, and one that satisfies $F = - \star F$ is called *anti-self-dual*. In a similar fashion, a connection D is called (anti-)self-dual precisely when its induced curvature 2-form is (anti-)self-dual. Vector potentials A that produce anti-self-dual curvatures will henceforth be referred to as *instantons*. The negative topological charge, or Pontryagin index, associated to an instanton is referred to as the *instanton number*.

Example 3.20. Using example 2.22, we can for a general curvature 2-form F on a flat space M write out the anti-self-duality equations. They are

$$F_{12} + F_{34} = 0, \quad F_{13} + F_{42} = 0, \quad \text{and} \quad F_{14} + F_{23} = 0.$$

Expanded just a tad further, we can write these in terms of the connection D as

$$[D_1, D_2] + [D_3, D_4] = 0, \quad [D_1, D_3] + [D_4, D_2] = 0, \quad \text{and} \quad [D_1, D_4] + [D_2, D_3] = 0.$$

This will prove to be useful in the justification of the ADHM-construction. △

A final result we will need in order to prove our main theorem about instantons, will be the existence of a particularly useful inner product.

Lemma 3.21. *We can define an inner product $\langle \alpha, \beta \rangle : M \to \mathbb{R}$ on End(E)-valued forms, by demanding that*

$$\mathrm{tr}(\alpha \wedge \star\beta) = \langle \alpha, \beta \rangle \mathrm{vol},$$

as an equality of differential n-forms.

It is worth remarking that the inner product defined above makes sense, since the space of differential n-forms on an n-dimensional manifold is 1-dimensional, which implies that all such forms are multiples of the volume form.

Remark 3.22. Note that we may rewrite the Yang-Mills action as

$$S = \int_M |F|^2 \text{ vol.}$$

Theorem 3.23. *Suppose that M is a flat 4-dimensional Riemannian manifold; that is, a manifold equipped with a metric that is positive definite. Then a curvature 2-form F is an* absolute minimum *of the Yang-Mills action precisely when F is self-dual or anti-self-dual.*

Proof. We first note that from remark 2.23 we have that $\star \circ \star : \Omega^2(M) \to \Omega^2(M)$ equals the identity. Namely, the permutations $\{1,2,3,4\} \to \{i_1, i_2, i_3, i_4\}$ and $\{1,2,3,4\} \to \{i_3, i_4, i_1, i_2\}$ differ by precisely two switches, so that they have the same sign. Therefore all possible eigenvalues of $\star$ are either 1 or -1, and $\star$ is diagonalizable in the sense that there exists a basis of eigenforms that spans the full space of $\text{End}(E)$-valued 2-forms. Note that from example 3.20, these eigenspaces are explicitly given by

$$\begin{cases} E_{-1} = \text{span}\{dx^1 \wedge dx^2 - dx^3 \wedge dx^4,\ dx^1 \wedge dx^3 - dx^4 \wedge dx^2,\ dx^1 \wedge dx^4 - dx^2 \wedge dx^3\}; \\ E_{+1} = \text{span}\{dx^1 \wedge dx^2 + dx^3 \wedge dx^4,\ dx^1 \wedge dx^3 + dx^4 \wedge dx^2,\ dx^1 \wedge dx^4 + dx^2 \wedge dx^3\}. \end{cases}$$

It can be verified through direct computation that for $\alpha \in E_{-1}$ and $\beta \in E_{+1}$, it holds that

$$\alpha \wedge \beta = 0, \quad \text{from which it follows that also} \quad \langle \alpha, \beta \rangle = 0.$$

Now we can define

$$F_+ = \frac{F + \star F}{2} \quad \text{and} \quad F_- = \frac{F - \star F}{2}, \quad \text{so that} \quad \star F_+ = F_+ \quad \text{and} \quad \star F_- = -F_-.$$

Since $F = F_+ + F_-$, orthogonality allows us to write

$$|F|^2 = |F_+|^2 + |F_-|^2,$$

so that

$$S = \int_M |F_+|^2 \text{ vol} + \int_M |F_-|^2 \text{ vol.}$$

Now using $F_+ \wedge F_- = F_- \wedge F_+ = 0$, we compute that

$$\begin{aligned} 8\pi^2 \text{n} &= \int_M \text{tr}(F \wedge F) \\ &= \int_M \text{tr}\left((F_+ + F_-) \wedge (F_+ + F_-)\right) \\ &= \int_M \text{tr}(F_+ \wedge \star F_+) - \int_M \text{tr}(F_- \wedge \star F_-) \\ &= \int_M |F_+|^2 - \int_M |F_-|^2. \end{aligned}$$

Combining this with the above expression for the action, we find

$$S \geq 8\pi^2 |\mathrm{n}|.$$

Since n is a fixed number given the vector bundle, independent of F, this gives us a lower bound for the action. It is now easy to see that equality holds precisely if either $F_- = 0$ or $F_+ = 0$. In other words, when F is either self-dual or anti-self-dual. This completes the proof. $\square$

This is an extremely important result. We stress that the criterion that M is Riemannian is crucial to the validity of the statement; for instance, on 4-dimensional Minkowski space, we have that $\star \circ \star = -\mathrm{id}$ on $\Omega^2(M)$, so that any (anti-)self-dual curvature satisfies $F = 0$; that is, there are no non-trivial instantons in that case.

It turns out that solving the (anti-)self-duality equations is a lot easier than solving the Yang-Mills equation. Restricting our focus to instatons will indeed not provide us with the full picture that is hidden in the local extrema of the Yang-Mills action, but it will be an interesting and accessible subclass to study. The fact that they are so well-behaved and easy to work with, has resulted in a great many studies on these particular set of curvatures alone, and it has provided physicists with intruiging new insights.

It should be clear that we are interested in a description of the set of all instantons in the case that non-trivial ones do exist. This is precisely what the ADHM-construction, which will be introduced momentarily, will provide us with.

4 The ADHM-construction

With most of the general preparations out of the way, we will for a while restrict our view solely to trying to understand the ADHM-construction and to see why it works the way it does. Most roughly, the ADHM-construction attemps to provide us with a complete description of all instantons on Euclidean space. Somewhat more precisely, it will, given a number of objects satisfying certain relations, construct a vector bundle over Euclidean space and an anti-self-dual connection thereon.

For the construction of the vector bundle and to make the process of understanding the construction a tad smoother, we will first introduce a few specific ideas that will be at the core of the construction. With these vital ingredients in hand, we will outline and motivate the premisses under which the ADHM-construction operates, before we finally move on to a complete description of the construction, accompanied by a proof of its validity. That section will be particularly computation heavy, but almost all will be straightforward in the language of linear algebra.

The main source of information used for the writing of this chapter was chapter 4 from [2].

4.1 Complexes, quaternions and the ADHM-data

We will start off this section with an applied definition of a very general concept in mathematics.

Definition 4.1. A collection $(V_1, V_2, V_3, \alpha, \beta)$ is called a *short complex* if V_1, V_2 and V_3 are complex vector spaces and $\alpha : V_1 \to V_2$ and $\beta : V_2 \to V_3$ are linear maps, so that $\mathrm{im}(\alpha) \subseteq \ker(\beta)$, or equivalently, if $\beta \circ \alpha = 0$.

At first sight, this may remind the reader of an exact sequence of modules, but we emphasize that the equality $\mathrm{im}(\alpha) = \ker(\beta)$ need *not* generally hold for a short complex. In fact, much of the interesting mathematics involving complexes is partially, if not fully, based on this distinction.

Example 4.2. For a given manifold M and some natural number p, we can consider the exterior derivative d, mapping

$$\Omega^{p-1}(M) \xrightarrow{d} \Omega^p(M) \xrightarrow{d} \Omega^{p+1}(M),$$

under the well-known identity that $d \circ d = 0$. An interesting space to study then would be the quotient space $\ker(d|_{\Omega^p(M)})/\mathrm{im}(d|_{\Omega^{p-1}(M)})$ which is in general non-trivial; indeed, this has been considered before, and these spaces nowadays go by the name of de Rham cohomology groups. The Chern classes we encountered in the previous chapter represent elements of these groups. △

Definition 4.3. Given a matrix $A = (a_{ij}) : V \to W$ for some complex inner product spaces V and W with the standard inner product, define the *hermitian transpose* of A, denoted by $A^\dagger$, by $A^\dagger = (b_{ij})$ where $b_{ij} = \overline{a_{ji}}$ for all i, j. That is, $A^\dagger$ is the complex conjugate of the transpose of A. For all $v \in V$, $w \in W$, we have that

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle, \quad \text{so that} \quad (\text{im } A)^\perp = \ker(A^\dagger).$$

Definition 4.4. Let V_1, V_2 and V_3 be complex, finite-dimensional vector spaces and consider two bundle maps $M \times V_1 \to M \times V_2$ and $M \times V_2 \to M \times V_3$. We can express these as $\alpha : M \to \text{Hom}(V_1, V_2)$ and $\beta : M \to \text{Hom}(V_2, V_3)$, and we will assume that α_x is injective and β_x is surjective for all $x \in M$. Furthermore, suppose that for any $x \in M$, we have that α_x and β_x form a short complex. We can then define a *vector bundle* E over M with fibres $E_x = \ker(\beta_x)/\text{im}(\alpha_x)$. Observe that the fibers of E are all of dimension $\dim(V_2) - [\dim(V_1) + \dim(V_3)]$. If in addition all vector spaces V_1, V_2 and V_3 allow Hermitian metrics, we may also write $E = (\text{im } \alpha_x)^\perp \cap \ker(\beta_x) = \ker(\alpha_x^\dagger) \cap \ker(\beta_x)$.

Since we will need the following result later on, we will state it here alongside the necessary definitions. It is by no means a deep theorem and can again be verified through direct computation.

Lemma 4.5. *For any $x \in M$, the orthogonal projection $P_x : V_2 \to E_x \subset V_2$ is explicitly given by*

$$P_x = \text{id}_{V_2} - \beta_x^\dagger(\beta_x\beta_x^\dagger)^{-1}\beta_x + \alpha_x(\alpha_x^\dagger\alpha_x)^{-1}\alpha_x^\dagger.$$

Proof. The proof can be found in [2]. □

Lastly, we note that there is a natural choice of a connection on E, that is indirectly defined by stating the effect of its E-valued derivative on sections $s : M \to E$.

Definition 4.6. We define the *induced connection* D on E by demanding that its E-valued derivative d_D satisfies

$$d_D s = \tilde{P} \circ d_0 s,$$

for all sections $s : M \to E$ as an equality of E-valued 1-forms, where d_0 denotes the E-valued derivative of the standard flat connection on V_2, say $\boldsymbol{D}$. Furthermore, $\tilde{P}_x$ acts identically to P_x, but maps to E_x as a space not considered as a subset of V_2.

Our new operator $\tilde{P}$ may seem like an unnecessary complication, but we will see later that it will be most convenient to consider $\tilde{P}$ instead of P, for it allows us to work with E and V_2 using different bases.

Before we give the set of data the ADHM-construction sets out with, we will motivate the expressions to a minute extent, by rewriting the anti-self-duality equations in yet another way. Suppose we are given a 4-dimensional, flat manifold M with a vector bundle

E over M. We can then introduce *complex coordinates* $z_1 = x^2 + ix^1$ and $z_2 = x^4 + ix^3$. One could try to introduce some sort of complex connections, via

$$\mathcal{D}_1 = iD_1 + D_2, \quad \text{such that} \quad \mathcal{D}_1^\dagger = iD_1 - D_2,$$

and also

$$\mathcal{D}_2 = iD_3 + D_4, \quad \text{such that} \quad \mathcal{D}_2^\dagger = iD_3 - D_4.$$

Note the peculiar way we defined the $\dagger$ on $\mathcal{D}_i$; contrary to ordinary conjugation, we add another minus sign. The reason for this will not be discussed here, since it makes no difference for the following two expressions, which can by direct computation be shown to be equivalent to the anti-self-duality equations:

$$[\mathcal{D}_1, \mathcal{D}_2] = 0 \quad \text{and} \quad [\mathcal{D}_1, \mathcal{D}_1^\dagger] + [\mathcal{D}_2, \mathcal{D}_2^\dagger] = 0.$$

Equations of this form will appear momentarily as well, albeit accompanied by some additional terms. The idea of the ADHM-construction is that the data defines a well-disguised complex, which will allow us to define a vector bundle E over the 4-dimensional manifold U. Futhermore, the data is chosen in such a way that it will allow us to easily contruct the projection map P as considered in the preceding section, which will allow us to define a vector potential A that, incredibly enough, will be anti-self-dual. Most remarkable of this construction is that most of it is done exclusively using the language of linear algebra.

The road that will take us to this point, which we have spent quite some time paving, will be quite long. Yet surprisingly, most of the results will still follow just by direct computation. The genius of the construction is not so much in the proofs of its correctness, but instead in the complicated data we will start working with and the operators that we will derive from it. A more detailed justification of the data the construction begins with will be given in chapter 6.

First we will have to introduce the space in which the ADHM-construction takes place, namely that of the quaternions.

Definition 4.7. The *algebra of quaternions* is defined as the set $\mathbb{H} = \{x^1 i + x^2 j + x^3 k + x^4\}$ where x^μ is a real number for all $\mu \in \{1, 2, 3, 4\}$, and i, j and k satisfy the relations $i^2 = j^2 = k^2 = ijk = -1$. Conjugation is defined as

$$\overline{x} = -x^1 i - x^2 j - x^3 k + x^4, \quad \text{and observe that} \quad x\overline{x} = \textstyle\sum_\mu (x^\mu)^2 \in \mathbb{R}_{\geq 0},$$

such that $x^{-1} = \overline{x}/(x\overline{x})$ is well-defined for all non-zero $x \in \mathbb{H}$. We set $z_1 = x^2 + ix^1$ and $z_2 = x^4 + ix^3$, which will again be referred to as complex coordinates. We denote $x = (z_1, z_2)$, and it is often convenient to write $(i, j, k, 1) = (\tau_1, \tau_2, \tau_3, \tau_4)$, so that we may also write $x = x^\mu \tau_\mu$.

Lemma 4.8. *The map given by*

$$\tau_1 \to \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \; \tau_2 \to \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \; \tau_3 \to \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad \textit{and} \quad \tau_4 \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

preserves multiplication in $\mathbb{H}$*. Therefore any quaternion can be viewed as a* 2×2*-matrix, via*

$$x = \tau_\mu x^\mu \equiv \begin{pmatrix} -ix^3 + x^4 & -ix^1 - x^2 \\ -ix^1 + x^2 & ix^3 + x^4 \end{pmatrix} = \begin{pmatrix} \overline{z_2} & -z_1 \\ \overline{z_1} & z_2 \end{pmatrix}.$$

Observe that we have $x^\dagger = \overline{x}$*.*

These two ways of looking at quaternions will be used interchangeably in the forthcoming.

