Electromagnetic Duality for Children

JM Figueroa-O'Farrill
J.M.Figueroa@ed.ac.uk

Version of 8 October 1998

Contents

I The Simplest Example: $SO(3)$ **11**

1 Classical Electromagnetic Duality **12**

1.1 The Dirac Monopole . . . 12

1.1.1 And in the beginning there was Maxwell... . . . 12

1.1.2 The Dirac quantisation condition . . . 14

1.1.3 Dyons and the Zwanziger–Schwinger quantisation condition . . . 16

1.2 The 't Hooft–Polyakov Monopole . . . 18

1.2.1 The bosonic part of the Georgi–Glashow model . . . 18

1.2.2 Finite-energy solutions: the 't Hooft–Polyakov Ansatz . 20

1.2.3 The topological origin of the magnetic charge . . . 24

1.3 BPS-monopoles . . . 26

1.3.1 Estimating the mass of a monopole: the Bogomol'nyi bound . . . 27

1.3.2 Saturating the bound: the BPS-monopole . . . 28

1.4 Duality conjectures . . . 30

1.4.1 The Montonen–Olive conjecture . . . 30

1.4.2 The Witten effect . . . 31

1.4.3 $SL(2,\mathbb{Z})$ duality . . . 33

2 Supersymmetry **39**

2.1 The super-Poincaré algebra in four dimensions . . . 40

2.1.1 Some notational remarks about spinors . . . 40

2.1.2 The Coleman–Mandula and Haag–Łopuszański–Sohnius theorems . . . 42

2.2 Unitary representations of the supersymmetry algebra . . . 44

2.2.1 Wigner's method and the little group . . . 44

2.2.2 Massless representations . . . 45

2.2.3 Massive representations . . . 47

No central charges . . . 48

Adding central charges . . . 49

2.3 N=2 Supersymmetric Yang-Mills 51
2.3.1 N=1 d=6 supersymmetric Yang–Mills 52
2.3.2 From N=1 in d=6 to N=2 in d=4 57
2.3.3 Higgsed N=2 supersymmetric Yang–Mills 59
2.3.4 N=2 avatar of the BPS-monopole 60
2.3.5 The supersymmetry bound is the Bogomol'nyi bound . 62
The supersymmetry algebra in six dimensions 63
The supersymmetry algebra in four dimensions 64
2.4 N=4 Supersymmetric Yang-Mills 66
2.4.1 N=1 d=10 supersymmetric Yang–Mills 67
2.4.2 Reduction to d=4: N=4 supersymmetric Yang–Mills . 69
2.4.3 Monopoles and gauge bosons in N=4 supersymmetric Yang–Mills 74
2.4.4 The mass bound for N=4 super Yang–Mills 76
The supersymmetry algebra in ten dimensions 77
The supersymmetry algebra in four dimensions 78

3 Collective Coordinates **80**
3.1 The metric on the moduli space 81
3.1.1 The physical configuration space 81
3.1.2 The metric on the physical configuration space 82
3.1.3 The metric on the moduli space 84
3.1.4 The 1-monopole moduli space 85
3.1.5 The quantisation of the effective action 87
3.1.6 Some general properties of the monopole moduli space 88
3.2 $\dim \mathcal{M}_k = 4k$ 89
3.2.1 The dimension as an index 90
3.2.2 Computing the index of $\mathcal{D}$ 92
3.2.3 Computing the current $J_i(x,x)$ 96
3.3 A quick motivation of hyperkähler geometry 99
3.3.1 Riemannian geometry 100
3.3.2 Kähler geometry 103
3.3.3 Ricci flatness 107
3.3.4 Hyperkähler geometry 109
3.4 $\mathcal{M}_k$ is hyperkähler 111
3.4.1 Symplectic quotients 111
3.4.2 Kähler quotients 113
3.4.3 Hyperkähler quotients 116
3.4.4 $\mathcal{M}_k$ as a hyperkähler quotient 116
3.4.5 Another proof that $\mathcal{M}_k$ is hyperkähler 119

4 The Effective Action for N=2 Supersymmetric Yang–Mills 122
4.1 Fermionic collective coordinates 123
4.1.1 Computing the index 124
4.1.2 Using supersymmetry 125
4.2 The effective action 127
4.3 N=4 supersymmetry of the effective action 130
4.3.1 N=4 supersymmetry in $\mathbb{R}^4$: a toy model 131
4.3.2 N=4 supersymmetry in hyperkähler manifolds 132
4.3.3 N=4 supersymmetry of L_{eff} 134
4.4 A brief review of harmonic theory 136
4.4.1 Harmonic theory for riemannian manifolds 136
4.4.2 Harmonic theory for Kähler manifolds 139
4.4.3 Explicit formulas for $\bar\partial$ and $\bar\partial^*$ 142
4.5 Quantisation of the effective action 144
4.5.1 Canonical analysis 144
4.5.2 The quantisation of the effective hamiltonian 146

5 The Effective Action for N=4 Supersymmetric Yang–Mills 149
5.1 Fermionic collective coordinates 150

II Arbitrary Gauge Groups 153

6 Monopoles for Arbitrary Gauge Groups 154
6.1 Topologically stable solutions 154
6.1.1 Some elements of homotopy 156
The fundamental group 157
Higher homotopy groups 162
6.1.2 Homotopy classification of finite-energy configurations . 166
Adding topological charges 171
6.2 The Dirac quantisation condition 174
6.3 Some facts about compact Lie groups and Lie algebras 177
6.3.1 Compact Lie groups 177
6.3.2 The Weyl group 180
6.3.3 Root systems and simple Lie algebras 182
Reconstructing the group 184
The centre of $\widetilde{G}$ 186
An example: $A_3 = D_3$ 188
Another example: D_4 189
All the connected compact simple Lie groups 190
6.3.4 Some simple examples 192

The simple root system A_2 192
The simple root systems $B_2 = C_2$ 194
The simple root system G_2 195
6.4 The magnetic dual of a compact Lie group 195
6.4.1 Some lattices and dual groups 197
H abelian 198
H simple 198
An example: *Spin*(8) and its quotients 200
Another example: *Spin*(12) and its quotients 202

List of Exercises

1.1 Angular momentum due to the electromagnetic field 17
1.2 The Zwanziger–Schwinger quantisation condition 17
1.3 Dyonic spectrum in CP non-violating theories 17
1.4 The spectrum of the model 19
1.5 Boundary conditions on H and K 22
1.6 The equations of motion for H and K 22
1.7 Asymptotic form of the electromagnetic field 23
1.8 Gauge field-strength in the Higgs vacuum 25
1.9 Additivity of the magnetic charge g_Σ 26
1.10 Dirac quantisation condition revisited 26
1.11 The Bogomol'nyi equation implies (1.10) 29
1.12 The BPS-monopole 29
1.13 The mass density at the origin is finite 29
1.14 $(P)SL(2,\mathbb{Z})$ and its action on the upper half-plane 34
1.15 $SL(2,\mathbb{Z})$-invariance of the mass formula 35
1.16 Orbifold points in the fundamental domain D 36
1.17 Properties of the mass matrix $A(\tau)$ 37

2.1 The little groups for positive-energy particles 45
2.2 Helicity content of massless multiplets 46
2.3 Highest spin in the multiplet 48
2.4 Massive N=2 multiplets with s=0 and s=1/2 48
2.5 Short N=2 multiplets with s=0 and s=1/2 50
2.6 Short N=4 multiplets with s=0 51
2.7 N=1 supersymmetric Yang–Mills 52
2.8 Supersymmetry variation of $\mathcal{L}$ 54
2.9 A Fierz rearrangement 55
2.10 Some Γ-matrix identities 56
2.11 ... and the proof of supersymmetry invariance 56
2.12 The supersymmetry multiplet 56
2.13 $\mathcal{L}$ in a Majorana basis 57

2.14 Explicit N=2 supersymmetry transformations 58
2.15 The $SO(2)$ Noether current 59
2.16 The perturbative spectrum of the model 59
2.17 Supersymmetric BPS-monopoles 61
2.18 More γ-matrix identities 61
2.19 Some euclidean γ-matrices 61
2.20 BPS-monopoles break one half of the supersymmetry 62
2.21 Supersymmetric variation of the supercurrent 63
2.22 The symmetric gauge-invariant energy momentum tensor . . . 64
2.23 The topological current 64
2.24 The "momenta" in the extra dimensions 64
2.25 The effective electromagnetic field strength 65
2.26 The space components of the topological charge 66
2.27 The Majorana condition 67
2.28 Properties of Majorana and Weyl fermions 68
2.29 Varying the lagrangian density 68
2.30 A ten-dimensional Fierz identity 69
2.31 Some more Γ-matrix identities 69
2.32 $so(4) \cong so(3) \times so(3)$ explicitly 70
2.33 Explicit realisation for α^I and β^J 71
2.34 The fundamental representation of $su(4)$ 71
2.35 The fermionic terms in the lagrangian 73
2.36 N=4 supersymmetry transformations 73
2.37 $su(4)$ invariance 73
2.38 The perturbative spectrum after higgsing 75
2.39 The supersymmetry algebra in ten dimensions 77
2.40 Another topological current 78
2.41 A self-dual 5-form 78
2.42 Momentum and topological charge in this background 79

3.1 The BPS-monopole as an instanton 82
3.2 D^2 has no normalisable zero modes 85
3.3 The electric charge 87
3.4 Two equations in one 90
3.5 $\mathcal{D}^\dagger$ has no normalisable zero modes 91
3.6 A formula for the index of $\mathcal{D}$ 92
3.7 Another formula for $I(M^2)$ 92
3.8 Some properties of kernels 93
3.9 Regularity properties of the propagator 94
3.10 The first term doesn't contribute 97
3.11 Another expression for the degree of the map $\boldsymbol{\phi}$ 99

3.12 A coordinate-free expression for ∇ . . . 100
3.13 The holonomy group of a connected manifold . . . 102
3.14 The holonomy group of a riemannian manifold . . . 102
3.15 The holonomy algebra of a riemannian manifold . . . 103
3.16 The Nijenhuis tensor . . . 104
3.17 The holonomy group of a complex connection . . . 104
3.18 The unitary group . . . 104
3.19 The Kähler form . . . 105
3.20 Kähler is Kähler is Kähler . . . 105
3.21 The determinant of $\mathbb{P}_\gamma$ . . . 107
3.22 The Ricci tensor is symmetric . . . 108
3.23 Curvature tensors in Kähler manifolds . . . 108
3.24 An equivalent expression for the Ricci curvature . . . 109
3.25 Hyperkähler implies integrability . . . 110
3.26 The holonomy group of a hyperkähler manifold . . . 110
3.27 The Lie derivative acting on forms . . . 111
3.28 The Poisson bracket . . . 112
3.29 Triviality of the normal bundle . . . 113
3.30 Integrability of the restricted complex structure . . . 115
3.31 O'Neill's formula . . . 115
3.32 Hyperkähler structure of $\mathbb{R}^4$ . . . 117
3.33 Equivariance of the moment mapping . . . 118
3.34 The Yang–Mills–Higgs effective action . . . 119
3.35 A somewhat more explicit formula for G_{ab} . . . 120
3.36 The Christoffel symbols . . . 120
3.37 Explicit expressions for the complex structures . . . 121

4.1 Fermion zero modes are chiral . . . 124
4.2 From bosonic to fermionic zero modes . . . 125
4.3 ... and back . . . 125
4.4 A complex structure . . . 126
4.5 iA lives on the sphere . . . 126
4.6 Adjoint transformation . . . 127
4.7 Computing $\boldsymbol{\psi}^\dagger \times \boldsymbol{\psi}$ . . . 128
4.8 The first kinetic term . . . 129
4.9 ... and the second kinetic term . . . 130
4.10 N=2 supersymmetry in flat space . . . 131
4.11 The N=4 supersymmetry algebra . . . 132
4.12 Another proof that $\delta^2\zeta^a = i\dot{\zeta}^a$ . . . 133
4.13 Commuting supersymmetries . . . 134
4.14 The supersymmetry charges . . . 135

4.15 The square of the Hodge $\star$ . . . 137
4.16 The pointwise metric on forms . . . 138
4.17 The adjoint de Rham operator . . . 138
4.18 Harmonic forms . . . 138
4.19 Poincaré duality . . . 139
4.20 The Hodge $\star$ and the complex structure . . . 140
4.21 The pointwise hermitian metric . . . 141
4.22 Serre duality . . . 142
4.23 An explicit expression for d . . . 142
4.24 An explicit expression for d^* . . . 143
4.25 The canonical momenta . . . 145
4.26 The effective hamiltonian . . . 145
4.27 The supersymmetry charges revisited . . . 145
4.28 Some checks . . . 146
4.29 Classical N=4 supersymmetry algebra . . . 146
4.30 Q^* is $\bar\partial^*$ . . . 148

5.1 Euclidean Clifford algebra . . . 150
5.2 $\bar\gamma_{\underline{i}}$ and the Majorana condition . . . 150
5.3 The charge conjugation matrix explicitly . . . 151
5.4 "Majorana–Weyl" spinors . . . 152
5.5 Counting zero modes . . . 152

6.1 Invariant metrics exist . . . 154
6.2 The temporal gauge . . . 155
6.3 $\Phi(S^2_\infty) \subset M_0$ lies in a single orbit . . . 156
6.4 $\pi_1(Y, y_0)$ is a group . . . 157
6.5 $\pi_1(Y, y_0) \cong \pi(Y, y_1)$ . . . 159
6.6 $SU(2)$ and $SO(3)$ . . . 161
6.7 $\pi_1(G)$ is abelian . . . 162
6.8 $\pi_2(X, x_0)$ is a group . . . 163
6.9 $\pi_2(X, x_0) \cong \pi_2(X, x_1)$ . . . 164
6.10 Gauge invariance of the homotopy class . . . 166
6.11 Solving for Φ . . . 168
6.12 The magnetic charge is covariantly constant . . . 174
6.13 The non-abelian Stoke's theorem . . . 175
6.14 The centre is the intersection of all the maximal tori . . . 186
6.15 The central and root lattices are dual . . . 187
6.16 Some facts about lattices . . . 187
6.17 The dual fundamental weights . . . 199

List of Figures

1.1 Fundamental domain (shaded) for the action of $PSL(2,\mathbb{Z})$ on the upper half plane, and some of its $PSL(2,\mathbb{Z})$ images. 36
1.2 Dyonic spectrum predicted by $SL(2,\mathbb{Z})$ duality. Dots indicate dyons, crosses indicates holes in the dyonic spectrum. Only dyons with non-negative magnetic charge are shown. 38

6.1 The two possible lifts to S^n of a loop in $\mathbb{RP}^n$. 161
6.2 A map $S^2 \to G/H$ as a loop of loops. 168
6.3 A monopole. 172
6.4 Patching two monopoles. 173
6.5 The root system A_2. 193
6.6 The root system B_2. 194
6.7 The root system G_2. 196

List of Tables

1.1 The perturbative spectrum after higgsing. 20

6.1 Allowed angles between roots in a root system. 183
6.2 Dynkin diagrams of the simple root systems. 185
6.3 Simple root systems, their Weyl groups, their complex Lie algebras, compact Lie groups and their centres. 191
6.4 The connected compact simple Lie groups and their duals. . . 200

Part I

The Simplest Example: $SO(3)$

Chapter 1

Classical Electromagnetic Duality

In this chapter we treat classical electromagnetic duality, and its manifestation (Montonen–Olive duality) in some spontaneously broken gauge theories. We start by reviewing the Dirac monopole, and then quickly move on to the 't Hooft–Polyakov monopole solution in the model described by the bosonic part of the $SO(3)$ Georgi–Glashow model. We focus on the monopole solution in the Prasad–Sommerfield limit and derive the Bogomol'nyi bound for the mass of the monopole. We show that the classical spectrum of the model is invariant under electromagnetic duality. This leads to the conjecture of Montonen and Olive. We then discuss "the Witten effect" and show that the $\mathbb{Z}_2$ electromagnetic duality extends to an $SL(2,\mathbb{Z})$ duality.

1.1 The Dirac Monopole

In this section we discuss the Dirac monopole and the Dirac(–Zwanziger–Schwinger) quantisation condition in the light of classical electromagnetic duality.

1.1.1 And in the beginning there was Maxwell...

Maxwell's equations in vacuo, given by

$$\begin{aligned} \vec{\boldsymbol{\partial}} \cdot \vec{\boldsymbol{E}} &= 0 & \qquad \vec{\boldsymbol{\partial}} \cdot \vec{\boldsymbol{B}} &= 0 \\ \vec{\boldsymbol{\partial}} \times \vec{\boldsymbol{E}} &= -\frac{\partial \vec{\boldsymbol{B}}}{\partial t} & \qquad \vec{\boldsymbol{\partial}} \times \vec{\boldsymbol{B}} &= \frac{\partial \vec{\boldsymbol{E}}}{\partial t} \end{aligned} \tag{1.1}$$

are highly symmetric. In fact, they are invariant under both Lorentz transformations (in fact, conformal) and under *electromagnetic duality*:

$$(\vec{\boldsymbol{E}}, \vec{\boldsymbol{B}}) \mapsto (\vec{\boldsymbol{B}}, -\vec{\boldsymbol{E}}) \; . \tag{1.2}$$

Lorentz invariance can be made manifest by introducing the field-strength $F_{\mu\nu}$, defined by[1]

$$F^{0i} = -F^{i0} = -E^i \qquad F^{ij} = -\epsilon_{ijk} B^k \; .$$

In terms of $F_{\mu\nu}$, Maxwell's equations (1.1) become

$$\partial_\nu F^{\mu\nu} = 0 \qquad \partial_\nu {}^\star F^{\mu\nu} = 0 \; , \tag{1.3}$$

where

$${}^\star F^{\mu\nu} = \tfrac{1}{2}\, \epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}$$

with $\epsilon^{0123} = +1$. This formulation has the added virtue that the duality transformation (1.2) is simply

$$F^{\mu\nu} \mapsto {}^\star F^{\mu\nu} \qquad {}^\star F^{\mu\nu} \mapsto -F^{\mu\nu} \; , \tag{1.4}$$

where the sign in the second equation is due to the fact that in Minkowski space $\star^2 = -1$.

In Minkowski space $\partial_\nu {}^\star F^{\mu\nu} = 0$ implies that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, for some electromagnetic potential A_μ. Similarly, $\partial_\nu F^{\mu\nu} = 0$ implies that ${}^\star F_{\mu\nu} = \partial_\mu \tilde{A}_\nu - \partial_\nu \tilde{A}_\mu$, for some dual electromagnetic field $\tilde{A}_\mu$. Notice however that the duality transformation relating A_μ and $\tilde{A}_\mu$ is nonlocal. It may be easier to visualise the following two-dimensional analogue, where the duality transformation relates functions ϕ and $\tilde{\phi}$ which satisfy $\epsilon_{\alpha\beta}\partial^\beta\phi = \partial_\alpha\tilde{\phi}$, where α, β take the values 0 and 1 now.

In the presence of sources, duality is preserved provided that we include both electric *and* magnetic sources:

$$\partial_\nu F^{\mu\nu} = j^\nu \qquad \partial_\nu {}^\star F^{\mu\nu} = k^\mu \; ,$$

and that we supplement the duality transformations (1.4) by a similar transformation of the sources:

$$j^\mu \mapsto k^\mu \qquad k^\mu \mapsto -j^\mu \; .$$

[1] In these lectures, we shall pretend to live in Minkowski space with signature $(+---)$. We will set $c = 1$ but will often keep $\hbar$ explicit.

A charged point-particle of in the presence of an electromagnetic field behaves according to the Lorentz force law. If the particle is also magnetically charge, the Lorentz law is then given by

$$m\,\frac{d^2x^\mu}{d\tau^2} = \left(qF^{\mu\nu} + g\,{}^\star F^{\mu\nu}\right)\frac{dx_\nu}{d\tau}$$

where τ is the proper time, and m, q and g are the mass, the electric and magnetic charges, respectively. This formula is also invariant under duality provided we interchange the electric and magnetic charges of the particle: $(q,g)\mapsto(g,-q)$.

Problem: Derive the above force law from a particle action.

Notice that in the presence of magnetic sources, $\partial_\nu {}^\star F^{\mu\nu} \neq 0$ whence there is no electromagnetic potential A_μ. Nevertheless if at any given moment in time, the magnetic sources are localised in space, one may define A_μ in those regions where $k^\mu = 0$. The topology of such regions is generically nontrivial and therefore a nonsingular A_μ need not exist throughout. Instead one solves for A_μ locally, any two solutions being related, in their common domain of definition, by a gauge transformation. We will see this explicitly for the magnetic monopole.

1.1.2 The Dirac quantisation condition

Whereas a particle interacting classically with an electromagnetic field does so solely via the field-strength $F^{\mu\nu}$, quantum mechanically the electromagnetic potential enters explicitly in the expression for the hamiltonian. Therefore the non-existence of the potential could spell trouble for the quantisation of, say, a charged particle interacting with the magnetic field of a monopole. In his celebrated paper of 1931, Dirac [Dir31] studied the problem of the quantum mechanics of a particle in the presence of a magnetic monopole and found that a consistent quantisation forced a relation between the electric charge of the particle and the magnetic charge of the monopole: the so-called Dirac quantisation condition. We will now derive this relation.

A magnetic monopole is a point-like source of magnetic field. If we place the source at the origin in $\mathbb{R}^3$, then the magnetic field is given by

$$\vec{\boldsymbol{B}}(\vec{\boldsymbol{r}}) = \frac{g}{4\pi}\frac{\vec{\boldsymbol{r}}}{r^3}\ , \tag{1.5}$$

where g is the magnetic charge. In these conventions, the magnetic charge is also the magnetic flux. Indeed, if Σ denotes the unit sphere in $\mathbb{R}^3$, then

$$g = \int_\Sigma \vec{\boldsymbol{B}}\cdot d\vec{\boldsymbol{S}}\ .$$

In the complement of the origin in $\mathbb{R}^3$, $\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{B}} = 0$, whence one can try to solve for a vector potential $\vec{\boldsymbol{A}}$ obeying $\vec{\boldsymbol{B}} = \vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}}$. For example, we can consider

$$\vec{\boldsymbol{A}}_+(\vec{\boldsymbol{r}}) = \frac{g}{4\pi r}\,\frac{1-\cos\theta}{\sin\theta}\,\hat{\boldsymbol{e}}_\phi\ ,$$

where (r,θ,ϕ) are spherical coordinates. $\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}}_+ = \vec{\boldsymbol{B}}$ everywhere but on the negative z-axis where $\theta=\pi$ and hence $\vec{\boldsymbol{A}}_+$ is singular. Similarly,

$$\vec{\boldsymbol{A}}_-(\vec{\boldsymbol{r}}) = -\frac{g}{4\pi r}\,\frac{1+\cos\theta}{\sin\theta}\,\hat{\boldsymbol{e}}_\phi$$

also obeys $\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}}_- = \vec{\boldsymbol{B}}$ everywhere but on the positive z-axis where $\theta = 0$ and $\vec{\boldsymbol{A}}_-$ is singular. It isn't that we haven't been clever enough, but that any $\vec{\boldsymbol{A}}$ which obeys $\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}} = \vec{\boldsymbol{B}}$ over some region will always be singular on some string-like region: the celebrated *Dirac string*.

Over their common domain of definition (the complement of the z-axis in $\mathbb{R}^3$) $\vec{\boldsymbol{\partial}}\times(\vec{\boldsymbol{A}}_+ - \vec{\boldsymbol{A}}_-) = 0$, whence one would expect that there exists a function χ so that $\vec{\boldsymbol{A}}_+ - \vec{\boldsymbol{A}}_- = \vec{\boldsymbol{\partial}}\chi$. However the complement of the z-axis is not simply-connected, and χ need only be defined locally. For example, restricting ourselves to $\theta = \frac{\pi}{2}$, we find that

$$\vec{\boldsymbol{A}}_+ - \vec{\boldsymbol{A}}_- = \frac{g}{2\pi r}\,\hat{\boldsymbol{e}}_\phi = \vec{\boldsymbol{\partial}}\left(\frac{g}{2\pi}\phi\right)\ ,$$

but notice that since ϕ is an angle, the function χ is not continuous. It couldn't possibly be continuous, for if it were there would be no flux. Indeed, if Σ again denotes the unit sphere in $\mathbb{R}^3$, $\Sigma^\pm$ the upper and lower hemispheres respectively, and E the equator, the flux can be computed in terms of χ as follows:

$$\begin{aligned} g &= \int_\Sigma \vec{\boldsymbol{B}}\cdot d\vec{\boldsymbol{S}} \\ &= \int_{\Sigma^+} (\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}}_+)\cdot d\vec{\boldsymbol{S}} + \int_{\Sigma^-} (\vec{\boldsymbol{\partial}}\times\vec{\boldsymbol{A}}_-)\cdot d\vec{\boldsymbol{S}} \\ &= \int_E \vec{\boldsymbol{A}}_+\cdot d\vec{\boldsymbol{\ell}} - \int_E \vec{\boldsymbol{A}}_-\cdot d\vec{\boldsymbol{\ell}} \\ &= \int_E \vec{\boldsymbol{\partial}}\chi\cdot d\vec{\boldsymbol{\ell}} \\ &= \chi(2\pi) - \chi(0)\ . \end{aligned}$$

Suppose now that we are quantising a particle of mass m and charge q in the field of a magnetic monopole. The Schrödinger equation satisfied by the wave-function is

$$-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi = i\hbar\frac{\partial\psi}{\partial t}$$

where $\vec{\boldsymbol{\nabla}} = \vec{\boldsymbol{\partial}} + ie\vec{\boldsymbol{A}}$, for $e = q/\hbar$. The Schrödinger equation is invariant under the gauge-transformations:

$$\vec{\boldsymbol{A}} \mapsto \vec{\boldsymbol{A}} + \vec{\boldsymbol{\partial}}\chi \quad \text{and} \quad \psi \mapsto \exp(-ie\chi)\psi \ .$$

This gauge invariance guarantees that solutions of the Schrödinger equation obtained locally with a particular $\vec{\boldsymbol{A}}$ will patch up nicely, *provided* that the wave-function be single-valued. This condition means that

$$\exp(-ie\chi) = \exp(-ieg\phi/2\pi)$$

must be a single-valued function, which is equivalent to the *Dirac quantisation condition*:

$$\boxed{eg = 2\pi\, n \qquad \text{for some} \quad n \in \mathbb{Z}.} \tag{1.6}$$

The Dirac quantisation condition has the following physical interpretation. Classically, there is not much of a distinction between a a magnetic monopole and a very long and very thin solenoid. The field inside the solenoid is of course, different, but in the limit in which the solenoid becomes infinitely long (on one end only) and infinitesimally thin, so that the inside of the solenoid lies beyond the probe of a classical experiment, the field at the end of the solenoid is indistinguishable from that of a magnetic monopole. Quantum mechanically, however, one can in principle detect the solenoid through the quantum interference pattern predicted by the Bohm-Aharanov effect. The condition for the absence of the interference is precisely the Dirac quantisation condition.

1.1.3 Dyons and the Zwanziger–Schwinger quantisation condition

A quicker, more heuristic derivation of the Dirac quantisation condition (1.6) follows by invoking the quantisation of angular momentum. The orbital angular momentum $\vec{\boldsymbol{L}} = \vec{\boldsymbol{r}} \times m\dot{\vec{\boldsymbol{r}}}$ of a particle of mass m and charge q in the presence of a magnetic monopole (1.5) is not conserved. Indeed, using the Lorentz force law,

$$\begin{aligned}
\frac{d\vec{\boldsymbol{L}}}{dt} &= \vec{\boldsymbol{r}} \times m\ddot{\vec{\boldsymbol{r}}} \\
&= \vec{\boldsymbol{r}} \times \left(q\dot{\vec{\boldsymbol{r}}} \times \vec{\boldsymbol{B}}\right) \\
&= \frac{qg}{4\pi r^3}\vec{\boldsymbol{r}} \times \left(\dot{\vec{\boldsymbol{r}}} \times \vec{\boldsymbol{r}}\right) \\
&= \frac{d}{dt}\left(\frac{qg}{4\pi}\frac{\vec{\boldsymbol{r}}}{r}\right) \ ,
\end{aligned}$$

whence the conserved quantity is instead

$$\vec{\boldsymbol{J}} \equiv \vec{\boldsymbol{L}} - \frac{qg}{4\pi}\frac{\vec{\boldsymbol{r}}}{r} \ ,$$

a result dating to 1896 and due to Poincaré.

Exercise 1.1 (Angular momentum due to the electromagnetic field)
Show that the correction term is in fact nothing else but the angular momentum of the electromagnetic field itself:

$$\vec{\boldsymbol{J}}_{\mathrm{em}} = \int_{\mathbb{R}^3} d^3r\, \vec{\boldsymbol{r}} \times \left(\vec{\boldsymbol{E}} \times \vec{\boldsymbol{B}}\right) \;,$$

where the $\vec{\boldsymbol{E}}$-field is the one due to the charged particle.

If we now assume that the electromagnetic angular momentum is separately quantised, so that

$$|\vec{\boldsymbol{J}}_{\mathrm{em}}| = \tfrac{1}{2} n\hbar \quad \text{for some} \quad n \in \mathbb{Z} \;,$$

we recover (1.6) again. The virtue of this derivation is that it provides a quick proof of the Zwanziger–Schwinger quantisation condition for dyons, as the following exercise asks you to show.

Exercise 1.2 (The Zwanziger–Schwinger quantisation condition)
A dyon is a particle which possesses both electric and magnetic charge. Consider two dyons of charges $(q{=}e\hbar, g)$ and $(q'{=}e'\hbar, g')$. Show that imposing the quantisation of the angular momentum of the resulting electromagnetic field yields the following condition:

$$eg' - e'g = 2\pi\, n \quad \textit{for some } n \in \mathbb{Z} \;. \tag{1.7}$$

Notice that the existence of the "electron" (that is, a particle with charges $(e, 0)$) does not tell us anything about the electric charge of a monopole (q, g); although it does tell us something about the difference between the electric charges of two such monopoles: (q, g) and (q', g). Indeed, (1.7) tells us immediately that $g(q-q') = 2\pi n$ for some integer n. If g has the minimum magnetic charge $g = 2\pi/e$, then the difference between the electric charges of the dyons (q, g) and (q', g) is an integer multiple of the electric charge of the electron: $q - q' = ne$ for some integer n. But we cannot say anything further about the absolute magnitude of either q or q'.

Exercise 1.3 (Dyonic spectrum in CP non-violating theories)
Prove that if CP is not *violated, then in fact there are only two (mutually exclusive) possibilities: either $q = ne$ or $q = ne + \frac{1}{2}e$.*
(Hint: use that under CP: $(q, g) \mapsto (-q, g)$. Why?)

We will see later when we discuss the so-called "Witten effect" that this gets modified in the presence of a CP-violating term, and the electric charge of the dyon will depend explicitly on the θ angle measuring the extent of the CP violation.

1.2 The 't Hooft–Polyakov Monopole

In 1974, 't Hooft [tH74] and Polyakov [Pol74] independently discovered that the bosonic part of the Georgi-Glashow model admits finite energy solutions that from far away look like Dirac monopoles. In contrast with the Dirac monopole, these solutions are everywhere regular and do not necessitate the introduction of a source of magnetic charge—this being due to the "twists" in (the vacuum expectation value of) the Higgs field.

1.2.1 The bosonic part of the Georgi–Glashow model

The Georgi–Glashow model was an early proposal to describe the electroweak interactions. We will be concerned here only with the bosonic part of the model which consists of an $SO(3)$ Yang–Mills field theory coupled to a Higgs field in the adjoint representation. The lagrangian density is given by

$$\mathcal{L} = -\tfrac{1}{4}\vec{\mathsf{G}}^{\mu\nu}\cdot\vec{\mathsf{G}}_{\mu\nu} + \tfrac{1}{2}D^\mu\vec{\boldsymbol{\phi}}\cdot D_\mu\vec{\boldsymbol{\phi}} - V(\phi)\ , \tag{1.8}$$

where

- the gauge field-strength $\vec{\mathsf{G}}_{\mu\nu}$ is defined by

  $$\vec{\mathsf{G}}_{\mu\nu} = \partial_\mu\vec{\mathsf{W}}_\nu - \partial_\nu\vec{\mathsf{W}}_\mu - e\vec{\mathsf{W}}_\mu\times\vec{\mathsf{W}}_\nu$$

  where $\vec{\mathsf{W}}_\mu$ are gauge potentials taking values in the Lie algebra of $SO(3)$, which we identify with $\mathbb{R}^3$ with the cross product for Lie bracket;

- the Higgs field $\vec{\boldsymbol{\phi}}$ is a vector in the (three-dimensional) adjoint representation of $SO(3)$, with components $\phi_a = (\phi_1,\phi_2,\phi_3)$ which is minimally coupled to the gauge field via the gauge-covariant derivative:

  $$D_\mu\vec{\boldsymbol{\phi}} = \partial_\mu\vec{\boldsymbol{\phi}} - e\vec{\mathsf{W}}_\mu\times\vec{\boldsymbol{\phi}}\ ;$$

- the Higgs potential $V(\phi)$ is given by

  $$V(\phi) = \tfrac{1}{4}\lambda\left(\phi^2 - a^2\right)^2$$

  where $\phi^2 = \vec{\boldsymbol{\phi}}\cdot\vec{\boldsymbol{\phi}}$ and λ is assumed non-negative.

The lagrangian density $\mathcal{L}$ is invariant under the following $SO(3)$ gauge transformations:

$$\begin{aligned}\vec{\boldsymbol{\phi}} &\mapsto \vec{\boldsymbol{\phi}}' = g(x)\,\vec{\boldsymbol{\phi}}\\ \vec{\mathsf{W}}_\mu &\mapsto \vec{\mathsf{W}}'_\mu = g(x)\vec{\mathsf{W}}_\mu g(x)^{-1} + \frac{1}{e}\partial_\mu g(x)\,g(x)^{-1}\ ,\end{aligned} \tag{1.9}$$

where $g(x)$ is a possibly x-dependent 3×3 orthogonal matrix with unit determinant.

The classical dynamics of the fields $\vec{\mathsf{W}}_\mu$ and $\vec{\boldsymbol{\phi}}$ are determined from the equations of motion

$$D_\nu \vec{\mathsf{G}}^{\mu\nu} = -e\vec{\boldsymbol{\phi}} \times D^\mu \vec{\boldsymbol{\phi}} \qquad D^\mu D_\mu \vec{\boldsymbol{\phi}} = -\lambda(\phi^2 - a^2)\vec{\boldsymbol{\phi}} \tag{1.10}$$

and by the Bianchi identity

$$D_\mu {}^\star\vec{\mathsf{G}}^{\mu\nu} = 0 \ , \tag{1.11}$$

where ${}^\star\vec{\mathsf{G}}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}\vec{\mathsf{G}}_{\lambda\rho}$.

The canonically conjugate momenta to the gauge field $\vec{\mathsf{W}}_\mu$ and the Higgs $\vec{\boldsymbol{\phi}}$ are given by

$$\vec{\mathsf{E}}^i = -\vec{\mathsf{G}}^{0i} \qquad \vec{\boldsymbol{\Pi}} = D_0 \vec{\boldsymbol{\phi}} \ . \tag{1.12}$$

Defining $\vec{\mathsf{B}}_i$ by

$$\vec{\mathsf{G}}_{ij} = -\epsilon_{ijk}\vec{\mathsf{B}}^k = +\epsilon_{ijk}\vec{\mathsf{B}}_k,$$

we can write the energy density as

$$\mathcal{H} = \tfrac{1}{2}\vec{\mathsf{E}}_i \cdot \vec{\mathsf{E}}_i + \tfrac{1}{2}\vec{\boldsymbol{\Pi}} \cdot \vec{\boldsymbol{\Pi}} + \tfrac{1}{2}\vec{\mathsf{B}}_i \cdot \vec{\mathsf{B}}_i + \tfrac{1}{2}D_i\vec{\boldsymbol{\phi}} \cdot D_i\vec{\boldsymbol{\phi}} + V(\phi) \tag{1.13}$$

which is manifestly positive-semidefinite and also gauge-invariant.

We define a *vacuum configuration* to be one for which the energy density vanishes. This means that

$$\vec{\mathsf{G}}_{\mu\nu} = 0 \qquad D^\mu \vec{\boldsymbol{\phi}} = 0 \qquad V(\phi) = 0 \ .$$

For example, $\vec{\boldsymbol{\phi}} = a\hat{\boldsymbol{e}}_3$ and $\vec{\mathsf{W}}_\mu = 0$ is such a configuration, where $(\hat{\boldsymbol{e}}_a)$ is an orthonormal basis for the three-dimensional representation space where the Higgs field takes values. We also define the *Higgs vacuum* as those configurations of the Higgs field which satisfy the latter two equations above. Notice that in the Higgs vacuum, the Higgs field obeys $\phi^2 = a^2$. Any such vacuum configuration is not invariant under the whole $SO(3)$, but only under an $SO(2) \cong U(1)$ subgroup, therefore this model exhibits spontaneous symmetry breaking.

Exercise 1.4 (The spectrum of the model)
Let $\vec{\boldsymbol{\phi}} = \vec{\boldsymbol{a}} + \vec{\boldsymbol{\varphi}}$ where $\vec{\boldsymbol{a}}$ is a constant vector obeying $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{a}} = a^2$. Expanding the lagrangian density in terms of $\vec{\boldsymbol{\varphi}}$, show that the model consists of a massless vector boson $A_\mu = \frac{1}{a}\vec{\boldsymbol{a}} \cdot \vec{\mathsf{W}}_\mu$ which we will identify with the photon, *a massive scalar field $\varphi = \frac{1}{a}\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{\phi}}$ and two massive vector bosons $W^\pm_\mu$ with the charge assignments given*

in Table 1.1.
(Hint: The masses are read off from the quadratic terms of the lagrangian density:

$$\mathcal{L} = \cdots + \tfrac{1}{2}\left(\frac{M_H}{\hbar}\right)^2 \varphi^2 + \tfrac{1}{2}\left(\frac{M_W}{\hbar}\right)^2 W_\mu^+ W^{\mu\,-} + \cdots$$

whereas the charges are read off from the coupling to the photon. The photon couples minimally via the covariant derivative $\nabla_\mu = \partial_\mu + iQ/\hbar A_\mu$. By examining how this covariant derivative embeds in the $SO(3)$ covariant derivative one can read off what Q are for the fields in the spectrum.)

Field	Mass	Charge
A_μ	0	0
φ	$M_H = a\sqrt{2\lambda}\,\hbar$	0
$W_\mu^\pm$	$M_W = ae\hbar$	$\pm e\hbar$

Table 1.1: The perturbative spectrum after higgsing.

1.2.2 Finite-energy solutions: the 't Hooft–Polyakov Ansatz

We now investigate the properties of finite-energy non-dissipative solutions to the equations of motion (1.10). But first let us remark a few properties of arbitrary finite-energy field configurations. The energy of a given field configuration is the spatial integral $E = \int d^3x\, \mathcal{H}$ of the energy density $\mathcal{H}$ given by equation (1.13). Finite energy means that the integral exists, hence the fields must approach a vacuum configuration asymptotically. In particular the Higgs field approaches the Higgs vacuum at spatial infinity. If we think of the Higgs potential V as a function $V : \mathbb{R}^3 \to \mathbb{R}$, let us define $\mathcal{M}_0 \subset \mathbb{R}^3$ as those points $\vec{\boldsymbol{x}} \in \mathbb{R}^3$ for which $V(\vec{\boldsymbol{x}}) = 0$. In the model at hand, $\mathcal{M}_0$ is the sphere of radius a, hence in any finite-energy configuration the Higgs field defines a function from the sphere at spatial infinity to $\mathcal{M}_0$:

$$\vec{\boldsymbol{\phi}}_\infty(\hat{\boldsymbol{r}}) \equiv \lim_{r\to\infty} \vec{\boldsymbol{\phi}}(\vec{\boldsymbol{r}}) \in \mathcal{M}_0 \ .$$

We will assume that the resulting function $\vec{\boldsymbol{\phi}}_\infty$ is actually continuous. This would follow from some uniformity property of the limit and such a property has been proven by Taubes [JT80].

It is well-known that the space of continuous functions from a sphere to a sphere is disconnected: it has an infinite number of connected components indexed by an integer called the *degree* of the map. A constant map has degree zero, whereas the identity map has degree 1. Heuristically, the degree is the number of times one sphere wraps around the other. It is the direct two-dimensional generalisation of the winding number for maps from a circle to a circle.

Taking these remarks into account it is not difficult to construct maps of arbitrary degree. Consider the map $f_n : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^3$ defined by

$$f_n(\vec{\boldsymbol{r}}) = (\sin\theta\,\cos n\varphi, \sin\theta\,\sin n\varphi, \cos\theta)\ , \tag{1.14}$$

where (r, θ, φ) are spherical coordinates. The map f_n restricts to a map from the unit sphere in $\mathbb{R}^3$ to itself which has degree n.

The *topological number* of a finite-energy configuration is defined to be the degree of the map $\vec{\boldsymbol{\phi}}_\infty$. The zero energy vacuum configuration $\vec{\mathsf{W}}_\mu = 0$, $\vec{\boldsymbol{\phi}} = a\hat{\boldsymbol{e}}_3$ has zero degree, since $\vec{\boldsymbol{\phi}}_\infty$ is constant. The topological number of a field configuration—being an integer—is invariant under any continuous deformation. In particular it is invariant under time evolution, and under gauge transformations, since the gauge group is connected. Hence if we set up a finite-energy field configuration at some moment in time whose topological number is different from zero, it will never dissipate; that is, it will never evolve in time towards a trivial solution. In other words, in a sense it will be stable.

We now investigate whether such stable solutions actually exist. We will narrow our search to spherically symmetric static solutions— a solution is defined to be *static* if it is time-independent and in addition the time-component of the gauge field $\vec{\mathsf{W}}_0$ vanishes.

One may be tempted to think that the this latter condition is simply a choice of gauge. Indeed it is easy to show that $\vec{\mathsf{W}}_0 = 0$ up to a gauge transformation, but the gauge transformation is actually time-dependent which is not allowed, since we are looking for time-independent solutions. *Coming soon: More details on the explicit time-dependent gauge transformation.* ✓

It follows from (1.12) that for static field configurations both $\vec{\mathsf{E}}$ and $\vec{\boldsymbol{\Pi}}$ vanish, and hence the energy agrees up to a sign with the lagrangian. This means that a field configuration will be a solution to the classical equations of motion if and only if it extremises the energy.

The 't Hooft–Polyakov Ansatz for the monopole is given by

$$\begin{aligned}\vec{\boldsymbol{\phi}}(\vec{\boldsymbol{r}}) &= \frac{\vec{\boldsymbol{r}}}{er^2} H(aer)\\ \mathsf{W}_a^i &= -\epsilon_{aij}\frac{r^j}{er^2}\left(1 - K(aer)\right)\\ \mathsf{W}_a^0 &= 0\ ,\end{aligned} \tag{1.15}$$

for some arbitrary functions H and K.

Exercise 1.5 (Boundary conditions on H and K)
Plugging the Ansatz into the expression for the energy density derive the following formula for the energy:

$$E = \frac{4\pi a}{e}\int_0^\infty \frac{d\xi}{\xi^2}\times \left(\xi^2\frac{dH}{d\xi} + \tfrac{1}{2}\left(\xi\frac{dH}{d\xi} - H\right)^2 + \tfrac{1}{2}\left(K^2-1\right)^2 + K^2H^2 + \frac{\lambda}{4e^2}\left(H^2-\xi^2\right)^2\right) . \tag{1.16}$$

Deduce that the integral exists provided that the following boundary conditions hold:

$$\begin{aligned} &K \to 0 \quad \text{and} \quad H/\xi \to 1 \quad \text{sufficiently fast as } \xi \to \infty \\ &K - 1 \leq O(\xi) \quad \text{and} \quad H \leq O(\xi) \quad \text{as } \xi \to 0 . \end{aligned} \tag{1.17}$$

This last equation means that H and K approach 0 and 1 respectively at least linearly in ξ as $\xi \to 0$.

Notice that with the above boundary conditions,

$$\vec{\boldsymbol{\phi}}_\infty(\hat{\boldsymbol{r}}) \equiv \lim_{r\to\infty} \frac{\vec{\boldsymbol{r}}}{er^2} H(aer) = a\hat{\boldsymbol{r}} ,$$

which is (homotopic to) the identity map, and hence has degree 1. In other words, the topological number of such a field configuration is 1. If such a solution exists it is therefore stable and non-dissipative.

Exercise 1.6 (The equations of motion for H and K)
Work out the equations of motion for the functions H and K in either of two ways: either plug the Ansatz into the equations of motion (1.10) *or else extremise the energy subject to the above boundary conditions. In either case you should get the following coupled nonlinear system of* ordinary *differential equations:*

$$\begin{aligned} \xi^2\frac{d^2K}{d\xi^2} &= KH^2 + K(K^2-1) \\ \xi^2\frac{d^2H}{d\xi^2} &= 2K^2H + \frac{\lambda}{e^2}H(H^2-\xi^2) . \end{aligned} \tag{1.18}$$

Initial numerical studies of the above differential equations for H and K together with the boundary conditions (1.17) suggested the existence of a solution. This was later proven rigourously by Taubes [JT80]. Notice that the asymptotic limit of the equations (1.18) in the limit $\xi \to \infty$ yields:

$$\begin{aligned} \frac{d^2K}{d\xi^2} &= K \\ \frac{d^2h}{d\xi^2} &= 2\frac{\lambda}{e^2}h , \end{aligned}$$

where $h \equiv H - \xi$. The above equations can be solved for at once and one finds that the solutions compatible with the boundary conditions are

$$\begin{aligned} K &\sim \exp(-\xi) = \exp(-M_W r/\hbar) \\ h &\sim \exp(-M_H r/\hbar) \ , \end{aligned}$$

where M_W and M_H were obtained in Exercise 1.4. This means that the solution describes an object of finite size given by the largest of the Compton wavelengths $\hbar/M_H$ or $\hbar/M_W$.

In order to identify the solution provided by the 't Hooft–Polyakov Ansatz we investigate the asymptotic electromagnetic field. Recall that the electromagnetic potential is identified with $A_\mu = \frac{1}{a}\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{W}}_\mu$, corresponding to the $U(1) \subset SO(3)$ defined as the stabiliser of $\vec{\boldsymbol{\phi}}$. The electromagnetic field can therefore be identified with $F_{\mu\nu} = \frac{1}{a}\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{G}}_{\mu,\nu}$. Because the 't Hooft–Polyakov Ansatz corresponds to a static solution, there is no electric field: $F_{0i} = 0$. However, as the next exercise shows, there is a magnetic field.

Exercise 1.7 (Asymptotic form of the electromagnetic field)
Show that the asymptotic form of $F_{ij} = \frac{1}{a}\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{G}}_{ij}$ *is given by*

$$F_{ij} = \epsilon_{ijk}\frac{r^k}{er^3} \ . \tag{1.19}$$

The form (1.19) of the electromagnetic field shows that the asymptotic magnetic field is that of a magnetic monopole:

$$\vec{\boldsymbol{B}} = -\frac{1}{e}\frac{\vec{\boldsymbol{r}}}{r^3} \ .$$

A quick comparison with equation (1.5) reveals that the magnetic charge of a 't Hooft–Polyakov monopole is (up to a sign) *twice* the minimum magnetic charge consistent with the electric charge e and the Dirac quantisation condition; that is, twice the *Dirac charge* corresponding to e. This follows from the fact that the electromagnetic $U(1)$ is embedded in $SO(3)$ in such a way that the electric charge is the eigenvalue of the T_3 isospin generator, which here is in the adjoint representation, which has integral isospin. The minimum electric charge is therefore $e_{\text{min}} = \frac{1}{2}e$, relative to which the charge of the 't Hooft–Polyakov monopole is indeed *one* Dirac charge, again up to a sign.

In fact, there is another solution with the opposite magnetic charge. It is obtained from the 't Hooft–Polyakov monopole by performing a parity transformation on the Ansatz.

One might wonder whether there also exist dyonic solutions. These solutions would not be static in that $\vec{\mathsf{W}}_0$ would be different from zero, but time-independent dyonic solutions have been found by Julia and Zee [JZ75] shortly after the results of 't Hooft and Polyakov.

In summary, we see that the 't Hooft–Polyakov solution describes an object of finite size which from far away cannot be distinguished from a Dirac monopole of charge $-4\pi/e$. In contrast with the Dirac monopole, the 't Hooft–Polyakov monopole is everywhere smooth—this being due to the massive fields which become relevant as we approach the "core" of the monopole.

1.2.3 The topological origin of the magnetic charge

Although the 't Hooft–Polyakov monopole is indistinguishable from far away from a Dirac monopole, as we approach its core the massive fields become relevant and the difference becomes evident. In contrast to the Dirac monopole, which necessitates a singular point-like magnetic source at the origin, the 't Hooft–Polyakov monopole is everywhere smooth and its magnetic charge is purely topological and, as we will see in this section, due completely to the behaviour of the Higgs field far away from the core.

Because of the exponential decay of the massive fields away from the core of the monopole, we notice that the Higgs field approaches the Higgs vacuum. In other words, a large distance away from the core of the monopole, the Higgs field satisfies

$$D_\mu \vec{\phi} = 0 \tag{1.20}$$

$$\vec{\phi} \cdot \vec{\phi} = a^2 \tag{1.21}$$

up to terms of order $O(\exp(-r/R))$ where R is the effective size of the monopole, which is governed by the mass of the heavy particles.

Notice that equation (1.20) already implies that $\vec{\phi} \cdot \vec{\phi}$ is a constant. Indeed,

$$\begin{aligned}\partial_\mu(\vec{\phi} \cdot \vec{\phi}) &= 2\vec{\phi} \cdot \partial_\mu \vec{\phi} \\ &= 2e\, \vec{\phi} \cdot (\vec{\mathsf{W}}_\mu \times \vec{\phi}) \quad \text{by (1.20)} \\ &= 0\ .\end{aligned}$$

What (1.21) tells us is that this constant is such that the potential attains its minimum.

It is therefore reasonable to assume that *any* finite-energy solution (not necessarily static or time-independent) of the Yang–Mills–Higgs system (1.10) satisfies equations (1.20) and (1.21) except in a finite number of well-separated compact localised regions in space, which we shall call *monopoles*. In other words, we are considering a "dilute gas of monopoles" surrounded by a Higgs vacuum.

Notice that in the Higgs vacuum, $\vec{\phi} \times \vec{\mathsf{W}}_\mu = -\frac{1}{e}\partial_\mu \vec{\phi}$, whence $\vec{\mathsf{W}}_\mu$ is fully determined except for the component in the $\vec{\phi}$-direction, which we denote by

A_μ. Computing the components perpendicular to $\vec{\boldsymbol{\phi}}$ we find that

$$\vec{\mathsf{W}}_\mu = \frac{1}{a^2 e}\vec{\boldsymbol{\phi}} \times \partial_\mu \vec{\boldsymbol{\phi}} + \frac{1}{a}\vec{\boldsymbol{\phi}}\, A_\mu \ .$$

Exercise 1.8 (Gauge field-strength in the Higgs vacuum)
Show that the field-strength in the Higgs vacuum points in the $\vec{\boldsymbol{\phi}}$-direction and is given by $\vec{\mathsf{G}}_{\mu\nu} = \frac{1}{a}\vec{\boldsymbol{\phi}}\, F_{\mu\nu}$ where

$$F_{\mu\nu} = \frac{1}{a^3 e}\vec{\boldsymbol{\phi}} \cdot \left(\partial_\mu \vec{\boldsymbol{\phi}} \times \partial_\nu \vec{\boldsymbol{\phi}}\right) + \partial_\mu A_\nu - \partial_\nu A_\mu \ .$$

Using the equations of motion (1.10) *and the Bianchi identity* (1.11) *prove that $F_{\mu\nu}$ satisfies Maxwell's equations* (1.3).

Now let Σ be a surface in the Higgs vacuum enclosing some monopoles in the volume it bounds. The magnetic flux through Σ measures the magnetic charge. Notice that A_μ doesn't contribute, and that we get:

$$\begin{aligned} g_\Sigma &\equiv \int_\Sigma \boldsymbol{\vec{B}} \cdot d\boldsymbol{\vec{S}} \\ &= -\frac{1}{2ea^2}\int_\Sigma \epsilon_{ijk}\vec{\boldsymbol{\phi}} \cdot \left(\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}}\right)\, dS_i \ . \end{aligned}$$

Notice that only the components of $\partial_i \vec{\boldsymbol{\phi}}$ tangential to Σ contribute to the integral and therefore the magnetic charge *only* depends on the behaviour of $\vec{\boldsymbol{\phi}}$ on Σ. Furthermore it only depends on the *homotopy* class of $\vec{\boldsymbol{\phi}}$ as map $\Sigma \to \mathcal{M}_0$; in other words, the above integral is invariant under deformations $\delta\vec{\boldsymbol{\phi}}$ of $\vec{\boldsymbol{\phi}}$ which preserve the Higgs vacuum:

$$D_\mu \delta\vec{\boldsymbol{\phi}} = 0 \qquad \text{and} \qquad \vec{\boldsymbol{\phi}} \cdot \delta\vec{\boldsymbol{\phi}} = 0 \ .$$

To see this, let's compute the variation of g_Σ under such a deformation of $\vec{\boldsymbol{\phi}}$. Notice first that

$$\begin{aligned} &\delta(\epsilon_{ijk}\vec{\boldsymbol{\phi}} \cdot (\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}})) = \\ &3\epsilon_{ijk}\delta\vec{\boldsymbol{\phi}} \cdot (\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}}) + 2\epsilon_{ijk}\partial_j(\vec{\boldsymbol{\phi}} \cdot (\delta\vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}})) \ . \end{aligned}$$

By Stokes' theorem, the second term in the right-hand-side integrates to zero. Now, because $\vec{\boldsymbol{\phi}} \cdot \partial_j \vec{\boldsymbol{\phi}} = 0$, $\vec{\boldsymbol{\phi}} \cdot (\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}}) = 0$, whence $\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}}$ is parallel to $\vec{\boldsymbol{\phi}}$. Hence, $\delta\vec{\boldsymbol{\phi}} \cdot (\partial_j \vec{\boldsymbol{\phi}} \times \partial_k \vec{\boldsymbol{\phi}}) = 0$. In other words, $\delta g_\Sigma = 0$. This means that g_Σ is invariant under arbitrary deformations of $\vec{\boldsymbol{\phi}}$ and hence under any deformation which can be achieved by iterating infinitesimal deformations: homotopies. Examples of homotopies are:

- time-evolution of $\vec{\phi}$;
- continuous gauge transformations on $\vec{\phi}$;
- continuous changes of Σ within the Higgs vacuum.

Exercise 1.9 (Additivity of the magnetic charge g_Σ)
Use the invariance of the magnetic charge under the last of the above homotopies, to argue that the magnetic charge is additive.
(Hint: Use a "contour" deformation argument.)

Notice that the magnetic charge can be written as $g_\Sigma = -\frac{4\pi}{e} N_\Sigma$, where

$$N_\Sigma = \frac{1}{8\pi a^3} \int_\Sigma dS_i \epsilon_{ijk} \vec{\phi} \times \left(\partial_j \vec{\phi} \times \partial_k \vec{\phi} \right) \ , \tag{1.22}$$

which as the next exercise asks you to show, is the degree of the map $\vec{\phi} : \Sigma \to \mathcal{M}_0$.

Exercise 1.10 (Dirac quantisation condition revisited)
Show that N_Σ is the integral of the jacobian of the map $\vec{\phi} : \Sigma \to \mathcal{M}_0$, which is the classical definition of the degree of the map. This means that N_Σ is an integer; a fact of which you may convince yourself by showing that if f_n is the map defined by (1.14), *then the value of N_Σ when Σ is, say, the unit sphere in $\mathbb{R}^3$, is equal to n. Taking this into account we recover again the Dirac quantisation condition:*

$$e\, g_\Sigma = -4\pi N_\Sigma \ , \tag{1.23}$$

with the same caveat as before about the fact that the minimum magnetic charge is twice the Dirac charge.

1.3 BPS-monopoles

Since the source for a Dirac monopole has to be put in by hand, its mass is a free parameter: it cannot be calculated. On the other hand, for the 't Hooft–Polyakov monopole there is no source, and the mass of the monopole is an intrinsic property of the Yang–Mills–Higgs system and as such it should be calculable. In the next section we derive a lower bound for its mass. A natural question to ask is whether there are solutions which saturate this bound, and in the section after that such a solution is found: the BPS-monopole.

1.3.1 Estimating the mass of a monopole: the Bogomol'nyi bound

In the centre of mass frame, all the energy of the monopole is concentrated in its mass. Therefore, taking equation (1.13) into account,

$$\begin{aligned} M &= \int_{\mathbb{R}^3} \left(\tfrac{1}{2}\vec{\mathsf{E}}_i \cdot \vec{\mathsf{E}}_i + \tfrac{1}{2}\vec{\mathsf{B}}_i \cdot \vec{\mathsf{B}}_i + \tfrac{1}{2}\vec{\mathbf{\Pi}} \cdot \vec{\mathbf{\Pi}} + \tfrac{1}{2} D_i\vec{\boldsymbol{\phi}} \cdot D_i\vec{\boldsymbol{\phi}} + V(\phi) \right) \\ &\geq \tfrac{1}{2} \int_{\mathbb{R}^3} \left(\vec{\mathsf{E}}_i \cdot \vec{\mathsf{E}}_i + \vec{\mathsf{B}}_i \cdot \vec{\mathsf{B}}_i + D_i\vec{\boldsymbol{\phi}} \cdot D_i\vec{\boldsymbol{\phi}} \right) , \end{aligned}$$

where we have dropped some non-negative terms. We now redistribute the last term as follows: we introduce an angular parameter θ and we add and subtract $\vec{\mathsf{E}}_i \cdot D_i\vec{\boldsymbol{\phi}} \sin\theta$ and $\vec{\mathsf{B}}_i \cdot D_i\vec{\boldsymbol{\phi}} \cos\theta$ to the integrand. This yields

$$\begin{aligned} M &\geq \tfrac{1}{2} \int_{\mathbb{R}^3} \left(\|\vec{\mathsf{E}}_i - D_i\vec{\boldsymbol{\phi}} \sin\theta\|^2 + \|\vec{\mathsf{B}}_i - D_i\vec{\boldsymbol{\phi}} \cos\theta\|^2 \right) \\ &\quad + \sin\theta \int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{E}}_i + \cos\theta \int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i \\ &\geq \sin\theta \int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{E}}_i + \cos\theta \int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i , \end{aligned}$$

where we have introduced the obvious shorthand $\|\mathsf{V}_i\|^2 = \mathsf{V}_i \cdot \mathsf{V}_i$. But now notice that

$$\begin{aligned} \int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i &= \int_{\mathbb{R}^3} \partial_i \left(\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i \right) \qquad \text{by the Bianchi identity (1.11)} \\ &= \int_{\Sigma_\infty} \vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i \, dS_i \qquad \text{by Stokes} \\ &= a \int_{\Sigma_\infty} \vec{\boldsymbol{B}} \cdot d\vec{\boldsymbol{S}} \equiv ag , \end{aligned} \tag{1.24}$$

where Σ_∞ is the sphere are spatial infinity and g is the magnetic charge of the solution. Notice that we have used the results of Exercise 1.8, which are valid since finite-energy demands that the sphere at spatial infinity be in the Higgs vacuum. Similarly, using the equations of motion this time instead of the Bianchi identity, one finds out that

$$\int_{\mathbb{R}^3} D_i\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{E}}_i = a \int_{\Sigma_\infty} \vec{\boldsymbol{E}} \cdot d\vec{\boldsymbol{S}} \equiv aq , \tag{1.25}$$

where q is the electric charge of the solution. Therefore for all angles θ we have the following bound on the mass:

$$M \geq ag\cos\theta + aq\sin\theta . \tag{1.26}$$

The sharpest bound occurs when the right hand side is a maximum, which happens for $q\cos\theta = g\sin\theta$. In other words, $\tan\theta = q/g$. Plugging this back into (1.26), we find the celebrated *Bogomol'nyi bound* for the mass of a monopole-like solution in terms of the electric and magnetic charges:

$$\boxed{M \geq a\sqrt{q^2+g^2}}\ , \tag{1.27}$$

derived for the first time in [Bog76] (see also [CPNS76]).

For the 't Hooft–Polyakov monopole, which is electrically neutral, the Bogomol'nyi bound yields $M \geq a|g| = 4\pi a/e$. But $a/e = M_W/e^2\hbar$, whence $M \gtrsim M_W/\alpha$, where $\alpha \simeq 1/137$ is the fine structure constant. If $M_W \simeq 90\text{GeV}$, say, then $M_W \gtrsim 12\text{TeV}$—beyond the present experimental range. This concludes the phenomenological part of these lectures!

1.3.2 Saturating the bound: the BPS-monopole

Having derived the Bogomol'nyi bound, it is natural to ask whether there exist solutions which saturate the bound. We will follow custom and call such states *BPS-states*. We incurred in the inequalities for the mass by discarding certain terms from the mass formula. To saturate the bound, these terms would have to be equal to zero. Since they are all integrals of non-negative quantities, we must impose that these quantities vanish *throughout space* and not just asymptotically as the weaker requirement of finite-energy would demand.

Let us concentrate on *static* solutions which saturate the bound. Static solutions satisfy $\vec{\mathsf{E}}_i = 0$ and $D_0\vec{\boldsymbol{\phi}} = 0$. In particular they have no electric charge, so that $\sin\theta = 0$. This means that $\cos\theta = \pm 1$ correlated to the sign of the magnetic charge. A quick inspection at the way we derived the bound reveals that for saturation we must also require that $V(\phi)$ should vanish and that in addition the *Bogomol'nyi equation* should hold:

$$\vec{\mathsf{B}}_i = \pm D_i\vec{\boldsymbol{\phi}}\ . \tag{1.28}$$

Now the only way to satisfy $V(\phi) = 0$ and yet obtain a solution with nonzero magnetic charge, is for λ to vanish. Why? Because for $\lambda \neq 0$, $\phi^2 = a^2$ throughout space, and in particular, $\vec{\boldsymbol{\phi}} \cdot D_i\vec{\boldsymbol{\phi}} = \vec{\boldsymbol{\phi}} \cdot \partial_i\vec{\boldsymbol{\phi}} = 0$. But using the Bogomol'nyi equation (1.28), this means that $\vec{\boldsymbol{\phi}} \cdot \vec{\mathsf{B}}_i = 0$, whence the solution carries no magnetic field. One way to understand the condition $\lambda = 0$ is as a limiting value. We let $\lambda \downarrow 0$, while at the same time retaining the boundary condition that at spatial infinity $\vec{\boldsymbol{\phi}}$ satisfies (1.21). This is known as the *Prasad–Sommerfield limit* [PS75].

Exercise 1.11 (The Bogomol'nyi equation implies (1.10)**)**
Show that the Bogomol'nyi equation together with the Bianchi identity (1.11) *implies the equations of motion* (1.10) *for the Yang–Mills–Higgs system with* $\lambda = 0$.

Of course, the advantage of the Bogomol'nyi equation lies in its simplicity. In fact, it is not hard to find an explicit solution to the Bogomol'nyi equation in the 't Hooft–Polyakov Ansatz, as the next exercise asks you to do.

Exercise 1.12 (The BPS-monopole)
Show that the Bogomol'nyi equation in the 't Hooft–Polyakov Ansatz yields the following systems of equations for the functions H *and* K*:*

$$\begin{aligned} \xi \frac{dK}{d\xi} &= -KH \\ \xi \frac{dH}{d\xi} &= H + 1 - K^2 \ . \end{aligned}$$

Show that the following is a solution with the right asymptotic boundary conditions:

$$\begin{aligned} H(\xi) &= \xi \coth \xi - 1 \\ K(\xi) &= \frac{\xi}{\sinh \xi} \ . \end{aligned}$$

Notice that the solution for the BPS-monopole is such that

$$H(\xi) - \xi = 1 + O(\exp(-\xi)) \ ,$$

which does not contradict (1.17) because for $\lambda = 0$ the Higgs field is massless. Its interactions are long range and hence the BPS-monopole can be distinguished from a Dirac monopole from afar.

One consequence of the Bogomol'nyi equation is that both the photon (through $\vec{\mathsf{B}}_i$) and the Higgs (via $D_i\vec{\phi}$) contribute equally to the mass density. One can show that the long-range force exerted by the Higgs is always attractive and for static monopoles, it is equal in magnitude to the $1/r^2$ magnetic force. Therefore the forces add for oppositely charged monopoles, yet they cancel for equally charged monopoles. This is as it should be if static multi-monopole solutions saturating the Bogomol'nyi bound are to exist. To see this, notice that the mass of a two-monopole system with charges g and g' (of the same sign) is precisely equal to the sum of the masses of each of the BPS-monopoles. Hence there can be no net force between them.

Exercise 1.13 (The mass density at the origin is finite)
Show that the mass density at the origin for a BPS-monopole is not merely integrable, but actually finite!
(Hint: Notice that the mass density is given by $\|D_i\vec{\phi}\|^2$*. Compute this for the BPS-monopole and expand as* $\xi \sim 0$*.)*

1.4 Duality conjectures

In this section we discuss the observed duality symmetries between perturbative and nonperturbative states in the Georgi–Glashow model and the conjectures that this observation suggests. We start with the Montonen–Olive conjecture and then, after introducing a CP-violating term in the theory, the Witten effect will suggest an improved $SL(2,\mathbb{Z})$ duality conjecture.

1.4.1 The Montonen–Olive conjecture

At $\lambda = 0$, the (bosonic) spectrum of the Georgi–Glashow model (including the BPS-monopoles) is the following:

Particle	Mass	Electric Charge	Magnetic Charge	Spin/ Helicity
Photon	0	0	0	± 1
Higgs	0	0	0	0
$W_\pm$ boson	aq	$\pm q$	0	1
$M_\pm$ monopole	ag	0	$\pm g$	0

where $q = e\hbar$. Two features are immediately striking:

- all particles satisfy the Bogomol'nyi bound; and
- the spectrum is invariant under electromagnetic $\mathbb{Z}_2$ duality: $(q, g) \mapsto (g, -q)$ *provided* that we also interchange the BPS-monopoles and the massive vector bosons.

The invariance of the spectrum under electromagnetic duality is a consequence of the fact that the formula for the Bogomol'nyi bound is invariant under electromagnetic duality and the fact that the spectrum saturates the bound. This observation prompted Montonen and Olive [MO77] to conjecture that there should be a dual ("magnetic") description of this gauge theory where the elementary gauge particles are the BPS-monopoles and where the massive vector bosons appear as "electric monopoles". This conjecture is reinforced by the fact that two very different calculations for the inter-particle force between the massive vector bosons (done by computing tree diagrams in the quantum field theory) and between the BPS-monopoles (a calculation due to Manton) yield identical answers. Notice, however, that because of the ✓ Dirac quantisation condition, if the coupling constant e of the original theory is small, the coupling constant g of the magnetic theory must be large, and viceversa. Hence the duality conjecture would imply that the strong coupling behaviour of a gauge theory could be determined by the weak coupling behaviour of its dual theory—a very attractive possibility.

The formula for the Bogomol'nyi bound is actually invariant under rotations in the (g, q)-plane. But the quantisation of the electric and magnetic charges, actually breaks this symmetry down to the $\mathbb{Z}_2$ duality symmetry. In their paper, Montonen and Olive speculate that the massless Higgs could play the role of a Goldstone boson associated to the breaking of this $SO(2)$ symmetry down to $\mathbb{Z}_2$. I am not aware of any further progress in this direction.

The Montonen–Olive conjecture suffers from several drawbacks:

- there is no reason to believe that the duality symmetry of the spectrum is not broken by radiative corrections through a renormalisation of the Bogomol'nyi bound;
- in order to understand the BPS-monopoles as gauge particles, we would expect that their spin be equal to one—yet it would seem naively that due to their rotational symmetry, they have spin zero; and
- the conjecture is untestable unless we get a better handle at strongly coupled theories—of course, this also means that it cannot be disproved!

We will see in the next chapter that supersymmetry solves the first two problems. The third problem is of course very difficult, but we will now see that by introducing a CP violating term in the action, the duality conjecture will imply a richer dyonic spectrum which can be tested in principle.

1.4.2 The Witten effect

Exercise 1.3 asked you to compute the dyonic spectrum consistent with the quantisation condition (1.7) in a CP non-violating theory. You should have found that the electric charge q of a dyon with minimal magnetic charge g could take one of two sets of mutually exclusive values: either $q = ne$ or $q = ne + \frac{1}{2}e$, where n is some integer. We will see that indeed it is the former case which holds.

Let N denote the operator which generates gauge transformations about the direction $\vec{\boldsymbol{\phi}}$:

$$\begin{aligned} \delta\vec{\boldsymbol{v}} &= \frac{1}{a}\vec{\boldsymbol{\phi}} \times \vec{\boldsymbol{v}} \\ \delta\vec{\mathsf{W}}_\mu &= -\frac{1}{ea} D_\mu \vec{\boldsymbol{\phi}} \ , \end{aligned} \tag{1.29}$$

where $\vec{\boldsymbol{v}}$ is any isovector. And consider the operator $\exp 2\pi i N$. In the background of a finite energy solution, the Higgs field is in the Higgs vacuum at spatial infinity, whence $\exp 2\pi i N$ generates the identity transformation. On isovectors it generates a rotation about $\vec{\boldsymbol{\phi}}$ of magnitude $2\pi|\vec{\boldsymbol{\phi}}|/a = 2\pi$, and on the gauge fields we notice that $D_\mu \vec{\boldsymbol{\phi}} = 0$ in the Higgs vacuum. Since $\exp 2\pi i N = \mathbb{1}$, the eigenvalues of N are integral. To see what this means, we compute N.

We can compute N since it is the charge of the Noether current associated with the transformations (1.29). Indeed,

$$N = \int_{\mathbb{R}^3} \left(\frac{\partial \mathcal{L}}{\partial\, \partial_0 \vec{\mathsf{W}}_\mu} \cdot \delta \vec{\mathsf{W}}_\mu + \frac{\partial \mathcal{L}}{\partial\, \partial_0 \vec{\boldsymbol{\phi}}} \cdot \delta \vec{\boldsymbol{\phi}} \right) .$$

Using equation (1.29), and hence that $\delta \vec{\boldsymbol{\phi}} = 0$, we can rewrite N as

$$N = -\frac{1}{ae} \int_{\mathbb{R}^3} \frac{\partial \mathcal{L}}{\partial\, \partial_0 \vec{\mathsf{W}}_i} \cdot D_i \vec{\boldsymbol{\phi}} .$$

Since the conjugate momentum to $\vec{\mathsf{W}}_i$ is $-\vec{\mathsf{G}}^{0i} = \vec{\mathsf{E}}^i = -\vec{\mathsf{E}}_i$, we find that

$$N = \frac{1}{ae} \int_{\mathbb{R}^3} \vec{\mathsf{E}}_i \cdot D_i \vec{\boldsymbol{\phi}} = \frac{q}{e} , \tag{1.30}$$

where we have used the expression (1.25) for the electric charge q of the configuration. The quantisation of N then implies that $q = ne$ for some integer n.

Let us now introduce a θ-term in the action:

$$\mathcal{L}_\theta = \tfrac{1}{2} \frac{e^2\theta}{32\pi^2} \epsilon^{\alpha\beta\mu\nu} \vec{\mathsf{G}}_{\alpha\beta} \cdot \vec{\mathsf{G}}_{\mu\nu} = -\frac{e^2\theta}{32\pi^2} {}^\star\vec{\mathsf{G}}^{\mu\nu} \cdot \vec{\mathsf{G}}_{\mu\nu} .$$

This term is locally a total derivative and hence does not contribute to the equations of motion. Its integral in a given configuration is an integral multiple (called the instanton number) of the parameter θ. θ is therefore an angular variable and parametrises inequivalent vacua. The Noether charge N gets modified in the presence of this term as follows:

$$N \mapsto N - \frac{1}{ae} \int_{\mathbb{R}^3} \frac{\partial \mathcal{L}_\theta}{\partial\, \partial_0 \vec{\mathsf{W}}_i} \cdot D_i \vec{\boldsymbol{\phi}} .$$

Computing this we find

$$\begin{aligned}\Delta N &= -\frac{e\theta}{16\pi^2 a}\int_{\mathbb{R}^3} \epsilon^{0i\alpha\beta}\vec{\mathsf{G}}_{\alpha\beta}\cdot D_i\vec{\boldsymbol{\phi}}\\ &= -\frac{e\theta}{16\pi^2 a}\int_{\mathbb{R}^3} \epsilon^{ijk}\vec{\mathsf{G}}_{jk}\cdot D_i\vec{\boldsymbol{\phi}}\\ &= \frac{e\theta}{8\pi^2 a}\int_{\mathbb{R}^3} \vec{\mathsf{B}}_i\cdot D_i\vec{\boldsymbol{\phi}}\\ &= \frac{e\theta}{8\pi^2}\, g\ ,\end{aligned}$$

where g, given by equation (1.24), is the magnetic charge of the configuration. In other words,

$$N = \frac{q}{e} + \frac{e\theta}{8\pi^2}\, g\ .$$

For the 't Hooft–Polyakov monopole, $eg = -4\pi$, hence the integrality of N means that

$$\boxed{q = ne + \frac{e\theta}{2\pi} \qquad \text{for some} \quad n\in\mathbb{Z}.} \tag{1.31}$$

This result, which was first obtained by Witten in [Wit79], is of course consistent with the quantisation condition (1.7) since for a fixed θ the difference between any charges is an integral multiple of e.

1.4.3 $SL(2,\mathbb{Z})$ duality

The action defined by $\mathcal{L} + \mathcal{L}_\theta$ depends on four parameters: e, θ, λ and a. The dependence on the first two can be unified into a complex parameter τ. To see this, let us first rescale the gauge fields $\vec{\mathsf{W}}_\mu \mapsto e\vec{\mathsf{W}}_\mu$. This has the effect of bringing out into the open all the dependence on e. The lagrangian is now

$$\mathcal{L} + \mathcal{L}_\theta = -\frac{1}{4e^2}\vec{\mathsf{G}}_{\mu\nu}\cdot\vec{\mathsf{G}}^{\mu\nu} + \frac{\theta}{32\pi^2}\vec{\mathsf{G}}_{\mu\nu}\cdot{}^\star\vec{\mathsf{G}}^{\mu\nu} + \tfrac{1}{2}D^\mu\vec{\boldsymbol{\phi}}\cdot D_\mu\vec{\boldsymbol{\phi}} - V(\phi)\ , \tag{1.32}$$

where all the (e,θ)-dependence is now shown explicitly. We now define a complex parameter

$$\tau \equiv \frac{\theta}{2\pi} + i\frac{4\pi}{e^2}\ ,$$

whose imaginary part is positive since e is real. To write the lagrangian explicitly in terms of τ it is convenient to introduce the following complex linear combination:

$$\vec{\mathcal{G}}_{\mu\nu} \equiv \vec{\mathsf{G}}_{\mu\nu} + i\,{}^\star\vec{\mathsf{G}}_{\mu\nu}\ . \tag{1.33}$$

It then follows that

$$\vec{\mathfrak{G}}_{\mu\nu} \cdot \vec{\mathfrak{G}}^{\mu\nu} = 2\vec{\mathsf{G}}_{\mu\nu} \cdot \vec{\mathsf{G}}^{\mu\nu} + 2i\vec{\mathsf{G}}_{\mu\nu} \cdot {}^\star\vec{\mathsf{G}}^{\mu\nu} ,$$

whence the first two terms in the lagrangian (1.32) can be written simply as

$$-\frac{1}{32\pi} \operatorname{Im}\left(\tau\, \vec{\mathfrak{G}}_{\mu\nu} \cdot \vec{\mathfrak{G}}^{\mu\nu}\right) . \tag{1.34}$$

Notice that because θ is an angular variable, it is only defined up to 2π. This means that physics is invariant under $\tau \mapsto \tau+1$. At $\theta = 0$, the conjecture of electromagnetic duality says that $e \mapsto g = -4\pi/e$ is a symmetry. But this duality transformation is just $\tau \mapsto -1/\tau$. We are therefore tempted to strengthen the conjecture of electromagnetic duality to say that for arbitrary θ, the physics should depend on τ only modulo the transformations:

$$T : \tau \mapsto \tau + 1$$
$$S : \tau \mapsto -\frac{1}{\tau} .$$

Exercise 1.14 ($(P)SL(2,\mathbb{Z})$ and its action on the upper half-plane)
The group $SL(2,\mathbb{Z})$ of all 2×2 matrices with unit determinant and with integer entries acts naturally on the complex plane:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d} .$$

Prove that this action preserves the upper half-plane, so that if $\operatorname{Im}\tau > 0$, so will its transform under $SL(2,\mathbb{Z})$. Prove that the matrices $\mathbb{1}$ and $-\mathbb{1}$ both act trivially (and are the only two matrices that do). Thus the action is not faithful, but it becomes faithful if we identify every matrix $M \in SL(2,\mathbb{Z})$ with $-M$. The resulting group is denoted $PSL(2,\mathbb{Z}) \equiv SL(2,\mathbb{Z})/\{\pm\mathbb{1}\}$.