Definition 4.9. Consider $\mathbb{H}$ with complex coordinates z_1 and z_2. Then an *ADHM-system* consists of complex vector spaces V and W with respective dimensions k and n, and two $k \times k$ matrices B_1 and B_2, a $k \times n$ matrix I and a $n \times k$ matrix J, all with complex entries. A tuple (V, W, B_1, B_2, I, J) is referred to as *ADHM-data* over $\mathbb{H}$ if in addition to the above, the following two properties hold:

1. The *ADHM-equations* are satisfied:

$$[B_1, B_2] + IJ = 0,$$
$$[B_1, B_1^\dagger] + [B_2, B_2^\dagger] + II^\dagger - J^\dagger J = 0.$$

2. For all $(x, y) \in \mathbb{H}^2 \setminus \{(0,0)\}$, when we write $x = (z_1, z_2)$ and $y = (w_1, w_2)$, the map defined by

$$\alpha_{(x,y)} : V \to W \oplus V \oplus V : v \mapsto \begin{pmatrix} w_2 J - w_1 I^\dagger \\ -w_2 B_1 - w_1 B_2^\dagger - z_1 \mathrm{id}_k \\ w_2 B_2 - w_1 B_1^\dagger + z_2 \mathrm{id}_k \end{pmatrix} v$$

is injective. Also, $\beta_{(x,y)} : W \oplus V \oplus V \to V$ given by

$$\beta_{(x,y)} : \begin{pmatrix} u \\ v_1 \\ v_2 \end{pmatrix} \mapsto \begin{pmatrix} w_2 I + w_1 J^\dagger & w_2 B_2 - w_1 B_1^\dagger + z_2 \mathrm{id}_k & w_2 B_1 + w_1 B_2^\dagger + z_1 \mathrm{id}_k \end{pmatrix} \begin{pmatrix} u \\ v_1 \\ v_2 \end{pmatrix}$$

must be surjective. Note that these two constraints can be more concisely described by demanding that

$$R_{(x,y)} : W \oplus V \oplus V \to V \times V : \begin{pmatrix} u \\ v_1 \\ v_2 \end{pmatrix} \mapsto \begin{pmatrix} \beta_{(x,y)} \\ \alpha^\dagger_{(x,y)} \end{pmatrix} \begin{pmatrix} u \\ v_1 \\ v_2 \end{pmatrix}$$

is surjective, where we use the fact that for any matrix O we have that O is injective precisely if $O^\dagger$ is surjective.

Remark 4.10. We remark that the ADHM-equations do in some sense resemble the expressions we derived for the anti-self-duality equations for the curvature, substituting $\mathcal{D}_1 \to B_1$ and $\mathcal{D}_2 \to B_2$, bar the terms involving I and J. Still, this should reassure the reader that the above definition is not completely arbitrary. It may also appear at first sight that the first and second constraints are disjoint in nature and have nothing to do with one another. However, as we will see in the next lemma, this could not be further from the truth. Also, in chapter 6 we will see that both ADHM-equations can be derived from similar concepts, namely that of a *moment map*. More on that later.

Lemma 4.11. *For any $(x, y) \in \mathbb{H}^2 \setminus \{(0,0)\}$, we have that $\alpha_{(x,y)} : V \to W \oplus V \oplus V$ and $\beta_{(x,y)} : W \oplus V \oplus V \to V$ from a complex; that is, $\beta_{(x,y)} \circ \alpha_{(x,y)} = 0$. Moreover, they form a complex for all (x, y) precisely when the ADHM-equations hold.*

Proof. A straightforward and tedious computation shows that

$$\begin{aligned}(\beta_{(x.y)} \circ \alpha_{(x,y)})(v) = &\left(w_2^2([B_1, B_2] + IJ) - w_1^2([B_1, B_2] + IJ)^\dagger\right) v \\ &- w_1 w_2 \left([B_1, B_1^\dagger] + [B_2, B_2^\dagger] + II^\dagger - J^\dagger J\right) v.\end{aligned}$$

For this to vanish for all non-zero choices of (x, y), the ADHM-equations must hold, as desired. $\square$

This shows that we can define a vector bundle E over $\mathbb{H}^2 \setminus \{(0,0)\}$ satisfying $E_{(x,y)} = \ker(\alpha^\dagger_{(x,y)}) \cap \ker(\beta_{(x,y)}) = \ker(R_{(x,y)})$. For the dimension of E, observe that $\ker(\beta_{(x,y)})$ has dimension $(n+k+k) - k = n+k$, since $\beta_{(x,y)}$ is assumed to be surjective. Similarly, the image of $\alpha_{(x,y)}$ has dimension k, since $\alpha_{(x,y)}$ is assumed to be injective. Therefore E has dimension $(n + k) - k = n$. Now that we have our vector bundle, we can finally work through the ADHM construction in its full glory. Without further interruptions, we will next discuss the process of the contruction in detail.

4.2 The ADHM-construction

Much of the work we will have to do before reaching our anti-self-dual connection, will involve the following operators.

Definition 4.12. Define the operator $\Delta_{(x,y)} : V \times V \to W \oplus V \oplus V$ by

$$\Delta_{(x,y)} = R^\dagger_{(x,y)} = \begin{pmatrix} \beta^\dagger_{(x,y)} & \alpha_{(x,y)} \end{pmatrix} = \begin{pmatrix} \overline{w_2} I^\dagger + \overline{w_1} J & w_2 J - w_1 I^\dagger \\ \overline{w_2} B_2^\dagger - \overline{w_1} B_1 + \overline{z_2}\mathrm{id}_k & -w_2 B_1 - w_1 B_2^\dagger - z_1 \mathrm{id}_k \\ \overline{w_2} B_1^\dagger + \overline{w_1} B_2 + \overline{z_1}\mathrm{id}_k & w_2 B_2 - w_1 B_1^\dagger + z_2 \mathrm{id}_k \end{pmatrix}.$$

Futhermore, if we define

$$a = \begin{pmatrix} I^\dagger & J \\ B_2^\dagger & -B_1 \\ B_1^\dagger & B_2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 & 0 \\ \mathrm{id}_k & 0 \\ 0 & \mathrm{id}_k \end{pmatrix},$$

then by direct computation one can verify that

$$\Delta_{(x,y)} = a(y \otimes \mathrm{id}_k) + b(x \otimes \mathrm{id}_k).$$

It should be noted that the vector bundle E over $\mathbb{H}^2\backslash\{(0,0)\}$ is not yet quite the vector bundle over S^4 we are looking for. This will be established by the following lemma.

Lemma 4.13. *For any non-zero quaternion q, we have that $E_{(x,y)} = E_{(xq,yq)}$.*

Proof. The proof is actually quite easy; first observe that for any matrices O_1 and O_2, we have that

$$(O_1O_2 \otimes \mathrm{id}_k)^\dagger = O_2^\dagger O_1^\dagger \otimes \mathrm{id}_k = (O_2^\dagger \otimes \mathrm{id}_k)(O_1^\dagger \otimes \mathrm{id}_k).$$

Now we calculate that

$$\begin{aligned}
R_{(xq,yq)} &= \Delta^\dagger_{(xq,yq)} \\
&= [a(yq \otimes \mathrm{id}_k) + b(xq \otimes \mathrm{id}_k)]^\dagger \\
&= (yq \otimes \mathrm{id}_k)^\dagger a^\dagger + (xq \otimes \mathrm{id}_k)^\dagger b^\dagger \\
&= (q^\dagger \otimes \mathrm{id}_k)[(y^\dagger \otimes \mathrm{id}_k)a^\dagger + (x^\dagger \otimes \mathrm{id}_k)b^\dagger] \\
&= (q^\dagger \otimes \mathrm{id}_k)[a(y \otimes \mathrm{id}_k) + b(x \otimes \mathrm{id}_k)]^\dagger \\
&= (q^\dagger \otimes \mathrm{id}_k)\Delta^\dagger_{(x,y)} \\
&= (q^\dagger \otimes \mathrm{id}_k)R_{(x,y)}.
\end{aligned}$$

Now since $q \neq 0$, it is invertible, so is $q^\dagger \otimes \mathrm{id}_k$. Thus we find that $E_{(x,y)} = \ker(R_{x,y}) = \ker(R_{(xq,yq)}) = E_{(xq,yq)}$, as desired. $\square$

This shows that we may freely scale our coordinates by some non-zero quaternion without the fiber of the vector bundle changing. Therefore, we have constructed a bundle E over

$$\big(\mathbb{H}^2 \setminus \{(0,0)\}\big)\Big/\big(\mathbb{H} \setminus \{(0,0)\}\big) = \mathrm{P}(\mathbb{H}) \cong S^4,$$

where $\mathrm{P}(\mathbb{H})$ denotes the projective space. Setting $y = 1 \equiv (0,1) \equiv \mathrm{id}_2$ has the effect so as to simplify our expressions to

$$\Delta_x := \Delta_{(x,1)} = \begin{pmatrix} I^\dagger & J \\ B_2^\dagger + \overline{z_2}\mathrm{id}_k & -B_1 - z_1\mathrm{id}_k \\ B_1^\dagger + \overline{z_1}\mathrm{id}_k & B_2 + z_2\mathrm{id}_k \end{pmatrix} = a + b(x \otimes \mathrm{id}_k) = a + b(\tau_\mu x^\mu \otimes \mathrm{id}_k).$$

Observe that by doing this we removed the point $y = (1,0)$ from our base space, so that now what we have left is a bundle over $\mathbb{H}$. Now we will make use of the ADHM-data in a clever way one final time, to note some nice property of Δ_x.

Lemma 4.14. *For any $x \in \mathbb{H}\backslash\{0\}$, there exists some invertible, self-adjoint $k \times k$ matrix f_x such that $\Delta_x^\dagger\Delta_x = \mathrm{id}_2 \otimes f_x^{-1}$.*

Proof. We will use that

$$\Delta_x^\dagger = \begin{pmatrix} \beta_x \\ \alpha_x^\dagger \end{pmatrix}, \quad \text{so that} \quad \Delta_x^\dagger\Delta_x = \begin{pmatrix} \beta_x\beta_x^\dagger & \beta_x\alpha_x \\ \alpha_x^\dagger\beta_x^\dagger & \alpha_x^\dagger\alpha_x \end{pmatrix}.$$

Recall Lemma 4.1. Since α_x and β_x form a complex, we have that $\alpha_x^\dagger \beta_x^\dagger = (\beta_x \alpha_x)^\dagger = 0$. Therefore the off-diagonal terms of the matrix vanish. It can also be shown by mere perseverance that the diagonal terms differ by precisely the second ADHM equation, which makes them equal. Now we note that since $\alpha_x^\dagger$ and α_x both have maximum rank k, their product is of full rank too; that is, it is invertible. Therefore we will set $f_x = (\alpha_x^\dagger \alpha_x)^{-1} = (\beta_x \beta_x^\dagger)^{-1}$, so that $\Delta_x^\dagger \Delta_x$ is of the desired from, since f_x is clearly self-adjoint. □

Now we will use this observation to note a very elegant way of writing the projection of $W \oplus V \oplus V$ onto the fiber E_x, which will be one of the final stepping stones before we can define our connection.

Lemma 4.15. *Let P_x be the projection of $W \oplus V \oplus V$ onto the fiber E_x as in Lemma 4.5 and define $Q_x = \Delta_x (\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger$. Then Q_x is an orthogonal projection, and we have that $P + Q = \mathrm{id}_{n+2k}$.*

Proof. For Q_x to be an orthogonal projection, it must be both self-adjoint, $Q_x = Q^\dagger$, and also idempotent, $Q_x^2 = Q_x$. The first statement follows directly, and for the second statement we compute that

$$\begin{aligned} Q_x^2 &= \Delta_x (\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger \cdot \Delta_x (\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger \\ &= \Delta_x (\mathrm{id}_2 \otimes f_x)(\mathrm{id}_2 \otimes f_x^{-1})(\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger = \Delta_x (\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger = Q_x, \end{aligned}$$

where we used Lemma 4.14. To see the validity of the second statement, we write out Q_x in terms of its building blocks, to find

$$\begin{aligned} Q_x = \Delta_x (\mathrm{id}_2 \otimes f_x) \Delta_x^\dagger &= \begin{pmatrix} \beta_x^\dagger & \alpha_x \end{pmatrix} \begin{pmatrix} f & 0 \\ 0 & f \end{pmatrix} \begin{pmatrix} \beta_x \\ \alpha_x^\dagger \end{pmatrix} \\ &= \beta_x^\dagger f \beta_x + \alpha_x f \alpha_x^\dagger = \beta_x^\dagger (\beta_x \beta_x^\dagger)^{-1} \beta_x + \alpha_x (\alpha_x^\dagger \alpha_x)^{-1} \alpha_x^\dagger \\ &= \mathrm{id}_{n+2k} - P_x, \end{aligned}$$

where we use both the proof of Lemma 4.14 and the result of Lemma 4.5. This completes the proof. □

Let us pause for a brief moment to consider the dimensions of the operators we are working with; Δ_x is a $(n+2k) \times (2k)$ matrix, so that $\Delta_x^\dagger$ is a $(2k) \times (n+2k)$ matrix. Recall that $\Delta_x^\dagger = R_{(x,1)}$ is assumed to be surjective, so that it is a matrix of maximal rank, $2k$. This means that, should we define a matrix M_x whose columns are comprised of an orthonormal basis of the kernel of $\Delta_x^\dagger$, this would be a $(n+2k) \times n$ matrix, in full accordance with our previously acquired result that E was an n-dimensional vector bundle. By construction, M_x satisfies

$$\Delta_x^\dagger M_x = 0 \quad \text{and} \quad M_x^\dagger M_x = \mathrm{id}_n.$$

We thus have that M_x maps a vector in E expressed in the basis that its columns constitute, into $W \oplus V \oplus V$ in the standard basis of a trivial bundle. Similarly, $M_x^\dagger$ maps a vector of $W \oplus V \oplus V$ back into E.

The following statement will not be fully proved here, as it relies on the fairly involved, yet most credible, result that any hermitian matrix O_1 can be written as $O_1 = O_2 O_2^\dagger$ for some matrix O_2. Despite this, this Lemma will be of crucial importance in order to study our connection.

Lemma 4.16. *The orthogonal projection P_x is explicitly given by $P_x = M_x M_x^\dagger$.*

Proof. This is just a sketch; the full proof can be found in [2]. The key idea is to show that $1 - Q_x - M_x M_x^\dagger$ is a projection operator, which is also hermitian. Writing $1 - Q_x - M_x M_x^\dagger = O_x O_x^\dagger$ as above, we can show that O_x consists of vectors in the kernel of $\Delta_x^\dagger$, which are linearly independent from the vectors in M. Since M consists of a basis of the kernel of $\Delta_x^\dagger$, we must have $O_x = 0$, from which the claim follows. □

Remark 4.17. We remark that $M_x M_x^\dagger$ is clearly self-adjoint, and also that

$$P_x^2 = M_x (M_x^\dagger M_x) M_x^\dagger = M_x M_x^\dagger = P_x,$$

so that this operator indeed satisfies the properties of an orthogonal projection. Furthermore, $\Delta_x^\dagger P_x = (\Delta_x^\dagger M_x) M_x^\dagger = 0$, which is to be expected of a map that projects onto the kernel of $\Delta_x^\dagger = R_x$. Hopefully this will convince the reader that the above claim is not completely non-sensical, and that it is acceptable to continue without presenting the full proof.