The operations S and T defined above are clearly invertible and hence generate a discrete group. Prove that they satisfy the following relations:

$$S^2 = \mathbb{1} \qquad \text{and} \qquad (ST)^3 = \mathbb{1} .$$

Prove that the group generated by S and T subject to the above relations is a subgroup of $PSL(2,\mathbb{Z})$, by exhibiting matrices $\hat{S}$ and $\hat{T}$ whose action on τ coincides with the action of S and T. These matrices are not unique, since in going from $PSL(2,\mathbb{Z})$ to $SL(2,\mathbb{Z})$ we have to choose a sign. Nevertheless, for any choice of $\hat{S}$ and $\hat{T}$, prove that the following matrix identities are satisfied:

$$\hat{S}^2 = -\mathbb{1} \qquad \text{and} \qquad (\hat{S}\hat{T})^6 = \mathbb{1} .$$

The matrices $\hat{S}$ and $\hat{T}$ thus generate a subgroup of $SL(2,\mathbb{Z})$. Prove that this subgroup is in fact the whole group, which implies that S and T generate all of $PSL(2,\mathbb{Z})$.
(Hint: if you get stuck look in [Ser73].)

Is physics invariant under $SL(2,\mathbb{Z})$? Clearly this would be a bold conjecture, but no bolder than the original Montonen–Olive $\mathbb{Z}_2$ conjecture, for in fact the evidence for both is more or less the same. Indeed, as we now show the mass formula for BPS-states is invariant under $SL(2,\mathbb{Z})$. The mass of a BPS-state with charges (q,g) is given by the equality in formula (1.27). From formula (1.23) it follows that the allowed magnetic charges of the form $g = n_m 4\pi/e$, for some $n_m \in \mathbb{Z}$. As a consequence of the Witten effect, the allowed electric charges are given by $q = n_e e + n_m e\theta/2\pi$. The mass of BPS-states is then given by

$$M^2 = 4\pi a^2\, \vec{\boldsymbol{n}}^t \cdot A(\tau) \cdot \vec{\boldsymbol{n}} \;, \tag{1.35}$$

where $\vec{\boldsymbol{n}} = (n_e, n_m)^t \in \mathbb{Z}\times\mathbb{Z}$ and where

$$A(\tau) = \frac{1}{\operatorname{Im}\tau}\begin{pmatrix} 1 & \operatorname{Re}\tau \\ \operatorname{Re}\tau & |\tau|^2\end{pmatrix}$$

Exercise 1.15 ($SL(2,\mathbb{Z})$-invariance of the mass formula)
Let

$$M = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in SL(2,\mathbb{Z}) \;.$$

First prove that

$$A(M\cdot\tau) = \left(M^{-1}\right)^t \cdot A(\tau)\cdot M^{-1} \;,$$

and as a consequence deduce that the mass formula is invariant provided that we also transform the charges:

$$\vec{\boldsymbol{n}} \mapsto M\cdot\vec{\boldsymbol{n}} \;.$$

The improved Montonen–Olive conjecture states that physics is $SL(2,\mathbb{Z})$-invariant. If this is true, this means that the theories defined by two values of τ related by the action of $SL(2,\mathbb{Z})$ are physically equivalent, provided that we are willing to relabel magnetic and electric charges by that same $SL(2,\mathbb{Z})$ transformation.

The action of $PSL(2,\mathbb{Z})$ on the upper half-plane is well-known (see for example Serre's book [Ser73]). There is a *fundamental domain* D defined by

$$D = \{\tau\in\mathbb{C} \mid \operatorname{Im}\tau > 0,\ |\operatorname{Re}\tau| \leq \tfrac{1}{2},\ |\tau|\geq 1\} \;, \tag{1.36}$$

which has the property that its orbit under $PSL(2,\mathbb{Z})$ span the whole upper half-plane and that *no* two points in its interior

$$\operatorname{Int} D = \{\tau\in\mathbb{C} \mid \operatorname{Im}\tau > 0,\ |\operatorname{Re}\tau| < \tfrac{1}{2},\ |\tau| > 1\} \;,$$

are related by the action of $PSL(2,\mathbb{Z})$.

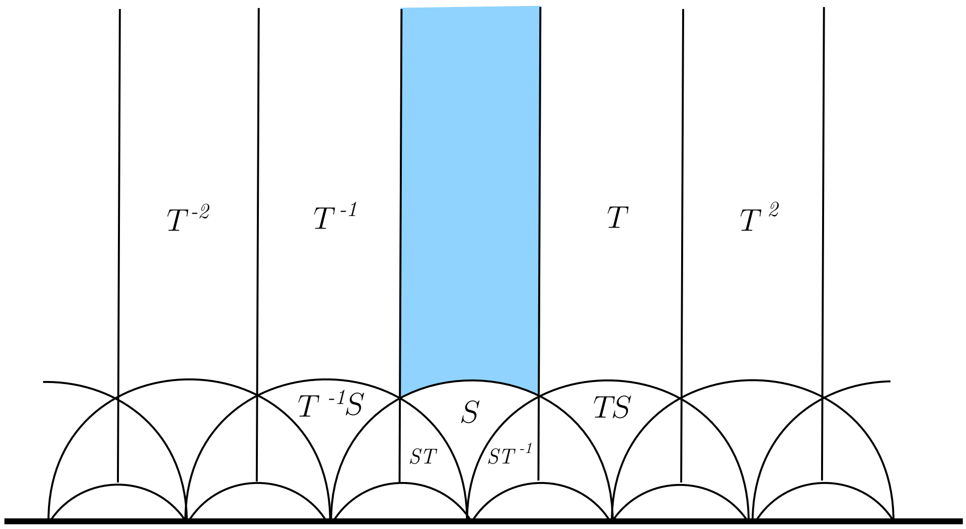


Figure 1.1: Fundamental domain (shaded) for the action of $PSL(2,\mathbb{Z})$ on the upper half plane, and some of its $PSL(2,\mathbb{Z})$ images.

Exercise 1.16 (Orbifold points in the fundamental domain D)
The fundamental domain D contains three "orbifold" points: i, $\omega = \exp(i\pi/3)$ and $-\bar{\omega} = \exp(2i\pi/3)$ which are fixed by some finite subgroup of $PSL(2,\mathbb{Z})$. Indeed, prove that i is fixed by the $\mathbb{Z}_2$-subgroup generated by S, whereas ω and $-\bar{\omega}$ are fixed respectively by the $\mathbb{Z}_3$-subgroups generated by TS and ST.

We end this section and this chapter with a discussion of the dyonic spectrum predicted by $SL(2,\mathbb{Z})$-duality. If we had believed in the electromagnetic $\mathbb{Z}_2$-duality, we could have predicted the existence of the BPS-monopoles from the knowledge of the existence of the massive vector bosons (and viceversa). But this is as far as we could have gone with $\mathbb{Z}_2$. On the other hand $SL(2,\mathbb{Z})$ has infinite order, and *assuming that for all values of τ there are massive vector bosons in the spectrum*, $SL(2,\mathbb{Z})$-duality predicts an infinite number of dyonic states. This assumption is not as innocent as it seems, as the Seiberg–Witten solution to pure N=2 supersymmetric Yang–Mills demonstrates; but it seems to hold if we have N=4 supersymmetry. But for now let us simply follow our noses and see what this assumption implies.

Let's assume then that for all values of τ there is a state with quantum numbers $\vec{n} = (1,0)^t$. The duality conjecture predicts the existence of one state each with quantum numbers in the $SL(2,\mathbb{Z})$-orbit of $\vec{\boldsymbol{n}}$:

$$M \cdot \vec{\boldsymbol{n}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} .$$

Because M has unit determinant, a and c are not arbitrary integers: there exist integers b and d such that $ad - bc = 1$. This means that a and c are *coprime*; that is, they don't have a common factor (other than 1). Indeed, if n were a common factor: $a = na'$ and $c = nc'$ for integers a' and c', and we would have that $n(a'd - bc') = 1$ which forces $n = 1$. We will now show that this arithmetic property of a and c actually translates into the stability of the associated dyonic state!

Exercise 1.17 (Properties of the mass matrix $A(\tau)$)
Notice that the matrix $A(\tau)$ in the mass formula (1.35) *enjoys the following properties for all τ in the upper half-plane:*

$$\det A(\tau) = 1 \qquad \textit{and} \qquad A(\tau) \textit{ is positive-definite.}$$

Prove that this latter property implies that the mass formula defines a distance function, so that in particular it obeys the triangle inequality. In other words, if we define $\|\vec{\boldsymbol{n}}\|^2 \equiv M^2_{\vec{\boldsymbol{n}}}$—that is, the Bogomol'nyi mass of a dyonic state with that charge assignment—then prove that

$$\|\vec{\boldsymbol{n}} + \vec{\boldsymbol{m}}\| \leq \|\vec{\boldsymbol{n}}\| + \|\vec{\boldsymbol{m}}\| . \tag{1.37}$$

Now let's consider a dyonic state $\vec{\boldsymbol{q}} = (a, c)^t$. The triangle inequality (1.37) says that for *any* two dyonic states $\vec{\boldsymbol{n}}$ and $\vec{\boldsymbol{m}}$ which obey $\vec{\boldsymbol{n}} + \vec{\boldsymbol{m}} = \vec{\boldsymbol{q}}$, the mass of the $\vec{\boldsymbol{q}}$ is less than or equal to the sum of the masses of $\vec{\boldsymbol{n}}$ and $\vec{\boldsymbol{m}}$. But we claim that when a and c are coprime, the inequality is actually strict! Indeed, the inequality is only saturated when $\vec{\boldsymbol{n}}$ and $\vec{\boldsymbol{m}}$, and hence $\vec{\boldsymbol{q}}$, are collinear. But if this is the case, a and c must have a common factor. Assume for a contradiction that they don't. If $\vec{\boldsymbol{n}} = (p, q)^t$ and $\vec{\boldsymbol{m}} = (r, s)^t$, then we must have that both p and r are smaller in magnitude to a, and that q and s are smaller in magnitude to c. But collinearity means that $pc = qa$. Since a and c are relatively prime, it must be that a divides p so that there is some integer n such that $p = an$, which contradicts that p is smaller in magnitude to a. This also follows pictorially from the fact that a and c are coprime if and only if in the straight line from the origin to $\vec{\boldsymbol{q}} \in \mathbb{Z}^2 \subset \mathbb{R}^2$, $\vec{\boldsymbol{q}}$ is the first integral point. Therefore the dyonic state represented by $\vec{\boldsymbol{q}}$ is is a genuine stable state which cannot be interpreted as a bound state of other dyonic states with "smaller" charges.

Contrast this with the case where the triangle inequality saturates. In this case, this means that the bound state of two dyons with "collinear" charges exhibits no net force between its constituents. Compare this with the discussion in section 1.3.2.

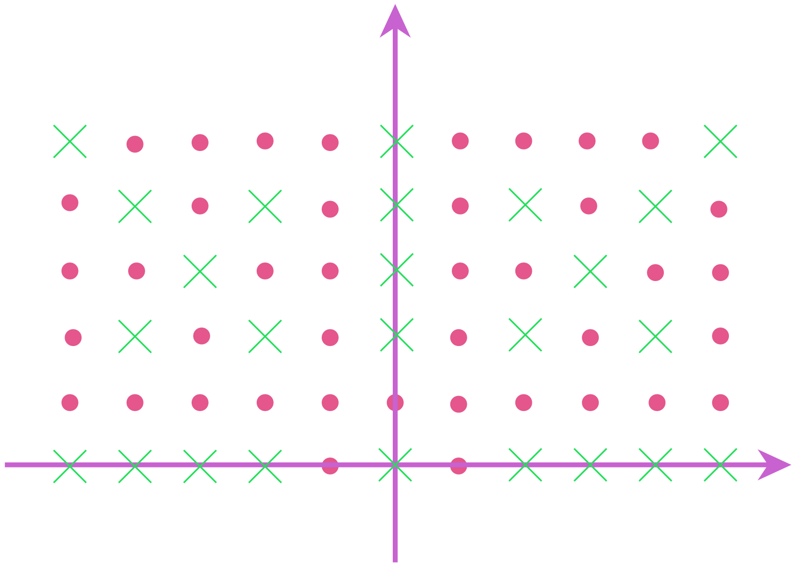


Figure 1.2: Dyonic spectrum predicted by $SL(2,\mathbb{Z})$ duality. Dots indicate dyons, crosses indicates holes in the dyonic spectrum. Only dyons with non-negative magnetic charge are shown.

The dyonic states in the $SL(2,\mathbb{Z})$-orbit of $(1,0)$ can be depicted as follows:

Notice that for $c{=}0$ we have the original state and its charge conjugate. For $c{=}1$ we have the Julia-Zee dyons but with quantised electric charge: a can be an arbitrary integer. For $c{=}2$, we have that a must be odd. Notice that in every rational direction (that is, every half-line with rational slope emanating from the origin) only the first integral point is present. As explained above these are precisely those points (m,n) whose coordinates are coprime.

As we will see in the context of $N{=}4$ supersymmetric Yang–Mills theory, the dyonic spectrum is in one-to-one correspondence with square-integrable harmonic forms on monopole moduli space. This is a fascinating prediction: it says that there is an action of the modular group on the (L^2) cohomology of monopole moduli space.

Chapter 2

Supersymmetry

In this chapter we discuss the intimate relation between supersymmetry and the Bogomol'nyi bound. The effect of supersymmetry is two-fold: first of all, it enforces the bound since this is a property of unitary representations of the supersymmetry algebra; but it also protects the bound against quantum corrections, guaranteeing that if a state saturates the bound classically, it does so quantum mechanically. This last assertion follows because, as we will see, supersymmetry multiplets corresponding to BPS-states are smaller than the multiplets of states where the Bogomol'nyi bound is not saturated.

We first discuss the supersymmetry algebra and its representations. For definiteness we shall work in four dimensions, but much of what we'll say can (and will) be used in dimensions other than four. It will be while studying (massive) representations with central charges that we will see the mechanism by which the Bogomol'nyi bound follows from the algebra. We then illustrate this fact by studying a particular example: $N{=}2$ supersymmetric Yang–Mills in four dimensions. We define this theory by dimensional reduction from $N{=}1$ supersymmetric Yang–Mills in six dimensions. This theory admits a Higgs mechanism by which the gauge symmetry is broken to $U(1)$ while preserving supersymmetry. The higgsed spectrum falls into a massless gauge multiplet corresponding to the unbroken $U(1)$ and two massive short multiplets. From the structure of the short $N{=}2$ multiplets we can deduce that the $N{=}2$ supersymmetry algebra admits central charges and, moreover, that the multiplets containing the massive vector bosons must saturate the mass bound. We will also see that this theory admits BPS-like solutions, which are shown to break one half of the supersymmetries. This implies that the BPS-monopole belongs to a short multiplet and suggests that the bound which follows abstractly from the supersymmetry algebra agrees with the Bogomol'nyi bound for dyons given by equation (1.27). This is shown to be case. Nevertheless the short multiplets containing the massive vector bosons and those containing the BPS-monopole have different spins, whence N=2 supersymmetric Yang–Mills does not yet seem to be a candidate for a theory which is (Montonen–Olive) self-dual. This problem will be solved for N=4 supersymmetric Yang–Mills, which we study as the dimensional reduction of ten-dimensional N=1 supersymmetric Yang–Mills. At a formal level, N=4 supersymmetric Yang–Mills is qualitatively very similar to the N=2 theory; except that we will see that the short multiplets which contain the solitonic and the fundamental BPS-states have the same spin. This prompts the question whether N=4 supersymmetric Yang–Mills is self-dual – a conjecture that we will have ample opportunity to test as the lectures progress.

2.1 The super-Poincaré algebra in four dimensions

In this section we will briefly review the supersymmetric extension of the four-dimensional Poincaré algebra. There are plenty of good references available so we will be brief. We will follow for the most part the conventions in [Soh85], to where we refer the reader for the relevant references on supersymmetry.

2.1.1 Some notational remarks about spinors

The Lorentz group in four dimensions, $SO(1,3)$ in our conventions, is not simply-connected and therefore, strictly speaking, has no spinorial representations. In order to consider spinorial representations we must look to the corresponding spin group $Spin(1,3)$ which happens to be isomorphic to $SL(2,\mathbb{C})$—the group of 2×2-complex matrices with unit determinant. From its very definition, $SL(2,\mathbb{C})$ has a natural two-dimensional complex representation, which we shall call $\mathbb{S}$. More precisely, $\mathbb{S}$ is the vector space $\mathbb{C}^2$ with the natural action of $SL(2,\mathbb{C})$. If $u \in \mathbb{S}$ has components $u_\alpha = (u_1, u_2)$ relative to some fixed basis, and $M \in SL(2,\mathbb{C})$, the action of M on u is defined simply by $(M\,u)_\alpha = M_\alpha{}^\beta u_\beta$. We will abuse the notation and think of the components u_α as the vector and write $u_\alpha \in \mathbb{S}$.

This is not the only possible action of $SL(2,\mathbb{C})$ on $\mathbb{C}^2$, though. We could also define an action by using instead of the matrix M, its complex conjugate $\bar{M}$, its inverse transpose $(M^t)^{-1}$ or its inverse hermitian adjoint $(M^\dagger)^{-1}$, since they all obey the same group multiplication law. These choices correspond, respectively to the *conjugate* representation $\bar{\mathbb{S}}$, the *dual* representation $\mathbb{S}^*$, and the *conjugate dual* representation $\bar{\mathbb{S}}^*$. We will use the following notation: if $u_\alpha \in \mathbb{S}$, then $u_{\dot\alpha} \in \bar{\mathbb{S}}$, $u^\alpha \in \mathbb{S}^*$ and $u^{\dot\alpha} \in \bar{\mathbb{S}}^*$. These representations are not all different, however. Indeed, we have that $\mathbb{S} \cong \mathbb{S}^*$ and $\bar{\mathbb{S}} \cong \bar{\mathbb{S}}^*$, which follows from the existence of $\epsilon_{\alpha\beta}$: an $SL(2,\mathbb{C})$-invariant tensor (since $\epsilon_{\alpha\beta} \mapsto M_\alpha{}^{\alpha'} M_\beta{}^{\beta'} \epsilon_{\alpha'\beta'} = (\det M)\epsilon_{\alpha\beta}$ and $\det M = 1$) which allows us to raise and lower indices in an $SL(2,\mathbb{C})$-covariant manner: $u^\alpha = \epsilon^{\alpha\beta} u_\beta$, and $u^{\dot\beta} = u_{\dot\alpha}\epsilon^{\dot\alpha\dot\beta}$. We use conventions where $\epsilon_{12} = 1$ and $\epsilon_{\dot 1\dot 2} = -1$.

Because both the Lie algebra $sl(2,\mathbb{C})$ (when viewed as a real Lie algebra) and $su(2) \times su(2)$ are real forms of the same complex Lie algebra, one often employs the notation (j, j') for representations of $SL(2,\mathbb{C})$, where j and j' are the spins of the two $su(2)$'s. In this notation the trivial one dimensional representation is denoted $(0,0)$, whereas $\mathbb{S} = (\frac{1}{2}, 0)$. The two $su(2)$'s are actually not independent but are related by complex conjugation, hence $\bar{\mathbb{S}} = (0, \frac{1}{2})$. In general, complex conjugation will interchange the labels. If the labels are the same, say $(\frac{1}{2}, \frac{1}{2})$, complex conjugation sends the representation to itself and it makes sense to restrict to the sub-representation which is fixed by complex conjugation. This is a real representation and in the case of the $(\frac{1}{2}, \frac{1}{2})$ representation of $SL(2,\mathbb{C})$, it coincides with the defining representation of the Lorentz group $SO(1,3)$: that is, the vector representation.

Indeed, given a 4-vector $P_\mu = (p_0, \vec{\mathbf{p}})$ we can turn it into a bispinor as follows:

$$\sigma \cdot P \equiv \sigma^\mu P_\mu = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix}$$

where $\sigma^\mu = (\mathbb{1}, \vec{\boldsymbol{\sigma}})$ with $\vec{\boldsymbol{\sigma}}$ the Pauli matrices:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Since the Pauli matrices are hermitian, so will be $\sigma \cdot P$ provided P_μ is real. The Pauli matrices have indices $\sigma^\mu_{\alpha\dot\alpha}$, which shows how $SL(2,\mathbb{C})$ acts on this space. If $M \in SL(2,\mathbb{C})$, then the action of M on such matrices is given by $\sigma \cdot P \mapsto M\, \sigma \cdot P\, M^\dagger$. This action is linear and preserves both the hermiticity of $\sigma \cdot P$ and the determinant $\det(\sigma \cdot P) = P^2 = p_0^2 - \vec{\mathbf{p}} \cdot \vec{\mathbf{p}}$, just as we expect of Lorentz transformations. We can summarise this discussion by saying that the $\sigma^\mu_{\alpha\dot\alpha}$ are Clebsch–Gordon coefficients intertwining between the "vector" and the $(\frac{1}{2}, \frac{1}{2})$ representations of $SL(2,\mathbb{C})$. Notice also that both M and $-M$ at the same way on bispinors, which reiterates the fact that $SL(2,\mathbb{C})$ is the double-cover of the Lorentz group $SO(1,3)$.

Finally we discuss the adjoint representation of the Lorentz group, which is generated by antisymmetric tensors $L_{\mu\nu} = -L_{\nu\mu}$. In terms of bispinors, such an $L_{\mu\nu}$ becomes a pair $(L_{\alpha\beta}, \bar{L}_{\dot\alpha\dot\beta})$ where $L_{\alpha\beta} = L_{\beta\alpha}$ and similarly for $\bar{L}_{\dot\alpha\dot\beta}$. In other words, $L_{\mu\nu}$ transforms as the $(1,0) \oplus (0,1)$ representation of $SL(2,\mathbb{C})$: notice that we need to take the direct sum because the representation is real.

2.1.2 The Coleman–Mandula and Haag–Łopuszański–Sohnius theorems

Back in the days when symmetry was everything, physicists spent a lot of time trying to unify the internal symmetries responsible for the observed particle spectrum and the Poincaré group into the same group: the holy grail being the so-called relativistic quark model. However their hopes were dashed by the celebrated no-go theorem of Coleman and Mandula. In a nutshell, this theorem states that the maximal Lie algebra of symmetries of the S-matrix of a unitary local relativistic quantum field theory obeying some technical but reasonable assumptions (roughly equivalent to demanding that the S-matrix be analytic), is a direct product of the Poincaré algebra with the Lie algebra of some *compact* internal symmetry group. Since Lie algebras of compact Lie groups are *reductive*: that is, the direct product of a semisimple and an abelian Lie algebras, the largest Lie algebra of symmetries of the S-matrix is a direct product: Poincaré × semisimple × abelian. In particular this implies that multiplets of the internal symmetry group consist of particles with the same mass and the same spin or helicity.

If all one-particle states are massless, then the symmetry is enhanced to conformal × semisimple × abelian; but the conclusions are unaltered: there is no way to unify the spacetime symmetries and the internal symmetries in a nontrivial way.

A wise person once said that inside every no-go theorem there is a "yes-go" theorem waiting to come out,[1] and the Coleman–Mandula theorem is no exception. The trick consists, not in trying to relax some of the assumptions on the S-matrix of the field theory, but in redefining the very notion of symmetry to encompass Lie superalgebras. In a classic paper Haag, Łopuzański and Sohnius re-examined the result of Coleman and Mandula in this new light and found the most general Lie superalgebra of symmetries of an S-matrix. The Coleman–Mandula theorem applies to the bosonic sector of the Lie superalgebra, so this is given again by Poincaré × reductive. In terms of representations of $SL(2,\mathbb{C})$, these generators transform according to the $(0,0)$, $(\frac{1}{2},\frac{1}{2})$, $(0,1)$ and $(1,0)$ representations. The singlets are the internal symmetry generators which we will denote collectively by B_ℓ. The $(\frac{1}{2},\frac{1}{2})$ generators correspond to the translations $P_{\alpha\dot\alpha}$, and the $(1,0)$ and $(0,1)$ generators are the Lorentz generators: $L_{\alpha\beta}$ and $\bar{L}_{\dot\alpha\dot\beta}$.

The novelty lies in the fermionic sector, which is generated by spinorial charges $Q_{\alpha I}$ in the $(\frac{1}{2},0)$ representation of $SL(2,\mathbb{C})$ and their hermitian adjoints $\bar{Q}^I_{\dot\alpha} = (Q_{\alpha I})^\dagger$ in the $(0,\frac{1}{2})$. Here I is a label running from 1 to some positive integer N. The Lie superalgebra generated by these objects is called

[1] and a wise guy said that we should call it a "go-go" theorem

the N-*extended super-Poincaré algebra*. The important Lie brackets are given by

$$\begin{array}{ll} [B_\ell, Q_{\alpha\, I}] = b_{\ell I}{}^J Q_{\alpha\, J} & [B_\ell, \bar{Q}^I_{\dot\alpha}] = -\bar{b}_\ell{}^I{}_J \bar{Q}^J_{\dot\alpha} \\ [P_{\alpha\dot\alpha}, Q_{\beta\, I}] = 0 & [P_{\alpha\dot\alpha}, \bar{Q}^I_{\dot\alpha}] = 0 \\ \{Q_{\alpha\, I}, Q_{\beta\, J}\} = 2\epsilon_{\alpha\beta} Z_{IJ} & \{\bar{Q}^I_{\dot\alpha}, \bar{Q}^J_{\dot\beta}\} = -2\bar\epsilon_{\dot\alpha\dot\beta} Z^{IJ} \\ \{Q_{\alpha\, I}, \bar{Q}^J_{\dot\alpha}\} = 2\delta_I{}^J P_{\alpha\dot\alpha} & [Z_{IJ}, \text{anything}] = 0 \end{array} \tag{2.1}$$

where $Z_{IJ} = z_{IJ}{}^m B_m$, $Z^{IJ} = (Z_{IJ})^\dagger$ and the coefficients $b_{\ell I}{}^J$ and $z_{IJ}{}^m$ must obey:

$$b_{\ell I}{}^K z_{KJ}{}^m + b_{\ell J}{}^K z_{IK}{}^m = 0 \ . \tag{2.2}$$

This last condition is nontrivial and constraints the structure of that part of the internal symmetry group which acts nontrivially on the spinorial charges of the supersymmetry algebra, what we will call the internal automorphism group of the supersymmetry algebra. In the absence of central charges, the internal automorphism group of the supersymmetry algebra is $U(N)$, but in the presence of the central charges, it gets restricted generically to $USp(N)$, since condition (2.2) can be interpreted as the invariance under the internal automorphism group of each of the antisymmetric forms $z_{IJ}{}^m$, for each fixed value of m. Notice that $Z_{IJ} = -Z_{JI}$, whence central charges requires $N \geq 2$.

The above Lie superalgebra is the most general symmetry of a local relativistic S-matrix in a theory describing point-particles. In the presence of extended objects: strings or, more generally, p-branes, the supersymmetry algebra receives extra terms involving topological conserved charges. These charges are no longer central since they fail to commute with the Lorentz generators; nevertheless they still commute with the spinorial charges and with the momentum generators. We will see an example of this later on when we discuss the six-dimensional N=1 supersymmetry algebra.

It is sometimes convenient, especially when considering supersymmetry algebras in dimensions other than 4, where there is no analogue to the isomorphism $Spin(1,3) \cong SL(2,\mathbb{C})$, to work with 4-spinors. We can assemble the spinorial changes $Q_{\alpha\, I}$ and $\bar{Q}^I_{\dot\alpha}$ into a Majorana spinor: $Q_I = (Q_{\alpha\, I}, \bar{Q}^{\dot\alpha\, I})^t$. The Dirac (=Majorana) conjugate is given by $\bar{Q}_I = (Q^\alpha{}_I, \bar{Q}^I_{\dot\alpha})$, and the relevant bit of the supersymmetry algebra is now given by

$$\{Q_I, \bar{Q}_J\} = 2\delta_{IJ}\gamma^\mu P_\mu + 2i(\operatorname{Im} Z_{IJ} + \gamma_5 \operatorname{Re} Z_{IJ}) \ , \tag{2.3}$$

where our conventions are such that

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar\sigma^\mu & 0 \end{pmatrix} \ , \tag{2.4}$$

and $\bar\sigma^\mu = (\mathbb{1}, -\vec{\boldsymbol{\sigma}})$.

2.2 Unitary representations of the supersymmetry algebra

The construction of unitary representations of the super-Poincaré algebra can be thought of as a mild extension of the construction of unitary representations of the Poincaré algebra. Because the Lorentz group is simple but noncompact, any nontrivial unitary representation is infinite-dimensional. The irreducible unitary representations are simply given by classical fields in Minkowski space subject to their equations of motion. Indeed the Klein–Gordon and Dirac equations, among others, can be understood as irreducibility constraints on the fields. The method of construction for the Poincaré algebra is originally due to Wigner and was greatly generalised by Mackey. The method consists of inducing the representation from a finite-dimensional unitary representation of some compact subgroup. Let us review this briefly.

2.2.1 Wigner's method and the little group

The Poincaré algebra has two casimir operators: P^2 and W^2, where $W^\mu = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}P_\nu L_{\lambda\rho}$ is the Pauli–Lubansky vector. By Schur's lemma, on an irreducible representation they must both act as multiplication by scalars. Let's focus on P^2. On an irreducible representation $P^2 = M^2$, where M is the "rest-mass" of the particle described by the representation. With our choice of metric, physical masses are real, whence $M^2 \geq 0$. We can thus distinguish two kinds of representations: *massless* for which $M^2 = 0$ and *massive* for which $M^2 > 0$.

Wigner's method starts by choosing a nonzero momentum k_μ on the mass-shell: $k^2 = M^2$. That is, this is a character (that is, a one-dimensional representation) of the translation subalgebra generated by the P_μ. We let G_k denote the subgroup of the Lorentz group (or rather of its double-cover $SL(2,\mathbb{C})$) which leaves k_μ invariant. G_k is known as the *little group*. Wigner's method, which we will not describe in any more detail than this, consists in inducing a representation of the Poincaré group from a finite-dimensional unitary representation of the little group. This is done by boosting the representation to fields on the mass shell and then Fourier transforming to yield fields on Minkowski space subject to their equations of motion.

In extending this method to the super-Poincaré algebra all that happens is that now the Lie algebra of the little group gets extended by the spinorial supersymmetry charges, since these commute with P_μ and hence stabilise the chosen 4-vector.

We will need to know about the structure of the little groups before introducing supersymmetry. The little group happens to be different for massive and for massless representations, as the next exercise asks you to show.

Exercise 2.1 (The little groups for positive-energy particles)
Let k_μ be a 4-vector obeying $k_0 > 0$, $k^2 = M^2 \geq 0$. Prove that the little group of k_μ is isomorphic to:

- *$SU(2)$, for $M^2 > 0$;*
- *$\widetilde{E}_2$, for $M^2 = 0$,*

where $E_2 \cong SO(2) \ltimes \mathbb{R}^2$, is the two-dimensional euclidean group and $\widetilde{E}_2 \cong Spin(2) \ltimes \mathbb{R}^2$ is its double cover.
(Hint: argue that two momenta k_μ which are Lorentz-related have isomorphic little groups. Then choose a convenient k_μ in each case, examine the action of $SL(2,\mathbb{C})$ on the bispinor $\sigma^\mu k_\mu$, and identify those $M \in SL(2,\mathbb{C})$ for which $M\sigma\cdot kM^\dagger = \sigma\cdot k$.)

The reason why we have restricted ourselves to positive-energy representations in this exercise, is that unitary representations of the supersymmetry algebra have non-negative energy. Indeed, for an arbitrary momentum k_μ, the supersymmetry algebra becomes

$$\{Q_{\alpha I}, \bar{Q}^J_{\dot\alpha}\} = 2\delta_I{}^J \begin{pmatrix} k_0 + k_3 & k_1 - ik_2 \\ k_1 + ik_2 & k_0 - k_3 \end{pmatrix} .$$

Therefore the energy k_0 of any state $|k\rangle$ with momentum k_μ can be written as follows (for a fixed but otherwise arbitrary I)

$$\begin{aligned} k_0 \| |k\rangle \|^2 &= \langle k|\, k_0 \,|k\rangle \\ &= \tfrac{1}{2}\langle k|\, \{Q_{1\,I}, \bar{Q}^I_{\dot 1}\}\, |k\rangle + \tfrac{1}{2}\langle k|\, \{Q_{2\,I}, \bar{Q}^I_{\dot 2}\}\, |k\rangle \\ &= \tfrac{1}{2}\|Q_{1\,I}|k\rangle\|^2 + \tfrac{1}{2}\|Q_{2\,I}|k\rangle\|^2 + \tfrac{1}{2}\|(Q_{1\,I})^\dagger|k\rangle\|^2 + \tfrac{1}{2}\|(Q_{2\,I})^\dagger|k\rangle\|^2 , \end{aligned}$$

whence k_0 is positive, unless $|k\rangle$ is annihilated by all the supersymmetry charges.

2.2.2 Massless representations

We start by considering massless representations. As shown in Exercise 2.1, the little group for the momentum k_μ of a massless particle is noncompact. Therefore its finite-dimensional unitary representations must all come from its maximal compact subgroup $Spin(2)$ and be trivial on the translation subgroup $\mathbb{R}^2$. The unitary representations of $Spin(2)$ are one-dimensional and indexed by a number $\lambda \in \frac{1}{2}\mathbb{Z}$ called the *helicity*. For CPT-invariance of the spectrum, it may be necessary to include both helicities $\pm\lambda$, but clearly all this does is double the states and we will not mention this again except to point out that some supersymmetry multiplets are CPT-self-conjugate.

Let's choose $k_\mu = (E, 0, 0, E)$, with $E > 0$. Then

$$\sigma^\mu k_\mu = \begin{pmatrix} 2E & 0 \\ 0 & 0 \end{pmatrix}$$

and the supersymmetry algebra becomes

$$\{Q_{\alpha\, I}, \bar{Q}^J_{\dot\alpha}\} = 4E\, \delta_I{}^J \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} .$$

In particular this means that $\{Q_{2\, I}, \bar{Q}^J_{\dot 2}\} = 0$. Because $\bar{Q}^J_{\dot 2} = (Q_{2\, J})^\dagger$, it follows that in a unitary representation $Q_{2\, I} = 0$ for all I. Indeed for any state $|\psi\rangle$,

$$0 = \langle\psi|\, \{Q_{2\, I}, (Q_{2\, I})^\dagger\}\, |\psi\rangle = \|Q_{2\, I}|\psi\rangle\|^2 + \|(Q_{2\, I})^\dagger|\psi\rangle\|^2 .$$

Plugging this back into the supersymmetry algebra (2.1) we see that $Z_{IJ} = \frac{1}{2}\{Q_{1\, I}, Q_{2\, J}\} = 0$, so that there are *no central charges for massless representations.*

Let us now introduce $q_I \equiv (1/2\sqrt{E})\, Q_{1\, I}$, in terms of which the supersymmetry algebra becomes

$$\{q_I, q_J^\dagger\} = \delta_{IJ} \qquad \{q_I, q_J\} = \{q_I^\dagger, q_J^\dagger\} = 0 .$$

We immediately recognise this is as a Clifford algebra corresponding to a $2N$-dimensional pseudo-euclidean space with signature (N, N). The irreducible representations of such Clifford algebras are well-known. We simply start with a *Clifford vacuum* $|\Omega\rangle$ satisfying

$$q_I\, |\Omega\rangle = 0 \qquad \text{for all } I = 1, \ldots, N ,$$

and we act repeatedly with the $q_I^\dagger$. Since $\{q_I^\dagger, q_J^\dagger\} = 0$, we obtain a 2^N-dimensional representation spanned by the vectors: $q_{I_1}^\dagger q_{I_2}^\dagger \cdots q_{I_p}^\dagger |\Omega\rangle$, where $1 \leq I_1 < I_2 < \cdots < I_p \leq N$, and $p = 0, \ldots, N$.

The Clifford vacuum actually carries quantum numbers corresponding to the momentum k and also to the helicity: $|\Omega\rangle = |k, \lambda\rangle$. It may also contain quantum numbers corresponding to the internal symmetry generators B_ℓ, but we ignore them in what follows.

Exercise 2.2 (Helicity content of massless multiplets)
Paying close attention to the helicity of the supersymmetry charges, prove that $Q_{1\, I}$ raises the helicity by $\frac{1}{2}$, whereas $Q_{2\, I}$ lowers it by the same amount. Deduce that the massless supersymmetry multiplet of helicity λ contains the following states:

States	*Helicity*	*Number*
$\|k,\lambda\rangle$	λ	*1*
$q_I^\dagger\|k,\lambda\rangle$	$\lambda+\frac{1}{2}$	N
$q_J^\dagger q_I^\dagger\|k,\lambda\rangle$	$\lambda+1$	$\binom{N}{2}$
$\vdots$	$\vdots$	$\vdots$
$q_{I_1}^\dagger q_{I_2}^\dagger\cdots q_{I_p}^\dagger\|k,\lambda\rangle$	$\lambda+p/2$	$\binom{N}{p}$
$\vdots$	$\vdots$	$\vdots$
$q_1^\dagger q_2^\dagger\cdots q_N^\dagger\|k,\lambda\rangle$	$\lambda+N/2$	*1*

Particularly interesting cases are the CPT-self-conjugate massless multiplets. First notice that CPT-self-conjugate multiplets can only exist for N even. For N=2 we have the helicity $\lambda=-\frac{1}{2}$ multiplet, whose spectrum consists of

Helicity	-1/2	0	1/2
Number	1	2	1

.

Then we have the N=4 gauge multiplet which has λ=-1 and whose spectrum is given by:

Helicity	-1	-1/2	0	1/2	1
Number	1	4	6	4	1

.

Pure (that is, without matter) N=4 supersymmetric Yang–Mills in four-dimensions consists of several of these multiplets—one for each generator of the gauge algebra. Finally, the third interesting case is the N=8 supergravity multiplet with λ=-2 and spectrum given by:

Helicity	-2	-3/2	-1	-1/2	0	1/2	1	3/2	2
Number	1	8	28	56	70	56	28	8	1

.

2.2.3 Massive representations

We now consider massive representations. As shown in Exercise 2.1, the little group for the momentum k_μ of a massive particle is $SU(2)$. Its finite-dimensional irreducible unitary representations are well-known: they are indexed by the *spin* s, where $2s$ is a non-negative integer, and have dimension $2s+1$.

A massive particle can always be boosted to its rest frame, so that we can choose a momentum $k_\mu=(M,0,0,0)$. Then

$$\sigma^\mu k_\mu = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}$$

and the supersymmetry algebra becomes

$$\{Q_{\alpha I}, \bar{Q}^J_{\dot\alpha}\} = 2M\, \delta_I{}^J\, \mathbb{1}_{\alpha\dot\alpha}\ .$$

No central charges

In the absence of central charges, $\{Q_{\alpha I}, Q_{\beta J}\} = 0$. Thus we can introduce $q_{\alpha I} \equiv (1/\sqrt{2M})Q_{\alpha,I}$, in terms of which the supersymmetry algebra is again a Clifford algebra:

$$\{q_{\alpha I}, q^\dagger_{\beta J}\} = \delta_{IJ}\delta_{\alpha\beta} \qquad \{q_{\alpha I}, q_{\beta J}\} = \{q^\dagger_{\alpha I}, q^\dagger_{\beta J}\} = 0\ ; \tag{2.5}$$

but where now the underlying pseudo-euclidean space is $4N$-dimensional with signature $(2N, 2N)$. The unique irreducible representation of such a Clifford algebra is now 2^{2N}-dimensional and it is built just as before from a Clifford vacuum by acting successively with the $q^\dagger_{\alpha I}$.

However unlike the case of massless representations, the Clifford vacuum is now degenerate since it carries spin: for spin s the Clifford vacuum is really a $(2s+1)$-dimensional $SU(2)$ multiplet. Notice that for fixed I, $q^\dagger_{\alpha I}$ transforms as a $SU(2)$-doublet of spin $\frac{1}{2}$. This must be taken into account when determining the spin content of the states in the supersymmetry multiplet. Instead of simply adding the helicities like in the massless case, now we must use the Clebsch–Gordon series to add the spins.

Exercise 2.3 (Highest spin in the multiplet)
Prove that the highest spin in the multiplet will be carried by states of the form $q^\dagger_{1\,1}q^\dagger_{1\,2}\cdots q^\dagger_{1\,N}$ acting on the Clifford vacuum, and that their spin is $s + N/2$.

For example, if $N = 1$ and $s = 0$, then we find the following spectrum: $|k,0\rangle$ with spin 0, $(q^\dagger_1|k,0\rangle, q^\dagger_2|k,0\rangle)$ with spin 1/2 and $q^\dagger_1 q^\dagger_2|k,0\rangle$ which has spin 0 too. The supersymmetric field theory describing this multiplet consists of a scalar field, a pseudo-scalar field, and a Majorana fermion: it is the celebrated Wess–Zumino model and the multiplet is known as the massive Wess–Zumino multiplet. Another example that will be important to us is the N=2 multiplets with spins s=0 and s=1/2, which we leave as an exercise.

Exercise 2.4 (Massive N=2 multiplets with s=0 and s=1/2)
Work out the spin content of the massive N=2 multiplets without central charges and with spins s=0 and s=1/2. Show that for s=0 the spin content is $(0^5, \frac{1}{2}^4, 1)$ in the obvious notation, and for s=1/2 it is given by $(3/2, 1^4, \frac{1}{2}^6, 0^4)$.

Adding central charges

Adding central charges changes the nature of the supersymmetry algebra, which now becomes

$$\{Q_{\alpha I}, \bar{Q}^J_{\dot\alpha}\} = 2M\, \delta_I{}^J\, \mathbb{1}_{\alpha\dot\alpha} \qquad \{Q_{\alpha I}, Q_{\beta J}\} = 2\epsilon_{\alpha\beta} Z_{IJ} \ .$$

Because Z_{IJ} is antisymmetric, we can rotate the $Q_{\alpha I}$ unitarily—which is an automorphism of the first of the above brackets—in such a way that Z_{IJ} takes a standard form:

$$Z_{IJ} = \begin{pmatrix} 0 & z_1 & & & & & & \\ -z_1 & 0 & & & & & & \\ & & 0 & z_2 & & & & \\ & & -z_2 & 0 & & & & \\ & & & & \ddots & & & \\ & & & & & \ddots & & \\ & & & & & & 0 & z_{N/2} \\ & & & & & & -z_{N/2} & 0 \end{pmatrix}$$

where the z_i can be chosen to be real and non-negative. To simplify the discussion we have assumed that N is even, but one should keep in mind that if N is odd, there will of course be a zero 1×1 block in the above normal form for Z_{IJ}.

Let us break up the index I into a pair (A,i) where $A = 1,2$ and $i = 1,\ldots,N/2$. In terms of these indices the supersymmetry algebra can be rewritten as

$$\{Q_{\alpha\, Ai}, Q^\dagger_{\beta\, Bj}\} = 2M\, \delta_{ij}\delta_{\alpha\beta}\delta_{AB} \qquad \{Q_{\alpha\, Ai}, Q_{\beta\, Bj}\} = 2\epsilon_{\alpha\beta}\epsilon_{AB}\delta_{ij} z_i \ .$$

Define the following linear combinations

$$S^\pm_{\alpha\, i} \equiv \tfrac{1}{2}\left(Q_{\alpha\, 1i} \pm \epsilon^{\dot\alpha\dot\beta} \bar{Q}^{2i}_{\dot\beta} \right)$$

where we have raised the spinor index in the second term in order to preserve covariance under the little group $SU(2)$. In terms of these generators, the algebra is now:

$$\{S^\pm_{\alpha\, i}, (S^\pm_{\beta\, j})^\dagger\} = \delta_{\alpha\beta}\delta_{ij}\,(M \pm z_i)$$

with all other brackets being zero. Notice that acting on any state $|\psi\rangle$,

$$\begin{aligned}(M \pm z_i)\||\psi\rangle\|^2 &= \langle\psi|(M \pm z_i)|\psi\rangle \\ &= \langle\psi|\{S^\pm_{1\, i}, (S^\pm_{1\, i})^\dagger\}|\psi\rangle \\ &= \|S^\pm_{1\, i}|\psi\rangle\|^2 + \|(S^\pm_{1\, i})^\dagger|\psi\rangle\|^2 \ ,\end{aligned}$$

from where it follows that $M \pm z_i \geq 0$ for all i, or

$$\boxed{M \geq |z_i| \qquad \text{for all } i = 1, \ldots, N/2}\ , \tag{2.6}$$

which is reminiscent of the Bogomol'nyi bound (1.27). Notice that this bound is an unavoidable consequence of having a unitary representation of the supersymmetry algebra. Therefore provided that supersymmetry is not broken quantum-mechanically, the bound will be maintained.

Suppose that $M > z_i$ for all i. Then we can define $q^{\pm}_{\alpha i} \equiv (1/\sqrt{M \pm z_i}) S^{\pm}_{\alpha i}$, in terms of which the supersymmetry algebra is again given by equation (2.5) once we recombine the indices $(\pm, i)$ into I. Therefore we are back in the case of massive representations without central charges, at least as far as the dimension of the representations is concerned.

Suppose instead that some of the z_i saturate the bound (2.6): $z_i = M$ for $i = 1, \ldots, q \leq N/2$. Then a similar argument as in the discussion of the massless representations allows us to conclude that the $2q$ generators $S^{-}_{\alpha i}$ for $i = 1, \ldots, q$ act trivially and can be taken to be zero. The remaining $2N - 2q$ generators obey a Clifford algebra whose unique irreducible representation has dimension 2^{2N-2q}. Notice that the smallest representation occurs when all central charges saturate the bound (2.6), in which case all the $S^{-}_{\alpha i} = 0$ and we are left only with 2^N states, just as in the case of a massless multiplet. These massive multiplets are known as *short multiplets*.

For example, in N=2 there is only one $z = z_1$. If $z < M$ the massive multiplet contains $2^4 = 16$ states, whereas if $z = M$ the short multiplet only contains $2^2 = 4$ states. For N=4, there are two z_i. If both $z_i < M$, then the massive multiplet has $2^8 = 256$ states, whereas if both $z_i = M$, then the short multiplet contains only $2^4 = 16$ states. Half-way we find the case when exactly one of the $z_i = M$, in which case the dimension of the multiplet is $2^6 = 64$. Strictly speaking we shouldn't call these numbers the dimension of the multiplet, but rather the *degeneracy*, since it may be that the Clifford vacuum is degenerate, in which case the dimension of the supersymmetry multiplet is the product of what we've been calling the dimension of the multiplet and that of the Clifford vacuum. Let us work out some examples. We first work out the case of N=2 and spins s=0 and s=$\frac{1}{2}$ in the following exercise.

Exercise 2.5 (Short N=2 multiplets with s=0 and s=1/2)
Prove that the spin contents of the short multiplet with s=0 is $(\frac{1}{2}, 0^2)$ and that of the short multiplet with s=1/2 is $(1, \frac{1}{2}^2, 0)$. Compare with the results of Exercise 2.4, which are the spin contents when the central charge does not saturate the bound. We will see that the s=0 multiplet contains the BPS-monopole, whereas the s=1/2 multiplet contains the massive vector bosons.

Next we take a look at the short N=4 multiplets with s=0. These will be the important ones when we discuss N=4 supersymmetric Yang–Mills.

Exercise 2.6 (Short N=4 multiplets with s=0)
Prove that the spin content of the N=4 short multiplet with s=0 is $(1, \frac{1}{2}^4, 0^5)$, which totals the expected 16 states. As we will see later, this will be the multiplet containing both the BPS-monopole and the massive vector boson.

This difference in the dimension of representations for which the bound (2.6) is saturated is responsible for the fact that if a multiplet saturates the bound classically, it will continue to do so when perturbative quantum corrections are taken into account. This is because perturbative quantum corrections do not alter the number of degrees of freedom, hence a short multiplet (that is, one which saturates the bound) cannot all of a sudden undergo the explosion in size required to obey the bound strictly.

2.3 N=2 Supersymmetric Yang-Mills

The supersymmetric bound (2.6) for massive representations with central charges may seem a little abstract, but it comes to life in particular field theoretical models, where we can explicitly calculate the central charges in terms of the field variables. We will see this first of all in pure N=2 supersymmetric Yang–Mills, which embeds the bosonic part of the Georgi–Glashow model. This result is due to Witten and Olive [WO78].

We could simply write the action down and compute the supersymmetry algebra directly as was done in [WO78], but it is much more instructive to derive it by dimensional reduction from the N=1 supersymmetric Yang–Mills action in six dimensions. This derivation of N=2 supersymmetric Yang–Mills by dimensional reduction was first done in [DHdV78], and the six-dimensional computation of the central charges was first done in [Oli79].

That there should be a N=1 supersymmetric Yang–Mills theory in six dimensions is not obvious: unlike its nonsupersymmetric counterpart, supersymmetric Yang–Mills theories only exist in a certain number of dimensions. Of course one can always write down the Yang–Mills action in any dimension and then couple it to fermions, but supersymmetry requires a delicate balance between the bosonic and fermionic degrees of freedom. A gauge field in d dimensions has $d-2$ physical degrees of freedom corresponding to the transverse polarisations. The number of degrees of freedom of a fermion field depends on what kind fermion it is, but it always a power of 2. An unconstrained Dirac spinor in d dimensions has $2^{d/2}$ or $2^{(d-1)/2}$ real degrees of freedom, for d even or odd respectively: a Dirac spinor has $2^{d/2}$ or $2^{(d-1)/2}$

complex components but the Dirac equation cuts this number in half. In even dimensions, one can further restrict the spinor by imposing that it be *chiral* or *Weyl*. This cuts the number of degrees of freedom by two. Alternatively, in some dimensions (depending on the signature of the metric) one can impose a *reality* or *Majorana* condition which also halves the number of degrees of freedom. For a lorentzian metric of signature $(1, d-1)$, Majorana spinors exist for $d \equiv 1, 2, 3, 4 \mod 8$. When $d \equiv 2 \mod 8$ one can in fact impose that a spinor be both Majorana and Weyl, cutting the number of degrees of freedom in four. The next exercise asks you to determine in which dimensions can supersymmetric Yang–Mills theory exist based on the balance between bosonic and fermionic degrees of freedom.

Exercise 2.7 (N**=1 supersymmetric Yang–Mills)**
Verify via a counting of degrees of freedom that N*=1 supersymmetric Yang–Mills can exist only in the following dimensions and with the following types of spinors:*

d	*Spinor*
3	*Majorana*
4	*Majorana or Weyl*
6	*Weyl*
10	*Majorana–Weyl*

It is a curious fact that these are precisely the dimensions in which the classical superstring exists. Unlike superstring theory, in which only the ten-dimensional theory survives quantisation, it turns out that supersymmetric Yang–Mills theory exists in each of these dimensions. Although we are mostly concerned with four-dimensional field theories in these notes, the six-dimensional and ten-dimensional theories are useful tools since upon dimensional reduction to four dimensions they yield N=2 and N=4 supersymmetric Yang–Mills, respectively.

2.3.1 N=1 d=6 supersymmetric Yang–Mills

We start by setting some conventions. We will let uppercase Latin indices from the beginning of the alphabet $A, B, \ldots$ take the values $0, 1, 2, 3, 5, 6$. Our metric η_{AB} is "mostly minus"; that is, with signature $(1, 5)$. We choose the following explicit realisation of the Dirac matrices:

$$\Gamma_\mu = \begin{pmatrix} 0 & \gamma_\mu \\ \gamma_\mu & 0 \end{pmatrix} \quad \Gamma_5 = \begin{pmatrix} 0 & \gamma_5 \\ \gamma_5 & 0 \end{pmatrix} \quad \Gamma_6 = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}$$

where $\mu = 0, 1, 2, 3$, and where γ_μ are defined in (2.4) and $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$. The Γ_A obey the Clifford algebra

$$\{\Gamma_A, \Gamma_B\} = 2\,\eta_{AB}\,\mathbb{1}\ .$$

Weyl spinors are defined relative to Γ_7, which is defined by

$$\Gamma_7 = \Gamma_0\Gamma_1\cdots\Gamma_6 = \begin{pmatrix} -\mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}\ .$$

We now write down the action for N=1 supersymmetric Yang–Mills. We will take the gauge group to be $SO(3)$ for definiteness, but it should be clear that the formalism is general. As before we will identify the Lie algebra $so(3)$ with $\mathbb{R}^3$ but we will now drop the arrows on the vectors to unclutter the notation, hoping it causes no confusion. The lagrangian density is given by

$$\mathcal{L} = -\tfrac{1}{4}\mathsf{G}^{AB}\cdot\mathsf{G}_{AB} + \tfrac{i}{2}\,\bar{\boldsymbol{\Psi}}\cdot\Gamma^A \overleftrightarrow{D}_A\boldsymbol{\Psi}\ , \tag{2.7}$$

where

$$\begin{aligned} \mathsf{G}_{AB} &= \partial_A\mathsf{W}_B - \partial_B\mathsf{W}_A - e\,\mathsf{W}_A\times\mathsf{W}_B \\ D_A\boldsymbol{\Psi} &= \partial_A\boldsymbol{\Psi} - e\,\mathsf{W}_A\times\boldsymbol{\Psi}\ , \end{aligned}$$

and where $\boldsymbol{\Psi}$ is a complex Weyl spinor obeying $\Gamma_7\boldsymbol{\Psi} = -\boldsymbol{\Psi}$. The Dirac conjugate spinor is defined by $\bar{\boldsymbol{\Psi}} = \boldsymbol{\Psi}^\dagger\,\Gamma_0$, and obeys $\bar{\boldsymbol{\Psi}}\Gamma_7 = \bar{\boldsymbol{\Psi}}$. Finally we have used the convenient shorthand $\overleftrightarrow{D}_A$ to mean

$$\bar{\boldsymbol{\Psi}}\cdot\Gamma^A\overleftrightarrow{D}_A\boldsymbol{\Psi} = \bar{\boldsymbol{\Psi}}\cdot\Gamma^A D_A\boldsymbol{\Psi} - D_A\,\bar{\boldsymbol{\Psi}}\cdot\Gamma^A\boldsymbol{\Psi}\ .$$

The action defined by (2.7) is manifestly gauge invariant, but it is also invariant under supersymmetry. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be two constant anticommuting Weyl spinors of the same chirality as $\boldsymbol{\Psi}$. Let us define the following transformations:

$$\begin{aligned} \delta\,\mathsf{W}_A &= i\bar{\boldsymbol{\alpha}}\Gamma_A\boldsymbol{\Psi} & \bar{\delta}\,\mathsf{W}_A &= -i\bar{\boldsymbol{\Psi}}\Gamma_A\boldsymbol{\beta} \\ \delta\,\boldsymbol{\Psi} &= 0 & \bar{\delta}\,\boldsymbol{\Psi} &= \tfrac{1}{2}\mathsf{G}^{AB}\Gamma_{AB}\boldsymbol{\beta} \\ \delta\,\bar{\boldsymbol{\Psi}} &= -\tfrac{1}{2}\bar{\boldsymbol{\alpha}}\mathsf{G}^{AB}\Gamma_{AB} & \bar{\delta}\,\bar{\boldsymbol{\Psi}} &= 0 \end{aligned} \tag{2.8}$$

where $\Gamma_{AB} = \frac{1}{2}(\Gamma_A\Gamma_B - \Gamma_B\Gamma_A)$. We should remark that there is only one supersymmetry in our theory: $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are chiral. That is, there is only one spinorial charge Q, in terms of which the transformations δ and $\bar{\delta}$ defined above can be understood as follows:

$$\delta\phi = [\bar{\boldsymbol{\alpha}}Q, \phi] \quad\text{and}\quad \bar{\delta}\phi = [\bar{Q}\boldsymbol{\beta}, \phi]\ ,$$

for any field ϕ. Notice that it follows from this that the action of $\bar{\delta}$ can be deduced from that of δ as follows: $\bar{\delta}\phi = (\delta\phi^\dagger)^\dagger$ (apart from the obvious change of $\boldsymbol{\alpha}$ to $\boldsymbol{\beta}$, of course). Keep in mind that we have chosen the Lie algebra structure constants to be real, whence the generators are antihermitian.

We claim that $\mathcal{L}$ is invariant under δ and $\bar{\delta}$ above up to a divergence. In order to derive the supersymmetry current, we will actually take $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to depend on the position and simply vary the lagrangian density. We expect a total divergence plus a term with the current multiplying the derivative of the parameter. The calculations will take us until the end of the section and are contained in the following set of exercises.

Exercise 2.8 (Supersymmetry variation of $\mathcal{L}$)
Prove first of all that for any derivation δ,

$$\delta\mathsf{G}_{AB} = D_A\delta\mathsf{W}_B - D_B\delta\mathsf{W}_A \ ,$$

and conclude that the variation of the bosonic part of the action $\mathcal{L}_\mathrm{b}$ is given by

$$\delta\mathcal{L}_\mathrm{b} = -i\mathsf{G}^{AB}\cdot D_A(\bar{\boldsymbol{\alpha}}\Gamma_B\boldsymbol{\Psi}) \qquad \textit{and} \qquad \bar{\delta}\mathcal{L}_\mathrm{b} = i\mathsf{G}^{AB}\cdot D_A(\bar{\boldsymbol{\Psi}}\Gamma_B\boldsymbol{\beta}) \ .$$

Next we tackle the fermions. Prove the following identities:

$$\begin{aligned}\delta\left(D_A\boldsymbol{\Psi}\right) &= -ie\left(\bar{\boldsymbol{\alpha}}\Gamma_A\boldsymbol{\Psi}\right)\times\boldsymbol{\Psi}\\ \delta\left(D_A\bar{\boldsymbol{\Psi}}\right) &= -\tfrac{1}{2}D_A\left(\bar{\boldsymbol{\alpha}}\mathsf{G}^{BC}\Gamma_{BC}\right) - ie\left(\bar{\boldsymbol{\alpha}}\Gamma_A\boldsymbol{\Psi}\right)\times\bar{\boldsymbol{\Psi}}\end{aligned}$$

and

$$\begin{aligned}\bar{\delta}\left(D_A\boldsymbol{\Psi}\right) &= \tfrac{1}{2}D_A\left(\mathsf{G}^{BC}\Gamma_{BC}\boldsymbol{\beta}\right) + ie\left(\bar{\boldsymbol{\Psi}}\Gamma_A\boldsymbol{\beta}\right)\times\boldsymbol{\Psi}\\ \bar{\delta}\left(D_A\bar{\boldsymbol{\Psi}}\right) &= ie\left(\bar{\boldsymbol{\Psi}}\Gamma_A\boldsymbol{\beta}\right)\times\bar{\boldsymbol{\Psi}} \ ,\end{aligned}$$

and conclude that the variation of the fermionic part of the action $\mathcal{L}_\mathrm{f}$ is given by

$$\delta\mathcal{L}_\mathrm{f} = \tfrac{i}{4}D_A\left(\bar{\boldsymbol{\alpha}}\mathsf{G}^{BC}\right)\cdot\Gamma_{BC}\Gamma^A\boldsymbol{\Psi} - \tfrac{i}{4}\bar{\boldsymbol{\alpha}}\mathsf{G}^{BC}\cdot\Gamma_{BC}\Gamma^A D_A\boldsymbol{\Psi} + e\bar{\boldsymbol{\Psi}}\cdot\left(\bar{\boldsymbol{\alpha}}\Gamma^A\boldsymbol{\Psi}\right)\times\Gamma_A\boldsymbol{\Psi}$$

and

$$\bar{\delta}\mathcal{L}_\mathrm{f} = \tfrac{i}{4}\bar{\boldsymbol{\Psi}}\Gamma^A\Gamma^{BC}\cdot D_A\left(\mathsf{G}_{BC}\boldsymbol{\beta}\right) - \frac{i}{4}D_A\bar{\boldsymbol{\Psi}}\Gamma^A\Gamma^{BC}\cdot\mathsf{G}_{BC}\boldsymbol{\beta} - e\bar{\boldsymbol{\Psi}}\cdot\left(\bar{\boldsymbol{\Psi}}\Gamma^A\boldsymbol{\beta}\right)\times\Gamma_A\boldsymbol{\Psi} \ .$$

Supersymmetry invariance demands, in particular, that the fermion trilinear terms in $\delta\mathcal{L}_\mathrm{f}$ should cancel. This requires a Fierz rearrangement, and this is as good a time as any to discuss this useful technique. Writing explicitly the Lie algebra indices on the fermions, the trilinear terms in $\delta\mathcal{L}_\mathrm{f}$ become

$$e\epsilon_{abc}\left(\bar{\boldsymbol{\alpha}}\Gamma^A\boldsymbol{\Psi}^a\right)\left(\bar{\boldsymbol{\Psi}}^c\Gamma_A\boldsymbol{\Psi}^b\right) \ . \tag{2.9}$$

Let us focus on the expression $\boldsymbol{\Psi}^a\,\bar{\boldsymbol{\Psi}}^c$. This is a bispinor. Since spinors in six dimensions have 8 components, bispinors form a 64-dimensional vector space spanned by the antisymmetrised products of Γ-matrices:

$$\mathbb{1},\quad \Gamma_A,\quad \Gamma_{AB},\quad \Gamma_{ABC},\quad \Gamma_{ABCD},\quad \Gamma_{ABCDE}\quad \text{and}\quad \Gamma_{ABCDEF}\ ,$$

or equivalently

$$\mathbb{1},\quad \Gamma_A,\quad \Gamma_{AB},\quad \Gamma_{ABC},\quad \Gamma_{AB}\Gamma_7,\quad \Gamma_A\Gamma_7\quad \text{and}\quad \Gamma_7\ .$$

(Notice that antisymmetrisation is defined by

$$\Gamma_{A_1A_2\cdots A_p} = \Gamma_{[A_1}\Gamma_{A_2}\cdots\Gamma_{A_p]} = \frac{1}{p!}\sum_{\sigma\in\mathfrak{S}_p}\operatorname{sign}\sigma\,\Gamma_{A_{\sigma(1)}}\Gamma_{A_{\sigma(2)}}\cdots\Gamma_{A_{\sigma(p)}}\ ,$$

so that it has "strength one.")

We will let $\{M_\Lambda\}$ denote collectively these matrices. The above basis is orthogonal relative to the inner product defined by the trace:

$$\operatorname{tr} M_\Lambda M_{\Lambda'} = c_\Lambda \delta_{\Lambda\Lambda'}\ ,$$

which allows us to expand

$$\boldsymbol{\Psi}^a\,\bar{\boldsymbol{\Psi}}^c = \sum_\Lambda b_\Lambda{}^{ac} M_\Lambda\ ,$$

and to compute the coefficients $b_\Lambda{}^{ac}$ simply by taking traces. Remembering that $\boldsymbol{\Psi}^a$ are anticommuting, we find

$$b_\Lambda{}^{ac} = -\frac{1}{c_\Lambda}\left(\bar{\boldsymbol{\Psi}}^c M_\Lambda \boldsymbol{\Psi}^a\right)\ . \tag{2.10}$$

Exercise 2.9 (A Fierz rearrangement)
Using the above formula and computing the relevant traces, prove that

$$\boldsymbol{\Psi}^a\,\bar{\boldsymbol{\Psi}}^c = -\tfrac{1}{8}\left(\bar{\boldsymbol{\Psi}}^c\Gamma_A\boldsymbol{\Psi}^a\right)\Gamma^A(\mathbb{1}+\Gamma_7) - \tfrac{1}{48}\left(\bar{\boldsymbol{\Psi}}^c\Gamma_{ABC}\boldsymbol{\Psi}^a\right)\Gamma^{ABC}\ .$$

(Hint: use the fact that $\Gamma_7\boldsymbol{\Psi} = -\boldsymbol{\Psi}$ to discard from the start many of the terms in the general Fierz expansion.)

We now use this Fierz rearrangement to rewrite the trilinear term (2.9) as follows:

$$-\tfrac{1}{8}e\epsilon_{abc}\left(\bar{\boldsymbol{\alpha}}\Gamma^A\Gamma^B(\mathbb{1}+\Gamma_7)\Gamma_A\boldsymbol{\Psi}^b\right)\left(\bar{\boldsymbol{\Psi}}^c\Gamma_B\boldsymbol{\Psi}^a\right) - \tfrac{1}{48}e\epsilon_{abc}\left(\bar{\boldsymbol{\alpha}}\Gamma^A\Gamma^{BCD}\Gamma_A\boldsymbol{\Psi}^b\right)\left(\bar{\boldsymbol{\Psi}}^c\Gamma_{BCD}\boldsymbol{\Psi}^a\right)\ ,$$

which using that $\boldsymbol{\Psi}$ is Weyl, can be simplified to

$$-\tfrac{1}{4}e\epsilon_{abc}\left(\bar{\boldsymbol{\alpha}}\Gamma^A\Gamma^B\Gamma_A\boldsymbol{\Psi}^b\right)\left(\bar{\boldsymbol{\Psi}}^c\Gamma_B\boldsymbol{\Psi}^a\right) - \tfrac{1}{48}e\epsilon_{abc}\left(\bar{\boldsymbol{\alpha}}\Gamma^A\Gamma^{BCD}\Gamma_A\boldsymbol{\Psi}^b\right)\left(\bar{\boldsymbol{\Psi}}^c\Gamma_{BCD}\boldsymbol{\Psi}^a\right)\ .$$

Exercise 2.10 (Some Γ-matrix identities)
Prove the following two identities:

$$\Gamma^A\Gamma^B\Gamma_A = -4\,\Gamma^B \qquad \text{and} \qquad \Gamma^A\Gamma^{BCD}\Gamma_A = 0\ , \tag{2.11}$$

and deduce that the trilinear terms cancel exactly. The above identities are in fact the minor miracle that makes supersymmetric Yang–Mills possible in six dimensions.

Up to a divergence, the remaining terms in the supersymmetric variation of the lagrangian density $\mathfrak{L}$ are then:

$$\delta\mathfrak{L} = -i\mathsf{G}^{AB}\cdot D_A\left(\bar{\boldsymbol{\alpha}}\Gamma_B\boldsymbol{\Psi}\right) - \frac{i}{2}\mathsf{G}^{BC}\cdot\bar{\boldsymbol{\alpha}}\Gamma_{BC}\Gamma^A D_A\boldsymbol{\Psi}$$

and

$$\bar{\delta}\mathfrak{L} = i\mathsf{G}^{AB}\cdot D_A\left(\bar{\boldsymbol{\Psi}}\Gamma_B\boldsymbol{\beta}\right) - \frac{i}{2}\mathsf{G}_{BC}\cdot D_A\bar{\boldsymbol{\Psi}}\Gamma^A\Gamma^{BC}\boldsymbol{\beta}\ .$$

Exercise 2.11 (... and the proof of supersymmetry invariance)
Prove the following identity between Dirac matrices

$$\Gamma_{AB}\Gamma_C = \Gamma_{ABC} + \eta_{BC}\,\Gamma_A - \eta_{AC}\,\Gamma_B\ ,$$

and use it to rewrite the supersymmetric variations of $\mathfrak{L}$ as

$$\delta\mathfrak{L} = \frac{i}{2}\partial_A\bar{\boldsymbol{\alpha}}\,\mathsf{G}^{BC}\cdot\Gamma_{BC}\Gamma^A\boldsymbol{\Psi} \qquad \text{and} \qquad \bar{\delta}\mathfrak{L} = \frac{i}{2}\bar{\boldsymbol{\Psi}}\Gamma^A\Gamma_{BC}\cdot\mathsf{G}^{BC}\partial_A\boldsymbol{\beta}\ , \tag{2.12}$$

again up to divergences and where we have used the Bianchi identity in the form $\Gamma_{ABC}D^A\mathsf{G}^{BC} = 0$. This proves the invariance of $\mathfrak{L}$ under the supersymmetry transformations (2.8).

From (2.12) we can read the expression for the supersymmetry currents:

$$J^A = \frac{i}{2}\mathsf{G}^{BC}\cdot\Gamma_{BC}\Gamma^A\boldsymbol{\Psi} \qquad \text{and} \qquad \bar{J}^A = \frac{i}{2}\bar{\boldsymbol{\Psi}}\Gamma^A\Gamma_{BC}\cdot\mathsf{G}^{BC}\ .$$

As usual the spinorial supersymmetry charge is the space integral of the zero component of the current. Provided we already knew that $\mathfrak{L}$ is supersymmetric, there is a more economical way to derive the expression of the supercurrent. This uses the fact that the supercurrent is part of a supersymmetry multiplet.

Exercise 2.12 (The supersymmetry multiplet)
Prove that the lagrangian density (2.7) *is invariant under the transformation $\boldsymbol{\Psi} \mapsto \exp(i\theta)\,\boldsymbol{\Psi}$, $\bar{\boldsymbol{\Psi}} \mapsto \exp(-i\theta)\,\bar{\boldsymbol{\Psi}}$, and that the corresponding Noether current is given by $j_A = \bar{\boldsymbol{\Psi}}\cdot\Gamma_A\boldsymbol{\Psi}$. Prove that*

$$\delta j_A = i\bar{\boldsymbol{\alpha}}J_A \qquad \text{and} \qquad \bar{\delta}j_A = -i\bar{J}_A\boldsymbol{\beta}\ .$$

The supersymmetry multiplet also contains the energy-momentum tensor, alone or in combination with other topological currents that may appear in the right hand side of $\{Q, \bar{Q}\}$ in the supersymmetry algebra. We will use this later to compute the supersymmetry algebra corresponding to six-dimensional supersymmetric Yang–Mills. But first we perform the dimensional reduction to four dimensions.

2.3.2 From N=1 in d=6 to N=2 in d=4

Let us single out two of the coordinates (x^5, x^6) in six dimensions and assume that none of our fields depend on them: $\partial_5 \equiv \partial_6 \equiv 0$. This breaks $SO(1,5)$ Lorentz invariance down to $SO(1,3) \times SO(2)$. Let us therefore decompose our six-dimensional fields in a way that reflects this. In fact, we will at first ignore the $SO(2)$ invariance and focus only on the behaviour of the components of the six-dimensional fields under the action of $SO(1,3)$. The gauge field W_A breaks up into a vector W_μ, a pseudo-scalar $\mathsf{P} = \mathsf{W}_5$ and a scalar $\mathsf{S} = \mathsf{W}_6$. In terms of these fields, the field-strength breaks up as $\mathsf{G}_{\mu\nu}$, $\mathsf{G}_{\mu 5} = D_\mu \mathsf{P}$, $\mathsf{G}_{\mu 6} = D_\mu \mathsf{S}$ and $\mathsf{G}_{56} = e\, \mathsf{S} \times \mathsf{P}$. Meanwhile, the Weyl spinor breaks up as $\boldsymbol{\Psi} = \binom{\boldsymbol{\psi}}{0}$, where $\boldsymbol{\psi}$ is an unconstrained (complex) Dirac spinor. The covariant derivative of the spinor then breaks up as $(D_\mu \boldsymbol{\psi}, -e\mathsf{P} \times \boldsymbol{\psi}, -e\mathsf{S} \times \boldsymbol{\psi})$.

The lagrangian density now becomes $\mathcal{L} = \mathcal{L}_{\mathrm{b}} + \mathcal{L}_{\mathrm{f}}$ where

$$\mathcal{L}_{\mathrm{b}} = -\tfrac{1}{4} \mathsf{G}_{\mu\nu} \cdot \mathsf{G}^{\mu\nu} + \tfrac{1}{2} D_\mu \mathsf{P} \cdot D^\mu \mathsf{P} + \tfrac{1}{2} D_\mu \mathsf{S} \cdot D^\mu \mathsf{S} - \tfrac{1}{2} e^2 \|\mathsf{P} \times \mathsf{S}\|^2$$

and

$$\mathcal{L}_{\mathrm{f}} = i\bar{\boldsymbol{\psi}} \cdot \gamma^\mu D_\mu \boldsymbol{\psi} + ie\bar{\boldsymbol{\psi}} \cdot \gamma_5 \mathsf{P} \times \boldsymbol{\psi} + ie\bar{\boldsymbol{\psi}} \cdot \mathsf{S} \times \boldsymbol{\psi} \ , \tag{2.13}$$

where we see that P is indeed as pseudo-scalar as claimed. $\mathcal{L}$ is the lagrangian density of N=2 supersymmetric Yang–Mills theory in four dimensions. The supersymmetry parameter $\boldsymbol{\alpha}$, which in the six-dimensional theory is a Weyl spinor, becomes upon dimensional reduction a Dirac spinor. But in four dimensions the supersymmetry parameters are Majorana, hence this gives rise to N=2 supersymmetry. One can see this explicitly by breaking up the supersymmetry parameter into its Majorana components: simply choose a Majorana representation and split it into its real and imaginary parts. Each of these spinors is Majorana and generates one supersymmetry.

Let us first do this with $\boldsymbol{\psi}$. The next exercise shows the resulting fermion action in a Majorana basis.

Exercise 2.13 ($\mathcal{L}$ in a Majorana basis)
In a Majorana basis, let us split $\boldsymbol{\psi}$ *as follows:*

$$\boldsymbol{\psi} = \frac{1}{\sqrt{2}} (\boldsymbol{\psi}_1 - i\boldsymbol{\psi}_2) \ .$$

Prove that relative to $\boldsymbol{\psi}_\alpha$, $\alpha = 1, 2$, *the fermionic part* $\mathcal{L}_\mathrm{f}$ *of the lagrangian density becomes (up to a total derivative)* ✓

$$\mathcal{L}_\mathrm{f} = \frac{i}{2}\bar{\boldsymbol{\psi}}_1 \cdot \gamma^\mu D_\mu \boldsymbol{\psi}_1 + \frac{i}{2}\bar{\boldsymbol{\psi}}_2 \cdot \gamma^\mu D_\mu \boldsymbol{\psi}_2 + e\bar{\boldsymbol{\psi}}_1 \cdot \gamma_5 \mathsf{P} \times \boldsymbol{\psi}_2 + e\bar{\boldsymbol{\psi}}_1 \cdot \mathsf{S} \times \boldsymbol{\psi}_2 \ . \quad (2.14)$$

(Hint: you may find useful the following identities for anticommuting Majorana spinors in four dimensions:

$$\bar{\chi}\lambda = \bar{\lambda}\chi \qquad \bar{\chi}\gamma_5\lambda = -\bar{\lambda}\gamma_5\chi \qquad \bar{\chi}\gamma_\mu\lambda = -\bar{\lambda}\gamma_\mu\chi \ , \quad (2.15)$$

which you are encouraged to prove!)

We can do the same with the supersymmetry transformations (2.8), as the next exercise asks you to show.