Lemma 4.18. *Let A be the vector potential associated with the induced connection D on E. Then we have that $A = M^\dagger d_{\mathcal{D}} M$, where $d_{\mathcal{D}}$ denotes the exterior covariant derivative of the End(E)-valued 0-form $M : x \mapsto M_x \in End(E)$.*

Proof. For the proof we will work locally, and it will be a brilliant show-off of the theory we developed in the introductory chapter. The crucial observation is to note that if we write

$$M = [v_1 | v_2 | \dots | v_n],$$

we have that $\{v_i\}$ is in fact a basis of sections of E, since the orthonormal basis of E_x depends on x. Therefore we will consider any section s of the vector bundle E and we will write it as $s = s^i v_i$ for certain smooth functions s^i. Now we desire to compute

$$(d_D s)(\partial_\mu) = D_\mu(s) = D_\mu^0(s^i v_i) + A_\mu s = (\partial_\mu s^i) v_i + A_\mu s.$$

We will do this using the definition of our connection. Namely, we have

$$(d_D s)(\partial_\mu) = \tilde{P} \circ d_0(s^i v_i)(\partial_\mu) = \tilde{P} \circ \boldsymbol{D}_\mu(s^i v_i) = \tilde{P}(\partial_\mu s^i) v_i + \tilde{P} s^i \boldsymbol{D}_\mu(v_i),$$

where we used the definition of a connection. Now it is crucial to note that since all v_i map to E and $\tilde{P}$ is a projection onto E, we have that $\tilde{P}(\partial_\mu s^i) v_i = (\partial_\mu s^i) v_i$, so that we may conclude that

$$A_\mu(v_i) = \tilde{P} \boldsymbol{D}_\mu(v_i),$$

since both $\tilde{P}$ and A_μ are $C(\mathbb{H} \setminus \{0\})$-linear. Now, let $\{e_j\}$ be the standard basis of the trivial bundle $(\mathbb{H} \setminus \{0\}) \times (W \oplus V \oplus V)$, so that we may write $v_i = M_i^k e_k$ by definition of M, where the M_i^k are functions. Then we can compute that on the one hand, since $\boldsymbol{D}$ is now flat,

$$\boldsymbol{D}_\mu(v_i) = \boldsymbol{D}_\mu(M_i^k e_k) = (\partial_\mu M_i^k) e_k,$$

whereas on the other, we have by definition

$$A_\mu(v_i) = A_{\mu i}^j v_j.$$

Now observe that since $P = MM^\dagger$, we have that $\tilde{P} = M^\dagger$; namely, M puts the vectors in E back into the trivial bundle it is contained in, but we do not want $\tilde{P}$ to do that. Therefore,

$$\tilde{P}\boldsymbol{D}_\mu(v_i) = M^\dagger(\partial_\mu M_i^k) e_k = (M^\dagger)_k^j (\partial_\mu M_i^k) v_j.$$

Hence we would have that $A_{\mu i}^j = (M^\dagger)_k^j(\partial_\mu M_i^k)$, so that $A_\mu = M^\dagger \partial_\mu M$. All directions considered, we find that $A = M^\dagger d_{\mathcal{D}} M$, as desired. □

Remark 4.19. We stress that from this, together with the fact that $d_{\mathcal{D}} M = \partial_\mu M \otimes dx^\mu$, we find that the components of the vector potential can be expressed as $A_\mu = M^\dagger \partial_\mu M$. Also note that since $M^\dagger M = \mathrm{id}_n$ is a constant, we have that

$$0 = \partial_\mu(M^\dagger M) = (\partial_\mu M^\dagger) M + M^\dagger \partial_\mu M,$$

where the validity of a Leibniz-like rule for matrices can be verified by direct computation. Lastly, it is worth noting that A defined in this way, will satisfy

$$A^\dagger = (M^\dagger d_{\mathcal{D}} M)^\dagger = (d_{\mathcal{D}} M^\dagger) M = -M^\dagger d_{\mathcal{D}} M = -A,$$

by the above. In a gauge theory where $U(k)$ and $SU(n)$ act on V and W respectively, this fact that A is *skew-Hermitian* can be used to show that A is (gauge equivalent to a connection that is) in fact $\mathfrak{su}(n)$-valued, which will be important in proving statements regarding the instantons obtainable with the ADHM-construction. We will not go further into the details here.

Now we are finally ready to state and prove the main result from this construction. Namely, that the induced connection above produces a curvature 2-form F that is anti-self-dual.

Theorem 4.20 (Atiyah, Drinfeld, Hitchin and Manin)**.** *The connection $A = M^\dagger d_{\mathcal{D}} M$ defines an anti-self-dual curvature.*

Proof. We start off by calculating $F_{\mu\nu}$ in terms of the operators we introduced along the way in the ADHM-construction. In fact, this will constitute most of the work. We find

$$\begin{aligned}
F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \\
&= \partial_\mu(M^\dagger \partial_\nu M) - \partial_\nu(M^\dagger \partial_\mu M) + (M^\dagger \partial_\mu M)(M^\dagger \partial_\nu M) - (M^\dagger \partial_\nu M)(M^\dagger \partial_\mu M) \\
&= \partial_\mu(M^\dagger \partial_\nu M) - \partial_\nu(M^\dagger \partial_\mu M) - (\partial_\mu M^\dagger M)(M^\dagger \partial_\nu M) + (\partial_\nu M^\dagger M)(M^\dagger \partial_\mu M) \\
&= \partial_\mu M^\dagger (1 - MM^\dagger)\partial_\nu M - \partial_\nu M^\dagger (1 - MM^\dagger)\partial_\mu M \\
&= \partial_\mu M^\dagger (1 - P)\partial_\nu M - \partial_\nu M^\dagger (1 - P)\partial_\mu M \\
&= \partial_\mu M^\dagger Q \partial_\nu M - \partial_\nu M^\dagger Q \partial_\mu M \\
&= (\partial_\mu M^\dagger)\Delta(\mathrm{id}_2 \otimes f)\Delta^\dagger(\partial_\nu M) - (\partial_\nu M^\dagger)\Delta(\mathrm{id}_2 \otimes f)\Delta^\dagger(\partial_\mu M),
\end{aligned}$$

where the observant reader may have noticed that in moving around some brackets, we used that $\partial_\mu\partial_\nu = \partial_\nu\partial_\mu$. Note that we have that $0 = \partial_\nu(\Delta^\dagger M) = (\partial_v \Delta^\dagger)M + \Delta^\dagger(\partial_\nu M)$, since $\Delta^\dagger M = 0$ by definition of M. Conjugating this expression, we also find that $M^\dagger(\partial_\nu \Delta) + (\partial_\nu M^\dagger)\Delta = 0$. Therefore, replacing all instances of the above combinations by their negative counterparts, we find that

$$F_{\mu\nu} = M^\dagger(\partial_\mu \Delta)(\mathrm{id}_2 \otimes f)(\partial_\nu \Delta^\dagger)M - M^\dagger(\partial_\nu \Delta)(\mathrm{id}_2 \otimes f)(\partial_\mu \Delta^\dagger)M.$$

Now recall that

$$\Delta_x = a + b(\tau_\mu x^\mu \otimes \mathrm{id}_k), \quad \text{so that} \quad \partial_\mu \Delta = b(\tau_\mu \otimes \mathrm{id}_k).$$

We may now proceed to compute

$$\begin{aligned}
F_{\mu\nu} &= M^\dagger b(\tau_\mu \otimes \mathrm{id}_k)(\mathrm{id}_2 \otimes f)(\tau_\nu^\dagger \otimes \mathrm{id}_k)b^\dagger M - M^\dagger b(\tau_\nu \otimes \mathrm{id}_k)(\mathrm{id}_2 \otimes f)(\tau_\mu^\dagger \otimes \mathrm{id}_k)b^\dagger M \\
&= M^\dagger b[(\tau_\mu \otimes \mathrm{id}_k)(\mathrm{id}_2 \otimes f)(\tau_\nu^\dagger \otimes \mathrm{id}_k) - (\tau_\nu \otimes \mathrm{id}_k)(\mathrm{id}_2 \otimes f)(\tau_\mu^\dagger \otimes \mathrm{id}_k)]b^\dagger M \\
&= M^\dagger b[(\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger) \otimes f]b^\dagger M.
\end{aligned}$$

From this form, it is not hard to show that F is anti-self-dual. Namely, for example

$$\tau_1\overline{\tau_2} - \tau_2\overline{\tau_1} = i(-j) - j(-i) = -2k = -[k1 - 1(-k)] = -[\tau_3\overline{\tau_4} - \tau_4\overline{\tau_3}],$$

implying that $F_{12} + F_{34} = 0$; one of the anti-self-duality equations. The other two are derived equally easily. This completes the proof. □

This shows that the ADHM-data are sufficient to define an anti-self-dual connection on $\mathbb{H} \setminus \{0\}$; that is, to find a solution to the Yang-Mills equations. It can also be shown, using more advanced mathematical techniques, that it is even true that all anti-self dual connections can be obtained in this way, when one considers the proper gauge theory. We will not show this here. We will present one last theorem we will also not prove, but which will be useful to us in the next chapter.

Theorem 4.21. *The instanton defined above has an instanton number equal to k.*

Proof. The proof can be found in [2]. □

5 Calculating instantons

In the previous chapter we outlined a way to construct instantons out of nothing more than the ADHM-data, which was just a set of matrices obeying two equations, one complex and one real. Of course the most natural way to proceed would be to wonder what these instantons would actually look like in practice, and how convenient a way the ADHM-construction actually is to explicitly calculate instantons. This will be done by first examining what happens for small values of k and n, in order to find expressions for instantons of low topological charge. In particular, the well-known BPST-instanton of unit charge will be thoroughly explored. Then we will generalize to a method of finding instantons of arbitrary topological charge. The main source used for the section about calculating the BPST-instanton was [2].

5.1 The $k = 1$ and $n = 1$ case

Most naively we could try to construct a 1-instanton picking $k = n = 1$. Then we would simply have that $B_1, B_2, I, J \in \mathbb{C}$, and their conjugate transpose is simply given by their complex conjugate. Since $\mathbb{C}$ is commutative, we would find that the ADHM-equations require that

$$IJ = [B_1, B_2] + IJ = 0 \quad \text{and} \quad |I|^2 - |J|^2 = [B_1, \overline{B_1}] + [B_2, \overline{B_2}] + I\overline{I} - \overline{J}J = 0.$$

Thus we find that $|I| = |J|$, and since their product vanishes, one of them must be zero. Therefore we find that $I = J = 0$. In addition, we must also have that the map given by

$$\alpha_{(x,y)} : V \cong \mathbb{C} \to W \oplus V \oplus V \cong \mathbb{C}^3 : v \mapsto \begin{pmatrix} 0 \\ -w_2 B_1 - w_1 \overline{B_2} - z_1 \\ w_2 B_2 - w_1 \overline{B_1} + z_2 \end{pmatrix} v$$

is injective for any $x = (z_1, z_2)$ and $y = (w_1, w_2)$ such that $(x, y) \neq (0, 0)$. However, should we choose $w_2 = -1$, $w_1 = 0$, $z_1 = B_1$ and $z_2 = B_2$ we would have that $\alpha_{(x,y)} = 0$, and therefore it is not injective. We find that no such ADHM-data exists in this case, and therefore we cannot construct any instantons in this way.

In fact, we could have paused for a second when we arrived at the fact that $I = J = 0$. Namely, it turns out that there exists no system of ADHM-data with the property that $I = 0$ or $J = 0$, regardless of the values of n and k. This shows that one should be very careful in choosing the ADHM-data, and that it is not allowed to set too many variables to zero at the same time.

Lemma 5.1. *Let S and T be commuting complex matrices. Then for any eigenvalue λ of T, there exists some eigenvector v of T with that eigenvalue that is also an eigenvector of A.*

Proof. Let E_λ be the eigenspace of T corresponding to the eigenvalue λ. Then we observe that for any $v \in E_\lambda$, we have that

$$TSv = STv = \lambda Sv, \quad \text{so that also} \quad Sv \in E_\lambda.$$

We therefore find a linear map $S|_{E_\lambda} : E_\lambda \to E_\lambda$ which, since we are working over $\mathbb{C}$, has some eigenvector $v \in E_\lambda$. Since this is also an eigenvector for S, the claim follows. $\square$

Proposition 5.2. *There exist no ADHM-data satisfying either $I = 0$ or $J = 0$.*

Proof. First suppose that $I = 0$, and consider the second ADHM-equation. As the trace of any commutator vanishes, since $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, we find that

$$\sum_{i=1}^{n} \sum_{j=1}^{k} |J_{ij}|^2 = \mathrm{tr}(J^\dagger J) = 0.$$

This shows that also $J = 0$. If one sets out with $J = 0$, then it follows by the same argument that also $I = 0$. The first ADHM equation now reduces to

$$[B_1, B_2] = 0, \quad \text{so that} \quad B_1 B_2 = B_2 B_1.$$

We will proceed to show that $\alpha_{(x,y)}$ is not injective for some particular $(x, y) \neq (0, 0)$. To this end, again choose $w_1 = 0$, and $w_2 = -1$.

Since B_1 and B_2 commute, by the previous lemma there exists some $v \neq 0$ that is an eigenvector of B_1 and B_2 with eigenvalues λ_1 and λ_2 respectively. Should we then choose $z_1 = \lambda_1$ and $z_2 = \lambda_2$, we find that

$$\alpha_{(x,y)}(v) = \begin{pmatrix} 0 \\ B_1 - \lambda_1 \mathrm{id}_k \\ -B_2 + \lambda_2 \mathrm{id}_k \end{pmatrix} v = 0,$$

so that $\alpha_{(x,y)}$ cannot be injective. This completes the proof. $\square$

Remark 5.3. Should one not immediately fix $y = (w_1, w_2)$, it is even possible to show that for *any* y, there exists some $x = (z_1, z_2)$ such that $\alpha_{(x,y)}$ is not injective. This can be seen by observing that both reduced ADHM-equations together imply that $w_2 B_2 - w_1 B_1^\dagger$ and $-w_2 B_1 - w_1 B_2^\dagger$ commute for any w_1 and w_2, after which the result follows by the same argument as in the proof above.