Exercise 2.14 (Explicit N=2 supersymmetry transformations)
Show that in a Majorana basis, the dimensional reduction of the supersymmetry transformations (2.8) *becomes:* ✓

$$\begin{aligned}
\delta_1 \mathsf{W}_\mu &= i\bar{\alpha}\gamma_\mu\boldsymbol{\psi}_1 + \bar{\alpha}\gamma_\mu\boldsymbol{\psi}_2 & \delta_2 \mathsf{W}_\mu &= -\bar{\alpha}\gamma_\mu\boldsymbol{\psi}_1 + i\bar{\alpha}\gamma_\mu\boldsymbol{\psi}_2\\
\delta_1 \mathsf{P} &= i\bar{\alpha}\gamma_5\boldsymbol{\psi}_1 + \bar{\alpha}\gamma_5\boldsymbol{\psi}_2 & \delta_2 \mathsf{P} &= -\bar{\alpha}\gamma_5\boldsymbol{\psi}_1 + i\bar{\alpha}\gamma_5\boldsymbol{\psi}_2\\
\delta_1 \mathsf{S} &= i\bar{\alpha}\boldsymbol{\psi}_1 + \bar{\alpha}\boldsymbol{\psi}_2 & \delta_2 \mathsf{S} &= -\bar{\alpha}\boldsymbol{\psi}_1 + i\bar{\alpha}\boldsymbol{\psi}_2\\
\delta_1\boldsymbol{\psi}_1 &= -D^\mu(\mathsf{S} + \mathsf{P}\gamma_5)\gamma_\mu\alpha + \tfrac{1}{2}e(\mathsf{S}\times\mathsf{P})\gamma_5\alpha + \tfrac{1}{2}\mathsf{G}^{\mu\nu}\gamma_{\mu\nu}\alpha & \delta_1\boldsymbol{\psi}_2 &= 0\\
\delta_2\boldsymbol{\psi}_1 &= 0 & \delta_2\boldsymbol{\psi}_2 &= -D^\mu(\mathsf{S} + \mathsf{P}\gamma_5)\gamma_\mu\alpha + \tfrac{1}{2}e(\mathsf{S}\times\mathsf{P})\gamma_5\alpha + \tfrac{1}{2}\mathsf{G}^{\mu\nu}\gamma_{\mu\nu}\alpha \ .
\end{aligned}$$

The $SO(2)$ invariance of (2.14) can be made manifest by rewriting $\mathcal{L}_\mathrm{f}$ explicitly in terms of the $SO(2)$ invariant tensors $\delta_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$. In fact, using the identities (2.15), one can rewrite (2.14) as:

$$\mathcal{L}_\mathrm{f} = \frac{i}{2}\delta^{\alpha\beta}\bar{\boldsymbol{\psi}}_\alpha \cdot \gamma^\mu D_\mu \boldsymbol{\psi}_\beta + \frac{e}{2}\epsilon^{\alpha\beta}\bar{\boldsymbol{\psi}}_\alpha \cdot (\gamma_5\mathsf{P} + \mathsf{S}) \times \boldsymbol{\psi}_\beta \ .$$

The $SO(2)$ transformation properties of the four-dimensional fields can be succinctly written as follows:

$$\begin{aligned}
\mathsf{S} + i\mathsf{P} &\mapsto e^{-i\mu}(\mathsf{S} + i\mathsf{P})\\
\boldsymbol{\psi} &\mapsto e^{\mu\gamma_5/2}\,\boldsymbol{\psi}\\
\mathsf{W}_\mu &\mapsto \mathsf{W}_\mu \ .
\end{aligned} \quad (2.16)$$

Exercise 2.15 (The $SO(2)$ Noether current)
Prove that the Noether current associated with the $SO(2)$ transformations (2.16) *is given by*

$$j^5_\mu = \mathsf{P}\cdot D_\mu\mathsf{S} - \mathsf{S}\cdot D_\mu\mathsf{P} + \tfrac{i}{2}\bar{\boldsymbol{\psi}}\cdot\gamma_5\gamma_\mu\boldsymbol{\psi}\ .$$

Notice that this current contains the axial current, hence the notation.
Problem: *Is it anomalous in this theory?*

2.3.3 Higgsed N=2 supersymmetric Yang–Mills

The hamiltonian density corresponding to the N=2 supersymmetric Yang–Mills theory defined by (2.13) is given by $\mathcal{H} = \mathcal{H}_b + \mathcal{H}_f$. We focus on the bosonic part:

$$\mathcal{H}_b = \tfrac{1}{2}\|\mathsf{E}_i\|^2 + \tfrac{1}{2}\|D_0\mathsf{S}\|^2 + \tfrac{1}{2}\|D_0\mathsf{P}\|^2 + \tfrac{1}{2}\|\mathsf{B}_i\|^2 + \tfrac{1}{2}\|D_i\mathsf{S}\|^2 + \tfrac{1}{2}\|D_i\mathsf{P}\|^2 + \tfrac{1}{2}\,e^2\|\mathsf{P}\times\mathsf{S}\|^2\ .$$

Demanding that the energy of a given field configuration be finite doesn't necessarily imply that P and S acquire non-zero vacuum expectation values—for the term $\|\mathsf{P}\times\mathsf{S}\|^2$ is already zero provided that $\mathsf{P}\times\mathsf{S}=0$, which for $so(3)$ means that they be parallel. Indeed, except for that term and the extra field, $\mathcal{H}_b$ is nothing but the energy density (1.13) of (the bosonic part of) the Georgi–Glashow model in the limit of vanishing potential. We could add a potential term $\lambda\left(\|\mathsf{P}\|^2 + \|\mathsf{S}\|^2 - a^2\right)^2$ to the lagrangian (2.13) to force S and P to acquire a nonzero vacuum expectation value, but such a term would break supersymmetry. Nevertheless we could then take the limit $\lambda \downarrow 0$ while keeping the nonzero vacuum expectation values of S and P. This restores the supersymmetry provided that $\langle\mathsf{S}\rangle$ and $\langle\mathsf{P}\rangle$ are parallel, which would be the supersymmetric version of the Prasad–Sommerfield limit. Since the potential depends only on the $SO(2)$ invariant combination $\|\mathsf{P}\|^2 + \|\mathsf{S}\|^2$, $SO(2)$ is preserved and we could use this symmetry to choose $\langle\mathsf{P}\rangle = 0$ and $\langle\mathsf{S}\rangle = \boldsymbol{a}$, where $\boldsymbol{a}$ is a fixed vector with $\|\boldsymbol{a}\|^2 = a^2$.

Exercise 2.16 (The perturbative spectrum of the model)
We can analyse the perturbative spectrum of the model around such a vacuum in exactly the same way as we did in Exercise 1.4. Choosing for example the unitary *gauge $\boldsymbol{a} = a\boldsymbol{e}_3$, show that there are now two massive multiplets $(\psi^\pm, W^\pm_\mu, P^\pm)$ of mass $M_W = ae\hbar$, and a massless gauge multiplet corresponding to the unbroken $U(1)$: $(\psi^3, W^3_\mu, S^3, P^3)$. Prove that the massless gauge multiplet is actually made out of two massless multiplets with helicities $\lambda = -1$ and $\lambda = 0$.*

Now *watch carefully*: something curious has happened. From the analysis in section 2.2.3, we know that the generic massive representations of N=2 supersymmetry are sixteen-fold degenerate, and from Exercise 2.3 we know that they must have a state with spin 3/2. Yet the massive multiplets which have arisen out of higgsing the model contain maximum spin 1 and are only four-fold degenerate. This is *only* possible if the N=2 supersymmetry algebra in this model has central charges and these charges saturate the bound! Indeed, the only way to reconcile the above spectrum with the structure of massive representations of the N=2 supersymmetry algebra studied in section 2.2.3 is if it corresponds to the short multiplet with spin s=1/2 studied in Exercise 2.5. In the next section we will actually compute the supersymmetry algebra for this model and we will see that the central charges are precisely the electric and magnetic charges relative to the unbroken $U(1)$. But before doing this let us check that the BPS-monopole is actually a solution of N=2 supersymmetric Yang–Mills.

2.3.4 N=2 avatar of the BPS-monopole

We now show that this N=2 supersymmetric Yang–Mills theory admits BPS-monopole solutions. We look for static solutions, so we put $\mathsf{W}_0 = 0$. Since the fermion equations of motion are linear, we can always put $\boldsymbol{\psi} = 0$ at the start. Applying supersymmetry transformations to such a solution, we will be able to generate solutions with nonzero fermions. Similarly, using the $SO(2)$ invariance we can look for a solution with $\mathsf{P} = 0$, and then obtain solutions with nonzero P by acting with $SO(2)$. Having made these choices, we are left with W_i and S, which is precisely the spectrum of the bosonic part of the Georgi–Glashow model provided we identify S and $\boldsymbol{\phi}$. Furthermore, not just the spectrum, but also the lagrangian density agrees, with potential set to zero, of course. Therefore the BPS-monopole given by (1.15) with H and K given in Exercise 1.12 is a solution of N=2 supersymmetric Yang–Mills. If we now apply an $SO(2)$ rotation to this solution, we find the following BPS-monopole solution:

$$
\begin{aligned}
\boldsymbol{\psi} &= \mathsf{W}_0 = 0 \\
\mathsf{S}^a &= \alpha \frac{r^a}{er^2} H(\xi) \\
\mathsf{P}^a &= \beta \frac{r^a}{er^2} H(\xi) \\
\mathsf{W}_i^a &= \epsilon_{aij} \frac{r^j}{er^2} \left(K(\xi) - 1 \right)
\end{aligned}
\tag{2.17}
$$

where as before $\xi = aer$, where H and K are the same functions in Exercise 1.12, and $\alpha^2 + \beta^2 = 1$. Putting $\beta = 0$ we recover the BPS-monopole and anti-monopole for $\alpha = \pm 1$, respectively—a result first obtained in [DHdV78].

Since (2.17) is a solution of the field equations of a supersymmetric theory, supersymmetry transformations map solutions to solutions. Hence starting with (2.17) we can try to generate solutions with nonzero fermions by performing a supersymmetry transformation. We will actually assume a more general solution than the one above.

Exercise 2.17 (Supersymmetric BPS-monopoles)
Prove that any BPS-monopole, *that is, any static solution* $(\mathsf{W}_i, \boldsymbol{\phi})$ *of the Bogomol'nyi equation* (1.28)*, can be thought of as an* $N{=}2$ *BPS-monopole by setting* $\mathsf{S} = \alpha\boldsymbol{\phi}$, $\mathsf{P} = \beta\boldsymbol{\phi}$ *and* $\boldsymbol{\psi} = 0$*, with* $\alpha^2 + \beta^2 = 1$*.*

We will then take one such $N{=}2$ BPS-monopole as our starting point and try to generate other solutions via supersymmetry transformations. The supersymmetry transformation laws on the four-dimensional fields can be read off from those given in (2.8) for the six-dimensional fields. Since we start with a background in which $\boldsymbol{\psi} = 0$, the bosonic fields are invariant under supersymmetry. The supersymmetry transformation law of the fermion $\boldsymbol{\psi}$ is given by

$$\delta\boldsymbol{\psi} = \left(\tfrac{1}{2}\mathsf{G}^{\mu\nu}\gamma_{\mu\nu} - D_\mu\boldsymbol{\phi}\gamma^\mu(\alpha + \beta\gamma_5)\right)\epsilon\ ,$$

where ϵ is an unconstrained (complex) Dirac spinor. Because the solution is static—$\mathsf{W}_0 = 0$ and all fields are time-independent—the above can be rewritten as

$$\delta\boldsymbol{\psi} = \left(\tfrac{1}{2}\mathsf{G}_{ij}\gamma_{ij} + D_i\boldsymbol{\phi}\gamma_i(\alpha + \beta\gamma_5)\right)\epsilon\ .$$

For definiteness we will assume that $(\mathsf{W}_i, \boldsymbol{\phi})$ describe a BPS-monopole (as opposed to an anti-monopole) so that $D_i\boldsymbol{\phi} = +\tfrac{1}{2}\epsilon_{ijk}\mathsf{G}_{jk}$. Then we can rewrite the above transformation law once more as

$$\delta\boldsymbol{\psi} = D_k\boldsymbol{\phi}\ \left(\tfrac{1}{2}\epsilon_{ijk}\gamma_{ij} + \gamma_k(\alpha + \beta\gamma_5)\right)\epsilon\ . \tag{2.18}$$

Exercise 2.18 (More γ-matrix identities)
Prove the following identity:

$$\tfrac{1}{2}\epsilon_{ijk}\gamma_{ij} = -\gamma_0\gamma_5\gamma_k\ . \tag{2.19}$$

Exercise 2.19 (Some euclidean γ-matrices)
Let $\bar\gamma_i \equiv \gamma_0\gamma_i$ *for* $i = 1, 2, 3$*, and let* $\bar\gamma_4 = \gamma_0(\alpha + \beta\gamma_5)$*. Prove that they generate a euclidean Clifford algebra. Define* $\bar\gamma_5 \equiv \bar\gamma_1\bar\gamma_2\bar\gamma_3\bar\gamma_4 = \gamma_0(\alpha\gamma_5 - \beta)$*. Prove that* $\bar\gamma_5$ *is hermitian and that* $\bar\gamma_5^2 = \mathbb{1}$*.*

In terms of these euclidean Clifford algebra, and using (2.19), we can rewrite (2.18) as

$$\delta\boldsymbol{\psi} = \gamma_5\bar\gamma_k D_k\boldsymbol{\phi}\ (\mathbb{1} - \bar\gamma_5)\,\epsilon\ .$$

Notice that $\frac{1}{2}(\mathbb{1} \pm \bar{\gamma}_5)$ is a projector. If we denote $\epsilon_\pm = \frac{1}{2}(\epsilon \pm \bar{\gamma}_5 \epsilon)$, then the supersymmetric variation of $\boldsymbol{\psi}$ in a BPS-monopole background is given simply by

$$\delta\boldsymbol{\psi} = 2\,\gamma_5 \bar{\gamma}_k D_k \boldsymbol{\phi}\,\epsilon_- \ .$$

This means that if ϵ has negative chirality relative to $\bar{\gamma}_5$, then we don't generate new solutions, yet if ϵ has positive chirality, then we do. Equivalently, supersymmetry transformations with negative chirality parameters preserve the solution, whereas those with positive chirality parameters break it.

Exercise 2.20 (BPS-monopoles break one half of the supersymmetry)
Prove that the (± 1)-eigenspaces of $\bar{\gamma}_5$ have the same dimension. Conclude that the projector $\frac{1}{2}(\mathbb{1} \pm \bar{\gamma}_5)$ projects out precisely one half of spinors.

As a corollary of the above exercise we see that *supersymmetric BPS-monopoles break half the supersymmetries.*

Notice that the parameter ϵ, being an unconstrained Dirac spinor has 4 complex (or 8 real) components, whereas $\epsilon_\pm$ only has 2 complex (or 4 real) components. Hence we expect that the BPS-monopole belongs to a fourfold degenerate multiplet. From our study in section 2.2.3 of massive representations of the N=2 supersymmetry algebra, we know that those massive multiplets preserving half the supersymmetries are necessarily short, and from Exercise 2.5 we see that the k=1 BPS-monopole given by (2.17) generates a short multiple with spin s=0. This multiplet contains two "particles" of spin 0 and one of spin 1/2, yet none of spin 1. Therefore although as we will see in the next section, N=2 supersymmetry solves the first of the problems with the Montonen–Olive conjecture mentioned at the end of section 1.4.1, it still does not address the second problem satisfactorily. As we will see later, the solution of this problem requires N=4 supersymmetry.

2.3.5 The supersymmetry bound is the Bogomol'nyi bound

The Bogomol'nyi bound (1.27) can be suggestively rewritten as

$$M^2 - (aq)^2 - (ag)^2 \geq 0 \ ,$$

which is begging us to add two *spatial* dimensions to our spacetime and interpret the above inequality as the positivity of mass. As explained in section 2.2.1, the positivity of the mass is a consequence of unitarity and the supersymmetry algebra. Therefore it would make sense to look for a six-dimensional supersymmetric explanation of the Bogomol'nyi bound. The explanation of [WO78] used the central charges in four-dimensional N=2 supersymmetric Yang–Mills theory, and as we have seen this theory comes induced from six-dimensional N=1 supersymmetric Yang–Mills via dimensional reduction. It would make sense therefore to look for a direct six-dimensional explanation. This was done to a large extent in [Oli79] and we will now review this.

The above heuristics suggest that we think of the electric and magnetic charges as *momenta* in the two extra spatial dimensions. However it isn't hard to see that this interpretation is not quite correct. If one computes the energy-momentum tensor T_{AB} of the six-dimensional supersymmetric Yang–Mills theory, and from there the momenta $P_A = T_{0A}$, then the positivity of mass formula in six-dimensions:

$$M^2 \geq P_5^2 + P_6^2 \ , \tag{2.20}$$

where $M^2 = P^\mu P_\mu$ is the four-dimensional mass, does not agree with the Bogomol'nyi bound (1.27). In fact one finds that the magnetic charge does not appear. *What is wrong then?* Simply that we have assumed that it is P_μ which appears in the right hand side of $\{Q, \bar{Q}\}$ in the supersymmetry algebra, when in fact it is $P_\mu + Z_\mu$, where Z_μ can be interpreted as the topological charge due to the presence of a string-like source in six-dimensions. We now find out what Z_μ is by computing the supersymmetry algebra. We first do this in six dimensions and then reduce down to four.

The supersymmetry algebra in six dimensions

We start by noticing that the space integral of $\delta\bar{\delta}j_0$ is equal to $\bar{\boldsymbol{\alpha}}\,\{Q, \bar{Q}\}\,\boldsymbol{\beta}$, whence it is enough to compute $\delta\bar{\delta}j_A$, which we naturally leave as an exercise.

Exercise 2.21 (Supersymmetric variation of the supercurrent)
Prove that

$$\begin{aligned}\delta\bar{\delta}j_A &= -i\delta\bar{J}_A\boldsymbol{\beta}\\ &= -\tfrac{1}{4}\mathsf{G}^{BC}\cdot\mathsf{G}^{EF}\bar{\boldsymbol{\alpha}}\Gamma_{BC}\Gamma_A\Gamma_{EF}\boldsymbol{\beta} + i\left(\bar{\boldsymbol{\alpha}}\Gamma_C D_B\boldsymbol{\Psi}\right)\cdot\left(\bar{\boldsymbol{\Psi}}\Gamma_A\Gamma^{BC}\boldsymbol{\beta}\right)\ .\end{aligned}$$

The fermion bilinear term has to be Fierzed, but we will not be concerned with the fermions in what follows: we are interested in computing the "momenta" in classical configurations like the BPS-monopole, where the fermions have been set to zero. Of course, it would be a good exercise in Γ-matrix algebra to compute the fermionic terms, not that there is little Γ-matrix algebra to be done. In fact, prove that setting $\boldsymbol{\Psi} = 0$, $\delta\bar{\delta}j_A$ is given by

$$\delta\bar{\delta}j_A = 2\bar{\boldsymbol{\alpha}}\left(-\tfrac{1}{8}\epsilon_{BCAEFD}\mathsf{G}^{BC}\cdot\mathsf{G}^{EF} + \mathsf{G}^{BC}\cdot\mathsf{G}_{CA}\eta_{BD} + \tfrac{1}{4}\mathsf{G}^{BC}\cdot\mathsf{G}_{BC}\eta_{AD}\right)\Gamma^D\boldsymbol{\beta}\ . \tag{2.21}$$

(Hint: use that $\Gamma_{ABCDE} = -\epsilon_{ABCDEF}\Gamma^F\Gamma_7$ (prove it!) *and use the fact that $\Gamma_7\boldsymbol{\beta} = -\boldsymbol{\beta}$.)*

We see that there are two very different tensors appearing in the right-hand-side of $\delta\bar{\delta}j_A$:

$$\Theta_{AB} = -\tfrac{1}{8}\,\epsilon_{ABCDEF}\,\mathsf{G}^{CD}\cdot\mathsf{G}^{EF} \tag{2.22}$$

$$T_{AB} = \mathsf{G}_A{}^C\cdot\mathsf{G}_{CB} + \tfrac{1}{4}\,\mathsf{G}^{CD}\cdot\mathsf{G}_{CD}\,\eta_{AB}\ . \tag{2.23}$$

Notice that T_{AB} is symmetric, whereas Θ_{AB} is antisymmetric. In fact, T_{AB} is (the bosonic part of) the energy-momentum tensor of the six-dimensional theory.

Exercise 2.22 (The symmetric gauge-invariant energy momentum tensor)
Prove that the energy-momentum tensor of the six-dimensional supersymmetric Yang–Mills theory is given by

$$T_{AB} + \tfrac{i}{2}\,\bar{\mathbf{\Psi}}\cdot\Gamma_{(A}\overleftrightarrow{D}_{B)}\mathbf{\Psi} - \eta_{AB}\,\tfrac{i}{2}\,\bar{\mathbf{\Psi}}\cdot\Gamma^C\overleftrightarrow{D}_C\mathbf{\Psi}\ .$$

Prove that T_{AB} is gauge-invariant and that it is conserved on-shell.
(Hint: Vary $\mathcal{L}_b$ with respect to an infinitesimal translation $x^A \mapsto x^A + \varepsilon^A(x)$ and determine the associated Noether current, which after symmetrisation is the energy-momentum tensor, by definition.)

Defining P_A to be the space integral of T_{0A} and Z_A the space integral of Θ_{0A}, we see that the supersymmetry algebra becomes $\{Q,\bar{Q}\}$:

$$\boxed{\{Q,\bar{Q}\} = 2\Gamma^A\,(P_A + Z_A)}\ .$$

Only the first of these terms is to be interpreted as the momentum, the other term is associated with a topologically conserved current.

Exercise 2.23 (The topological current)
Prove that Θ_{AB} is gauge invariant and that it is conserved off-shell*, that is, without imposing the equations of motion. This means that it is a* topological *current.*
(Hint: show that $\Theta_{AB} = \partial^C\Xi_{ABC}$ where Ξ_{ABC} is totally antisymmetric, though not gauge invariant.)

The supersymmetry algebra in four dimensions

It is now time to dimensionally reduce the supersymmetry algebra. The following exercise asks you to compute P_5, P_6, Z_5, and Z_6 (with fermions put to zero) after dimensional reduction.

Exercise 2.24 (The "momenta" in the extra dimensions)
Prove that

$$\begin{aligned}
T_{05} &= -D_i\mathsf{P}\cdot\mathsf{G}_{0i} - e\,(\mathsf{P}\times\mathsf{S})\cdot D_0\mathsf{S}\\
T_{06} &= -D_i\mathsf{S}\cdot\mathsf{G}_{0i} + e\,(\mathsf{P}\times\mathsf{S})\cdot D_0\mathsf{P}\\
\Theta_{05} &= \tfrac{1}{2}\epsilon_{ijk}\mathsf{G}_{ij}\cdot D_k\mathsf{S}\\
\Theta_{05} &= -\tfrac{1}{2}\epsilon_{ijk}\mathsf{G}_{ij}\cdot D_k\mathsf{P}\ .
\end{aligned}$$

Using the Bianchi identity $\epsilon_{ijk}D_i\mathsf{G}_{jk}=0$, we can rewrite Θ_{05} and Θ_{06} as follows:

$$\Theta_{05} = \tfrac{1}{2}\partial_i\left(\epsilon_{ijk}\mathsf{G}_{jk}\cdot\mathsf{S}\right)\quad\text{and}\quad \Theta_{05} = -\tfrac{1}{2}\partial_i\left(\epsilon_{ijk}\mathsf{G}_{jk}\cdot\mathsf{P}\right)$$

whereas using the equations of motion (for zero fermions)

$$-D_i\mathsf{G}_{0i} + e\,\mathsf{P}\times D_0\mathsf{P} + e\,\mathsf{S}\times D_0\mathsf{S} = 0\ ,$$

we can rewrite T_{05} and T_{06} as follows:

$$T_{05} = -\partial_i\left(\mathsf{G}_{0i}\cdot\mathsf{P}\right)\quad\text{and}\quad T_{05} = -\partial_i\left(\mathsf{G}_{0i}\cdot\mathsf{S}\right)\ .$$

We see that all the densities are divergences, whence their space integrals only receive contribution from spatial infinity:

$$\begin{aligned}
P_5 + Z_5 &= \int_{\Sigma_\infty}\left(-\mathsf{P}\cdot\mathsf{G}_{0i} + \tfrac{1}{2}\epsilon_{ijk}\mathsf{S}\cdot\mathsf{G}_{jk}\right)d\Sigma_i\\
P_6 + Z_6 &= \int_{\Sigma_\infty}\left(-\mathsf{S}\cdot\mathsf{G}_{0i} - \tfrac{1}{2}\epsilon_{ijk}\mathsf{P}\cdot\mathsf{G}_{jk}\right)d\Sigma_i\ .
\end{aligned}$$

To interpret these integrals we can proceed in either of two ways. The fastest way is to use the $SO(2)$ invariance of the theory to choose $\mathsf{P}=0$ and $\|\mathsf{S}\|^2 = a^2$ at spatial infinity. Comparing with (1.24) and (1.25), we see that $P_5+Z_5 = ag$ and $P_6+Z_6=-aq$. The same reasoning follows without having to use $SO(2)$ invariance, as the next exercise shows.

Exercise 2.25 (The effective electromagnetic field strength)
Define the following field strength:

$$F_{\mu\nu} \equiv \frac{1}{a}\left(\mathsf{S}\cdot\mathsf{G}_{\mu\nu} + \mathsf{P}\cdot{}^\star\mathsf{G}_{\mu\nu}\right)\ . \tag{2.24}$$

Prove that in the "Higgs vacuum" it obeys Maxwell's equations, and deduce that $P_5+Z_5=ag$ *and* $P_6+Z_6=-aq$ *where* g *and* q *are, respectively, the magnetic and electric charges of this electromagnetic field.*
(Hint: Compare with Exercise 1.8).

To prove that (2.20) is the Bogomol'nyi bound (1.27), we can proceed in two ways. We can exploit the $SO(2)$ invariance of the supersymmetry algebra in order to set $\mathsf{P} = 0$, and then notice that $Z_\mu = 0$. Using the fact that P_μ is indeed the honest momentum of the theory, namely the space integral of $T_{0\mu}$, and plugging the expressions for $P_A + Z_A$ into (2.20), we finally arrive at the Bogomol'nyi bound (1.27)!

Alternatively we can deduce that $Z_\mu = 0$ without having to set $\mathsf{P} = 0$. This is the purpose of the following exercise.

Exercise 2.26 (The space components of the topological charge)
Prove that Θ_{0i} *is given by*

$$\Theta_{0i} = \epsilon_{ijk}\partial_j \left(\mathsf{P} \cdot D_k\mathsf{S}\right) \ ,$$

whence Z_i *is given by*

$$Z_i = \epsilon_{ijk} \int_{\Sigma_\infty} \left(\mathsf{P} \cdot D_k\mathsf{S}\right) d\Sigma_j \ .$$

Prove that this vanishes for a finite-energy configuration.
(Hint: Notice that for a solution of the Bogomol'nyi equation $\mathsf{S} = \alpha\boldsymbol{\phi}$, $\mathsf{P} = \beta\boldsymbol{\phi}$ *with* $\alpha^2 + \beta^2 = 1$, $\mathsf{P} \cdot D_k\mathsf{S} = \frac{1}{2}\alpha\beta\partial_k\|\boldsymbol{\phi}\|^2$, *and that the derivative* ∂_k *is tangential to* Σ_∞ *due to the* ϵ_{ijk}. *Since* $\|\boldsymbol{\phi}\|^2 = a^2$ *on* Σ_∞, *its tangential derivative vanishes.)*

Define the following complex linear combinations of fields (*cf.* (1.33)):

$$\mathcal{G}_{\mu\nu} = \mathsf{G}_{\mu\nu} + i\,{}^\star\mathsf{G}_{\mu\nu}$$
$$\boldsymbol{\Phi} = \mathsf{S} + i\,\mathsf{P} \ ,$$

in terms of which the effective electromagnetic field strength defined in (2.24), becomes

$$F_{\mu\nu} = \frac{1}{a}\,\mathrm{Re}\left(\bar{\boldsymbol{\Phi}} \cdot \mathcal{G}_{\mu\nu}\right) \ .$$

Under an infinitesimal $SO(2)$ transformation, $\delta\bar{\boldsymbol{\Phi}} = i\bar{\boldsymbol{\Phi}}$, and because $i\,\mathcal{G}_{\mu\nu} = -{}^\star\mathcal{G}_{\mu\nu}$, we can write

$$\delta F_{\mu\nu} = -{}^\star F_{\mu\nu} \ .$$

In other words, $SO(2)$ transformations become infinitesimal duality transformations in the effective electromagnetic theory.
Problem: Are anomalies responsible for the breaking of this symmetry in the quantum theory?

2.4 N=4 Supersymmetric Yang-Mills

We saw in Exercise 2.7 that 10 is the largest dimension in which N=1 supersymmetric Yang–Mills theory can exist and that for it to exist we must impose that the spinors be both Weyl and Majorana—conditions which, luckily for us, can be simultaneously satisfied in ten-dimensional Minkowski space. In this section we will prove that this theory exists and that upon dimensional reduction to four dimensions yields a gauge theory with $N{=}4$ supersymmetry. This theory admits Higgs phenomena and has room to embed the BPS-monopole and indeed, any solution of the Bogomol'nyi equation, just as in the $N{=}2$ theory discussed in the previous section. We will see that both the massive fundamental states (*e.g.*, vector bosons) and the solitonic states (*e.g.*, BPS-monopoles) belong to isomorphic (short) multiplets saturating the supersymmetry mass bound which once again will be shown to agree with the Bogomol'nyi bound for dyons.

2.4.1 $N{=}1$ $d{=}10$ supersymmetric Yang–Mills

We start by setting up some conventions. We will let indices $A, B, \ldots$ from the start of the Latin alphabet run from 0 to 9. (No confusion should arise from the fact that in the previous section the very same indices only reached 6.) The metric η_{AB} is mostly minus and the 32×32 matrices $\{\Gamma_A\}$ obey the Clifford algebra $\{\Gamma_A, \Gamma_B\} = 2\,\eta_{AB}\,\mathbb{1}$. We let $\Gamma_{11} \equiv \Gamma_0\Gamma_1\cdots\Gamma_9$; it obeys $\Gamma_{11}^2 = \mathbb{1}$. We shall also need the charge conjugation matrix $\mathcal{C}$, which obeys $\mathcal{C}^t = -\mathcal{C}$ and $(\mathcal{C}\Gamma_A)^t = \mathcal{C}\Gamma_A$, from where it follows that $(\Gamma_A)^t = -\mathcal{C}\Gamma_A\mathcal{C}^{-1}$.

$N{=}1$ supersymmetric Yang–Mills theory is defined by the following lagrangian density:

$$\mathcal{L} = -\tfrac{1}{4}\mathsf{G}^{AB}\cdot\mathsf{G}_{AB} + \tfrac{i}{2}\,\bar{\mathbf{\Psi}}\cdot\Gamma^A D_A\mathbf{\Psi}\ , \tag{2.25}$$

where

$$\begin{aligned} \mathsf{G}_{AB} &= \partial_A\mathsf{W}_B - \partial_B\mathsf{W}_A - e\,\mathsf{W}_A\times\mathsf{W}_B\ ,\\ D_A\mathbf{\Psi} &= \partial_A\mathbf{\Psi} - e\,\mathsf{W}_A\times\mathbf{\Psi}\ , \end{aligned}$$

and where $\mathbf{\Psi}$ is a complex Majorana–Weyl spinor obeying $\Gamma_{11}\mathbf{\Psi} = -\mathbf{\Psi}$ (Weyl) and $\bar{\mathbf{\Psi}} = \mathbf{\Psi}^\dagger\,\Gamma_0 = \mathbf{\Psi}^t\,\mathcal{C}$ (Majorana).

Exercise 2.27 (The Majorana condition)
Prove that the Majorana condition above relates $\mathbf{\Psi}$ and its complex conjugate $\mathbf{\Psi}^*$:

$$\mathbf{\Psi}^* = \mathcal{C}\Gamma_0\,\mathbf{\Psi}\ ,$$

whence it can be considered a reality condition on the spinor.

The action defined above is clearly gauge invariant. We claim that it is also invariant under the following supersymmetry:

$$\begin{aligned} \delta\,\mathsf{W}_A &= i\bar{\boldsymbol{\alpha}}\,\Gamma_A\mathbf{\Psi} = -i\bar{\mathbf{\Psi}}\,\Gamma_A\boldsymbol{\alpha}\\ \delta\,\mathbf{\Psi} &= \tfrac{1}{2}\mathsf{G}^{AB}\Gamma_{AB}\boldsymbol{\alpha}\\ \delta\,\bar{\mathbf{\Psi}} &= -\tfrac{1}{2}\bar{\boldsymbol{\alpha}}\Gamma_{AB}\mathsf{G}^{AB}\ , \end{aligned} \tag{2.26}$$

where $\boldsymbol{\alpha}$ is a constant anticommuting Majorana–Weyl spinor of the same chirality as $\boldsymbol{\Psi}$, and where the second and third relations above imply each other.

The proof that the action is invariant under the supersymmetry transformations (2.26) is very similar to the analogous statement for the six-dimensional theory, so we will not be as verbose.

We start by varying the action with respect to (2.26). We don't take $\boldsymbol{\alpha}$ to be constant in order to be able to read off the form of the supersymmetry current from the variation of the lagrangian density. We will have proven invariance if we can show that up to a divergence, the variation of (2.25) is proportional to the derivative of $\bar{\boldsymbol{\alpha}}$—the coefficient being the supersymmetry current. Varying the lagrangian density we encounter two kinds of terms: terms linear in the fermions, and a term trilinear in the fermions and without derivatives, coming from the variation of the gauge field inside the covariant derivative acting on the fermions.

Before getting into the computation, it is useful to derive some properties of Majorana–Weyl fermions, which are left as an instructive exercise.

Exercise 2.28 (Properties of Majorana and Weyl fermions)
Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be anticommuting Majorana fermions in ten dimensions. Prove that

$$\bar{\boldsymbol{\alpha}}\,\Gamma_{A_1A_2\cdots A_k}\,\boldsymbol{\beta} = (-)^{k(k+1)/2}\bar{\boldsymbol{\beta}}\,\Gamma_{A_1A_2\cdots A_k}\,\boldsymbol{\alpha}\ . \tag{2.27}$$

If, in addition, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are Weyl and of the same chirality, then prove that

$$\bar{\boldsymbol{\alpha}}\,(\text{even number of}\ \Gamma\text{s})\,\boldsymbol{\beta} = 0\ .$$

(Hint: It may prove useful to first prove the identity

$$(\mathcal{C}\Gamma_{A_1A_2\cdots A_k})^t = -(-)^{k(k+1)/2}\mathcal{C}\Gamma_{A_1A_2\cdots A_k}\ , \tag{2.28}$$

which will play a role also later on.)

We now vary the lagrangian density.

Exercise 2.29 (Varying the lagrangian density)
Prove that supersymmetric variation of the lagrangian density $\mathcal{L}$ is given, up to a divergence, by:

$$\delta\mathcal{L} = \tfrac{i}{2}D_C\mathsf{G}_{AB}\cdot\bar{\boldsymbol{\alpha}}\,\Gamma^{AB}\Gamma^C\,\boldsymbol{\Psi} + iD^A\mathsf{G}_{AB}\cdot\bar{\boldsymbol{\alpha}}\,\Gamma^B\boldsymbol{\Psi} \\ + \tfrac{i}{2}\mathsf{G}_{AB}\cdot\partial_C\bar{\boldsymbol{\alpha}}\,\Gamma^{AB}\Gamma^C\boldsymbol{\Psi} + \tfrac{1}{2}e\bar{\boldsymbol{\Psi}}\,\Gamma^A\cdot((\bar{\boldsymbol{\alpha}}\,\Gamma_A\boldsymbol{\Psi})\times\boldsymbol{\Psi})\ .$$

(Hint: Integrate by parts and use the identity (2.27) repeatedly.)

Using the Bianchi identity in the form $\Gamma^{ABC} D_C \mathsf{G}_{AB} = 0$, it is easy to prove that the first two terms in the above expression for $\delta\mathcal{L}$ cancel out, leaving the trilinear terms and the term involving the supersymmetry current:

$$\delta\mathcal{L} = \partial_A \bar{\boldsymbol{\alpha}} J^A + \tfrac{1}{2} e \bar{\boldsymbol{\Psi}} \Gamma^A \cdot ((\bar{\boldsymbol{\alpha}} \Gamma_A \boldsymbol{\Psi}) \times \boldsymbol{\Psi}) \ ,$$

where the supersymmetry current J^A is given by

$$J^A = \tfrac{i}{2} \mathsf{G}_{BC} \cdot \Gamma^{BC} \Gamma^A \boldsymbol{\Psi} \ . \tag{2.29}$$

Finally we tackle the trilinear terms, which as usual are the trickier ones. Just as in the six-dimensional theory, their vanishing will be seen to be a property of some identities between the Γ-matrices. Writing the SO(3) indices explicitly, we find that these terms are given by

$$\tfrac{1}{2} e \epsilon_{abc} \bar{\boldsymbol{\alpha}} \Gamma_A \boldsymbol{\Psi}^a \, \bar{\boldsymbol{\Psi}}^c \Gamma^A \times \boldsymbol{\Psi}^b \ , \tag{2.30}$$

and we once again must use the Fierz identities to expand the bi-spinor $\boldsymbol{\Psi}^a \, \bar{\boldsymbol{\Psi}}^c$.

Exercise 2.30 (A ten-dimensional Fierz identity)
Prove that

$$\begin{aligned} \boldsymbol{\Psi}^a \, \bar{\boldsymbol{\Psi}}^c = &-\tfrac{1}{32} \bar{\boldsymbol{\Psi}}^c \Gamma_A \boldsymbol{\Psi}^a \, \Gamma^A (\mathbb{1} + \Gamma_{11}) \\ &+ \tfrac{1}{32 \cdot 3!} \bar{\boldsymbol{\Psi}}^c \Gamma_{ABC} \boldsymbol{\Psi}^a \, \Gamma^{ABC} (\mathbb{1} + \Gamma_{11}) - \tfrac{1}{32 \cdot 5!} \bar{\boldsymbol{\Psi}}^c \Gamma_{ABCDE} \boldsymbol{\Psi}^a \, \Gamma^{ABCDE} \ . \end{aligned}$$

Using the results of Exercise 2.28—in particular equation (2.27)—and taking into account the antisymmetry of ϵ_{abc} we see that only the first and last terms on the right-hand side of the above Fierz identity contribute to (2.30).

Exercise 2.31 (Some more Γ-matrix identities)
Prove the following identities between ten-dimensional Γ*-matrices:*

$$\Gamma^A \Gamma^B \Gamma_A = -8\, \Gamma^B \qquad \text{and} \qquad \Gamma^F \Gamma^{ABCDE} \Gamma_F = 0 \ , \tag{2.31}$$

and use them to deduce that the trilinear terms (2.30) *cancel exactly.*
(Compare these identities with those in Exercise 2.10.)

2.4.2 Reduction to d=4: N=4 supersymmetric Yang–Mills

We now dimensionally reduce the d=10 N=1 supersymmetric Yang–Mills theory described in the previous section down to four dimensions. From now on we will let uppercase indices from the middle of the Latin alphabet: $I, J, K, \ldots$ run from 1 to 3 inclusive. It will be convenient to break up the ten-dimensional coordinates as $x^A = (x^\mu, x^{3+I}, x^{6+J})$, and by dimensional reduction we simply mean that we drop the dependence of the fields on (x^{3+I}, x^{6+J}): $\partial_{3+I} \equiv \partial_{6+J} \equiv 0$.

We also need to decompose the ten-dimensional Γ-matrices. This is done as follows:

$$\begin{aligned} \Gamma^\mu &= \gamma^\mu \otimes \mathbb{1}_4 \otimes \sigma_3 & \qquad \mathcal{C} &= C \otimes \mathbb{1}_4 \otimes \mathbb{1}_2 \\ \Gamma^{3+I} &= \mathbb{1}_4 \otimes \alpha^I \otimes \sigma_1 & \qquad \Gamma^{6+J} &= i\gamma_5 \otimes \beta^J \otimes \sigma_3 \ , \end{aligned} \tag{2.32}$$

where C is the charge conjugation matrix in four-dimensional Minkowski space obeying $C^t = -C$ and $(C\gamma_\mu)^t = C\gamma_\mu$; and $\{\alpha^I\}$ and $\{\beta^J\}$ are 4×4 real antisymmetric matrices satisfying the following algebraic relations:

$$\begin{aligned} [\alpha^I, \alpha^J] &= -2\epsilon^{IJK}\alpha^K & \qquad \{\alpha^I, \alpha^J\} &= -2\delta^{IJ}\, \mathbb{1}_4 \\ [\beta^I, \beta^J] &= -2\epsilon^{IJK}\beta^K & \qquad \{\beta^I, \beta^J\} &= -2\delta^{IJ}\, \mathbb{1}_4 \\ [\alpha^I, \beta^J] &= 0 \ ; \end{aligned}$$

and where $\mathbb{1}_n$ denotes the $n \times n$ unit matrix. (From now on we will drop the subscript when the dimension is clear from the context.) In the above decomposition, Γ_{11} takes the form:

$$\Gamma_{11} = -\mathbb{1} \otimes \mathbb{1} \otimes \sigma_2 \ .$$

We can find an explicit realisation for the matrices α^I and β^J as follows. Because they are real antisymmetric 4×4 matrices, they belong to $so(4)$. Their commutation relations say that they each generate an $so(3)$ subalgebra and moreover that these two $so(3)$ subalgebras commute. Happily $so(4) \cong so(3) \times so(3)$, so that all we have to find is an explicit realisation of this isomorphism. This is found as follows. We say that a matrix A in $so(4)$ is *self-dual* (respectively *antiselfdual*), if its entries obey $A_{ij} = \frac{1}{2}\epsilon_{ijkl}A_{kl}$ (respectively, $A_{ij} = -\frac{1}{2}\epsilon_{ijkl}A_{kl}$). The next exercise asks you to show that the subspaces of $so(4)$ consisting of (anti)self-dual matrices define commuting subalgebras.

Exercise 2.32 ($so(4) \cong so(3) \times so(3)$ explicitly)
Prove that the commutator of two (anti)self-dual matrices in $so(4)$ is (anti)self-dual, and that the commutator of a self-dual matrix and an antiselfdual matrix in $so(4)$ vanishes.
(Hint: Either compute this directly or use the fact that the "duality" operation is $so(4)$ invariant since ϵ_{ijkl} is an $so(4)$-invariant tensor, whence its eigenspaces are ideals.)

Using this result we can now find a explicit realisation for the α^I and the β^J: we simply find a basis for the (anti)self-dual matrices in $so(4)$. This is the purpose of the next exercise.

Exercise 2.33 (Explicit realisation for α^I and β^J)
Prove that a matrix A in $so(4)$ is (anti)self-dual if its entries are related in the following way:

$$A_{12} = \pm A_{34} \qquad A_{13} = \mp A_{24} \qquad A_{14} = \pm A_{23} \ ,$$

where the top signs are for the self-dual case and the bottom signs for the antiself-dual case. Conclude that explicit bases for the (anti)self-dual matrices are given by:

$$\begin{aligned}
e_1^+ &= i\sigma_2 \otimes \mathbb{1} = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} & e_1^- &= \mathbb{1} \otimes i\sigma_2 = \begin{pmatrix} i\sigma_2 & 0 \\ 0 & i\sigma_2 \end{pmatrix} \\
e_2^+ &= \sigma_3 \otimes i\sigma_2 = \begin{pmatrix} i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} & e_2^- &= i\sigma_2 \otimes \sigma_3 = \begin{pmatrix} 0 & \sigma_3 \\ -\sigma_3 & 0 \end{pmatrix} \\
e_3^+ &= \sigma_1 \otimes i\sigma_2 = \begin{pmatrix} 0 & i\sigma_2 \\ i\sigma_2 & 0 \end{pmatrix} & e_3^- &= i\sigma_2 \otimes \sigma_1 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix} \ ,
\end{aligned}$$

where the $\{e_I^+\}$ are self-dual and the $\{e_I^-\}$ are antiselfdual. Prove that $\alpha^I = e_I^\pm$ and $\beta^J = e_J^\mp$ is a valid realisation.

Exercise 2.34 (The fundamental representation of $su(4)$)
Either abstractly or using the above explicit realisation, prove that the fifteen (4×4)-matrices:

$$A^{IJ} = \epsilon^{IJK}\alpha^K \quad B^{IJ} = \epsilon^{IJK}\beta^K \quad C^{IJ} = i\{\alpha^I, \beta^J\} \tag{2.33}$$

are antihermitian and generate the $su(4)$ Lie algebra. This is the fundamental representation of $su(4)$.

The result of the above exercise and the above decomposition of the ten-dimensional Γ-matrices mean that we have broken up a ten-dimensional spinor index (running from 1 to 32) into three indices: a four-dimensional spinor index (running from 1 to 4), an internal $su(4)$ index in the fundamental representation (*i.e.*, also running from 1 to 4), and an internal $su(2)$ index also in the fundamental representation. We have chosen the above decomposition of the Γ-matrices because it possesses two immediate advantages:

1. Because of the form of Γ_{11}, a Weyl spinor in ten-dimensions gives rise to a unconstrained Dirac spinor in four-dimensions with values in the fundamental representation of $su(4)$. In other words, the chirality condition only affects the internal $su(2)$ space and does not constrain the other degrees of freedom; and

2. Because of the form of the charge conjugation matrix, the Majorana condition in ten dimensions becomes the Majorana condition in four dimensions.

Thus we see immediately that a Majorana–Weyl spinor in ten-dimensions yields a quartet of Majorana spinors in four-dimensions, or equivalently a Majorana spinor in four-dimensions with values in the fundamental representation of $su(4)$.

This $su(4)$ is a "flavour" index of the four-dimensional theory; that is, $su(4)$ is a global symmetry of N=4 supersymmetric Yang–Mills theory, not a gauge symmetry. Of course, this flavour symmetry is nothing but the residual Lorentz symmetry in ten-dimensions which upon dimensional reduction to four-dimensions breaks down to $SO(1,3) \times SO(6)$. The Lie algebras of $SO(6)$ and $SU(4)$ are isomorphic. In fact, $SU(4) \cong Spin(6)$, the universal covering group of $SO(6)$; and the four dimensional representations of $SU(4)$ are precisely the spinorial representations of $Spin(6)$ under which the supersymmetric charges transform.

We now want to write down the four-dimensional action obtained by the above dimensional reduction. We define the scalar fields $\mathsf{S}_I = \mathsf{W}_{3+I}$ and pseudoscalar fields $\mathsf{P}_J = \mathsf{W}_{6+J}$. Together with the four-dimensional gauge fields W_μ they comprise the bosonic field content of the four-dimensional theory. As mentioned above, the chirality condition on a ten-dimensional spinor can be easily imposed. Let us write

$$\boldsymbol{\Psi} = \boldsymbol{\psi} \otimes \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} ,$$

where $\boldsymbol{\psi}$ is a quartet of unconstrained Dirac spinors in four dimensions. From the form of Γ_{11} it is easy so see that $\Gamma_{11}\boldsymbol{\Psi} = -\boldsymbol{\psi} \otimes \frac{1}{\sqrt{2}}\sigma_2 \begin{pmatrix} 1 \\ i \end{pmatrix} = -\boldsymbol{\Psi}$. The Majorana condition says that $\boldsymbol{\psi}$ is Majorana in four dimensions. Naturally, all fields are in the adjoint representation of the gauge group $SO(3)$.

In order to write down the action we need to dimensionally reduce the Dirac operator and the gauge field-strength. We find that G_{AB} breaks up as $\mathsf{G}_{\mu\nu}$, $\mathsf{G}_{\mu,3+I} = D_\mu \mathsf{S}_I$, $\mathsf{G}_{\mu,6+I} = D_\mu \mathsf{P}_J$, $\mathsf{G}_{3+I,3+J} = -e\mathsf{S}_I \times \mathsf{S}_J$, $\mathsf{G}_{3+I,6+J} = -e\mathsf{S}_I \times \mathsf{P}_J$, and $\mathsf{G}_{6+I,6+J} = -e\mathsf{P}_I \times \mathsf{P}_J$. This allows us to write the bosonic part of the lagrangian density immediately:

$$\begin{aligned} \mathcal{L}_{\mathrm{b}} = -\tfrac{1}{4}\mathsf{G}^{\mu\nu} \cdot \mathsf{G}_{\mu\nu} + \tfrac{1}{2} D_\mu \mathsf{S}_I \cdot D^\mu \mathsf{S}_I + \tfrac{1}{2} D_\mu \mathsf{P}_J \cdot D^\mu \mathsf{P}_J \\ - \tfrac{1}{4} e^2 \|\mathsf{S}_I \times \mathsf{S}_J\|^2 - \tfrac{1}{4} e^2 \|\mathsf{P}_I \times \mathsf{P}_J\|^2 - \tfrac{1}{2} e^2 \|\mathsf{S}_I \times \mathsf{P}_J\|^2 \ . \quad (2.34) \end{aligned}$$

The fermionic part of the action requires a bit more work, but it is nevertheless straightforward, and is left as an exercise.

Exercise 2.35 (The fermionic terms in the lagrangian)
Using the explicit form of the Γ*-matrices, prove that the term* $\frac{i}{2}\bar{\boldsymbol{\Psi}} \cdot \Gamma^A D_A \boldsymbol{\Psi}$ *in equation* (2.25)*, becomes*

$$\mathcal{L}_{\mathrm{f}} = \tfrac{i}{2}\bar{\boldsymbol{\psi}} \cdot \gamma^\mu D_\mu \boldsymbol{\psi} + \tfrac{e}{2}\bar{\boldsymbol{\psi}} \cdot \left((\alpha^I \mathsf{S}_I + \beta^J \mathsf{P}_J \gamma_5) \times \boldsymbol{\psi} \right) \ , \tag{2.35}$$

from where we see that indeed we were justified in calling S_I *scalars and* P_J *pseudoscalars.*

We now write down the supersymmetry transformations. In ten dimensions, the parameter of the supersymmetry transformation is a Majorana–Weyl spinor. As we have seen, upon dimensional reduction, such a spinor yields a quartet of Majorana spinors in four-dimensions. Therefore the four-dimensional theory will have N=4 supersymmetry. Indeed the lagrangian density $\mathcal{L}_{\mathrm{b}} + \mathcal{L}_{\mathrm{f}}$, understood in four dimensions, defines N=4 *supersymmetric Yang–Mills theory*.

Since the supersymmetry parameter $\boldsymbol{\alpha}$ is a Majorana–Weyl spinor obeying $\Gamma_{11}\boldsymbol{\alpha} = -\boldsymbol{\alpha}$, we can write it as $\boldsymbol{\alpha} = \boldsymbol{\epsilon} \otimes \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$, where $\boldsymbol{\epsilon}$ is a quartet of four-dimensional anticommuting Majorana spinors.

Exercise 2.36 (N=4 supersymmetry transformations)
Expand equation (2.26) *in this reparametrisation to obtain the following supersymmetry transformations for the four-dimensional fields:*

$$\begin{aligned}
\delta \mathsf{W}_\mu &= i\bar{\boldsymbol{\epsilon}}\gamma_\mu \boldsymbol{\psi} \\
\delta \mathsf{S}_I &= \bar{\boldsymbol{\epsilon}}\alpha^I \boldsymbol{\psi} \\
\delta \mathsf{P}_J &= \bar{\boldsymbol{\epsilon}}\gamma_5 \beta^J \boldsymbol{\psi} \\
\delta \boldsymbol{\psi} &= \tfrac{1}{2}\mathsf{G}^{\mu\nu}\gamma_{\mu\nu}\boldsymbol{\epsilon} + iD_\mu \mathsf{S}_I \gamma^\mu \alpha^I \boldsymbol{\epsilon} + iD_\mu \mathsf{P}_J \gamma^\mu \gamma_5 \beta^J \boldsymbol{\epsilon} \\
&\quad - e(\mathsf{S}_I \times \mathsf{P}_J)\gamma_5 \alpha^I \beta^J \boldsymbol{\epsilon} + \tfrac{1}{2} e \epsilon_{IJK} (\mathsf{S}_I \times \mathsf{S}_J) \alpha^K \boldsymbol{\epsilon} + \tfrac{1}{2} e \epsilon_{IJK} (\mathsf{P}_I \times \mathsf{P}_J) \beta^K \boldsymbol{\epsilon} \ .
\end{aligned}$$

Finally we have the $su(4)$ invariance of the action.

Exercise 2.37 ($su(4)$ invariance)
Prove that, for every choice of constant parameters (a_{IJ}, b_{IJ}, c_{IJ}) *where* $a_{IJ} = -a_{JI}$ *and* $b_{IJ} = -b_{JI}$*, the following transformations are a symmetry of* N=4 *supersymmetric Yang–Mills theory:*

$$\begin{aligned}
\delta \mathsf{W}_\mu &= 0 \\
\delta \mathsf{S}_I &= 2a_{IJ}\mathsf{S}_J + 2c_{IJ}\mathsf{P}_J \\
\delta \mathsf{P}_I &= 2b_{IJ}\mathsf{P}_J - 2c_{JI}\mathsf{S}_J \\
\delta \boldsymbol{\psi} &= -\tfrac{1}{2}a_{IJ}A^{IJ}\boldsymbol{\psi} - \tfrac{1}{2}b_{IJ}B^{IJ}\boldsymbol{\psi} + \tfrac{i}{2}c_{IJ}C^{IJ}\gamma_5\boldsymbol{\psi} \ ,
\end{aligned}$$

where A^{IJ}, B^{IJ} and C^{IJ} are the $su(4)$ generators in the fundamental representation given by equation (2.33).
(Hint: You may save some time by first showing that these transformations are induced from Lorentz transformations in ten dimensions, and then using the Lorentz invariance of the ten-dimensional action.)

2.4.3 Monopoles and gauge bosons in N=4 supersymmetric Yang–Mills

In section 2.3.4, we saw how any BPS-monopole could be thought of as a solution to the equations of motion of N=2 supersymmetric Yang–Mills by setting the fermions to zero and aligning the scalar fields properly. Moreover we saw that such solutions break one half of the supersymmetry, so that these N=2 BPS-monopoles naturally belong to a short multiplet. In fact, they belong to the short multiplet with spin s=0. On the other hand, we had seen in section 2.3.3 that after higgsing, the perturbative spectrum of the theory arranged itself in a massless vector multiplet corresponding to the unbroken $U(1)$ and massive short multiplets with spin $s=\frac{1}{2}$ containing the massive vector bosons. It therefore seemed unlikely that N=2 super Yang–Mills would be self-dual, since the perturbative spectrum of the dual theory (i.e., the monopoles) now live in a different supersymmetry multiplet. And in fact, we now know from the results of Seiberg and Witten, that this theory is not self-dual. In this section we will see that this obstacle is overcome in N=4 supersymmetric Yang–Mills theory. The discussion is very similar to that of sections 2.3.3 and 2.3.4, with the important distinction that the short multiplet containing the BPS-monopole and the one containing the massive vector boson are now isomorphic, being the one with spin s=0. This section and the next are based on the work of Osborn [Osb79].

The bosonic part of the hamiltonian density corresponding to the N=4 supersymmetric Yang–Mills theory defined by (2.34) and (2.35) is given by:

$$\mathcal{H}_{\rm b} = \tfrac{1}{2}\|\mathsf{E}_i\|^2 + \tfrac{1}{2}\|D_0\mathsf{S}_I\|^2 + \tfrac{1}{2}\|D_0\mathsf{P}_J\|^2 + \tfrac{1}{2}\|\mathsf{B}_i\|^2 + \tfrac{1}{2}\|D_i\mathsf{S}_I\|^2 + \tfrac{1}{2}\|D_i\mathsf{P}_J\|^2 + \tfrac{1}{2}e^2\|\mathsf{S}_I\times\mathsf{P}_J\|^2 + \tfrac{1}{4}e^2\|\mathsf{S}_I\times\mathsf{S}_J\|^2 + \tfrac{1}{4}e^2\|\mathsf{P}_I\times\mathsf{P}_J\|^2\ .$$

Demanding that the energy of a given field configuration be finite doesn't necessarily imply that all the scalars P_I and S_J acquire non-zero vacuum expectation values at spatial infinity. Indeed, looking at the potential terms it is sufficient (for $so(3)$) that they be parallel. This defines the supersymmetric Prasad–Sommerfield limit as in N=2. In more detail, we add a potential term $\lambda\left(\|\mathsf{S}_I\|^2+\|\mathsf{P}_J\|^2-a^2\right)^2$ to the lagrangian (2.34) to force S_I and P_J to acquire a nonzero vacuum expectation value, but since such a term would break supersymmetry, we take the limit $\lambda \downarrow 0$ while keeping the nonzero vacuum expectation values of S_I and P_J. This restores the supersymmetry provided that $\langle \mathsf{S}_I \rangle$ and $\langle \mathsf{P}_J \rangle$ are parallel. We could choose $\mathsf{S}_I = a_I \boldsymbol{\phi}$ and $\mathsf{P}_J = b_J \boldsymbol{\phi}$ where $\sum_I (a_I^2 + b_I^2) = 1$, and where $\langle \boldsymbol{\phi} \rangle$ has length a at infinity. Since the potential depends only on the $SO(6)$ invariant combination $\|\mathsf{S}_I\|^2 + \|\mathsf{P}_J\|^2$, we could use this symmetry to choose, say, $b_J = a_2 = a_3 = 0$, $a_1 = 1$ and $\langle \boldsymbol{\phi} \rangle = \boldsymbol{a}$, where $\boldsymbol{a}$ is a fixed vector with $\|\boldsymbol{a}\|^2 = a^2$.

Exercise 2.38 (The perturbative spectrum after higgsing)
We can analyse the spectrum of the model around such a vacuum in exactly the same way as we did in Exercise 1.4. Choosing for example the unitary *gauge $\boldsymbol{a} = a\boldsymbol{e}_3$, show that there is now a massless gauge multiplet with helicity $\lambda = -1$, corresponding to the unbroken $U(1)$: $(W_\mu^3, \boldsymbol{\psi}^3, S_I^3, P_J^3)$; and two massive multiplets $(\boldsymbol{\psi}^\pm, W_\mu^\pm, P_I^\pm, S_{1,2}^\pm)$ of mass $M_W = ae\hbar$. Conclude that these massive multiplets are actually short multiplets of spin s=0.*

Now let's see to what kind of multiplets the N=4 BPS-monopoles belong. Let $(\mathsf{W}_i, \boldsymbol{\phi})$ be a BPS-monopole and let us set $\mathsf{W}_0 = \boldsymbol{\psi} = 0$, $\mathsf{S}_I = a_I \boldsymbol{\phi}$, and $\mathsf{P}_J = b_J \boldsymbol{\phi}$, where a_I and b_J are real numbers satisfying $\sum_I (a_I^2 + b_I^2) = 1$. Because the fermions are zero, only the bosonic part of the lagrangian is nonzero. Plugging in these field configurations into (2.34), we find

$$\mathcal{L} = -\tfrac{1}{4} G_{ij} G_{ij} - \tfrac{1}{2} \|D_i \boldsymbol{\phi}\|^2 ,$$

after using that the fields are static and that S_I and P_J are all collinear. But this is precisely the action for static solutions to the bosonic Yang–Mills–Higgs theory, hence it is minimised by BPS-monopoles. Therefore the above field configurations minimise the equations of motion of N=4 supersymmetric Yang–Mills. In other words, we have shown that any BPS-monopole can be embedded as a solution of the N=4 supersymmetric Yang–Mills theory. (Compare with Exercise 2.17.)

Now we will prove that such a solution breaks one half of the supersymmetry and hence it lives in a short multiplet. Because the fermions are put to zero, the supersymmetry transformation of the bosonic fields is automatically zero. From the results of Exercise 2.36 we can read off the expression for the supersymmetry transformation of the spinors in this background:

$$\delta \boldsymbol{\psi} = \left(\tfrac{1}{2} G_{ij} \gamma_{ij} - i D_i \boldsymbol{\phi} \gamma_i (a_I \alpha^I + b_J \beta^J \gamma_5) \right) \boldsymbol{\epsilon} .$$

If we now use equation (2.19), and the Bogomol'nyi equation in the form $G_{ij} = \epsilon_{ijk} D_k \boldsymbol{\phi}$, $\delta \boldsymbol{\psi}$ takes the form:

$$\begin{aligned} \delta \boldsymbol{\psi} &= \gamma_k D_k \boldsymbol{\phi} \left(\gamma_5 \gamma_0 - i (a_I \alpha^I + b_J \beta^J \gamma_5) \right) \boldsymbol{\epsilon} \\ &= \gamma_5 \bar{\gamma}_k D_k \boldsymbol{\phi} \left(\mathbb{1} - \bar{\gamma}_5 \right) \boldsymbol{\epsilon} , \end{aligned}$$

where $\bar{\gamma}_i \equiv \gamma_0\gamma_i$, $\bar{\gamma}_4 \equiv -i\gamma_0(a_I\alpha^I + b_J\beta^J\gamma_5)$, and $\bar{\gamma}_5 \equiv \bar{\gamma}_1\bar{\gamma}_2\bar{\gamma}_3\bar{\gamma}_4$. As expected, the $\bar{\gamma}_{\underline{i}}$ generate a euclidean Clifford algebra, and it follows that $\frac{1}{2}(\mathbb{1} - \bar{\gamma}_5)$ projects out one half of the states: those which have positive chirality with respect to $\bar{\gamma}_5$. Hence we conclude that N=4 BPS-monopoles break one half of the supersymmetry.

From the analysis of N=4 multiplets in section 2.2.3 and, in particular, from Exercise 2.6, we see that a multiplet where states break one half of the supersymmetries are short; and looking at the spectrum, we see that it is the short multiplet with spin s=0, which is the only short multiplet with spins not exceeding 1. Therefore the BPS-monopole and the massive vector boson belong to isomorphic multiplets. This solves the second problem with the Montonen–Olive duality conjecture alluded to in section 1.4.1—for, certainly, if N=4 supersymmetric Yang–Mills is to be self-dual, the BPS-monopole and the massive vector boson should belong to isomorphic multiplets.

Finally, we come to a minor point. The supersymmetry parameter $\boldsymbol{\epsilon}$, just like the fermion $\boldsymbol{\psi}$, is a quartet of Majorana spinors. The additional condition for the parameter to preserve the supersymmetry is that it be chiral with respect to $\bar{\gamma}_5$. One might be tempted to think that there is a problem since in four-dimensions (either with euclidean or lorentzian signature) there are no Majorana–Weyl spinors. However the Majorana condition is a condition in Minkowski spacetime, whereas the chirality condition is a condition relative to the euclidean $\bar{\gamma}_5$. We will see this more explicitly later on when we consider the effective action for the collective coordinates, but for now let us simply state without proof that these two conditions are indeed simultaneously realisable.

2.4.4 The mass bound for N=4 super Yang–Mills

We end this chapter with a derivation of the mass bound for N=4 super Yang–Mills. Keeping in mind the similar calculation for N=2 super Yang–Mills, it should come as no surprise that the mass bound coincides once again with the Bogomol'nyi bound. In order to derive the mass bound, we will first write down the algebra obeyed by the supersymmetry charges in d=10 N=1 super Yang–Mills. After dimensional reduction this will give us an explicit expression for the central charges appearing in the four-dimensional supersymmetry algebra. Naturally one could compute the supersymmetry algebra directly in four dimensions, but we find it simple to dimensionally reduce the algebra in ten dimensions.

The supersymmetry algebra in ten dimensions

The supersymmetry algebra can be derived by varying the supersymmetry current (2.29). Indeed, the supersymmetry algebra will be read off from the space integral of the supersymmetry variation of the timelike (zeroth) component of the supersymmetry current. Explicitly, if $\boldsymbol{\epsilon}$ is a Majorana–Weyl spinor just like $\boldsymbol{\Psi}$, then

$$\bar{\boldsymbol{\alpha}}\{Q,\bar{Q}\}\boldsymbol{\epsilon} = -i\int_{\text{space}} \delta\bar{J}^0\boldsymbol{\epsilon} \ ,$$

where the integral is over a spacelike hypersurface. We can get an idea of what to expect in the right-hand side of the supersymmetry algebra purely from the fact that Q is an anticommuting Majorana–Weyl spinor. From Exercise 2.28 we see that in the right-hand side of the supersymmetry algebra, we expect only terms consisting of an odd number of Γ matrices and moreover only those bispinors Γ for which $\mathcal{C}\Gamma$ is symmetric, since so is the left-hand side of the supersymmetry algebra. Using equation (2.28), we see that only those terms with 1 and 5 Γ matrices survive. We now turn to the computation, which is left as an exercise. We will only be interested in terms which survive in a BPS-monopole monopole background in which the fermions have been put to zero.

Exercise 2.39 (The supersymmetry algebra in ten dimensions)
Prove that up to terms involving the fermions, the variation of the supersymmetry current is given by

$$-i\delta\bar{J}^E\boldsymbol{\epsilon} = -\tfrac{1}{4}\bar{\boldsymbol{\alpha}}\Gamma^{AB}\Gamma^E\Gamma^{CD}\boldsymbol{\epsilon}\,\mathsf{G}_{AB}\cdot\mathsf{G}_{CD} \ .$$

Perform the Γ matrix algebra and, taking into account that $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ are Majorana–Weyl, show that

$$-i\delta\bar{J}^E\boldsymbol{\epsilon} = \bar{\boldsymbol{\alpha}}\left(-\tfrac{1}{4}\Gamma^{ABCDE}\mathsf{G}_{AB}\cdot\mathsf{G}_{CD} + 2\mathsf{G}^{EA}\cdot\mathsf{G}_{AB}\Gamma^B + \tfrac{1}{2}\mathsf{G}^{AB}\cdot\mathsf{G}_{AB}\Gamma^E\right)\boldsymbol{\epsilon} \ .$$

Prove the identity

$$\Gamma^{ABCDE} = -\tfrac{1}{5!}\epsilon^{ABCDEFGHIJ}\Gamma_{FGHIJ}\Gamma_{11} \ ;$$

and using the fact that $\Gamma_{11}\boldsymbol{\epsilon} = -\boldsymbol{\epsilon}$, conclude that

$$-i\delta\bar{J}^E\boldsymbol{\epsilon} = 2\bar{\boldsymbol{\alpha}}\left(\tfrac{1}{8\cdot 5!}\epsilon^{ABCDEFGHIJ}\mathsf{G}_{AB}\cdot\mathsf{G}_{CD}\Gamma_{FGHIJ} + \left(\mathsf{G}^{EA}\cdot\mathsf{G}_{AB} + \tfrac{1}{4}\mathsf{G}^{CD}\cdot\mathsf{G}_{CD}\delta^E_B\right)\Gamma^B\right)\boldsymbol{\epsilon} \ .$$

We now define the following tensors

$$\begin{aligned} T^{AB} &= \mathsf{G}^{AC} \cdot \mathsf{G}_C{}^B + \tfrac{1}{4}\eta^{AB}\mathsf{G}^{CD} \cdot \mathsf{G}_{CD} \\ \Theta^{ABCDEF} &= \tfrac{1}{8}\epsilon^{ABCDEFGHIJ}\mathsf{G}_{GH} \cdot \mathsf{G}_{IJ} \ . \end{aligned}$$

We recognise T as the bosonic part of the (improved) energy-momentum tensor of the super Yang–Mills theory. The momentum is then given by the space integral of T^{0A}:

$$P^A = \int_{\text{space}} T^{0A} \ .$$

How about Θ? Just as in the case of N=2, it is a topological current.

Exercise 2.40 (Another topological current)
Prove that Θ_{ABCDEF} is gauge invariant and that it is conserved without imposing the equations of motion.
(Hint: Compare with Exercise 2.23.)

We define the topological charge associated to Θ as the space integral of Θ^{0ABCDE}:

$$Z^{ABCDE} = \int_{\text{space}} \Theta^{0ABCDE} \ .$$

In summary, the supersymmetry algebra remains as follows:

$$\boxed{\{Q, \bar{Q}\} = 2P^A\Gamma_A + \tfrac{2}{5!}Z^{ABCDE}\Gamma_{ABCDE}} \ . \tag{2.36}$$

In 10 dimensions, the 5-form Z^{ABCDE} can be decomposed into a self-dual and an antiself-dual part. The next exercise asks you to show that only the self-dual part contributes to the algebra.

Exercise 2.41 (A self-dual 5-form)
Using the fact that the supersymmetry charge has negative chirality, show that we can for free project onto the self-dual part of Z^{ABCDE} in the left-hand side of the supersymmetry algebra.

This 5-form belies the existence of a 5-brane solution of ten dimensional supersymmetric Yang–Mills. Under double dimensional reduction, it gives rise to the string-like solution of six-dimensional supersymmetric Yang–Mills briefly alluded to in section 2.3.4.

The supersymmetry algebra in four dimensions

In order to write down the supersymmetry algebra in four dimensions, we need to dimensionally reduce both the momenta and the topological charge appearing in the ten-dimensional algebra (2.36). We will assume from the start a BPS-monopole background where the fermions are put to zero. We will not demand that the solutions be static, since that is the only way we can generate electric charge. Moreover we will exploit the internal $SO(6)$ symmetry to choose $\mathsf{P}_J = \mathsf{S}_{2,3} = 0$ and $\mathsf{S}_1 = \boldsymbol{\phi}$. In such a background, the only nonzero components of the field strength G_{AB} are $\mathsf{G}_{\mu\nu}$, $\mathsf{G}_{\mu 4}$. This limits considerably the nonzero terms of the momentum P^A and the topological charge Z^{ABCDE}, as the next exercise shows.

Exercise 2.42 (Momentum and topological charge in this background) *Prove that in the background chosen above, the only nonzero components of the momentum and topological charge densities are the following: $T^{0\mu}$, T^{04}, Θ^{056789}. The first term is of course simply the four-momentum density, whereas the other two terms are given by:*

$$\begin{aligned} T^{04} &= -\mathsf{G}_{0i} \cdot D_i \boldsymbol{\phi} = -\partial_i \left(\mathsf{G}_{0i} \cdot \boldsymbol{\phi}\right) \\ \Theta^{056789} &= -\tfrac{1}{2}\epsilon_{ijk} \mathsf{G}_{ij} \cdot D_k \boldsymbol{\phi} = -\tfrac{1}{2}\partial_k \left(\epsilon_{ijk}\mathsf{G}_{ij} \cdot \boldsymbol{\phi}\right) \ . \end{aligned}$$

(Hint: In order to rewrite the right-hand sides of the equations, use the equations of motion in this background, and the Bianchi identity. (Compare with the discussion following Exercise 2.24.))

Taking into account the results of the previous exercise we can rewrite the supersymmetry algebra in four dimensions as follows:

$$\{Q, \bar{Q}\} = 2\Gamma_\mu P^\mu - 2\Gamma_4 \int_{\Sigma_\infty} \mathsf{G}_{0i} \cdot \boldsymbol{\phi}\, d\Sigma_i - \Gamma_{56789}\epsilon_{ijk} \int_{\Sigma_\infty} \mathsf{G}_{ij} \cdot \boldsymbol{\phi}\, d\Sigma_k \ . \quad (2.37)$$

But now notice that Γ_μ, Γ_4 and $\Gamma_{56789} = \Gamma_5\Gamma_6\Gamma_7\Gamma_8\Gamma_9$ generate a lorentzian Clifford algebra of signature (1,5), hence the supersymmetry algebra (2.37) is formally identical to one in six-dimensional Minkowski spacetime where the momenta in the extra two dimensions are given by:

$$\begin{aligned} P'_5 &= \int_{\Sigma_\infty} \mathsf{G}_{0i} \cdot \boldsymbol{\phi}\, d\Sigma_i = -aq \\ P'_6 &= \tfrac{1}{2}\epsilon_{ijk} \int_{\Sigma_\infty} \mathsf{G}_{ij} \cdot \boldsymbol{\phi}\, d\Sigma_k = ag \ , \end{aligned}$$

and where we have used equations (1.24) and (1.25) to rewrite the extra momenta in terms of the electric and magnetic charges. Finally, as in the $N{=}2$ case, the mass bound is simply the positivity of the six-dimensional "mass" given by equation (2.20). Plugging in the expression for the extra momenta, we once again recover the Bogomol'nyi bound (1.27).

Chapter 3

Collective Coordinates

BPS-monopoles are static: any motion, however small, increases their kinetic energy and makes their total energy strictly greater than the Bogomol'nyi bound. Nevertheless, if we keep the velocity small *and* if the motion starts off *tangent* to the space of static BPS-monopoles, energy conservation will prevent the motion from taking the monopoles very far away from this space. Much like a point-particle moving slowly near the bottom of a potential well, the motion of slow BPS-monopoles may be approximated by motion on the space of static BPS-monopoles (*i.e.*, along the flat directions of the potential) and small oscillations in the transverse directions. We can trade the limit of velocities going to zero, for a limit in which the potential well becomes infinitely steep. This suppresses the oscillations in the transverse directions (which become increasingly expensive energetically) and motion is effectively constrained to take place along the flat directions, since this motion costs very little energy. Manton [Man82] showed that the motion along the flat directions is *geodesic* relative to a metric on the moduli space of BPS-monopoles, which is induced naturally from the Yang–Mills–Higgs action functional. Expanding the action functional around a BPS-monopole gives rise to an effective theory in terms of *collective coordinates*. These are the coordinates on the moduli space of BPS-monopoles and the effective action is nothing but a (1+0)-dimensional σ-model with target space the moduli space.

In this chapter we will study the moduli space $\mathcal{M}$ of BPS-monopoles. Our aim is to prove that it is a hyperkähler manifold which, for a given magnetic charge, is finite-dimensional, and to compute its (formal) dimension. For the simplest case of magnetic charge k=1, we will also work out the metric explicitly from the field theory and hence the effective action. We quantise the effective action following the review in the introduction of [GM86]. This is as much as can be done directly with field-theoretical methods. The only other case in which the metric on moduli space is known exactly is the $k=2$ monopole sector. This metric was constructed by Atiyah and Hitchin [AH85] by indirect methods. We will eventually review their construction as well.

3.1 The metric on the moduli space

We start by constructing the metric on the true physical configuration space of the Yang–Mills–Higgs theory. This will induce a metric on the moduli space of BPS-monopoles, which is a submanifold.

3.1.1 The physical configuration space

Let $\mathcal{A}'$ denote the space of configurations $(\mathsf{W}_\mu, \boldsymbol{\phi})$ of Yang–Mills–Higgs fields of finite energy in the Prasad–Sommerfield limit. Recall that in this limit, the energy density is given by (1.13) setting $V(\boldsymbol{\phi}) = 0$. Configurations which are related by a *short-range* gauge transformation—that is, gauge transformations which tend to the identity at infinity—are to be thought of as physically indistinguishable. Hence if we let $\mathcal{G}'$ denote the group of short-range gauge transformations, the true configuration space of the Yang–Mills–Higgs system is the quotient

$$\mathcal{C} \cong \mathcal{A}'/\mathcal{G}' \ .$$

It is convenient to fix the gauge partially by setting $\mathsf{W}_0 = 0$, that is by going to the *temporal* gauge. This still leaves the freedom of performing time-independent gauge transformations, since these are the gauge transformations which preserve the temporal gauge:

$$e\,\delta_{\boldsymbol{\epsilon}} \mathsf{W}_0 = D_0 \boldsymbol{\epsilon} = \dot{\boldsymbol{\epsilon}} = 0 \ .$$

We will therefore let $\mathcal{G} \subset \mathcal{G}'$ denote the group of time-independent short-range gauge transformations.

The temporal gauge $\mathsf{W}_0 = 0$ is preserved by the dynamics provided we impose its equation of motion, that is, *Gauss's law*:

$$D_i \dot{\mathsf{W}}_i - e\boldsymbol{\phi} \times \dot{\boldsymbol{\phi}} = 0 \ . \tag{3.1}$$

Therefore if we let $\mathcal{A}$ denote the space of finite-energy configurations $(\mathsf{W}_i, \boldsymbol{\phi})$ subject to Gauss's law, then the configuration space $\mathcal{C}$ also admits the description

$$\mathcal{C} \cong \mathcal{A}/\mathcal{G} \ .$$

Notice that setting $\mathsf{W}_0 = 0$ means that there will be no static dyons (*cf.* the Julia–Zee Ansatz). We will see that dyonic solutions emerge when we consider moving monopoles.

We can simplify many of the calculations by describing the space $\mathcal{A}$ in a different way. We introduce a fourth spatial coordinate x^4 and interpret the Higgs field $\boldsymbol{\phi}$ as the fourth-component W_4 of the gauge field. Notice however that we must impose that nothing depends on the new coordinate: $\partial_4 \equiv 0$. We will write $\mathsf{W}_{\underline{i}} = (\mathsf{W}_i, \mathsf{W}_4 = \boldsymbol{\phi})$, where the underlined indices run from 1 to 4. Notice that the field-strength has components $\mathsf{G}_{\underline{ij}} = (\mathsf{G}_{ij}, \mathsf{G}_{i4} = D_i\boldsymbol{\phi})$. In this new notation, gauge transformations, Gauss's law and the Bogomol'nyi equation all have natural and simple descriptions.

Exercise 3.1 (The BPS-monopole as an instanton)
Prove that infinitesimal gauge transformations on the Yang–Mills–Higgs system now take the form

$$\delta_\epsilon \mathsf{W}_{\underline{i}} = \frac{1}{e} D_{\underline{i}} \boldsymbol{\epsilon} \ ;$$

that Gauss's law (3.1) *becomes simply*

$$D_{\underline{i}} \dot{\mathsf{W}}_{\underline{i}} = 0 \ ; \tag{3.2}$$

and that the Bogomol'nyi equation (1.28) *is nothing but the (anti)self-duality equation*

$$\mathsf{G}_{\underline{ij}} = \pm \tfrac{1}{2} \epsilon_{\underline{ijk\ell}} \mathsf{G}_{\underline{k\ell}} \ . \tag{3.3}$$

In summary, this proves that BPS-monopoles in 3+1 *dimensions are in one-to-one correspondence with static instantons in* 4+1 *dimensions which are translationally invariant in the fourth spatial direction.*

Therefore in this description, the space $\mathcal{A}$ is given as those gauge fields $\mathsf{W}_{\underline{i}}$ in 4+1 dimensions, independent of x^4, of finite energy per unit length in the x^4-direction, and whose time-dependence is subject to (3.2).