5.2 The $k = 1$ and $n = 2$ case

A natural way to proceed would be to ask whether or not it is also impossible for B_1 and B_2 to be identically zero. It turns out that this is not the case, as we will be able to construct an instanton when $k = 1$ and $n = 2$ for any values of $B_1 = b_1 \in \mathbb{C}$ and $B_2 = b_2 \in \mathbb{C}$; even when they are zero. As a result we will have constructed the most fundamental instanton of all; the BPST-instanton. We start off by writing out what the ADHM-equations tell us about I and J. Again since $\mathbb{C}$ is commutative, we find that

$$IJ = 0 \quad \text{and} \quad II^\dagger = J^\dagger J.$$

Therefore we find that setting

$$I = \begin{pmatrix} \rho & 0 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 \\ \rho \end{pmatrix}$$

for some $\rho \in \mathbb{R} \setminus \{0\}$ is a valid solution to the ADHM-equations, regardless of the values of b_1 and b_2. We can explicitly verify that now $\alpha_{(x,y)}$ is injective for all $(x, y) \neq (0, 0)$. Namely,

$$\alpha_{(x,y)} : \mathbb{C} \to \mathbb{C}^4 : \begin{pmatrix} w_2 \begin{pmatrix} 0 \\ \rho \end{pmatrix} - w_1 \begin{pmatrix} \rho \\ 0 \end{pmatrix} \\ -w_2 b_1 - w_1 \overline{b_2} - z_1 \\ w_2 b_2 - w_1 \overline{b_1} + z_2 \end{pmatrix}.$$

The first row does not yield an injective function if and only if $w_1 = w_2 = 0$. Then the second row does not yield an injective function only if $z_1 = 0$, and similarly for the last row. Therefore $\alpha_{(x,y)}$ is always injective, bar for $(0, 0)$. By the same line of reasoning we can show that $\beta_{(x,y)}$ is surjective for all $(x, y) \neq (0, 0)$. This shows that the above described set (B_1, B_2, I, J) indeed forms valid ADHM-data.

From here on out it is just calculating. We have that

$$\Delta_x^\dagger = \begin{pmatrix} \rho & 0 & b_2 + z_2 & b_1 + z_1 \\ 0 & \rho & -\overline{b_1} - \overline{z_1} & \overline{b_2} + \overline{z_2} \end{pmatrix},$$

and it is easy to see that

$$\begin{aligned} M &= \frac{1}{\sqrt{\rho^2 + |z_1 + b_1|^2 + |z_2 + b_2|^2}} \begin{pmatrix} -z_2 - b_2 & -z_1 - b_1 \\ \overline{z_1} + \overline{b_1} & -\overline{z_2} - \overline{b_2} \\ \rho & 0 \\ 0 & \rho \end{pmatrix} \\ &= \frac{1}{\sqrt{\rho^2 + |x + b|^2}} \begin{pmatrix} -\overline{x + b} \\ \rho \mathrm{id}_2 \end{pmatrix} \end{aligned}$$

consists of an orthonormal basis of the kernel of $\Delta_x^\dagger$, where we have identified the quaternions x and b with their respective matrices. Since $\partial_\mu x = \tau_\mu$ for x viewed as a matrix, and $\partial_\mu |x + b|^2 = 2x^\mu + 2b^\mu$, it follows that

$$\partial_\mu M = \begin{pmatrix} \frac{-\tau_\mu^\dagger}{\sqrt{\rho^2+|x+b|^2}} - \frac{-\overline{x+b}(x^\mu+b^\mu)\mathrm{id}_2}{(\rho^2+|x+b|^2)^{3/2}} \\ -\frac{\rho(x^\mu+b^\mu)\mathrm{id}_2}{(\rho^2+|x+b|^2)^{3/2}} \end{pmatrix}.$$

Thus we can compute that

$$\begin{aligned}
A_\mu &= M^\dagger \partial_\mu M \\
&= \frac{\tau_\mu^\dagger (x+b)}{\rho^2 + |x+b|^2} - \frac{|x+b|^2 (x^\mu + b^\mu)\mathrm{id}_2}{(\rho^2+|x+b|^2)^2} - \frac{\rho^2 (x^\mu + b^\mu)\mathrm{id}_2}{(\rho^2+|x+b|^2)^2} \\
&= \frac{\tau_\mu^\dagger (x+b) - (x^\mu + b^\mu)\mathrm{id}_2}{\rho^2 + |x+b|^2}.
\end{aligned}$$

This is the famous BPST-instanton, and the usual physical interpretation is that it is an instanton *centered at* $-b$ with *size* ρ. As an example, we will briefly verify a number of properties of this instanton, starting off with the fact that it is $\mathfrak{su}(2)$-valued. To this end, we will write out our concise expression for A_1, using $y = x + b$ for simplicity. We compute that

$$\begin{aligned}
\tau_1^\dagger y - y^1 \mathrm{id}_2 &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \begin{pmatrix} -iy^3 + y^4 & -iy^1 - y^2 \\ -iy^1 + y^2 & iy^3 + y^4 \end{pmatrix} - \begin{pmatrix} y^1 & 0 \\ 0 & y^1 \end{pmatrix} \\
&= \begin{pmatrix} iy^2 & -y^3 + iy^4 \\ y^3 + iy^4 & -iy^2 \end{pmatrix},
\end{aligned}$$

so that indeed $A_1^\dagger = -A_1$, so that A_1 is skew-hermitian, and that also $\mathrm{tr}(A_1) = 0$. As shown in proposition 3.8, this implies that A_1 is in fact $\mathfrak{su}(2)$-valued, as desired. It can be checked by a similar computation that the other three components of A are $\mathfrak{su}(2)$-valued as well.

It might also be illustrative to calculate the accompanied curvature, and to use it to calculate its instanton number.

Recall that from the proof that F was anti-self-dual, we found a particularly simple expression for $F_{\mu\nu}$. First observe that from

$$\Delta_x^\dagger \Delta_x = \begin{pmatrix} \rho^2 + |x+b|^2 & 0 \\ 0 & \rho^2 + |x+b|^2 \end{pmatrix},$$

we find that $f_x = 1/(\rho^2 + |x+b|^2)$. Now we can calculate that

$$\begin{aligned}
F_{\mu\nu} &= M^\dagger b[(\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger) \otimes f] b^\dagger M \\
&= \frac{1}{(\rho^2 + |x+b|^2)^2} \begin{pmatrix} -(x+b) & \rho \mathrm{id}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} (\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger) \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\overline{x+b} \\ \rho \mathrm{id}_2 \end{pmatrix} \\
&= \frac{\rho^2}{(\rho^2 + |x+b|^2)^2} (\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger).
\end{aligned}$$

Explicitely writing out all the commutators, we find that

$$\begin{aligned}
\frac{(\rho^2 + |x+b|^2)^2}{\rho^2} F &= 2(dx^1 \wedge dx^4 - dx^2 \wedge dx^3) \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} \\
&\quad + 2(dx^2 \wedge dx^4 - dx^1 \wedge dx^3) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + 2(dx^3 \wedge dx^4 - dx^1 \wedge dx^2) \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.
\end{aligned}$$

Now if we observe that all matrices square to $-\mathrm{id}_2$, we obtain, carefully keeping track of the correct signs in all six terms, that

$$F \wedge F = -\frac{\rho^4}{(\rho^2 + |x+b|^2)^4} 24\mathrm{id}_2 \text{ vol}, \quad \text{so that} \quad \mathrm{tr}(F \wedge F) = -\frac{48\rho^4}{(\rho^2 + |x+b|^2)^4}\text{vol}.$$

Therefore we can at last compute that

$$\mathrm{n} = \frac{1}{8\pi^2}\int_{\mathbb{H}} -\frac{48\rho^4}{(\rho^2 + |x+b|^2)^4}\text{vol} = -\frac{1}{8\pi^2}\int_{\mathbb{R}^4} \frac{48\rho^4}{(\rho^2 + |x|^2)^4} dx^1 dx^2 dx^3 dx^4 = -1,$$

where the integral can be evaluated by switching to polar coordinates. This confirms that we have indeed constructed a 1-instanton on $\mathbb{R}^4$ with an $\mathfrak{su}(2)$-valued vector potential, precisely as the ADHM-construction claimed.

5.3 The $k = 1$ and $n = 3$ case

Feeling confident in our abilities, we could attempt to describe an even more general 1-instanton, by increasing our number of free variables n with 1. For example, we can completely analogously to before choose in this case that

$$I = \begin{pmatrix} \rho & 0 & 0 \end{pmatrix}, \quad \text{and} \quad J = \begin{pmatrix} 0 \\ j_2 \\ j_3 \end{pmatrix},$$

subject to the constraint that $|j_2|^2 + |j_3|^2 = \rho^2$. Subsequently we find that

$$\Delta_x^\dagger = \begin{pmatrix} \rho & 0 & 0 & b_2 + z_2 & b_1 + z_1 \\ 0 & \overline{j_2} & \overline{j_3} & -\overline{b_1} - \overline{z_1} & \overline{b_2} + \overline{z_2} \end{pmatrix}.$$

Now, it is easy to find two orthogonal vectors that are in the kernel of $\Delta_x^\dagger$, namely

$$\begin{pmatrix} 0 \\ -\overline{j_3} \\ \overline{j_2} \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} -|b_1 + z_1|^2 - |b_2 + z_2|^2 \\ 0 \\ 0 \\ \rho(\overline{b_2} + \overline{z_2}) \\ \rho(\overline{b_1} + \overline{z_1}) \end{pmatrix}.$$

If we then pick another vector in the kernel at random, say

$$\begin{pmatrix} \overline{j_2}(b_2 + z_2) \\ -\rho(\overline{b_1} + \overline{z_1}) \\ 0 \\ -\rho\overline{j_2} \\ 0 \end{pmatrix},$$

then the Gram-Schmidt orthogonalization process gives us that

$$M = \begin{pmatrix} 0 & \frac{-|x+b|}{\sqrt{\rho^2+|x+b|^2}} & 0 \\ -\overline{j_3}/\rho & 0 & \frac{-|j_2|(\overline{b_1}+\overline{z_1})|x+b|}{\rho|b_1+z_1|\sqrt{\rho^2+|x+b|^2}} \\ \overline{j_2}/\rho & 0 & \frac{-\overline{j_2}j_3(\overline{b_1}+\overline{z_1})|x+b|}{|j_2|\rho|b_1+z_1|\sqrt{\rho^2+|x+b|^2}} \\ 0 & \frac{\rho(\overline{b_2}+\overline{z_2})}{|x+b|\sqrt{\rho^2+|x+b|^2}} & \frac{-\rho\overline{j_2}|b_1+z_1|}{|j_2||x+b|\sqrt{\rho^2+|x+b|^2}} \\ 0 & \frac{\rho(\overline{b_1}+\overline{z_1})}{|x+b|\sqrt{\rho^2+|x+b|^2}} & \frac{\rho\overline{j_2}(\overline{b_1}+\overline{z_1})(b_2+z_2)}{|b_1+z_1||j_2||x+b|\sqrt{\rho^2+|x+b|^2}} \end{pmatrix}.$$

Now by mere perseverance and the aid of a computer, using $A_\mu = M^\dagger \partial_\mu M$, we obtain after up to five hours of computation time for some entries:

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\frac{i\rho^2(b_{1r}+x_2)}{|x+b|^2(\rho^2+|x+b|^2)} & \frac{-i\rho^2(\overline{b_1}+\overline{z_1})(b_2+z_2)\overline{j_2}}{|b_1+z_1||x+b|^2(\rho^2+|x+b|^2)|j_2|} \\ 0 & \frac{-i\rho^2(b_1+z_1)(\overline{b_2}+\overline{z_2})j_2}{|b_1+z_1||x+b|^2(\rho^2+|x+b|^2)|j_2|} & \frac{-i(b_{1r}+x_2)}{|b_1+z_1|^2} - \frac{-i(b_{1r}+x_2)}{|x+b|^2} + \frac{-i(b_{1r}+x_2)}{\rho^2+|x+b|^2} \end{pmatrix};$$

$$A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{i\rho^2(b_{1i}+x_1)}{|x+b|^2(\rho^2+|x+b|^2)} & \frac{-\rho^2(\overline{b_1}+\overline{z_1})(b_2+z_2)\overline{j_2}}{|b_1+z_1||x+b|^2(\rho^2+|x+b|^2)|j_2|} \\ 0 & \frac{\rho^2(b_1+z_1)(\overline{b_2}+\overline{z_2})j_2}{|b_1+z_1||x+b|^2(\rho^2+|x+b|^2)|j_2|} & \frac{i(b_{1i}+x_1)}{|b_1+z_1|^2} - \frac{i(b_{1i}+x_1)}{|x+b|^2} + \frac{i(b_{1i}+x_1)}{\rho^2+|x+b|^2} \end{pmatrix};$$

$$A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{-i\rho^2(b_{2r}+x_4)}{|x+b|^2(\rho^2+|x+b|^2)} & \frac{i\rho^2|b_1+z_1|\overline{j_2}}{|x+b|^2(\rho^2+|x+b|^2)|j_2|} \\ 0 & \frac{i\rho^2|b_1+z_1|j_2}{|x+b|^2(\rho^2+|x+b|^2)|j_2|} & \frac{i\rho^2(b_{2r}+x_4)}{|x+b|^2(\rho^2+|x+b|^2)} \end{pmatrix};$$

$$A_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{i\rho^2(b_{2i}+x_3)}{|x+b|^2(\rho^2+|x+b|^2)} & \frac{\rho^2|b_1+z_1|\overline{j_2}}{|x+b|^2(\rho^2+|x+b|^2)|j_2|} \\ 0 & \frac{-\rho^2|b_1+z_1|j_2}{|x+b|^2(\rho^2+|x+b|^2)|j_2|} & \frac{-i\rho^2(b_{2i}+x_3)}{|x+b|^2(\rho^2+|x+b|^2)} \end{pmatrix}.$$

Remark 5.4. It should not come as a surprise that we found so many zeroes. Namely, one of the vectors we used to span the kernel of $\Delta_x^\dagger$ was constant, so that its derivative vanishes. Therefore the first column of any A_μ is most certainly zero. To see why the first row vanishes as well, observe that if of some vector the second and third entry have a ratio of j_2 to j_3 so as to ensure orthogonality with the constant vector, so do their derivatives, and hence those are still orthogonal. This argument trivially extends to the case in which $k = 1$ and n is arbitrary; then we can find $n-2$ constant vectors in the kernel of $\Delta_x^\dagger$, so that again only some 2×2 block in the lower right corner of A_μ will be non-zero. A second reason for this fenomenon will be explained in chapter 6.

It is worth to note that all matrices are indeed skew-Hermitian, and that A_3 and A_4 indeed have vanishing trace, so that they are $\mathfrak{su}(3)$-valued. To get the trace of A_1 and A_2 to vanish as well, a proper gauge transformation must be applied first. A lot less computation-heavy would be to calculate the curvature, which does not depend on derivatives of M at all. We again have that $f = 1/(\rho^2 + |x+b|^2)$. Therefore we calculate that

$$\begin{aligned}(\rho^2+|x+b|^2)^2|x+b|^2 F_{\mu\nu} &= (\rho^2+|x+b|^2)^2|x+b|^2 M^\dagger b[(\tau_\mu\tau_\nu^\dagger - \tau_\nu\tau_\mu^\dagger)\otimes f]b^\dagger M \\ &= \begin{pmatrix} 0 & 0 \\ \rho(b_2+z_2) & \rho(b_1+z_1) \\ \frac{-\rho j_2|b_1+z_1|}{|j_2|} & \frac{\rho j_2(b_1+z_1)(\overline{b_2}+\overline{z_2})}{|b_1+z_1||j_2|} \end{pmatrix} (\tau_\mu\tau_\nu^\dagger - \tau_\nu\tau_\mu^\dagger) \begin{pmatrix} 0 & \rho(\overline{b_2}+\overline{z_2}) & \frac{-\rho\overline{j_2}|b_1+z_1|}{|j_2|} \\ 0 & \rho(\overline{b_1}+\overline{z_1}) & \frac{\rho\overline{j_2}(\overline{b_1}+\overline{z_1})(b_2+z_2)}{|b_1+z_1||j_2|} \end{pmatrix}.\end{aligned}$$

It turns out however that in this case there are not really any terms that cancel each other out, so that the above expression is about as insightful as we can have it. It should be noted that this calculation will result in a set of 3×3 matrices that have both their first row and first column identically zero, as was to be expected from the expressions of the A_μ.

5.4 The $k = 2$ and $n = 1$ case

We now turn to the question of constructing 2-instantons. Again, most naively, one could try to analyse the case with the fewest free variables; that is, $k = 2$ and $n = 1$, in which case there are 12. As is clear from the preceding section, it is imperative for any decently computable instanton to start out with ADHM-data containing a great many zeroes. An obvious choice for I and J would be similar to the $k = 1$ and $k = 2$ case. Recall that looking at the trace of the second ADHM-equation gave us that I and J have the same Frobenius-norm, so that we may try to choose

$$I = \begin{pmatrix} \rho \\ 0 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 & \tilde\rho \end{pmatrix}, \quad \text{or} \quad I = \begin{pmatrix} 0 \\ \rho \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} \tilde\rho & 0 \end{pmatrix},$$

where $|\rho| = |\tilde\rho|$. Since both B_1 and B_2 still contain 4 free variables each, it might be desirable to set some more entries to zero. Since products of upper-triangular matrices are again upper-triangular, an obvious choice would be to choose either B_1 or B_2 of this form. It turns out however, that this is not possible, and the proof heavily relies on the fact that $I, J \neq 0$, as proved in proposition 5.2.

Proposition 5.5. *Suppose that $k = 2$ and $n = 1$. If either I or J contains a zero, then neither B_1 nor B_2 is upper triangular.*

Proof. Write

$$I = \begin{pmatrix} i_1 \\ i_2 \end{pmatrix}, \quad J = \begin{pmatrix} j_1 & j_2 \end{pmatrix}, \quad B_1 = \begin{pmatrix} s_1 & s_2 \\ s_3 & s_4 \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} t_1 & t_2 \\ t_3 & t_4 \end{pmatrix}.$$

Consider the trace of the first ADHM-equation. We find that $i_1 j_1 + i_2 j_2 = 0$. Therefore if $i_2 = 0$, we have, by proposition 5.2, that $i_1 \neq 0$. Therefore it must follow that $j_1 = 0$. Similarly, if some entry of J is zero, we find that some entry of I is zero, so we restrict our view to the case where I constains a zero.

Suppose that $i_2 = 0$, so that also $j_1 = 0$, and that B_1 is upper-triangular, so that $s_3 = 0$. The case that B_2 is upper-triangular will go analogously. Observe that the only non-zero entry of IJ is the upper-right, since $i_1, j_2 \neq 0$. If we then consider the upper-left entry of the first ADHM-equation, we find that $s_2 t_3 = 0$. If $t_3 = 0$, then B_2 is also upper-triangular. But then the upper-left entry of the second ADHM-equation reduces to $|s_2|^2 + |t_2|^2 + |i_2|^2 = 0$, so that $i_2 = 0$. But then $I = 0$, contradicting proposition 5.2. Therefore $t_3 \neq 0$ and $s_2 = 0$, so that B_1 is diagonal. The lower-left entry of the same equation gives us $(s_4 - s_1)t_3 = 0$. Thus $s_4 = s_1$, so that B_1 is a multiple of the identity, thus $[B_1, B_2] = 0$. But then the first ADHM-equation reduces to $IJ = 0$, so that either I or J vanishes; this again contradicts proposition 5.2.

Now if we would have had that $i_1 = 0$ and $i_2 \neq 0$, by the same reasoning we would have had that $s_2 t_3 = 0$. If $t_3 = 0$, then B_2 is upper-triangular as well. Since IJ now only has a nonzero entry in the lower-left corner, we find that $IJ = 0$, so that either I or J vanishes, once more contradicting proposition 5.2. Therefore $t_3 \neq 0$ and $s_2 = 0$, so that B_1 is diagonal. The upper-right corner in the first ADHM-equation then gives us that $(s_1 - s_4)t_2 = 0$. Since $s_1 \neq s_4$ as before, we find that $t_2 = 0$, so that B_2 is lower-triangular. Now since $[B_1, B_1^\dagger] = 0$, the lower-right entry of the second ADHM-equation reduces to $|t_3|^2 + |i_2|^2 = 0$, so that $i_2 = 0$. But now $I = 0$, contradicting proposition 5.2 one final time. $\square$

Perhaps a better idea would be to abandon the nice forms of I and J, and to just focus on choosing B_1 and B_2 to make our lives a little easier. Unfortunately, the options for this are also very limited, as can be quite easily shown.

Proposition 5.6. *These exist no ADHM-data for $k = 2$ and $n = 1$ for which B_1 or B_2 is diagonal, or for which B_1 and B_2 are both upper-triangular.*

Proof. If B_1 is diagonal, then we find that the diagonal-entries in $[B_1, B_2]$ vanish. Diagonal-entries in the first ADHM-equation therefore give us that $i_1 j_1 = i_2 j_2 = 0$, so that we are in the situation of the preceding proposition. However, it was shown that B_1 cannot be upper-triangular, which diagonal matrices most certainly are. From an identical argument it follows that B_2 cannot be diagonal either.

If B_1 and B_2 are both upper-triangular, then $[B_1, B_2]$ is upper-triangular as well. Therefore, the lower-left corner of the first ADHM-equation gives us that $i_2 j_1 = 0$, so that we again find ourselves in the situation of the preceding proposition. $\square$

It is not hard to see that the same claims extend to lower-triangular matrices as well. We conclude that in order to construct 2-instantons using the ADHM-construction, we will have to go about a tad more cleverly than just setting many variables to zero to ease the calculations. This will be achieved in the next section.

5.5 The $k = 2$ and $n = 4$ case

We might have more luck if we allow ourselves some more freedom. So instead of limiting our matrices I and J to be so narrow as to represent vectors, we expand to $n = 4$, allowing two of our simplest solutions to exist alongside each other. More concretely, we choose

$$B_1 = \begin{pmatrix} b_1 & 0 \\ 0 & b_3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} b_2 & 0 \\ 0 & b_4 \end{pmatrix}, \quad I = \left(\begin{array}{cc|cc} \rho & 0 & 0 & 0 \\ \hline 0 & 0 & \mu & 0 \end{array}\right), \quad \text{and} \quad J = \left(\begin{array}{c|c} 0 & 0 \\ \rho & 0 \\ \hline 0 & 0 \\ 0 & \mu \end{array}\right),$$

for certain $\rho, \mu \in \mathbb{R}$. The effect of choosing our matrices like this, is that we essentially have two disjoint 1-instantons sitting next to each other. Then we find that

$$\Delta_x^\dagger = \begin{pmatrix} \rho & 0 & 0 & 0 & b_2 + z_2 & 0 & b_1 + z_1 & 0 \\ 0 & 0 & \mu & 0 & 0 & b_4 + z_2 & 0 & b_3 + z_1 \\ 0 & \rho & 0 & 0 & -\overline{b_1} - \overline{z_1} & 0 & \overline{b_2} + \overline{z_2} & 0 \\ 0 & 0 & 0 & \mu & 0 & -\overline{b_3} - \overline{z_1} & 0 & \overline{b_4} + \overline{z_2} \end{pmatrix},$$

and it is easy to see that an orthonormal basis of the kernel is given by

$$\begin{aligned} M &= \begin{pmatrix} \frac{-b_2 - z_2}{\sqrt{\rho^2 + |b^1 + x|^2}} & \frac{-b_1 - z_1}{\sqrt{\rho^2 + |b^1 + x|^2}} & 0 & 0 \\ \frac{\overline{b_1} + \overline{z_1}}{\sqrt{\rho^2 + |b^1 + x|^2}} & \frac{-\overline{b_2} - \overline{z_2}}{\sqrt{\rho^2 + |b^1 + x|^2}} & 0 & 0 \\ 0 & 0 & \frac{-b_4 - z_2}{\sqrt{\mu^2 + |b^2 + x|^2}} & \frac{-b_3 - z_1}{\sqrt{\mu^2 + |b^2 + x|^2}} \\ 0 & 0 & \frac{\overline{b_3} + \overline{z_1}}{\sqrt{\mu^2 + |b^2 + x|^2}} & \frac{-\overline{b_4} - \overline{z_2}}{\sqrt{\mu^2 + |b^2 + x|^2}} \\ \frac{\rho}{\sqrt{\rho^2 + |b^1 + x|^2}} & 0 & 0 & 0 \\ 0 & 0 & \frac{\mu}{\sqrt{\rho^2 + |b^2 + x|^2}} & 0 \\ 0 & \frac{\rho}{\sqrt{\rho^2 + |b^1 + x|^2}} & 0 & 0 \\ 0 & 0 & 0 & \frac{\mu}{\sqrt{\mu^2 + |b^2 + x|^2}} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{\overline{b^1 + x}}{\sqrt{\rho^2 + |b^1 + x|^2}} & 0 \\ 0 & -\frac{\overline{b^2 + x}}{\sqrt{\mu^2 + |b^2 + x|^2}} \\ \frac{\rho}{\sqrt{\rho^2 + |b^1 + x|^2}} E_{11} & \frac{\mu}{\sqrt{\mu^2 + |b^2 + x|^2}} E_{21} \\ \frac{\rho}{\sqrt{\rho^2 + |b^1 + x|^2}} E_{12} & \frac{\mu}{\sqrt{\mu^2 + |b^2 + x|^2}} E_{22} \end{pmatrix}, \end{aligned}$$

where we set the quaternions $x = (z_1, z_2)$, $b^1 = (b_1, b_2)$ and $b^2 = (b_3, b_4)$, and introduced the elementary 2×2 matrices E_{ij} defined by $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$. Therefore, just as before,

$$\partial_\mu M = \begin{pmatrix} \frac{-\tau_\mu^\dagger}{\sqrt{\rho^2+|b^1+x|^2}} - \frac{-\overline{b^1+x}(x^\mu+b^{1\mu})\mathrm{id}_2}{(\rho^2+|b^1+x|^2)^{3/2}} & 0 \\ 0 & \frac{-\tau_\mu^\dagger}{\sqrt{\mu^2+|b^2+x|^2}} - \frac{-\overline{b^2+x}(x^\mu+b^{2\mu})\mathrm{id}_2}{(\mu^2+|b^2+x|^2)^{3/2}} \\ -\frac{\rho(x^\mu+b^{1\mu})\mathrm{id}_2}{(\rho^2+|b^1+x|^2)^{3/2}}E_{11} & -\frac{\mu(x^\mu+b^{2\mu})\mathrm{id}_2}{(\mu^2+|b^2+x|^2)^{3/2}}E_{21} \\ -\frac{\rho(x^\mu+b^{1\mu})\mathrm{id}_2}{(\rho^2+|b^1+x|^2)^{3/2}}E_{12} & -\frac{\mu(x^\mu+b^{2\mu})\mathrm{id}_2}{(\mu^2+|b^2+x|^2)^{3/2}}E_{22} \end{pmatrix}.$$

Thus we can calculate, using certain multiplicative properties of the E_{ij}, that

$$A_\mu = M^\dagger \partial_\mu M = \begin{pmatrix} \frac{\tau_\mu^\dagger(b^1+x)-(x^\mu+b^{1\mu})\mathrm{id}_2}{\rho^2+|b^1+x|^2} & 0 \\ 0 & \frac{\tau_\mu^\dagger(b^2+x)-(x^\mu+b^{2\mu})\mathrm{id}_2}{\mu^2+|b^2+x|^2} \end{pmatrix}.$$

We conclude that the instanton we have constructed is nothing more than the old BPST-instanton, but now we have two of them, both with their own size and location. Now if we denote by F^{ρ,b^1} and F^{μ,b^2} the respective curvatures, it is not hard to see that now

$$F_{\mu\nu} = \begin{pmatrix} F^{\rho,b^1} & 0 \\ 0 & F^{\mu,b^2} \end{pmatrix}, \quad \text{so that} \quad \mathrm{tr}(F \wedge F) = \mathrm{tr}\left(F^{\rho,b^1} \wedge F^{\rho,b^1}\right) + \mathrm{tr}\left(F^{\mu,b^2} \wedge F^{\mu,b^2}\right).$$

Therefore it is clear that

$$\begin{aligned} \mathrm{n} &= \frac{1}{8\pi^2}\int_{\mathbb{H}} \mathrm{tr}(F \wedge F)\ \mathrm{vol} \\ &= \frac{1}{8\pi^2}\int_{\mathbb{H}} \mathrm{tr}\left(F^{\rho,b^1} \wedge F^{\rho,b^1}\right)\ \mathrm{vol} + \frac{1}{8\pi^2}\int_{\mathbb{H}} \mathrm{tr}\left(F^{\mu,b^2} \wedge F^{\mu,b^2}\right)\ \mathrm{vol} = -1 - 1 = -2. \end{aligned}$$

Therefore we have indeed constructed a 2-instanton, and since we already checked for the BPST-instanton that the vector potential is $\mathfrak{su}(2)$-valued, it follows immediately that the above acquired vector potential is $\mathfrak{su}(4)$-valued. It should of course not really come as a surprise that placing two BPST-instantons on the diagonal of a larger matrix will yield a 2-instanton. The reason for going through the above calculation, is that we have verified that by only using what we have learned before, we will not be constructing any *new* instantons anytime soon. Indeed, this procedure immediately generalizes to arbitrary $n = 2m$ to find an m-instanton, by placing m BPST-instantons next to eachother on the diagonal of some large matrix. Therefore the fact that instantons of arbitrary topological charge exist, follows somewhat trivially from the existence of the BPST-instanton. The price we had to pay however was that we abandoned the $SU(2)$ theory which was of particular interest, and that we introduced $SU(2m)$ theories for $m > 1$.