3.1.2 The metric on the physical configuration space

In the temporal gauge, the Yang–Mills–Higgs lagrangian in the Prasad–Sommerfield limit is given as a difference of two terms: $L = T - V$, where the kinetic term T is the 3-space integral of

$$\tfrac{1}{2} \|\dot{\mathsf{W}}_{\underline{i}}\|^2 = \tfrac{1}{2} \|\dot{\mathsf{W}}_i\|^2 + \tfrac{1}{2} \|\dot{\boldsymbol{\phi}}\|^2 \ , \tag{3.4}$$

and the potential term V is the 3-space integral of

$$\tfrac{1}{2} \|\mathsf{B}_i\|^2 + \tfrac{1}{2} \|D_i\boldsymbol{\phi}\|^2 = \tfrac{1}{2} \|\mathsf{B}_i \mp D_i\boldsymbol{\phi}\|^2 \pm \partial_i \left(\boldsymbol{\phi} \cdot \mathsf{B}_i\right) \ . \tag{3.5}$$

We will now show that the lagrangian is well-defined in the true configuration space $\mathcal{C}$. The kinetic term will induce a metric.

Suppose we would like to compute the value of the potential on some point in $\mathcal{C}$. Points in $\mathcal{C}$ are equivalence classes $[\mathsf{W}_{\underline{i}}]$ of points $\mathsf{W}_{\underline{i}}$ in $\mathcal{A}$: two points in $\mathcal{A}$ belong to the same equivalence class if and only if they are related by a gauge transformation in $\mathcal{G}$; that is, if they lie on the same $\mathcal{G}$-orbit. To define a potential on $\mathcal{C}$ we can simply use the potential term (3.5) on $\mathcal{A}$ as follows: to find out the value of the potential on a point $[\mathsf{W}_{\underline{i}}]$ in $\mathcal{C}$, we choose some point $\mathsf{W}_{\underline{i}}$ in the same equivalence class, and we evaluate the potential (3.5) on it. This will only make sense if the value of the potential doesn't depend on which element in the equivalence class we have chosen; that is, if the potential is gauge-invariant. More formally, a function on $\mathcal{A}$ will induce a function on $\mathcal{C} = \mathcal{A}/\mathcal{G}$ if and only if it is $\mathcal{G}$-invariant. Luckily this is the case, since in the temporal gauge, the potential and kinetic terms are separately invariant under time-independent gauge-transformations.

The kinetic term is trickier since it is not strictly speaking a function on $\mathcal{A}$: it requires not just knowledge of $\mathsf{W}_{\underline{i}}$ but also of its time-derivative $\dot{\mathsf{W}}_{\underline{i}}$. In other words, it is a function on the tangent bundle $T\mathcal{A}$. The typical fibre at a point $[\mathsf{W}_{\underline{i}}]$ of the tangent bundle is spanned by the velocities of all smooth curves passing through that point. We may lift such curves to curves in $\mathcal{A}$, but this procedure is not unique. First we have to choose a point $\mathsf{W}_{\underline{i}}$ in $\mathcal{A}$ in the equivalence class $[\mathsf{W}_{\underline{i}}]$. Just as before, this ambiguity is immaterial since the kinetic term is invariant under time-independent gauge transformations. But now we also have to choose a tangent vector $\dot{\mathsf{W}}_{\underline{i}}$. Clearly adding to a tangent vector a vector tangent to the orbits of $\mathcal{G}$ does not change the curve in $\mathcal{C}$ since every $\mathcal{G}$-orbit in $\mathcal{A}$ is identified with a single point in $\mathcal{C}$. Hence the kinetic term should be impervious to such a change. Tangent vectors to $\mathcal{G}$ are infinitesimal gauge-transformations whose parameters go to zero at spatial infinity, therefore the kinetic term (3.4) defines a kinetic energy on $\mathcal{C}$ provided that $\dot{\mathsf{W}}_{\underline{i}}$ and $\dot{\mathsf{W}}_{\underline{i}} + D_{\underline{i}}\boldsymbol{\epsilon}$ have the same kinetic energy. Integrating by parts we see that this is a consequence of Gauss's law (3.2).

In summary, the Yang–Mills–Higgs lagrangian induces a lagrangian in the true configuration space $\mathcal{C}$, whose energy is of course given by $E = T + V$. The kinetic energy term T defines a metric on $\mathcal{C}$. Motion on $\mathcal{C}$ is not "free" of course, since there is also a potential term, but for motion along the flat directions at the bottom of the potential well, this will be a good approximation. We turn to this now.

3.1.3 The metric on the moduli space

Let $\mathcal{M}$ denote the subspace of $\mathcal{C}$ where the energy E attains its minimum. From the explicit expression for T and V given in (3.4) and (3.5), we see that the minimum of the energy is given by

$$\left| \int_{\mathbb{R}^3} d^3x \, \partial_i \left(\boldsymbol{\phi} \cdot \mathsf{B}_i \right) \right| = a|g| = \frac{4\pi a}{e} |k| \ , \tag{3.6}$$

where g is the magnetic charge, and the integer k is the topological or monopole number. Clearly the minimum is attained by those configurations corresponding to static solutions on the Bogomol'nyi equation: BPS-monopoles, and where any two such solutions which are gauge-related are identified. In other words, $\mathcal{M}$ is the *moduli space* of static BPS-monopoles.

The monopole number labels different connected components of the space $\mathcal{A}$, so that

$$\mathcal{A} = \bigcup_k \mathcal{A}_k \ ,$$

but the gauge group $\mathcal{G}$ preserves each component. Therefore we can also decompose the true configuration space $\mathcal{C}$ as

$$\mathcal{C} = \bigcup_k \mathcal{C}_k \qquad \text{where} \quad \mathcal{C}_k = \mathcal{A}_k / \mathcal{G} \ .$$

Finally, let $\mathcal{M}_k = \mathcal{M} \cap \mathcal{C}_k$. This is then the moduli space of static BPS-monopoles of monopole number k, or BPS-k-monopoles, for short.

By definition, the potential is constant on $\mathcal{M}_k$, so that the Yang–Mills–Higgs lagrangian is given by

$$L = T - \frac{4\pi a}{e} |k| \ . \tag{3.7}$$

Therefore $\mathcal{M}_k$ corresponds to the manifold of flat directions of the potential. Manton's argument given at the beginning of this chapter, can now be proven. The motion of slow monopoles which start off tangent to $\mathcal{M}_k$ will consists of the superposition of two kinds of motions: motions along the flat directions $\mathcal{M}_k$ and small oscillations in the directions normal to $\mathcal{M}_k$. In the limit of zero velocity, the oscillatory motion is suppressed and we are left with motion on $\mathcal{M}_k$. But this motion is governed by the lagrangian (3.7) which only has a kinetic term—whence the motion is free, or in other words *geodesic* relative to the metric on $\mathcal{M}_k$ defined by T.

3.1.4 The 1-monopole moduli space

Let us study the moduli space $\mathcal{M}_1$ in the 1-monopole sector. The coordinates for $\mathcal{M}_1$ can be understood as parameters on which the BPS-monopole solution depends. In the 't Hooft–Polyakov Ansatz (1.15), the monopole is centred at the origin in $\mathbb{R}^3$, but the invariance under translations of the Yang–Mills–Higgs lagrangian (1.8) means that we can put the centre of the monopole where we please. This introduces three moduli parameters: $\boldsymbol{X}$. The time evolution of these parameters corresponds to the BPS-monopole moving as if it were a particle with mass $4\pi a/e$. The effective lagrangian for these *collective coordinates* is then

$$L_{\text{eff}} = \frac{2\pi a}{e}\dot{\boldsymbol{X}}^2 \ .$$

There is a fourth, more subtle, collective coordinate. Consider a one-parameter family $\mathsf{W}_{\underline{i}}(t)$ of gauge fields, but where the t-dependence is pure gauge:

$$\dot{\mathsf{W}}_{\underline{i}} = \frac{1}{e} D_{\underline{i}}\boldsymbol{\epsilon}(t) \ . \tag{3.8}$$

Since the potential is gauge invariant, this corresponds to a flat direction. But one might think that it is not a physical flat direction since it is tangent to the $\mathcal{G}$-orbits—it is an infinitesimal gauge transformation, after all. But recall that $\mathcal{G}$ is the group of short-range gauge transformations, whence for $D_{\underline{i}}\boldsymbol{\epsilon}$ to be tangent to the orbits, $\boldsymbol{\epsilon}$ has to tend to 0 as we approach infinity. Indeed, $D_{\underline{i}}\boldsymbol{\epsilon}$ would represent a physical deformation of the BPS-monopole if it would obey Gauss's law (3.2), which implies

$$D^2\boldsymbol{\epsilon} = 0 \ . \tag{3.9}$$

Exercise 3.2 (D^2 has no normalisable zero modes)
Prove that acting on square-integrable functions $D^2 = -D_{\underline{i}}^\dagger D_{\underline{i}}$ is a negative-definite operator. Deduce from this that any normalisable zero mode must be a zero mode of $D_{\underline{i}}$ for each $\underline{i}$, and deduce from this that the only square-integrable solution to (3.9) *is the trivial solution $\boldsymbol{\epsilon} = 0$. In other words, there exist no normalisable solutions.*
(Hint: If $D_{\underline{i}}\boldsymbol{\epsilon} = 0$, then $\|\epsilon\|^2$ is constant.)

This discussion suggests that we look for a gauge parameter $\boldsymbol{\epsilon}$ which does not tend to zero asymptotically. For example, let $\boldsymbol{\epsilon}(t) = f(t)\boldsymbol{\phi}$, where $f(t)$ is an arbitrary function. In the 1-monopole sector, $\boldsymbol{\phi}$ defines a map of degree 1 at infinity, hence it certainly does not go to zero. Moreover, using the Bogomol'nyi equation, it follows at once that $f(t)\boldsymbol{\phi}$ is a (un-normalisable) zero mode of D^2. It is clearly a true moduli parameter because it costs energy to excite it:

$$T = \frac{1}{2e^2} f^2 \int_{\mathbb{R}^3} \|D_i \boldsymbol{\phi}\|^2 > 0 \; . \tag{3.10}$$

We can understand this as follows. Let $g = \exp(\chi\boldsymbol{\phi}/a)$ be a time-dependent gauge transformation, where all the time-dependence resides in χ. Such a gauge transformation will move us away from the temporal gauge, but assume that at time t=0, say, we start from a configuration $\mathsf{W}_{\underline{i}}$ in the temporal gauge, and suppose that $g(t{=}0) = 1$. Then from (1.9),

$$\mathsf{W}_{\underline{i}}(t) = g\mathsf{W}_{\underline{i}}g^{-1} + \frac{1}{e}\partial_{\underline{i}} g\, g^{-1} \; ,$$

whence

$$\dot{\mathsf{W}}_{\underline{i}}(t) = \frac{1}{ae}\dot{\chi} D_{\underline{i}}\boldsymbol{\phi} \; .$$

Comparing with (3.8), we see that $f = \dot{\chi}/a$. Using (3.10), the kinetic energy of such a configuration is given by

$$T = \frac{1}{2a^2e^2}\dot{\chi}^2 \int_{\mathbb{R}^3} \|D_i\boldsymbol{\phi}\|^2 \; .$$

But notice that since $D_i\boldsymbol{\phi} = \mathsf{B}_i$,

$$\begin{aligned} T &= \frac{1}{2a^2e^2}\dot{\chi}^2 \int_{\mathbb{R}^3} \left(\tfrac{1}{2}\|D_i\boldsymbol{\phi}\|^2 + \tfrac{1}{2}\|\mathsf{B}_i\|^2\right) \\ &= \frac{1}{2a^2e^2}\dot{\chi}^2 \left(\frac{4\pi a}{e}\right) \\ &= \frac{2\pi}{ae^3}\dot{\chi}^2 \; . \end{aligned}$$

Notice that χ is an angular variable. To see this, let us define $g(\chi) = \exp(\chi\boldsymbol{\phi}/a)$. Then it is easy to see that $g(\chi)$ and $g(\chi{+}2\pi)$ are gauge-related in $\mathcal{G}$, that is, via short-range gauge transformations. Indeed, recall that $\|\boldsymbol{\phi}\| \to a$ at infinity in the Prasad–Sommerfield limit, whence $g(2\pi) = \exp(2\pi\boldsymbol{\phi}/a)$ tends to 1 at infinity. Since $g(\chi + 2\pi) = g(\chi)g(2\pi)$, we are done.

Assuming for the moment (we will prove this later) that there are no other collective coordinates in the 1-monopole sector, we have proven that the moduli space of BPS-1-monopoles is given by

$$\boxed{\mathcal{M}_1 \cong \mathbb{R}^3 \times S^1} \; ,$$

and the metric can be read off from the expression for the effective action

$$L_{\text{eff}} = \tfrac{1}{2} g_{ab} \dot{X}^a \dot{X}^b - \frac{4\pi a}{e} \tag{3.11}$$

$$= \frac{2\pi a}{e} \dot{\boldsymbol{X}}^2 + \frac{2\pi}{ae^3} \dot{\chi}^2 - \frac{4\pi a}{e} \ ; \tag{3.12}$$

that is

$$g_{ab} = \frac{4\pi a}{e} \begin{pmatrix} \mathbb{1}_3 & 0 \\ 0 & e^{-2} \end{pmatrix} \ ,$$

from where we can see that the radius of the circle is inversely proportional to the electric charge.

3.1.5 The quantisation of the effective action

The effective action (3.12) corresponds to a particle moving freely in $\mathbb{R}^3 \times S^1$ with the flat metric. The quantisation of this effective action is straightforward. The canonical momenta $(\boldsymbol{P}, Q)$ given by

$$\boldsymbol{P} = \frac{4\pi a}{e} \dot{\boldsymbol{X}} \qquad \text{and} \qquad Q = \frac{4\pi}{ae^3} \dot{\chi} \ ,$$

are conserved, and the hamiltonian is given by

$$H = \frac{e}{8\pi a} \boldsymbol{P}^2 + \frac{ae^3}{8\pi} Q^2 + \frac{4\pi a}{e} \ . \tag{3.13}$$

Bound states of minimum energy are given by those eigenstates of the hamiltonian for which $\boldsymbol{P} = 0$. Since χ is angular with period 2π, the eigenvalues of Q are quantised in units of $\hbar$, whence the spectrum looks like

$$E_n = \frac{4\pi a}{e} + \frac{ae^3}{8\pi} (n\hbar)^2 \ . \tag{3.14}$$

Notice that this energy spectrum has the standard form which corresponds to perturbative states around a nonperturbative vacuum. If we think of e as the coupling constant, then the zero-point energy is not analytic in e, hence it corresponds to a non-perturbative state in the theory: the BPS-monopole in this case. The second term in the energy corresponds to excitations around the monopole, which are clearly perturbative since their energy goes to zero as we let the coupling tend to zero.

Exercise 3.3 (The electric charge)
Prove that the electric field for classical configurations in which $\boldsymbol{P} = 0$ is given by

$$\mathsf{E}_i = -\mathsf{G}_{0i} = -\dot{\mathsf{W}}_i = -e^2 Q \mathsf{B}_i / 4\pi \ .$$

Conclude that $e\,Q$ can be interpreted as the electric charge.

Taking the above exercise into consideration, we see that the spectrum of the quantum effective theory corresponds to dyons of magnetic charge $-4\pi/e$ and electric charge $ne\hbar$, for $n \in \mathbb{Z}$. According to the classical BPS formula (1.27), the rest mass of such a dyon would be equal to

$$M_n = a\sqrt{\left(\frac{4\pi}{e}\right)^2 + (ne\hbar)^2} = \frac{4\pi a}{e}\sqrt{1 + \left(\frac{ne^2\hbar}{4\pi}\right)^2}$$

which, if we expand the square root assuming that e is small, becomes

$$= E_n + O(e^5) \ ,$$

where we've used (3.14).

In summary, the energy spectrum obtained from quantising the effective action of the collective coordinates is a small-coupling approximation to the expected BPS energy spectrum. However, even if their energy is only approximately correct, the multiplicity of bound states can be read accurately from the effective action. This is one of the important lessons to be drawn from the collective coordinate expansion.

In principle one can repeat this analysis in the k-monopole sector provided that one knows the form of the metric. But at the present moment this is only the case for k=1 and k=2. We will discuss the effective theory for k=2 later on in the lectures in the context of N=4 supersymmetric Yang–Mills.

3.1.6 Some general properties of the monopole moduli space

Quite a lot is known about the properties of the k-monopole[1] moduli space $\mathcal{M}_k$, even though its metric (and hence the effective action) is known explicitly only for $k = 1, 2$. As we saw in the previous section, the metric on $\mathcal{M}_1$ can be computed directly from the field theory. On the other hand, the metric on $\mathcal{M}_2$ can only be determined via indirect means. This result as well as much else of what is known about $\mathcal{M}_k$ is to be found either explicitly or referenced in the book [AH88] by Atiyah and Hitchin (see also [AH85]) to where we refer the reader for details.

We will now state some facts about $\mathcal{M}_k$. Some of them we will be able to prove later with field-theoretical means, but proving some others would take us too far afield. The following properties of $\mathcal{M}_k$ are known [AH88]:

[1] We will only concern ourselves with positive k: $\mathcal{M}_{-k}$ is naturally isomorphic to $\mathcal{M}_k$ by performing a parity transformation on the solutions.

1. $\mathcal{M}_k$ is a $4k$-dimensional (non-compact) complete riemannian manifold;

2. The natural metric on $\mathcal{M}_k$ is hyperkähler;

3. $\mathcal{M}_k \cong \tilde{\mathcal{M}}_k/\mathbb{Z}_k$ where $\tilde{\mathcal{M}}_k \cong (\mathbb{R}^3 \times S^1) \times \tilde{\mathcal{M}}_k^0$ as hyperkähler spaces.

4. $\tilde{\mathcal{M}}_k^0$ is a $4(k-1)$-dimensional, irreducible, simply-connected, hyperkähler manifold admitting an action of $SO(3)$ by isometries which rotates the three complex structures;

5. Asymptotically $\tilde{\mathcal{M}}_k \to \underbrace{\mathcal{M}_1 \times \mathcal{M}_1 \times \cdots \times \mathcal{M}_1}_{k \text{ times}} \equiv \mathcal{M}_1^k$, and $\mathcal{M}_k \to \mathcal{M}_1^k/\mathbb{Z}_k$. Physically this means that a configuration of well-separated BPS-k-monopoles can be considered as k 1-monopole configurations. The fact that BPS-monopoles are classically indistinguishable is responsible for the $\mathbb{Z}_k$-quotient.

3.2 $\dim \mathcal{M}_k = 4k$

In this section we compute the dimension of the moduli space of static BPS-monopoles. The strategy is typical of this kind of problems. We fix a reference BPS-monopole and ask in how many directions can we deform the solution infinitesimally and still remain with a BPS-monopole. Most of these directions will be unphysical: corresponding to infinitesimal gauge transformations. Discarding them leaves us with a finite number of physical directions along which to deform the BPS-monopole. In other words, we are computing the dimension of the tangent space at a particular point in the moduli space. If the point is regular (and generic points usually are) then this is the dimension of the moduli space itself. This number is in any case called the *formal dimension* of the moduli space.

Since the number of all deformations and of infinitesimal gauge transformations are both infinite, it is better to fix the gauge before counting: this eliminates the gauge-redundant deformations and leaves us with only a finite formal dimension. With a little extra argument, the counting can then be done via an index theorem. In the case of BPS-monopoles, the relevant index theorem is that of Callias [Cal78] (slightly modified by Weinberg [Wei79]) which is valid for open spaces and for operators with suitable decay properties at infinity. Weinberg's calculation contains steps which from a strictly mathematical point of view may be deemed unjustified. The necessary analytic details have been sorted out by Taubes [Tau83], but we will be following Weinberg's heuristic calculation in any case.

3.2.1 The dimension as an index

First let set up the problem. We want to find out in how many physically different ways can one deform a given BPS-monopole. It will turn out that these are given by zero modes of a differential operator. Asking for the number of zero modes will be the same as asking for the dimension of the tangent space at a given BPS-monopole solution. Hence let $t \rightsquigarrow (\mathsf{W}_i(t), \boldsymbol{\phi}(t))$ be a family of static BPS-monopoles (here t is an abstract parameter which has nothing to do with time). This means that $(\dot{\mathsf{W}}_i, \dot{\boldsymbol{\phi}}) \equiv (\dot{\mathsf{W}}_i(0), \dot{\boldsymbol{\phi}}(0))$ is a tangent vector to the moduli space at the point $(\mathsf{W}_i, \boldsymbol{\phi}) \equiv (\mathsf{W}_i(0), \boldsymbol{\phi}(0))$. Taking the t-derivative of the Bogomol'nyi equation (1.28), we find that $(\dot{\mathsf{W}}_i, \dot{\boldsymbol{\phi}})$ satisfies the linearised Bogomol'nyi equation:

$$D_i\dot{\boldsymbol{\phi}} + e\boldsymbol{\phi} \times \dot{\mathsf{W}}_i = \epsilon_{ijk} D_j \dot{\mathsf{W}}_k \ . \tag{3.15}$$

However, not every solution of the linearised Bogomol'nyi equation need be a physical deformation: it could be an infinitesimal gauge transformation. To make sure that it isn't, it is necessary to impose in addition Gauss's law (3.1). In other words, the dimension of the tangent space of the moduli space of BPS-monopoles is given by the maximum number of linearly independent solutions of *both* (3.15) and (3.1).

In order to count these solutions it will be convenient to rewrite both of these equations in terms of a single matrix-valued equation. We will define the following 2×2 complex matrix:

$$\boldsymbol{\Psi} = \dot{\boldsymbol{\phi}}\, \mathbb{1} + i\dot{\mathsf{W}}_j\, \sigma_j \ , \tag{3.16}$$

and the following linear operator:

$$\mathcal{D} = e\boldsymbol{\phi}\, \mathbb{1} + iD_j\, \sigma_j \ , \tag{3.17}$$

where we follow the convention that all fields which appear in operators are in the adjoint representation; that is, $\boldsymbol{\phi}$ really stands for $\operatorname{ad}\boldsymbol{\phi} = \boldsymbol{\phi} \times -$, etc.

Exercise 3.4 (Two equations in one)
Prove that the linearised Bogomol'nyi equation (3.15) *and Gauss's law* (3.1) *together are equivalent to the equation* $\mathcal{D}\boldsymbol{\Psi} = 0$.

We want to count the number of linearly independent real normalisable solutions to $\mathcal{D}\boldsymbol{\Psi} = 0$. It is easier to compute the index of the operator $\mathcal{D}$. By definition, the index of $\mathcal{D}$ is difference between the number of its normalisable zero modes and the number of normalisable zero modes of its hermitian adjoint $\mathcal{D}^\dagger$ relative to the inner product:

$$\int d^3x \ \operatorname{tr} \boldsymbol{\Psi}^*\boldsymbol{\Psi} = \int d^3x \ \left(\dot{\boldsymbol{\phi}}^* \cdot \dot{\boldsymbol{\phi}} + \dot{\mathsf{W}}_i^* \cdot \dot{\mathsf{W}}_i\right) \ ,$$

where * denotes complex conjugation of the fields and hermitian conjugation on the 2×2 matrices, and where tr denotes the 2×2 matrix trace.

The expression for the index of $\mathcal{D}$

$$\operatorname{ind} \mathcal{D} = \dim \ker \mathcal{D} - \dim \ker \mathcal{D}^\dagger$$

can be turned into an inequality

$$\dim \ker \mathcal{D} \geq \operatorname{ind} \mathcal{D} \ ,$$

which saturates precisely when $\mathcal{D}^\dagger$ has no normalisable zero modes. Happily this is the case, as the next exercise asks you to show.

Exercise 3.5 ($\mathcal{D}^\dagger$ has no normalisable zero modes)
Prove that $\mathcal{D}\mathcal{D}^\dagger$ is a positive-definite operator, whence it has no normalisable zero modes. Compare with Exercise 3.2. Also prove that for antimonopoles (i.e., the other sign in the Bogomol'nyi equation, it is $\mathcal{D}^\dagger\mathcal{D}$ that is the positive operator. (Hint: Use the fact that both $\boldsymbol{\phi}$ and D_j are antihermitian operators to prove that $\mathcal{D}^\dagger = -e\boldsymbol{\phi}\,\mathbb{1} + i\sigma_j\, D_j$, and that $\mathcal{D}\mathcal{D}^\dagger = -e^2\boldsymbol{\phi}^2 - (D_j)^2$. Deduce that this operator is positive-definite.

Therefore the number of normalisable zero modes of $\mathcal{D}$ equals the index of the operator $\mathcal{D}$. In the following sections we will compute the index of $\mathcal{D}$ acting on two-component complex vectors; that is, on functions $\mathbb{R}^3 \to \mathbb{C}^2$. However the (formal) dimension of monopole moduli space is given by the number of normalisable zero modes of $\mathcal{D}$ acting on matrices of the form (3.16). To a deformation $(\dot{\boldsymbol{\phi}}, \dot{\mathsf{W}}_i)$ there corresponds a matrix

$$\boldsymbol{\Psi} = \begin{pmatrix} \dot{\boldsymbol{\phi}} + i\dot{\mathsf{W}}_3 & \dot{\mathsf{W}}_2 + i\dot{\mathsf{W}}_1 \\ -\dot{\mathsf{W}}_2 + i\dot{\mathsf{W}}_1 & \dot{\boldsymbol{\phi}} - i\dot{\mathsf{W}}_3 \end{pmatrix} \ .$$

Clearly the first column of the above matrix $\boldsymbol{\Psi}$ determines the matrix. Moreover this first column is a normalisable zero mode of $\mathcal{D}$ acting on vectors if and only if $\boldsymbol{\Psi}$ is a normalisable zero mode of $\mathcal{D}$ acting on matrices. This would seems to indicate that there is a one-to-one correspondence between the normalisable zero modes of $\mathcal{D}$ acting on vectors and of $\mathcal{D}$ acting on matrices, but notice that $\mathcal{D}$ is a complex linear operator in a complex vector space, hence the space of its normalisable zero modes is complex, of complex dimension $\operatorname{ind}\mathcal{D}$. However the matrices $\boldsymbol{\Psi}$ and $i\boldsymbol{\Psi}$ determine linearly independent tangent vectors to monopole moduli space, hence it is its real dimension which equals the (formal) dimension of monopole moduli space $\mathcal{M}_k$. In other words,

$$\boxed{\dim \mathcal{M}_k = 2 \operatorname{ind} \mathcal{D}} \ .$$

3.2.2 Computing the index of $\mathcal{D}$

Our purpose is then to compute $\operatorname{ind}\mathcal{D}$. To this effect consider the following expression:

$$I(M^2) = \operatorname{Tr}\left(\frac{M^2}{\mathcal{D}^\dagger\mathcal{D} + M^2}\right) - \operatorname{Tr}\left(\frac{M^2}{\mathcal{D}\mathcal{D}^\dagger + M^2}\right) , \tag{3.18}$$

where Tr is the operatorial trace.

Exercise 3.6 (A formula for the index of $\mathcal{D}$)
Prove that the index of $\mathcal{D}$ is given by

$$\begin{aligned}\operatorname{ind}\mathcal{D} &= \dim\ker\mathcal{D}^\dagger\mathcal{D} - \dim\ker\mathcal{D}\mathcal{D}^\dagger\\ &= \lim_{M^2\to 0} I(M^2) .\end{aligned}$$

(Hint: Prove this assuming that there is a gap in the spectrum of these operators. This is not the case, but as argued in [Wei79] the conclusion is unaltered.)

In order to manipulate equation (3.18) it is again convenient to use the reformulation of the BPS-monopole as an instanton, in terms of $\mathsf{W}_{\underline{i}} = (\mathsf{W}_i, \phi)$, and to define the following four-dimensional euclidean Dirac matrices:

$$\bar\gamma_k = \begin{pmatrix} 0 & -i\sigma_k \\ i\sigma_k & 0\end{pmatrix} \quad \bar\gamma_4 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0\end{pmatrix} \quad \bar\gamma_5 = \bar\gamma_1\bar\gamma_2\bar\gamma_3\bar\gamma_4 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1}\end{pmatrix} \tag{3.19}$$

obeying

$$\{\bar\gamma_{\underline{i}}, \bar\gamma_{\underline{j}}\} = 2\delta_{\underline{ij}} .$$

Letting $D_{\underline{i}}$ denote the gauge covariant derivative corresponding to $\mathsf{W}_{\underline{i}}$, and remembering that $\partial_4 \equiv 0$, we find that

$$\bar\gamma \cdot D \equiv \bar\gamma_{\underline{i}} D_{\underline{i}} = \begin{pmatrix} 0 & -\mathcal{D} \\ \mathcal{D}^\dagger & 0\end{pmatrix} ,$$

whence

$$-(\bar\gamma\cdot D)^2 = \begin{pmatrix} \mathcal{D}\mathcal{D}^\dagger & 0 \\ 0 & \mathcal{D}^\dagger\mathcal{D}\end{pmatrix} .$$

Exercise 3.7 (Another formula for $I(M^2)$)
Prove that

$$I(M^2) = -\operatorname{Tr}\bar\gamma_5 \frac{M^2}{-(\bar\gamma\cdot D)^2 + M^2} ,$$

where Tr *now also includes the spinor trace. More generally, if f is any function for which the traces* $\operatorname{Tr} f(\mathcal{D}^\dagger\mathcal{D})$ *and* $\operatorname{Tr} f(\mathcal{D}\mathcal{D}^\dagger)$ *exist, prove that*

$$\operatorname{Tr}\bar\gamma_5 f(-(\bar\gamma\cdot D)^2) = \operatorname{Tr} f(\mathcal{D}\mathcal{D}^\dagger) - \operatorname{Tr} f(\mathcal{D}^\dagger\mathcal{D}) .$$

Let K be any operator acting on square-integrable (matrix-valued) functions $\psi(x)$. K is defined uniquely by its *kernel* $K(x,y)$:

$$(K\,\psi)(x) = \int d^3y\, K(x,y)\psi(y)\ .$$

If we rewrite this equation using Dirac's "ket" notation, so that $\psi(x) = \langle x|\psi\rangle$, then we see that the above equation becomes

$$\langle x|\, K\,|\psi\rangle = \int d^3y\, \langle x|\, K\,|y\rangle\langle y|\psi\rangle\ ,$$

whence we can think of the kernel $K(x,y)$ as $\langle x|\,K\,|y\rangle$. We will often use this abbreviation for the kernel of an operator. In particular, its trace is given by

$$\mathrm{Tr}\, K = \int d^3x\ \mathrm{tr}\,\langle x|\,K\,|x\rangle\ ,$$

where tr stands for the matrix trace, if any.

The rest of this section will concern the calculation of the following expression

$$I(M^2) = -\int d^3x\ \mathrm{tr}\,\bar\gamma_5\,\langle x|\frac{M^2}{-(\bar\gamma\cdot D)^2 + M^2}|x\rangle\ , \tag{3.20}$$

where tr now stands for both the spinor and matrix traces. Let's focus on the kernel

$$I(x,y) = -\,\mathrm{tr}\,\bar\gamma_5\,\langle x|\frac{M^2}{-(\bar\gamma\cdot D)^2 + M^2}|y\rangle\ .$$

Exercise 3.8 (Some properties of kernels)
Let A and B be operators acting on (matrix-valued) square-integrable functions. Let A be a differential operator. Then prove the following identities:

$$\begin{aligned} A(x)\cdot B(x,y) &= (AB)(x,y)\\ B(x,y)\cdot \overleftarrow{A}^\dagger(y) &= (BA)(x,y)\end{aligned}$$

where $\cdot$ means action of differential operators, the label on a differential operator denotes on which variable it acts, and the arrow on $A^\dagger$ in the second equation means that the derivatives act on B.

Using the results of this exercise and the fact that the trace of an even number of $\bar\gamma$-matrices vanishes, we can rewrite $I(x,y)$ slightly. Writing $-(\bar\gamma\cdot D)^2 + M^2 = (\bar\gamma\cdot D + M)(-\bar\gamma\cdot D + M)$, we have that

$$I(x,y) = -\,\mathrm{tr}\,\bar\gamma_5\,\langle x|\frac{M}{\bar\gamma\cdot D + M}|y\rangle = -M\,\mathrm{tr}\,\bar\gamma_5\,\Delta(x,y)\ , \tag{3.21}$$

where we have introduced the propagator

$$\Delta(x,y) = \langle x|\frac{1}{\bar\gamma\cdot D + M}|y\rangle\ .$$

Using once again the results of Exercise 3.8, one immediately deduces the following identities:

$$\begin{aligned}\left(\bar\gamma_i\frac{\partial}{\partial x^i} - e\bar\gamma_{\underline{i}}\mathsf{W}_{\underline{i}}(x) + M\right)\Delta(x,y) &= \delta(x-y)\\ \Delta(x,y)\left(-\bar\gamma_i\frac{\overleftarrow{\partial}}{\partial y^i} - e\bar\gamma_{\underline{i}}\mathsf{W}_{\underline{i}}(y) + M\right) &= \delta(x-y)\ ;\end{aligned}$$

and from them:

$$\begin{aligned}I(x,y) = \tfrac{1}{2}\left(\frac{\partial}{\partial x^i} + \frac{\partial}{\partial y^i}\right)&\operatorname{tr}\bar\gamma_5\bar\gamma_i\,\Delta(x,y)\\ &- \tfrac{e}{2}\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\left(\mathsf{W}_{\underline{i}}(x) - \mathsf{W}_{\underline{i}}(y)\right)\Delta(x,y)\ ,\end{aligned}\tag{3.22}$$

which can be understood as a "conservation law" for the bi-local current

$$J_i(x,y) \equiv \operatorname{tr}\bar\gamma_5\bar\gamma_i\,\Delta(x,y)\ .$$

In order to compute $I(M^2)$ we have to first take the limit $y\to x$ of $I(x,y)$. From equation (3.22) we find that

$$I(x,x) = \tfrac{1}{2}\frac{\partial}{\partial x^i}\,J_i(x,x) - \lim_{y\to x}\tfrac{e}{2}\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\left(\mathsf{W}_{\underline{i}}(x) - \mathsf{W}_{\underline{i}}(y)\right)\Delta(x,y)\ .\tag{3.23}$$

Although the last term has a $\mathsf{W}_{\underline{i}}(x) - \mathsf{W}_{\underline{i}}(y)$ which vanishes as $y\to x$, the propagator is singular in this limit and we have to pay careful attention to the nature of these singularities in order to conclude that this term does not contribute. Clearly we can admit at most a logarithmic singularity. The purpose of the following (long) exercise is to show that nothing more singular than that occurs.

Exercise 3.9 (Regularity properties of the propagator)
Prove that the following limit has at most a logarithmic singularity:

$$\lim_{y\to x}\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\,\Delta(x,y)\ ,$$

where tr *now only denotes the spinor trace.*
(Hint: Notice that $\Delta(x,y)$ is the propagator of a three-dimensional spinor in the

presence of a background gauge field. First let us approximate $\Delta(x,y)$ perturbatively in the coupling constant e:

$$\Delta(x,y) = \sum_{n=0}^{\infty} e^n \Delta_n(x,y) \ .$$

Imposing the equation

$$(\bar\gamma \cdot D + M)(x)\Delta(x,y) = \delta(x-y)$$

order by order in e, we find that

$$\Delta(x,y) = \sum_{n=0}^{\infty} e^n \int \prod_{i=1}^{n} d^3 z_i \Delta_0(x,z_1) \times \left[\prod_{i=1}^{n-1} (\bar\gamma \cdot \mathsf{W})(z_i)\Delta_0(z_i, z_{i+1})\right] (\bar\gamma \cdot \mathsf{W})(z_n)\Delta_0(z_n,y) \ ,$$

where $\Delta_0(x,y)$ is the free propagator:

$$\Delta_0(x,y) = \langle x|\frac{1}{(\bar\gamma\cdot\partial)+M}|y\rangle = \int \frac{d^3p}{(2\pi)^3}\frac{e^{ip\cdot(x-y)}}{i(\bar\gamma\cdot p)+M} \ .$$

Since we are interested in the behaviour of $\Delta(x,y)$ as $|x-y| \to 0$, we need to make some estimates. Prove that $\Delta_n(x-y) \sim |x-y|^{-2+n}$ in this limit, whence we the potentially singular contributions come from $n=0$ and $n=1$. Prove that $\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\,\Delta_0(x,y) = 0$ using the facts that $\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}} = \operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\bar\gamma_{\underline{j}} = 0$. These same identities reduce the computation of $\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\,\Delta_1(x,y)$ to

$$-\operatorname{tr}\bar\gamma_5\bar\gamma_{\underline{i}}\bar\gamma_{\underline{j}}\bar\gamma_k\bar\gamma_\ell \int \frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{p_k\, q_\ell\, e^{ip\cdot x}e^{-iq\cdot y}}{(p^2+M^2)(q^2+M^2)} \times \int d^3z\, e^{-i(p-q)\cdot z}\mathsf{W}_{\underline{j}}(z) \ .$$

Compute this (introducing Feynman parameters,...) and show that it vanishes.)

It is instructive to compare this with the calculation of the axial anomaly in four dimensions. The same calculation in four dimensions would have yielded a singularity $\sim |x-y|^{-1}$ in the n=2 term of the above calculation. A bit of familiar algebra would then have yielded a multiple of the Pontrjagin density for the second term in equation (3.23). In the four dimensional problem, the gauge fields go to zero at infinity and the integral of $I(x,x)$ would have received contributions only from the Pontrjagin term, since the $\partial_i J_i(x,x)$ would give a vanishing boundary term. In our case, though, the situation is different. The second term in (3.23) vanishes, whereas the boundary term coming from $\partial_i J_i(x,x)$ is not zero due to the nontrivial behaviour of the Higgs field at infinity.