5.6 The ‘t Hooft instantons

We will conclude by constructing a k-instanton for $n = 2$ for any k. This will result in a family of $\mathfrak{su}(2)$-valued vector potentials that were first discovered by ‘t Hooft, and are therefore called the ‘t Hooft k-instanton solutions. The key idea is that in order to construct a matrix in $\mathfrak{su}(2)$, it suffices to construct a well-chosen quaternion, since we may regard that as a 2×2 matrix itself. We will set out with the $(k+1) \times k$ matrix M_x that has its first row equal to $M_{x,1j} = \lambda_j > 0$ with real entries, and the remaining k rows satisfy $M_{x,i+1,j} = \delta_{ij}(y_i - x)$, where the quaternion x denotes the position in 4-space, and where the y_i are just constant quaternions. Now let N_x be a $(k+1)$-column vector consisting of quaternions defined by

$$N_{x1} = \frac{1}{\rho} \quad \text{and} \quad N_{xi} = -\frac{1}{\rho}\frac{y_{i-1} - x}{|y_{i-1} - x|^2}\lambda_{i-1} \quad \text{for } i > 1\text{, where} \quad \rho = 1 + \sum_{i=1}^{k} \frac{\lambda_i^2}{|y_i - x|^2}.$$

Observe that it now holds that

$$N_x^\dagger M_x = 0 \quad \text{and} \quad N_x^\dagger N_x = 1.$$

Introducing the ‘t Hooft symbols,

$$\overline{\eta}^1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \overline{\eta}^2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \overline{\eta}^3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \overline{\eta}^4 = 0,$$

we will calculate our instanton explicitly by

$$A_\mu = N^\dagger \partial_\mu N = -\frac{i}{\rho}\overline{\eta}^a_{\mu\nu} \sum_{i=1}^{k} \frac{\lambda_i^2 \tau_a (x - y_i)^\nu}{|x - y_i|^4},$$

where there is an implicit summation over both a and ν. A more comprehensive study of the $SU(2)$-case can be found in [7]. Quite frankly, this formula is extremely messy and not at all easy to work with, but it is a set of k-instantons in an $SU(2)$-gauge theory.

It is worth remarking that [7] concluded that the most general form of a 2-instanton in an $SU(2)$-gauge theory is given by choosing

$$M = \begin{pmatrix} q_1 & q_2 \\ y_1 - x & b \\ b & y_2 - x \end{pmatrix}, \quad \text{where} \quad b = \frac{y_1 - y_2}{2|y_1 - y_2|^2}(q_2 q_1 - q_1 q_2).$$

Since all the numbers appearing in the equation above have a total of 4 parameters describing them, as they are all quaternions, actually calculating what instanton this will result in is nearly undoable, and the authors of [7] thoughtfully refrained from doing so themselves. Yet it is interesting to see that these descriptions of all instantons of a certain type, albeit indirect, have actually been established and that finding these descriptions was a topic of active contemporary research not too long ago.

6 Representations of quivers

The aim of this concluding section is to motivate the origin of the ADHM-equations, which appeared to magically define an anti-self-dual curvature on 4-dimensional space. Of course, Atiyah, Drinfeld, Hitchin and Manin, after whom the ADHM-construction is named, did not stumble upon their equations accidentally. It turned out that their constraints followed quite naturally from studying mathematical objects called quivers, and the extensive theory involving representations of quivers will eventually hand us the correct constraints.

The main source used for this chapter was chapter 1 and 2 of [11] and for the final section also [14].

6.1 Quivers and path algebras

We will start off by introducing these objects formally.

Definition 6.1. A *quiver* Q is a finite directed multigraph, with vertex set I and arrows $a : i \to j$ for some $i, j \in I$. One denotes $h(a) = j$ for the head of the arrow a and $t(a) = i$ for its tail. A *path* p in a quiver is a sequence of arrows $a_1 a_2 \cdots a_n$ so that $h(a_i) = t(a_{i-1})$ for all $1 < i \leq n$. Observe that we write the path as if the arrows were functions, so we start at walking along a_n and end at a_1. We also allow for paths e_i, which are stationary paths at the vertex i.

Definition 6.2. A *representation* of a quiver Q is a set of finite dimensional vector spaces $(V_i)_{i \in I}$ over some field K, which we take to have characteristic 0, and for each $a : i \to j$ a linear map, or matrix, $X_a : V_i \to V_j$. The *dimension vector* α is defined by $\alpha = (\dim V_1, \dim V_2, \ldots) \in \mathbb{N}_0^I$, and we write $\mathrm{Rep}(Q, \alpha)$ for the space of representations of Q with dimension vector α. A *homomorphism* φ between two representations is a set of linear maps $V_i \to V_i'$ such that the diagram below commutes for all arrows a. This

$$\begin{tikzcd} V_i \arrow[r, "X_a"] \arrow[d, "\varphi_i"'] & V_j \arrow[d, "\varphi_j"] \\ V_i' \arrow[r, "X_a'"] & V_j' \end{tikzcd}$$

defines a category of representations of a given quiver, and it turns out that *isomorphisms* are precisely those homomorphisms for which each ϕ_i is an isomorphism. Observe that two isomorphic representations must necessarily have the same dimension vector.

Remark 6.3. A trivial yet crucial observation is that upon identifying $V_i \cong V_i' \cong K^{\alpha_i}$, the constraint for an isomorphism of quivers can be translated to viewing the φ_i as basis transformations of the vector spaces V_i, so that X_a and X_a' represent the same linear map.

A priori there are no constraints on the maps that represent the arrows of a given quiver; two different paths that have equal starting and ending points, need upon composition not yield the same function. Also, loops from one point to itself may be represented by a non-trivial endomorphism.

Lemma 6.4. *Consider the group* $\mathrm{GL}_\alpha = \prod_I \mathrm{GL}(\alpha_i)$ *and let it act on* $\mathrm{Rep}(Q, \alpha)$ *by conjugation on each component. Then* $G_\alpha = \mathrm{GL}_\alpha / K^*$ *acts as well, and isomorphism classes of respresentations correspond directly to orbits under these group actions.*

Proof. Since non-zero constant factors are annihilated upon conjugation, it is obvious that the action of G_α is well-defined. By the previous remark, the second claim follows immediately. ☐

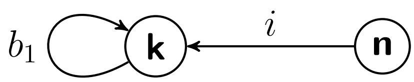
Recall that an *algebra* is a ring equipped with an action of some field K, called scalar multiplication. It turns out that a quiver naturally defines an algebra, which is most often at the center of attention.

Definition 6.5. The *path algebra* of a quiver Q over some field K is defined as the set of formal linear combinations of $\{p \mid p \text{ is a path}\}$, with the multiplication defined by

$$p \cdot q = \begin{cases} p \circ q & \text{if } h(q) = t(p) \\ 0 & \text{otherwise,} \end{cases}$$

for any two paths p and q, where $h(q)$ denotes the head of the last arrow of q and $t(p)$ the tail of the first arrow of p. The path algebra is usually denoted by KQ. Observe that the stationary paths e_i act trivially on compatible paths, but annihilate all others.

Example 6.6. As an easy example of a quiver, consider the picture below. It consists



of two vertices, labeled k and n, and two arrows, where arrow i goes between the two vertices and b_1 goes from k to itself. Its path algebra consists of expressions of the form

$$ke_n + \sum_{m=0}^{\infty} k_m (b_1)^m + \sum_{m'=0}^{\infty} k'_{m'} (b_1)^{m'} \circ i,$$

with all but finitely many $k_m, k'_{m'} \neq 0$, and where we set $(b_1)^0 = e_k$. Multiplication is generated by

$$i \cdot (b_1)^m = (b_1)^m \cdot e_n = e_n \cdot (b_1)^m = e_n \cdot i = i \cdot i = 0,$$

and in all other cases just by concatenation. △

Remark 6.7. The path algebra KQ is generally an infinte dimensional algebra, except for when Q has no cycles. It then follows immediately that the number of paths in Q is finite, and hence its path algebra is finitely generated over K.

In order to be able to give our last introductory definition, we will first need a small result that highlights why the path algebra is an interesting object to study.

Proposition 6.8. *There is an equivalence between the category of Q-representations and the category of KQ-modules.*

Proof. Given a representation of Q, we can define $M = \bigoplus_I V_i$. This is naturally a KQ module, where e_i acts as projection on the i-th component and any arrow $a : i \to j$ acts via projection on V_i, followed by the linear map associated with a and then inclusion back in M into the j-th component. Conversely, for any KQ-module M, we let $V_i = \{e_i \cdot m \mid m \in M\}$ and $V_i \to V_j$ be given by $e_i \cdot m \mapsto a \cdot m$, for $a : i \to j$. This defines a representation of Q. It is easy to check that the two-sided compositions of these functors are isomorphic to the identity, which concludes the proof. □

Definition 6.9. A representation of a quiver Q is said to be *indecomposable* if the associated module defined above is indecomposable; that is, there do not exist two proper submodules X and Y such that $M \cong X \oplus Y$. A representation of a quiver Q is said to be *simple* if the associated module defined above is simple; that is, it has no proper submodule. Observe that any simple module is trivially indecomposable.

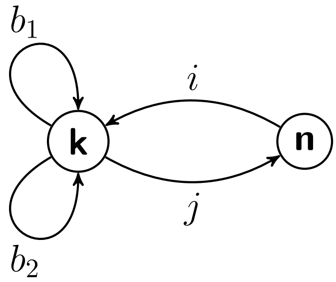
6.2 Preprojective algebras and the complex moment map

In order to proceed we will first need a very fundamental operation on quivers that allows our quiver to forget in a sense about the fact that it is a *directed* graph, by simply adding arrows opposite to the existent ones.

Definition 6.10. The *double* of a quiver Q is defined to be the quiver with the same vertices as Q, but for any arrow $a : i \to j$ in Q, we have an additional arrow $a^* : j \to i$. We denote this as $\overline{Q}$.

Example 6.11. Consider the quiver introduced in the previous example. We will construct its double, by denoting $b_1^* = b_2$ and $i^* = j$. The result is drawn in the picture below. This is the most fundamental example we will concern ourselves with, for it is crucial to observe that a representation of this quiver with, say, dimension vector $\alpha = (k, n)$, is precisely given by the objects appearing in the definition for the ADHM-data, if we associate $(b_1, b_2, i, j) \sim (B_1, B_2, I, J)$. For this reason, this quiver is generally referred to as the *ADHM-quiver.* △

Now we introduce another very fundamental object that will combine the notion of the double of a quiver and its path algebra.



Definition 6.12. Given a quiver Q and its double $\overline{Q}$, for any $\lambda \in K^I$ define the *deformed preprojective algebra of weight* λ by

$$\Pi^\lambda(Q) = K\overline{Q} \Big/ \Big(\sum_{a \in Q} [a, a^*] - \sum_{i \in I} \lambda_i e_i \Big).$$

For $\lambda = 0$ the object $\Pi^0(Q)$ is called the *preprojective algebra.*

Lemma 6.13. *Let M be a $\Pi^\lambda(Q)$-module and let X be the corresponding representation of $\overline{Q}$. Then for any vertex i we have that*

$$\sum_{h(a)=i} X_a X_{a^*} - \sum_{t(a)=i} X_{a^*} X_a = \lambda_i \mathrm{id}_{\alpha_i}.$$

Conversely, if the above equations hold for all vertices $i \in I$, then we have that the corresponding KQ-module M is also a $\Pi^\lambda(Q)$-module.

Proof. First let M be a $\Pi^\lambda(Q)$-module, and write r for the defining relation of the deformed preprojective algebra. Observe that

$$0 = e_i r e_i = \sum_{h(a)=i} a a^* - \sum_{t(a)=i} a^* a - \lambda_i e_i,$$

so that the claim follows immediately. For the reverse implication, note that if $e_i r e_i = 0$ for all i, then also $r = \sum_I e_i r e_i = 0$, so that M is a $\Pi^\lambda(Q)$-module. □

Corollary 6.14. *Any representation of $\overline{Q}$ corresponding to a $\Pi^\lambda(Q)$-module has a dimension vector α that satisfies $\langle \lambda, \alpha \rangle = 0$, where here $\langle ., . \rangle$ denotes the standard inner product.*

Proof. Take the trace and sum over all the above constraints; it follows that

$$\sum_{a \in Q} \mathrm{tr}(X_a X_{a^*}) - \sum_{a \in Q} \mathrm{tr}(X_{a^*} X_a) = \sum_i \mathrm{tr}(\lambda_i \mathrm{id}_{\alpha_i}) = \langle \lambda, \alpha \rangle.$$

Since the trace of any commutator vanishes, the claim follows. □

Example 6.15. Again consider the quiver from example 6.6, and its double, the ADHM-quiver, from example 6.11. For a given representation to be a $\Pi^0(Q)$-module, for vertex k we obtain the constraint

$$B_1B_2 + IJ - B_2B_1 = 0, \quad \text{or more succinctly,} \quad [B_1, B_2] + IJ = 0,$$

which is precisely the first ADHM-equation. $\triangle$

Remark 6.16. For a representation of the ADHM-quiver to be a $\Pi^0(Q)$-module, it must also hold that $JI = 0$. Observe that this must not necessarily hold for any set of ADHM-data, except for when $n = 1$. Namely, consider the trace of the first ADHM-equation. Since the trace of any commutator vanishes, we find that $\mathrm{tr}(IJ) = 0$, and since $n = 1$ and I and J are both vectors, it follows that $i_1j_1 + \ldots i_kj_k = 0$, or $JI = 0$, as desired. The BPST-instanton is an example of a set of ADHM-data which does not correspond to a $\Pi^0(Q)$-module, since $JI \neq 0$ in that case.

This shows that we are interested in representations of the ADHM-quiver that are also $\Pi^0(Q)$-modules, for all such representations automatically satisfy the first ADHM-equation. We now turn to a different concept that originated from algebraic geometry, which we will not generally introduce, but its application to quivers will be discussed alongside its connection to the ADHM-equations. For this we observe that the following function

$$\omega : \mathrm{Rep}(Q,\alpha) \times \mathrm{Rep}(Q,\alpha) \to K : (X,Y) \mapsto \sum_{a\in Q} \mathrm{tr}(X_{a^*}Y_a) - \sum_{a\in Q} \mathrm{tr}(X_aY_{a^*})$$

is a skew-symmetric, bilinear form that is also non-degenerate, so that if for some representation X it holds that $\omega(X,Y) = 0$ for all representations Y, then $X = 0$. For short, one can also say that ω is a *symplectic* form. We set $K = \mathbb{C}$ in the forthcoming.

Observe that the group action of G_α preserves ω, since for any $g \in G_\alpha$ and $a : i \to j$ we have that

$$\mathrm{tr}\left(g_iX_ag_j^{-1} \cdot g_jY_{a^*}g_i^{-1}\right) = \mathrm{tr}\left(g_iX_aY_{a^*}g_i^{-1}\right) = \mathrm{tr}(X_aY_{a^*}),$$

by the cyclic property of the trace.

Also observe that G_α is a Lie-group. Generally, an element of the Lie-algebra $\mathfrak{g}$ of some Lie-group G can be described by a path $\gamma : (-\epsilon, \epsilon) \to G$ such that $\gamma(0) = e$, so that we identify $\gamma'(0)$ with the associated element of $\mathfrak{g}$. Observe that γ, via the group-action, induces a path in $\mathrm{Rep}(Q,\alpha)$ for any representation X, via $\gamma(t) \cdot X$. Since $\mathrm{Rep}(Q,\alpha)$ is a vector space, which we may identify with its tangent space, we may associate a group action of $\mathfrak{g}$ onto $\mathrm{Rep}(Q,\alpha)$ by $\gamma'(0) \cdot X = \frac{d}{dt}\left(\gamma(t) \cdot X\right)|_{t=0}$.

It is a standard result from differential geometry that the Lie-algebra of $\mathrm{GL}(n)$ is given by $\mathrm{Mat}(n)$, and it is not a strech to accept that the Lie-algebra of G_α is given by End_α/K, where $\mathrm{End}_\alpha = \prod_I \mathrm{Mat}(\alpha_i)$. Observe that here we identify elements that differ additively by some tuple of $\lambda\mathrm{id}$ for some $\lambda \in K$, and not by scalar multiplication. This originates from the fact that the exponential map $\exp : \mathfrak{g} \to G$ turns scalar multiplication into addition. Namely,

$$\exp(x) \sim \exp(x) \cdot \exp(\lambda \mathrm{id}) = \exp(x + \lambda \mathrm{id}), \quad \text{hence} \quad x \sim x + \lambda \mathrm{id}.$$

Performing the above calculation we find that the action of the Lie-algebra is given by

$$(\theta \cdot X)_a = \theta_i X_a - X_a \theta_j,$$

where $a : i \to j$. Now observe that there exists an isomorphism $\mathrm{End}_\alpha^0 \cong (\mathrm{End}_\alpha / K)^*$, where

$$\mathrm{End}_\alpha^0 = \Big\{ \phi \in \mathrm{End}_\alpha \mid \sum_{i \in I} \mathrm{tr}(\phi_i) = 0 \Big\},$$

which is explicitly given by the *trace pairing*

$$\phi \mapsto \Big(\theta \mapsto \sum_I \mathrm{tr}(\phi_i \theta_i) \Big).$$

Indeed, adding $\lambda\mathrm{id}$ to each component of θ leaves $\sum \mathrm{tr}(\phi_i \theta_i)$ invariant, precisely since the traces of the components of ϕ add to 0, so that that map is well-defined.

Definition 6.17. The complex *moment map* $\mu_{\alpha,\mathbb{C}}$ is defined by

$$\mu_{\alpha,\mathbb{C}} : \mathrm{Rep}(Q, \alpha) \to (\mathrm{End}_\alpha / K)^* : X \mapsto \big(\theta \mapsto \frac{1}{2} \omega(X, \theta X) \big).$$

Lemma 6.18. *Using the isomorphism, we may write*

$$\mu_{\alpha,\mathbb{C}} : \mathrm{Rep}(Q, \alpha) \to \mathrm{End}_\alpha^0 : X \mapsto \Big(\sum_{h(a)=i} X_a X_{a^*} - \sum_{t(a)=i} X_{a^*} X_a \Big)_{i \in I}.$$

Proof. We straightforwardly compute

$$\begin{aligned}
\frac{1}{2}\omega(X, \theta X) &= \frac{1}{2} \sum_{a \in Q} \mathrm{tr}(X_{a^*}(\theta X)_a) - \frac{1}{2} \sum_{a \in Q} \mathrm{tr}(X_a (\theta X)_{a^*}) \\
&= \frac{1}{2} \sum_{a \in Q} \mathrm{tr}(X_{a^*}(\theta_{h(a)} X_a - X_a \theta_{t(a)}) - \frac{1}{2} \sum_{a \in Q} \mathrm{tr}(X_a (\theta_{t(a)} X_{a^*} - X_{a^*} \theta_{h(a)})) \\
&= \sum_{a \in Q} \mathrm{tr}(X_a X_{a^*} \theta_{h(a)}) - \sum_{a \in Q} \mathrm{tr}(X_{a^*} X_a \theta_{t(a)}),
\end{aligned}$$

where we used the cyclic property of the trace twice. Now as a final step it follows that by regrouping the terms

$$\frac{1}{2}\omega(X, \theta X) = \sum_I \mathrm{tr}\Big(\Big(\sum_{h(a)=i} X_a X_{a^*} - \sum_{t(a)=i} X_{a^*} X_a \Big) \theta_i \Big),$$

as desired. We remark that the image of $\mu_{\alpha,\mathbb{C}}$ is indeed in End_α^0, since summing over all components yields the sum over all commutators $[X_a, X_{a^*}]$, the trace of which vanishes. □

Remark 6.19. Observe that $(\lambda_1 \mathrm{id}_{\alpha_1}, \lambda_2 \mathrm{id}_{\alpha_2}, \ldots) \in \mathrm{End}^0_\alpha$ precisely when $\langle \lambda, \alpha \rangle = 0$, for the sum of the traces must vanish. Identifying this tuple with $\lambda \in K^I$, we obtain that $\mu^{-1}_{\alpha,\mathbb{C}}(\lambda)$ precisely equals the set of all representations of $\overline{Q}$ that correspond to $\Pi^\lambda(Q)$-modules.

6.3 Kac's Theorem

Given a quiver of finite size, it is not hard to imagine that it is not always possible to find an indecomposable representation with a given dimension vector. Fortunately, it turns out that there is a fairly easy way to check precisely which dimension vectors admit indecomposable representations, which we will briefly outline below.

Definition 6.20. Define the *Ringel form* for a given quiver Q by $\langle ., . \rangle : \mathbb{Z}^I \times \mathbb{Z}^I \to \mathbb{Z}$:

$$\langle \alpha, \beta \rangle = \sum_{i \in I} \alpha_i \beta_i - \sum_{a\,:\,i \to j} \alpha_i \beta_j,$$

and denote the corresponding symmetric bilinear form

$$(\alpha, \beta) = \langle \alpha, \beta \rangle + \langle \beta, \alpha \rangle.$$

Definition 6.21. For any loopfree vertex i, define the *reflection operator* by

$$s_i : \mathbb{Z}^I \to \mathbb{Z}^I : s_i(\alpha) = \alpha - (\alpha, e_i) e_i,$$

where $e_i \in \mathbb{Z}^I$ denotes the i-th standard basis vector. The group generated by the s_i upon function composition is called the *Weyl group.*

Definition 6.22. Define the *fundamental region* by

$$F = \{\alpha \in \mathbb{N}_0^I \mid \alpha \neq 0,\ \alpha \text{ has connected support, and } (\alpha, e_i) \leq 0 \text{ for all } i \in I\}.$$

Then we say that some $\alpha \in \mathbb{Z}^I$ is a *real root* if it is in the orbit of some e_i under the action of the Weyl group, and an *imaginary root* if it is in the orbit of some element from the fundamental region (or its negative) under the same action. If all the entries of some root are non-negative, we call it a *positive root.*

A priori it is not clear why these operations are interesting to study and what they would have to do with indecomposable representations. We shall state but not prove the following theorem that combines these two notions, for its proof is rather advanced.

Theorem 6.23 (Kac's Theorem)**.** *Consider a quiver Q. Then there exists an indecomposable representation of Q of dimension α if and only if α is a positive root. Futhermore, if α is a positive real root, there is a unique indecomposable representation of Q (up to isomorphism) with that dimension.*

Proof. The proof can be found in [12]. □

Example 6.24. Again consider the ADHM-quiver, and consider the vertex k to be the first and n to be the second. We then see that the Ringel form is given by

$$\langle \alpha, \beta \rangle = \alpha_1\beta_1 + \alpha_2\beta_2 - 2\alpha_1\beta_1 - \alpha_1\beta_2 - \alpha_2\beta_1,$$

so that after simplifying,

$$(\alpha, \beta) = 2(\alpha_2\beta_2 - \alpha_1\beta_1 - \alpha_1\beta_2 - \alpha_2\beta_1).$$

Now, since the first vertex has a loop, the only reflection is defined at vertex 2, and we find that

$$s_2 : \begin{pmatrix} k \\ n \end{pmatrix} \mapsto \begin{pmatrix} k \\ n \end{pmatrix} - (2n - 2k)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} k \\ 2k - n \end{pmatrix},$$

when we suggestively write $\alpha = \begin{pmatrix} k & n \end{pmatrix}^{\mathrm{T}}$. Observe that $s_2^2 = \mathrm{id}$, so that the Weyl group is simply given by $\{\mathrm{id}, s_2\}$. This allows us to calculate the real roots,

$$e_1 \xrightarrow{s_2} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \xrightarrow{s_2} e_1 \quad \text{and} \quad e_2 \xrightarrow{s_2} -e_2 \xrightarrow{s_2} e_2,$$

so we conclude that the positive roots are given by e_1, e_2 and $\begin{pmatrix} 1 & 2 \end{pmatrix}^{\mathrm{T}}$. Kac's theorem therefore gives us that there is a unique indecomposable representation of the ADHM-quiver, such that $k = 1$ and $n = 2$. Recall that this case led us to the BPST-instanton. This result tells us that we did not lose any generality in our choices for I and J at the time, so that we constructed the *unique* 1-instanton in an $SU(2)$ gauge theory. Now we proceed to compute the fundamental region, for which the following two constraints must hold:

$$0 \geq -k - n = (\alpha, e_1) \quad \text{and} \quad 0 \geq 2n - 2k = (\alpha, e_2).$$

Since the first condition trivially holds, we conclude that

$$F = \{\alpha \in \mathbb{N}_0^2 \setminus \{(0,0)\} \mid k \geq n\}.$$

Now it is not hard to see from the action of s_2 that the imaginary roots are given by

$$\{\alpha \in \mathbb{N}_0^2 \setminus \{(0,0)\} \mid 2k \geq n\}.$$

Therefore there can only exist indecomposable representations of the ADHM-quiver if $2k \geq n$. Recall section 5.3, in which we discussed the $k = 1$ and $n = 3$ case. We found that our results seemed to split up in a 1-dimensional part that was identically zero, and a 2-dimensional part that was non-zero. From the above it follows that indeed no representation with these dimensions could be indecomposable, so that these results perfectly align. Even more precisely, since

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and since any arrow either starts or ends at the k-vertex, we see that any representation with $k = 0$ must necessarily vanish completely, what explains the zero in the upper-left corner. $\triangle$

It is clear that for small quivers, Kac's theorem provides a highly convenient way of determining the possible dimension vectors of indecomposable representations. Classifying possible dimension vectors of simple modules is a bit trickier, since the criterion the following theorem provides regarding $\Pi^0(Q)$-modules, presented in [12], is a bit more convoluted to check.

Theorem 6.25. *Some $\alpha \in \mathbb{N}_0^I$ is the dimension vector of some simple $\Pi^0(Q)$-module if and only if α is a positive root and for any tuple $(\beta_1, \dots \beta_i) \subset (\mathbb{N}_0^I)^i$ such that $\alpha = \beta_1 + \dots + \beta_i$, it holds that*

$$1 - \langle \alpha, \alpha \rangle \geq i - \langle \beta_1, \beta_1 \rangle - \dots - \langle \beta_i, \beta_i \rangle.$$

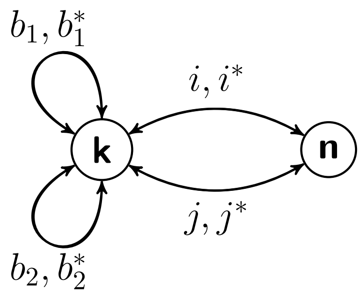
We will conclude our brief discussion on decomposability and simpleness and advance to the final topic; the justification of the second ADHM-equation, and we will derive some very useful properties of representations of quivers in the process.

6.4 Semisimple representations and the real moment map

Naively, one could try to explain the second ADHM-equation as follows. One could consider the double of the ADHM-quiver Q, depicted below, and start to work out what the constraints for a representation being a $\Pi^0(Q)$-module are. The first vertex then yields the constraint

$$B_1 B_1^* + B_2 B_2^* + II^* - B_1^* B_1 - B_2^* B_2 - J^* J = 0, \quad \text{or} \quad [B_1, B_1^*] + [B_2, B_2^*] + II^* - J^* J = 0.$$

By severe abuse of notation, interpreting the $*$ as hermitian transposition $\dagger$, this is precisely the second ADHM-equation. Of course there is a much more satisfying reason as to why the hermitian transpose appears in the second ADHM-equation. To this end, we will delve a little deeper into the theory. We remark that in this section we will restrict our focus to $K = \mathbb{C}$, since that is the sole field over which the ADHM-construction makes sense.



Definition 6.26. A representation of a quiver Q is called *semisimple* if its corresponding KQ-module is semisimple; that is, if it can be written as the direct sum of simple submodules. Define gr X to be the direct sum of the composition factors of X, which is then obviously semisimple.

Recall that the group GL_α acts on representations of a given quiver.

Proposition 6.27. *Consider the space of representations corresponding to* $\Pi^0(Q)$*-modules and let* X *be such a representation of dimension vector* α*. Then the closure of any orbit* $O(X)$ *contains a unique closed orbit,* $O(\mathrm{gr}\ X)$*. As a result, the closed orbits under the action of* GL_α *are precisely those of semisimple* $\Pi^0(Q)$*-representations.*

Proof. The proof can be found in [12]. □

It follows that studying the semisimple representations of some quiver is equivalent to studying the closed orbits under the action of GL_α. Fortunately, there is a widely celebrated result that characterizes precisely these closed orbits. To formulate this result, we will first need another moment map that is different from the complex moment map discussed before, called the *real moment map*.

We will define an inner product on the space of representations of a given quiver Q with fixed dimension vector α, by identiying all vector spaces $V_i \cong \mathbb{C}^{\alpha_i}$, and equipping all these vector spaces with their standard hermitian inner product. For all $a : i \to j$, this induces a natural inner product on $\mathrm{Hom}(V_i, V_j)$ by $\langle X_a, Y_a\rangle = \mathrm{tr}(X_a Y_a^\dagger)$, since $Y_a^\dagger$ is the adjoint map of Y_a in the standard inner product. Hence we may define

$$\langle X, Y\rangle = \sum_{a\in Q} \mathrm{tr}(X_a Y_a^\dagger),$$

for any two such representations of Q.

Lemma 6.28. *The maximal compact subgroup of* GL_α *that preserves the inner product is* U_α*, which is the subgroup of* GL_α *consisting of unitary matrices in each component.*

Proof. Observe that

$$\mathrm{tr}\big(UX_aV(UY_aV)^\dagger\big) = \mathrm{tr}\big(UX_a(VV^{-1})Y_a^\dagger U^{-1}\big) = \mathrm{tr}(X_aY_a^\dagger),$$

using the cyclicity of the trace. It is not hard to show that there are no other matrices leaving the trace invariant upon conjugation. Lastly, U_α is compact since it is a bounded set, for all its elements have unit norm, and also clearly closed in a finite dimensional space; hence the Heine-Borel theorem applies. □

Let $\mathfrak{k}$ denote the Lie-algebra of U_α. Then we have, just like before, that the action of the Lie-algebra is given by $(AX)_a = A_iX_a - X_aA_j$ for $a : i \to j$.

Definition 6.29. Observe that $\omega(X, Y) = 2\langle Y, X\rangle$ is a symplectic form. We then define the *real moment map*

$$\mu_{\alpha,\mathbb{R}} : \mathrm{Rep}(Q, \alpha) \to \mathfrak{k}^* : X \mapsto (A \mapsto \langle AX, X\rangle).$$

By a similar computation as in proving lemma 6.18 we may also write this as

$$\mu_{\alpha,\mathbb{R}}(X)(A) = \sum_{i\in I} \mathrm{tr}\Bigg(A_i\Bigg[\sum_{h(a)=i} X_aX_a^\dagger - \sum_{t(a)=i} X_a^\dagger X_a\Bigg]\Bigg).$$

We are now ready to formulate the connecting theorem, of which we state only a very special case, since it holds much more generally for group actions on vector spaces.

Theorem 6.30 (Kempf-Ness)**.** *The set $\mu_{\alpha,\mathbb{R}}^{-1}(0)$ intersects precisely those GL_α-orbits which are closed, and it does so in precisely one U_α orbit.*

Proof. The more general proof can be found in [14]. $\square$

Combining this theorem with proposition 6.27, we can state our final result.

Corollary 6.31. *For any semisimple $\Pi^0(Q)$-representation X, there exists some $g \in \mathrm{GL}_\alpha$ such that gX satisfies $\mu_{\alpha,\mathbb{R}}(gX) = 0$. Since the image of any $A \in \mathfrak{k}$ must vanish, we find that*

$$\sum_{h(a)=i} (gX)_a (gX)_a^\dagger - \sum_{t(a)=i} (gX)_a^\dagger (gX)_a = 0$$

must hold for all $i \in I$, and multiplication with some element from U_α will not change the validity of the above equation; hence it is U_α invariant.

Applying this to the ADHM-quiver, we see that the condition $\mu_{\alpha,\mathbb{R}}(X) = 0$ for a representation X gives us for the vertex k that

$$B_1B_1^\dagger + B_2B_2^\dagger + II^\dagger - B_1^\dagger B_1 - B_2^\dagger B_2 - J^\dagger J = 0, \quad \text{or} \quad [B_1, B_1^\dagger] + [B_2, B_2^\dagger] + II^\dagger - J^\dagger J = 0,$$

which is in fact precisely the second ADHM-equation. Again, the constraint we get for the other vertex, $JJ^\dagger - I^\dagger I = 0$, need not necessarily hold for a set of ADHM-data, as the BPST-instanton shows.

In summary, Atiyah, Drinfeld, Hitchin and Manin were interested in representations of the ADHM-quiver, and it turns out that focussing on semisimple $\Pi^0(Q)$-representations was sufficient to construct anti-self-dual connections. As a result, we can most succinctly write the ADHM-equations as

$$\mu_{\alpha,\mathbb{C}}(X) = 0 \quad \text{and} \quad \mu_{\alpha,\mathbb{R}}(X) = 0,$$

for some representation $X = (B_1, B_2, I, J)$. Unanswered questions remain as to why the ADHM-equations only focus on the constraints imposed by the above two equations at the k-vertex, and ignore those on the n-vertex. In addition, precisely what the connection is between representations of quivers and the hidden mechanisms that eventually provide the anti-self-dual connection, is beyond the scope of this thesis, though hopefully this concluding section did provide some insight as to where the ADHM-equations originated from, and what kind of theoretical concepts formed the basis of the theory behind the celebrated construction of instantons.

7 Conclusion

We have seen how the mathematical theory of differential geometry proved to be imperative to understanding physical gauge theories formally. What started out as a desire to differentiate sections, naturally brought us to the concept of a connection D and its accompanied curvature F. This allowed us to extend the notion of differential forms to vector potentials A, which classified all the different connections on a given vector bundle. This led us to the exterior covariant derivative of $\mathrm{End}(E)$-valued forms, which shared many useful properties with the ordinary exterior derivative.

Classical electromagnetism turned out to be the simplest case of a gauge-theory, and we analysed Maxwell's equations in the previously developed language. We introduced G-bundles for a gauge group G and defined the Yang-Mills action, which required a metric dependent operator called the Hodge-star operator $\star$, and the notion of the trace of an $\mathrm{End}(E)$-valued form. The convergence of this action resulted in the notion of the topological charge of a vector bundle over Euclidean space, and it turned out that the curvature provided a useful way of computing this invariant. We took a step back to investigate the simpler case of the magnetic monopole, and we found another topological invariant, what led us to more general Chern classes and forms. Continuing, we used the invariance of the topological charge to prove that instantons minimize the Yang-Mills action, and are therefore of particular interest.

The ADHM-construction made use of the algebra of quaternions, which we introduced and related via the Pauli-matrices to a group consisting of 2×2-matrices. We stated the ADHM-data and rigorously walked through the entire construction to see that the outcome of the ADHM-construction was indeed an anti-self-dual curvature, making it an instanton.

We continued to explicitly calculate instantons, varying the instanton number and the dimension of the gauge group. We found that it is not possible for either I or J to be zero, and the simplest non-trivial case resulted in the BPST-instanton, which is indeed anti-self-dual and has unit topological charge. We performed a big calculation to explore the $SU(3)$-case to find a different 1-instanton, but it turned out that we only got a non-zero 2×2 minor and we saw that this would happen for any $n > 2$. Subsequently, we motivated that finding explicit solutions to the ADHM-equations is a tough job, since our first attempt at constructing a 2-instanton did not allow for many entries of the ADHM-data to be set to zero. We cunningly found a family of instantons of arbitrary topological charge by using block matrices, and concluded by giving the expressions for the 't Hooft instantons, which are instantons of arbitrary topological charge in an $SU(2)$-theory.

Lastly, we explored quivers and their representations. We saw that such a representation could be viewed as a module over the path algebra, and we restricted our view to modules over the preprojective algebra, which gave us the first ADHM-equation. Kac's theorem allowed us to understand why the big calculation in the previous chapter yielded so many zeroes, and lastly we constructed both a complex and a real moment map, the zeroes of which via the Kempf-Ness theorem provided us with both the ADHM-equations.

Unexplored topics are the applications of instantons to quantum tunneling, as alluded to in the introduction, which can be found more in-depth in [10]. Also the underlying mechanism that ties together the moment maps and the success of the ADHM-construction is still a grey area, as a priori these notions appear unrelated. It could be an interesting topic of study to unearth this connection and to fully understand what the four gentlemen the ADHM-construction is named after were thinking about before publishing their renowned paper, which inspired hundreds, if not thousands of others to delve deeper into the study of Yang-Mills theory.

Bibliography

[1] Baez, John; Muniain, Javier P. "Gauge Fields, Knots and Gravity" (1994): 212-283.

[2] Lindenhovius, AJ. "Instantons and the ADHM-construction" (2011): 65-92

[3] Antanasov, Alex. "Instantons and the ADHM-construction" (2016): 1-10

[4] de Groot, Jim. "Lecture notes of differential geometry" (2014): 12-16

[5] Atiyah, M.F. "Geometry of Yang-Mills Fields" (1979): 12-33

[6] Griffiths, David. J. "Introduction to Quantum Mechanics, Second Edition" (2014): 387-393

[7] Christ, Norman H.; Weinberg, Erick J.; Stanton, Nancy K. "General self-dual Yang-Mills solutions" (1978): 1-8

[8] Liu, Zheng-Wen; "An introduction to Chern-Simons Theory" (2014): 1-25

[9] Tong, David; "TASI Lectures on Solitons" (2005): 2-7, 16-24

[10] Coleman, Sidney; "The uses of instantons" (1977): 1-47

[11] Le Bruyn, Lieven; "Noncommutative Geometry and Cayley-smooth orders" (2008): 1-15

[12] Crawley-Boevey, William; "DMV Lectures on Representations of quivers, preprojective algebras and deformations of quotient singularities" (2000): 3-31

[13] Haiden, Fabian; "Course Notes for Math 256: Dynamics, Stability, and Noncommutative Algebra" (2015): 51-65

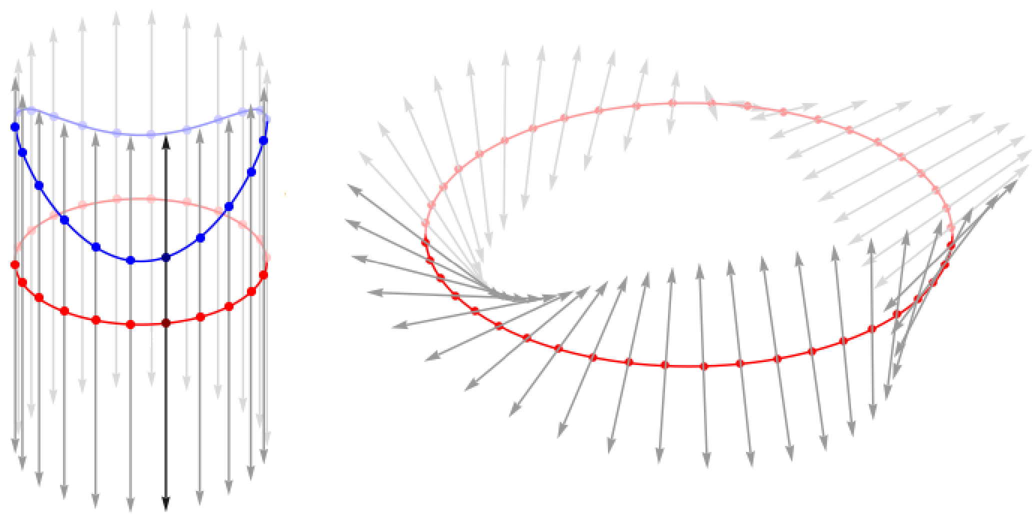
[14] King. A. D. "Moduli of representations of finite dimensional algebras" (1993): 1-16

[15] Egan, Greg; "Orthogonal: Geometry and Waves [Extra]" (2011)

Populaire samenvatting

We leven allemaal in dezelfde 3-dimensionale *ruimte*; het universum. Je kunt erin in drie richtingen bewegen: omhoog en omlaag, naar links en naar rechts, en naar voren en naar achteren; vandaar dat we het universum 3-dimensionaal noemen. Het oppervlak van een tafel of een bureau is een plat *vlak*. Met je pen op het papier kun je precies twee kanten uit; dit is dus een 2-dimensionaal oppervlak. Natuurlijk zijn niet alle 2-dimensionale oppervlakken *plat*; denk maar aan het oppervlak van een bol, of dat van een donut. Dit soort gladde oppervlakken noemen wiskundigen 2-dimensionale *manifolds*. Er bestaan manifolds van alle mogelijke dimensies; hoewel dimensies hoger dan 3 niet meer voor te stellen zijn, kunnen wiskundigen er nog zonder problemen mee werken. De meest eenvoudige manifolds zijn 1-dimensionale manifolds; dit zijn simpelweg gekromde lijntjes.

Een van de interessante dingen die je met zo'n manifold kunt doen, is aan elk punt van je manifold iets vastplakken, bijvoorbeeld een lijn. Twee mogelijke manieren om dat te doen bij het 1-dimensionale manifold de *cirkel* zijn weergegeven in onderstaande figuur. Deze twee mogelijke manieren zijn wezenlijk verschillend van elkaar; in de linkerfiguur



Figuur: Links is een *triviale* lijnbundel over de cirkel weergegeven en rechts een niet-triviale, genaamd de Möbiusbundel. De blauwe lijn in de linkerfiguur is de grafiek van een sectie. Bron: [15]

staan alle lijntjes in dezelfde richting, terwijl de rechterfiguur een soort *draai* heeft. Het vastplakken van een lijn aan elk punt heet het maken van een *lijnbundel*. Natuurlijk kunnen we ook vlakken of zelfs hogerdimensionale ruimtes vastplakken aan elk punt. In het algemeen heet het resultaat dan een *vectorbundel*. Wiskundigen zijn erin geïnteresseerd op welke manieren je vectorbundels kunt maken gegeven een bepaald manifold.

Normaliter pakt een *functie* een getal, past daar een formule op toe en geeft dan weer een getal; bijvoorbeeld $f(x) = x^2$ of $f(x) = \sin(x)$. Gegeven zo'n vectorbundel over een manifold is het mogelijk om een ander soort functie te maken. Deze pakt niet een getal maar een punt op ons manifold en geeft dan als output een punt in de ruimte die we aan dat punt hadden geplakt. Een voorbeeld is te zien als de blauwe lijn in de vorige figuur. Zo'n functie noemt men een *sectie*.

Herinner dat de *afgeleide* van een gewone functie in elk punt op de grafiek van die functie de helling van de grafiek aangeeft. Van gewone functies is het vaak mogelijk om de afgeleide te berekenen, maar het is niet heel duidelijk of dat voor secties ook altijd mogelijk is. Het blijkt dat dat wel het geval is, en een operator die een sectie pakt en ons dan zijn afgeleide geeft, heet een *connectie*. Aan de hand van een connectie kunnen we een andere operator definiëren, genaamd de *kromming*. De waarheid wordt wat geweld aangedaan als we deze kromming vergelijken met het soort kromming dat een plat vlak als een tafeloppervlak onderscheidt van een gekromd oppervlak als van een bol, maar zo kan men erover denken.

De manier om aan een sectie een afgeleide toe te kennen is helaas niet zo uniek als de afgeleide vinden van een gewone functie; de connectie is dus niet uniek. Daarom is de bijbehorende kromming ook niet uniek. Sommige krommingen hebben speciale eigenschappen. Een daarvan is het *anti-zelf-duaal* zijn, en dit soort krommingen is voor natuurkundige doeleinden erg interessant. Derhalve hebben ze een speciale naam gekregen: *instantons*. Deze kunnen worden opgevat als oplossingen van zekere natuurkundige bewegingsvergelijkingen die voorkomen bij het *quantum-tunneling effect*; dit is het effect dat bepaalde deeltjes door kleine barrières kunnen tunnelen; dat wil zeggen, door de barrière heen gaan terwijl ze eigenlijk geblokkeerd zouden moeten zijn.

Natuurkundigen raakten derhalve geïnteresseerd in het construeren van deze instantons en voor enige tijd was dit een lastig probleem. Totdat vier heren, genaamd Michael Atiyah, Vladimir Drinfeld, Nigel Hitchin en Yuri I. Manin, een methode vonden om instantons expliciet uit te rekenen. Verrassend genoeg was de berekening in theorie erg eenvoudig; in de praktijk echter bleek het nog knap wat werk. Deze constructie gaat onder de naam de ADHM-constructie.

Deze thesis behandelt op betrekkelijk formele wijze de wiskundige en natuurkundige theorie achter de connecties, krommingen en instantons. Ook behandelt het de ADHM-constructie in detail en rekent het een aantal instantons middels deze constructie uit. Ten slotte motiveren we de ideeën achter de constructie vanuit een wiskundige invalshoek.