From the results of the above exercise, the second term in (3.23) vanishes, and using the expression (3.20) for $I(M^2)$, we find that

$$I(M^2) = \tfrac{1}{2}\int d^3x\,\partial_i J_i(x,x) = \tfrac{1}{2}\int_{\Sigma_\infty} dS_i J_i(x,x) \ ,$$

where Σ_∞ is the sphere at spatial infinity. The (formal) dimension of the moduli space of BPS-k-monopole will then be given by

$$\dim \mathcal{M}_k = \lim_{M^2\to 0} \int_{\Sigma_\infty} dS_i J_i(x,x) \ . \tag{3.24}$$

In the remainder of this section, we will compute this integral and show that it is related to the magnetic number of the monopole.

3.2.3 Computing the current $J_i(x,x)$

We start by rewriting $J_i(x,x)$. Inserting 1 in the form $(-\bar\gamma\cdot D+M)^{-1}(-\bar\gamma\cdot D+M)$ into the definition of $J_i(x,x)$, we get

$$J_i(x,x) = \operatorname{tr}\bar\gamma_5\bar\gamma_i\,\langle x|\frac{1}{-(\bar\gamma\cdot D)^2+M^2}(-\bar\gamma\cdot D+M)|x\rangle \ .$$

Using the fact that the trace of an odd number of $\bar\gamma$-matrices vanishes, we remain with

$$J_i(x,x) = \operatorname{tr}\bar\gamma_5\bar\gamma_i\,\langle x|\frac{1}{-(\bar\gamma\cdot D)^2+M^2}(-\bar\gamma_j D_j + e\bar\gamma_4\boldsymbol{\phi})|x\rangle \ , \tag{3.25}$$

where now

$$-(\bar\gamma\cdot D)^2 = -(D_i)^2 - e^2\boldsymbol{\phi}^2 + \tfrac{1}{2}e\bar\gamma_{ij}\,\mathsf{G}_{ij} + e\bar\gamma_i\bar\gamma_4 D_i\boldsymbol{\phi} \ .$$

We proceed by treating the last terms $\frac{1}{2}e\bar\gamma_{ij}\,\mathsf{G}_{ij}+e\bar\gamma_i\bar\gamma_4 D_i\boldsymbol{\phi}$ as a perturbation and expanding

$$\frac{1}{-(\bar\gamma\cdot D)^2+M^2} = \frac{1}{-(D_i)^2-e^2\boldsymbol{\phi}^2+M^2} - \frac{1}{-(D_i)^2-e^2\boldsymbol{\phi}^2+M^2}\left(\tfrac{1}{2}e\bar\gamma_{ij}\,\mathsf{G}_{ij}+e\bar\gamma_i\bar\gamma_4 D_i\boldsymbol{\phi}\right)\frac{1}{-(D_i)^2-e^2\boldsymbol{\phi}^2+M^2}+\cdots \ .$$

It is now time to use the fact that $\mathsf{W}_{\underline{i}}$ corresponds to a monopole background. For such a background $\mathsf{G}_{\underline{ij}} = O(|x|^{-2})$ asymptotically as $|x|\to\infty$. Because we are integrating $J_i(x,x)$ on Σ_∞, we are free to discard terms which decay faster than $O(|x|^{-2})$ at infinity, hence no further terms other than those shown in the above perturbative expansion contribute. Plugging the remaining two terms of the expansion into (3.25), we notice that the first term vanishes due to the trace identity $\operatorname{tr}\bar\gamma_5\bar\gamma_i\bar\gamma_{\underline{j}}=0$. Similar identities leave only the following terms:

$$J_i(x,x) = e\operatorname{tr}\bar\gamma_5\bar\gamma_i\bar\gamma_j\bar\gamma_4\bar\gamma_k\,\langle x|\,\frac{1}{K}\,D_j\boldsymbol{\phi}\,\frac{1}{K}\,D_k\,|x\rangle - \tfrac{1}{2}e^2\,\operatorname{tr}\bar\gamma_5\bar\gamma_i\bar\gamma_{jk}\bar\gamma_4\langle x|\,\frac{1}{K}\,\mathsf{G}_{jk}\,\frac{1}{K}\,\boldsymbol{\phi}\,|x\rangle + O(|x|^{-3}) \ ,$$

where we have introduced the shorthand $K = -(D_i)^2 - e^2\boldsymbol{\phi}^2 + M^2$. Using the trace identity

$$\operatorname{tr} \bar{\gamma}_5\bar{\gamma}_i\bar{\gamma}_j\bar{\gamma}_k\bar{\gamma}_4 = 4\epsilon_{ijk}$$

we can rewrite the above equation as

$$\begin{aligned} J_i(x,x) = -4e\epsilon_{ijk} \ \operatorname{tr} \ \langle x| \frac{1}{K} D_j\boldsymbol{\phi} \frac{1}{K} D_k \,|x\rangle \\ - 2e^2\epsilon_{ijk} \ \operatorname{tr} \langle x| \frac{1}{K} \mathsf{G}_{jk} \frac{1}{K} \boldsymbol{\phi}\,|x\rangle + O(|x|^{-3}) \ , \end{aligned} \quad (3.26)$$

where the trace now refers only to the $SO(3)$ adjoint representation.

The propagators K^{-1} are not yet those of a free spinor, thanks to their dependence on $\boldsymbol{\phi}$ and W_i, thus we must treat them perturbatively as well. Since W_i decays at infinity, we can effectively put $\mathsf{W}_i = 0$ in the propagators in the above expressions which are already $O(|x|^{-2})$. On the other hand, the perturbative treatment of $\boldsymbol{\phi}$ is a bit more subtle, since it doesn't decay at infinity but rather behaves as a homogeneous function of degree zero; that is, its behaviour on the radius $|x|$ is constant at infinity, but not so its angular dependence, which gives rise to the topological stability of the BPS-monopole. First we notice that in the adjoint representation,

$$\boldsymbol{\phi}^2\boldsymbol{v} = \boldsymbol{\phi} \times (\boldsymbol{\phi} \times \boldsymbol{v}) = -a^2\boldsymbol{v} + (\boldsymbol{\phi}\cdot\boldsymbol{v})\boldsymbol{\phi} \ ,$$

where we have recalled that $\boldsymbol{\phi}\cdot\boldsymbol{\phi} = a^2$ at infinity. Hence on Σ_∞ we can put $K = Q + e^2\Omega$ where $Q = -\partial^2 + M^2 + a^2e^2$ and Ω is (up to a factor) the projector onto the $\boldsymbol{\phi}$ direction: $\Omega(\boldsymbol{v}) = (\boldsymbol{\phi}\cdot\boldsymbol{v})\boldsymbol{\phi}$. Because $[Q,\Omega] = O(|x|^{-1})$ asymptotically, we can effectively treat these two operators as commuting, whence we can write

$$\frac{1}{K} = \frac{1}{Q} + \sum_{n\geq 1} \frac{1}{Q^{n+1}}(e^2\Omega)^n + O(|x|^{-1}) \ .$$

Notice moreover that as operators in the adjoint representation of $SO(3)$,

$$(\Omega\circ\boldsymbol{\phi})\boldsymbol{v} = \Omega(\boldsymbol{\phi}\times\boldsymbol{v}) = \boldsymbol{\phi}\cdot(\boldsymbol{\phi}\times\boldsymbol{v})\boldsymbol{\phi} = 0 \ . \quad (3.27)$$

We are now in a position to prove that the first term in equation (3.26) doesn't contribute.

Exercise 3.10 (The first term doesn't contribute)
Prove that

$$\operatorname{tr} \ \langle x| \frac{1}{K} D_j\boldsymbol{\phi} \frac{1}{K} D_k \,|x\rangle = O(|x|^{-3})$$

whence it doesn't contribute to the integral over Σ_∞.
(Hint: First notice that $D_k = \partial_k + O(|x|^{-1})$, whence up to $O(|x|^{-3})$ we can simply substitute ∂_k for D_k in the above expression. Now use equation (3.27) and the fact that $\boldsymbol{\phi}$ and $D_j\phi$ are parallel in Σ_∞, to argue that one can substitute K for Q in the above expression (again up to terms of order $O(|x|^{-3})$). Then simply take the trace to obtain the result.)

Hence we are left with

$$J_i(x,x) = -2e^2\epsilon_{ijk}\ \mathrm{tr}\, \langle x|\, \frac{1}{K}\, \mathsf{G}_{jk}\, \frac{1}{K}\, \boldsymbol{\phi}\, |x\rangle + O(|x|^{-3})\ . \tag{3.28}$$

Since G_{jk} is parallel to $\boldsymbol{\phi}$ on Σ_∞, $\Omega \circ \mathsf{G}_{jk} = 0$ by (3.27). Thus we are free to substitute the free propagator Q^{-1} for K^{-1} in the above expression, to obtain:

$$J_i(x,x) = -2e^2\epsilon_{ijk} \int d^3x'\ \mathrm{tr}\, \boldsymbol{\phi}(x)\mathsf{G}_{jk}(x')\, Q^{-1}(x,x')Q^{-1}(x',x) + O(|x|^{-3})\ ,$$

where the free propagator is given by

$$Q^{-1}(x,y) = \int \frac{d^3p}{(2\pi)^3} \frac{e^{ip\cdot(x-y)}}{p^2+M^2+a^2e^2}\ .$$

Changing variables $x' \mapsto y = x - x'$, and using

$$\mathrm{tr}\, \boldsymbol{\phi}(x)\mathsf{G}_{jk}(x-y) = \mathrm{tr}\, \boldsymbol{\phi}(x)\mathsf{G}_{jk}(x) + O(|x|^{-3})\ ,$$

we remain with

$$\begin{aligned} J_i(x,x) = -2e^2\epsilon_{ijk}\, \mathrm{tr}\, \mathsf{G}_{jk}(x)\boldsymbol{\phi}(x) &\int d^3y \frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3} e^{i(p-q)\cdot y} \\ &\times \frac{1}{(p^2+M^2+a^2e^2)(q^2+M^2+a^2e^2)} + O(|x|^{-3})\ . \end{aligned}$$

The y-integral gives $2\pi^3\delta(p-q)$, which gets rid of the q-integral and we remain with

$$J_i(x,x) = -2e^2\epsilon_{ijk}\, \mathrm{tr}\, \boldsymbol{\phi}(x)\mathsf{G}_{jk}(x) \int \frac{d^3p}{(2\pi)^3} \frac{1}{(p^2+M^2+a^2e^2)^2} + O(1/|x|^3)\ .$$

The p-integral is readily evaluated

$$\int \frac{d^3p}{(2\pi)^3} \frac{1}{(p^2+M^2+a^2e^2)^2} = \frac{1}{8\pi} \frac{1}{\sqrt{M^2+a^2e^2}}\ ,$$

whence

$$\int_{\Sigma_\infty} dS_i\, J_i(x,x) = \frac{1}{2\pi}\frac{e^2}{\sqrt{M^2+a^2e^2}}\int_{\Sigma_\infty} dS_i\, \epsilon_{ijk}\boldsymbol{\phi}\cdot \mathsf{G}_{jk}\ ,$$

where have also used that the trace in the adjoint representation is normalised so that $\operatorname{tr} AB = -2A\cdot B$.

Exercise 3.11 (Another expression for the degree of the map ϕ)
Prove that on Σ_∞,

$$\boldsymbol{\phi}\cdot\mathsf{G}_{jk} = \frac{1}{ea^2}\boldsymbol{\phi}\cdot(\partial_j\boldsymbol{\phi}\times\partial_k\boldsymbol{\phi})\ ,$$

and, comparing with equation (1.22)*, deduce that the degree of the map $\boldsymbol{\phi}$ from Σ_∞ to the sphere of radius a in $\mathbb{R}^3$ is given by*

$$\deg\boldsymbol{\phi} = \frac{e}{8\pi a}\int_{\Sigma_\infty} dS_i\epsilon_{ijk}\boldsymbol{\phi}\cdot\mathsf{G}_{jk}\ .$$

(Hint: Use that $\boldsymbol{\phi}\times D_j\boldsymbol{\phi} = 0$ on Σ_∞, and expand $0 = \boldsymbol{\phi}\cdot(D_j\boldsymbol{\phi}\times D_k\boldsymbol{\phi})$.)

From the results of this exercise and the fact that for a k-monopole solution, the degree of $\boldsymbol{\phi}$ is k, we can write

$$\int_{\Sigma_\infty} dS_i\, J_i(x,x) = \frac{4aek}{\sqrt{M^2+a^2e^2}}\ ,$$

whence plugging this into the equation (3.24) for the formal dimension of $\mathcal{M}_k$, we find that

$$\boxed{\dim\mathcal{M}_k = 4k}\ .$$

If you are familiar with the calculation of the number of instanton parameters, you may be surprised by the explicit M^2 dependence in the expression $\int_{\Sigma_\infty} dS_i\, J_i(x,x)$. This is due to the asymptotic behaviour of $\boldsymbol{\phi}$, which prevents the above calculation from being tackled by the methods usually applied to index theorems on compact spaces. For the index of an operator on a compact space, or similarly for fields which decay at infinity, one can prove that the result of the above integral is actually independent of M^2, hence one can compute the integral for already in the limit $M^2 \to 0$ (*cf.* the Witten index). Here we are in fact faced essentially with the calculation of the index of an operator on a manifold with boundary, for which a satisfactory Witten-index treatment is lacking, to the best of my knowledge.

3.3 A quick motivation of hyperkähler geometry

In the next section we will prove that the natural metric on $\mathcal{M}_k$ induced by the Yang–Mills–Higgs functional is hyperkähler; but first we will briefly review the necessary notions from riemannian geometry leading to hyperkähler manifolds. The reader familiar with this topic can easily skip this section.

Hyperkähler geometry is probably best understood from the point of view of holonomy groups in riemannian geometry. In this section we review the basic notions. Sadly, the classic treatises on the holonomy approach to riemannian geometry [KN63, KN69, Lic76] stop just short of hyperkähler geometry; but two more recent books [Bes86, Sal89] on the subject do treat the hyperkähler case, albeit from a slightly different point of view than the one adopted here. We direct the mathematically inclined reader to the classics for the basic results on riemannian and Kähler geometry, which we will only have time to review ever so briefly in these notes; and to the newer references for a more thorough discussion of hyperkähler manifolds. All our manifolds will be assumed differentiable, as will be any geometric object defined on them, unless otherwise stated.

3.3.1 Riemannian geometry

Any manifold M admits a riemannian metric. Fix one such metric g. On the riemannian manifold (M, g) there exists a unique linear connection ∇ which is torsion-free

$$[X, Y] = \nabla_X Y - \nabla_Y X \qquad \text{for any vector fields } X, Y \text{ on } M \tag{3.29}$$

and preserves the metric $\nabla g = 0$. It is called the Levi-Civita connection and relative to a local chart x^a, it is defined by the Christoffel symbols $\Gamma_{ab}{}^c$ which in turn are defined by

$$\nabla_a \partial_b = \Gamma_{ab}{}^c \partial_c \ ,$$

where we have used the shorthand $\nabla_a = \nabla_{\partial_a}$. The defining properties of the Levi-Civita connection are sufficient to express the Christoffel symbols in terms of the components g_{ab} of the metric:

$$\Gamma_{ab}{}^c = \tfrac{1}{2} g^{cd} \left(\partial_a g_{db} + \partial_b g_{ad} - \partial_d g_{ab} \right) \ , \tag{3.30}$$

which proves the uniqueness of the Levi-Civita connection.

Exercise 3.12 (A coordinate-free expression for ∇)
Using the defining conditions of the Levi-Civita connection ∇, prove that

$$\begin{aligned} 2\langle Z, \nabla_X Y \rangle = X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle \\ - \langle [X,Z], Y\rangle - \langle [Y,Z], X\rangle - \langle [X,Y], Z\rangle \ , \end{aligned} \tag{3.31}$$

where we have used the notation $\langle X, Y\rangle = g(X, Y)$. Equation (3.30) *follows after substituting ∂_a, ∂_b and ∂_c for X, Y, and Z respectively.*

With ∇ we can give meaning to the notion of parallel transport. Given a curve $t \rightsquigarrow \gamma(t)$ on M with velocity vector $\dot\gamma$, we say that a vector field X is *parallel along* γ if $\nabla_{\dot\gamma} X = 0$. Relative to a local coordinate chart x^a, we can write this equation as

$$\frac{DX^b}{dt} \equiv \dot\gamma^a \nabla_a X^b = \dot\gamma^a \left(\partial_a X^b + \Gamma_{ac}{}^b X^c\right) = \dot X^b + \Gamma_{ac}{}^b \dot\gamma^a X^c = 0 \ . \tag{3.32}$$

A curve γ is a *geodesic* if its velocity vector is self-parallel: $\nabla_{\dot\gamma}\dot\gamma = 0$. In terms of (3.32) we arrive at the celebrated geodesic equation:

$$\ddot\gamma^c + \Gamma_{ab}{}^c \dot\gamma^a \dot\gamma^b = 0 \ . \tag{3.33}$$

This equation follows by extremising the action with lagrangian

$$L(x,\dot x) = \tfrac{1}{2} g_{ab}(x)\, \dot x^a \dot x^b \ ,$$

whence our claim at the end of section 3.1.3 that free motion on a riemannian manifold is geodesic.

We can integrate equation (3.32) and arrive at the concept of parallel-transport. More concretely, associated with any curve $\gamma : [0,1] \to M$ there is a linear map $\mathbb{P}_\gamma : T_{\gamma(0)}M \to T_{\gamma(1)}M$ taking vectors tangent to M at $\gamma(0)$ to vectors tangent to M at $\gamma(1)$. If $X \in T_pM$ is a tangent vector to M at $p \equiv \gamma(0)$ we define its *parallel transport* $\mathbb{P}_\gamma(X)$ *relative to* γ by first extending X to a vector field along γ in a way that solves (3.32), and then simply evaluating the vector field at $\gamma(1)$.

Now fix a point $p \in M$ and let γ be a piecewise differentiable loop based at p, that is, a piecewise differentiable curve which starts and ends at p. Then $\mathbb{P}_\gamma$ is a linear map $T_pM \to T_pM$. We can compose these maps: if γ and γ' are two loops based at p, then $\mathbb{P}_{\gamma'} \circ \mathbb{P}_\gamma$ is the linear map corresponding to parallel transport on the loop based at p obtained by first tracing the path γ and then γ'. (Notice that this new loop may not be differentiable even if γ and γ' are: but it is certainly piecewise differentiable, hence the need to consider such loops from the outset.) Also $\mathbb{P}_\gamma$ is invertible: simply trace the path γ backwards in time. Therefore the transformations $\{\mathbb{P}_\gamma\}$ form a group. If we restrict ourselves to loops which are contractible, the group of linear transformations:

$$H(p) = \{\mathbb{P}_\gamma \mid \gamma \text{ a contractible loop based at } p\}$$

is called the *(restricted) holonomy group at* p of the connection ∇. It can be shown that to be a Lie group.

One should hasten to add that there is no reason to restrict ourselves to the Levi-Civita connection. We will be mostly interested in the classical case, where ∇ is the Levi-Civita connection, but these definitions make sense in more generality.

Exercise 3.13 (The holonomy group of a connected manifold)
Prove that if two points p and q in M can be joined by a path in M, their holonomy groups $H(p)$ and $H(q)$ are conjugate and therefore isomorphic.
(Hint: Use parallel transport along the path joining p and q to provide the conjugation.)

Hence it makes sense to speak of the holonomy group of a connected manifold M. From now on all we will only concern ourselves with connected manifolds. A further useful restriction that one can impose on the type of manifolds we consider is that of irreducibility. A manifold is said to be *(ir)reducible* relative to a linear connection ∇ if the tangent space at any point is an (ir)reducible representation of the holonomy group. Clearly the holonomy group (relative to the Levi-Civita connection) of product manifold $M \times M'$ with the product metric acts reducibly. A famous theorem of de Rham's provides a converse. This theorem states that if a simply-connected complete riemannian manifold M is reducible relative to the Levi-Civita connection, then $M = M' \times M''$ isometrically. We will restrict ourselves in what follows to irreducible manifolds.

For a generic linear connection on an irreducible manifold M, the holonomy group is (isomorphic to) $GL(m)$, where $m = \dim M$. However, the Levi-Civita connection is far from generic as the following exercise shows.

Exercise 3.14 (The holonomy group of a riemannian manifold)
Prove that the holonomy group of an m-dimensional riemannian manifold (relative to the Levi-Civita connection) is actually in $SO(m)$.
(Hint: Show that $\nabla g = 0$ implies that the parallel transport operation $\mathbb{P}_\gamma$ preserves the norm of the vectors, whence the holonomy group is in $O(m)$. Argue that since we consider only contractible loops, the holonomy group is connected and hence it must be in $SO(m)$. By the way, the same would hold for orientable manifolds even if considering non-contractible loops.)

A celebrated theorem of Ambrose and Singer tells us that the Lie algebra of the holonomy group is generated by the Riemann curvature tensor in the following way. Recall that the Riemann curvature tensor is defined as follows. Fix vector fields X and Y on M, and define a linear map from vector fields to vector fields as follows:

$$R(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]} \ .$$

It is easy to prove that this map is actually tensorial in X and Y. Indeed, relative to a coordinate basis, it may be written out as a tensor ${R_{abc}}^d$ defined by

$$R(\partial_a, \partial_b)\, \partial_c = R_{abc}{}^d\, \partial_d\ ,$$

and therefore has components

$$R_{abc}{}^d = \partial_a \Gamma_{bc}{}^d + \Gamma_{bc}{}^e \Gamma_{ae}{}^d - \partial_b \Gamma_{ac}{}^d - \Gamma_{ac}{}^e \Gamma_{be}{}^d\ . \tag{3.34}$$

Then the Lie algebra of the holonomy group is the Lie subalgebra of $gl(m)$ spanned by the *curvature operators* $R_{ab} : \partial_c \mapsto R_{abc}{}^d \partial_d$.

Exercise 3.15 (The holonomy algebra of a riemannian manifold)
Using the Ambrose–Singer theorem this time, prove a second time that the holonomy group of a riemannian manifold lies in $SO(m)$, by showing that its Lie algebra lies in $so(m)$. In other words, prove that each curvature operator R_{ab} (for fixed a and b) is antisymmetric:

$$R_{abcd} = -R_{abdc} \qquad \textit{where} \quad R_{abcd} = R_{abc}{}^e g_{ed}\ .$$

Adding more structure to a riemannian manifold in a way that is consistent with the metric restricts the holonomy group further. Next we will discuss what happens when we add a complex structure.

3.3.2 Kähler geometry

An almost complex structure is a linear map $I : TM \to TM$ which obeys $I^2 = -\mathbb{1}$. This gives each tangent space T_pM the structure of a complex vector space, since we can multiply a tangent vector X by a complex number $z = x + i\,y$ simply by $z \cdot X = x\,X + y\,I(X)$. In particular, it means that the (real) dimension of each T_pM and hence of M must be even: $2n$, say. We will also assume that the complex structure I is compatible with the metric in the sense that $g(IX, IY) = g(X, Y)$ for all vector fields X and Y. Another way to say this is that the metric g is hermitian relative to the complex structure I.

If we complexify the tangent space, we can diagonalise the complex structure. Clearly the eigenvalues of I are $\pm i$. Complex vector fields Z for which $I\,Z = iZ$ are said to be of type $(1,0)$, whereas those for which $I\,Z = -iZ$ are of type $(0,1)$. If we can introduce local complex coordinates $(z^\alpha, \bar{z}^{\bar{\alpha}})$, $\alpha, \bar{\alpha} = 1, \ldots n$, relative to which a basis for the $(1,0)$ (resp. $(0,1)$) vector fields is given by ∂_α (resp. $\partial_{\bar{\alpha}}$) *and* if when we change charts the local complex coordinates are related by biholomorphic transformations, then we say that I is integrable.

A hard theorem due to Newlander and Nirenberg translates this into a beautiful local condition on the complex structure. According to the

Newlander–Nirenberg theorem, an almost complex structure I is integrable if and only if the Lie bracket of any two vector fields of type $(1,0)$ is again of type $(1,0)$. This in turns translates into the vanishing of a tensor.

Exercise 3.16 (The Nijenhuis tensor)
Using the Newlander–Nirenberg theorem prove that I is integrable if and only if the following tensor vanishes:

$$N_I(X,Y) = I[IX,IY] + [X,IY] + [IX,Y] - I[X,Y] \ .$$

N_I is known as the Nijenhuis tensor *of the complex structure I. It is easy to prove that the Nijenhuis tensor N_I vanishes in a complex manifold (do it!)—it is the converse that is hard to prove.*

Now suppose that ∇ is a linear connection relative to which I is parallel: $\nabla I = 0$. Let's call this a *complex connection*.

Exercise 3.17 (The holonomy group of a complex connection)
Let H denote the holonomy group of a complex connection on a complex manifold M. Prove that $H \subseteq GL(n,\mathbb{C}) \subset GL(2n,\mathbb{R})$.
(Hint: It's probably easiest to prove the equivalent statement that the holonomy algebra is a subalgebra of $gl(n,\mathbb{C})$. Choose a basis for $T_pM \cong \mathbb{R}^{2n}$ in which the complex structure I has the form

$$I = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \ ,$$

where $\mathbb{1}$ is the $n \times n$ unit matrix. Argue that the curvature operators R_{ab} commute with I, whence in this basis, they are of the form

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \ ,$$

where A and B are arbitrary real $n \times n$ matrices. This then corresponds to the real $2n$-dimensional representation of the matrix $A + i\,B \in gl(n,\mathbb{C})$.)

If ∇ is the Levi-Civita connection, then the holonomy lies in the intersection $GL(n,\mathbb{C}) \cap SO(2n) \subset GL(2n,\mathbb{R})$.

Exercise 3.18 (The unitary group)
Prove that $GL(n,\mathbb{C}) \cap SO(2n) \subset GL(2n,\mathbb{R})$ is precisely the image of the real $2n$-dimensional representation of the unitary group $U(n)$.
(Hint: Prove the equivalent statement for Lie algebras. In the basis of the previous exercise, prove that a matrix in $so(2n)$ has the form

$$\begin{pmatrix} A & B \\ -B^t & D \end{pmatrix} \ ,$$

where $A^t = -A$, $D^t = -D$ and B are otherwise arbitrary real $n \times n$ matrices. If the matrix is also in $gl(n,\mathbb{C})$, we know that $A = D$ and that $B = B^t$. Thus matrices in $gl(n,\mathbb{C}) \cap so(2n) \subset gl(2n,\mathbb{R})$ are of the form

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} ,$$

where $A^t = -A$ and $B^t = B$, which corresponds to the complex matrix $A + i\,B \in gl(n,\mathbb{C})$. Prove that this matrix is anti-hermitian, whence in $u(n)$.)

If the Levi-Civita connection is complex, so that the holonomy lies in $U(n)$, the manifold (M, g, I) is said to be *Kähler*. In other words, Kähler geometry is the intersection, so to speak, of riemannian and complex geometries.

There is another perhaps more familiar definition of Kähler manifolds, involving the Kähler form.

Exercise 3.19 (The Kähler form)
Given a complex structure I relative to which g is hermitian, we define a 2-form ω by

$$\omega(X, Y) = g(X, IY) \qquad \text{or equivalently} \qquad \omega_{ab} = I_a{}^c g_{bc} \;.$$

Prove that $\omega(X, Y) = -\omega(Y, X)$, so that it is in fact a form. Prove that it is actually of type $(1, 1)$.

An equivalent definition of a Kähler manifold is that (M, g, I) is Kähler if and only if ω is closed. These two definitions can of course be reconciled. We review this now.

Exercise 3.20 (Kähler is Kähler is Kähler)
Let (M, g, I) be a complex riemannian manifold, where g is hermitian relative to I. Let ω be the corresponding Kähler form and ∇ the Levi-Civita connection. Prove that the following three conditions are equivalent (and are themselves equivalent to (M, g, I) being a Kähler manifold):

(a) $\nabla I = 0$;

(b) $\nabla\omega = 0$; *and*

(c) $d\omega = 0$.

(Hint: $(a) \Leftrightarrow (b)$ is obvious. For $(b) \Rightarrow (c)$ simply antisymmetrise and use the fact that ∇ is torsion-less, which implies the symmetry of the Christoffel symbols in the lower two indices. The trickiest calculation is $(c) \Rightarrow (b)$, and we break this up into several steps:

1. *From (3.29), deduce that I is integrable if and only if*

$$\nabla_a \omega_{bc} + I_c{}^d I_b{}^e \nabla_e \omega_{ad} - (a \leftrightarrow b) = 0 \;.$$

2. *From the fact that g is hermitian relative to I, deduce that*

$$I_c{}^d \nabla_a \omega_{bd} = -I_b{}^d \nabla_a \omega_{cd} \ .$$

3. *Using the previous two steps, show that*

$$\nabla_a \omega_{bc} = -I_b{}^d I_c{}^e \nabla_a \omega_{de} \ .$$

4. *Finally, use these formulae to show that* $d\omega = 0 \Rightarrow \nabla\omega = 0$.*)*

Conversely, one can show that if the holonomy group of a $2n$-dimensional riemannian manifold is contained in $U(n)$, then the manifold is Kähler. The proof is paradigmatic of the more algebraic approach to the study of holonomy, which has begotten some of the more remarkable results in this field. We will therefore allow ourselves a brief digression. We urge the reader to take a look at the books [Bes86, Sal89] for a more thorough treatment.

For simplicity, we start with a torsionless connection ∇. The fundamental elementary observation is that there is a one-to-one correspondence between covariantly constant tensors and singlets of the holonomy group. (Clearly if $\nabla t = 0$, then t is invariant under the holonomy group; conversely, if t is invariant under the holonomy group, taking the derivative of the parallel transport of t along the path is zero, but to first order this is precisely ∇t.) In turn, singlets of the holonomy group determine to a large extent the geometry of M. For example, suppose that M is an m-dimensional irreducible manifold with holonomy group $G \subset GL(m, \mathbb{R})$. Irreducibility means that the fundamental m-dimensional representation of $GL(m, \mathbb{R})$ remains irreducible under G. Let us call this representation T—the "T" stands for tangent space. Under the action of G, tensors on M will transform according to tensor powers of the representation T. For example, 1-forms will transform according to the dual representation T^*, symmetric rank p tensors will transform as $S^p T^*$, whereas p-forms will transform as $\bigwedge^p T^*$, and so on. We can then break up all these tensorial representations in terms of irreducibles and, in particular, exhibit all the singlets. These singlets will correspond, by the observation made above, in a one-to-one fashion with covariantly constant tensors. Let us run through some examples.

Suppose that $G = SO(m)$. Then T is the fundamental m-dimensional representation of $SO(m)$. We know that, in particular, there is a singlet $\bar{g} \in S^2 T^*$ and moreover that the map $T \to T^*$ defined by this $\bar{g}$ is non-degenerate. Hence by the fundamental observation, there exists a covariantly constant tensor g which can be thought of as a riemannian metric. By the uniqueness of the Levi-Civita connection, it follows that ∇ is the Levi-Civita connection associated to g. In other words, manifolds with $SO(m)$ holonomy relative to a torsionless connection are simply riemannian manifolds. Well, not just any riemannian manifold. There is another $SO(m)$-invariant tensor $\bar{\Omega} \in \bigwedge^m T^*$. The covariantly constant m-form Ω defines an orientation on M. In fact, one can show that there are no other invariant tensors which are algebraic independent from these ones, so that manifolds with $SO(m)$ holonomy (again, relative to a torsionless connection) are precisely orientable riemannian manifolds.

Now suppose that the dimension of M is even: $m = 2n$, say, and that $G = U(n) \subset GL(2n, \mathbb{R})$. Then T is the real $2n$ irreducible representation of $U(n)$, whose complexification breaks up as $T^{\mathbb{C}} = T' \oplus T''$, where T' is the complex n-dimensional (fundamental) representation of $U(n)$, and $T'' = \overline{T'}$ is its conjugate. Since $G \subset SO(2n)$, we know from the previous paragraph that M is riemannian and orientable, and that we can think of ∇ as the Levi-Civita connection of this metric. However there is also a singlet $\bar{\omega} \in \bigwedge^2 T^*$. The resulting covariantly constant 2-form ω is precisely the Kähler form. Hence manifolds with $U(n)$-holonomy are precisely the Kähler manifolds.

3.3.3 Ricci flatness

We can now restrict the holonomy of a Kähler manifold a little bit further by imposing constraints on the curvature: namely that it be Ricci-flat. As we will see, this is equivalent to demanding that the holonomy lie in $SU(n) \subset U(n)$. As Lie groups, $U(n) = U(1) \times SU(n)$. If we think of $U(n)$ as unitary matrices, the $U(1)$ factor is simply the determinant. Hence the manifold will have $SU(n)$ holonomy provided that the determinant of every parallel transport operator $\mathbb{P}_\gamma$ is equal to 1.

Geometrically, the determinant can be understood as follows. Suppose that M is a Kähler manifold and let's look at how forms of type $(n,0)$ (or $(0,n)$) transform under parallel transport. At a fixed point p in M, the space of such forms is 1-dimensional. Hence if θ is an $(n,0)$-form, then $\mathbb{P}_\gamma\theta = \lambda_\gamma\theta$ where λ_γ is a complex number of unit norm.

Exercise 3.21 (The determinant of $\mathbb{P}_\gamma$)
Prove that λ_γ is the determinant of the linear map $\mathbb{P}_\gamma : T_pM \to T_pM$.

Therefore the holonomy lies in $SU(n)$ if and only if $\lambda_\gamma = 1$ for all γ. By our previous discussion, it means that there is a nonzero parallel $(n,0)$ form θ. Since parallel forms have constant norm, this form if nonzero at some point is nowhere vanishing, hence the bundle of $(n,0)$-forms is trivial. Equivalently this means that the first Chern class of the manifold vanishes. Such Kähler manifolds are known as Calabi–Yau manifolds, after the celebrated conjecture of Calabi, proven by Yau. Calabi's conjecture stated that given a fixed Kähler manifold with vanishing first Chern class, there exists a unique Ricci-flat Kähler metric in the same Kähler class. The Calabi conjecture (now theorem) allows us to construct manifolds admitting Ricci-flat Kähler metrics, by the simpler procedure of constructing Kähler manifolds with vanishing first Chern class. Algebraic geometry provides us with many constructions of such manifolds: as algebraic varieties of complex projective space, for example. The catch is that the Ricci-flat Kähler metric is most definitely *not* the induced metric. In fact, the form of the metric is very difficult to determine. Even for relatively simple examples like $K3$, the metric is not known.

It follows from this conjecture (now theorem) that an irreducible Kähler manifold has $SU(n)$ holonomy if and only if it is Ricci-flat. We don't need to appeal to the Calabi conjecture to prove this result, though, as we now begin to show.

Let's first recall how the Ricci tensor is defined. If X and Y are vector fields on M, $\mathrm{Ric}(X,Y)$ is defined as the trace of the map $V \mapsto R(V,X)\,Y$, or relative to a local chart

$$S_{ab} \equiv \mathrm{Ric}(\partial_a, \partial_b) = R_{cab}{}^{c} \ .$$

Exercise 3.22 (The Ricci tensor is symmetric)
Prove that $S_{ab} = S_{ba}$.

In a Kähler manifold and relative to complex coordinates adapted to I, many of the components of the Ricci and Riemann curvature tensors are zero.

Exercise 3.23 (Curvature tensors in Kähler manifolds)
Let $(z^\alpha, \bar z^{\bar\alpha})$ be complex coordinates adapted to the complex structure I; that is, the corresponding vector fields are of type $(1,0)$ and $(0,1)$ respectively: $I(\partial_\alpha) = i\partial_\alpha$ and $I(\partial_{\bar\alpha}) = -i\partial_{\bar\alpha}$. Prove that the metric has components $g_{\alpha\bar\beta}$, and that the Christoffel symbols have components $\Gamma_{\alpha\beta}{}^{\gamma}$ and $\Gamma_{\bar\alpha\bar\beta}{}^{\bar\gamma}$. Prove that the only nonzero components of the Riemann curvature are $R_{\alpha\bar\beta\gamma}{}^{\delta} = -R_{\bar\beta\alpha\gamma}{}^{\delta}$ and $R_{\alpha\bar\beta\bar\gamma}{}^{\bar\delta} = -R_{\bar\beta\alpha\bar\gamma}{}^{\bar\delta}$. Finally deduce that the Ricci tensor has components $S_{\alpha\bar\beta}$, so that $S_{\alpha\beta} = S_{\bar\alpha\bar\beta} = 0$.

The holonomy algebra of a Kähler manifold is $u(n)$, whence for fixed a and b, the curvature operator R_{ab} belongs to $u(n) = u(1) \times su(n)$. How can one extract the $u(1)$-component? For this we need to recall Exercise 3.18. If $A + i\,B \in u(n)$, $\exp(A + iB) \in U(n)$ and the $U(1)$ component is the determinant: $\det\exp(A + i\,B) = \exp(\mathrm{tr}(A + i\,B))$. Since $A^t = -A$, it is traceless, whence $\det\exp(A + i\,B) = \exp(i\,\mathrm{tr}\,B)$. Hence if we let $i\mathbb{1} \in u(n)$ be the generator of the $u(1)$ subalgebra, the $u(1)$ component of a matrix in $u(n)$ is just its trace. In a Kähler manifold, the holonomy representation of

$U(n)$ is real and $2n$-dimensional, which means that a matrix $A + i\,B \in u(n)$ is represented by a real $2n \times 2n$ matrix

$$Q = \begin{pmatrix} A & B \\ -B & A \end{pmatrix} ,$$

and therefore its $u(1)$-component, $\operatorname{tr} B$, is simply given by

$$\operatorname{tr} B = -\tfrac{1}{2} \operatorname{tr} \left[\begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \right] = -\tfrac{1}{2} \operatorname{tr}(I\,Q) .$$

From this it follows that the $u(1)$-component F_{ab} of the curvature operator is given by $F_{ab} = -\frac{1}{2} \operatorname{tr}(I \circ R_{ab}) = -\frac{1}{2} R_{abc}{}^{d} I_{d}{}^{c}$. The next exercise asks you to show that this is essentially the Ricci curvature, from where it follows that Ricci-flat Kähler manifolds have $SU(n)$ holonomy and viceversa.

Exercise 3.24 (An equivalent expression for the Ricci curvature)
Prove that the Ricci curvature on a Kähler manifold can be also be defined by

$$\operatorname{Ric}(X, Y) = \tfrac{1}{2} \operatorname{tr}(V \mapsto I \circ R(X, IY)\,V) ;$$

or equivalently,

$$S_{ac} I_{b}{}^{c} = -\tfrac{1}{2} \operatorname{tr}(I \circ R_{ab}) = F_{ab} ,$$

which, relative to complex coordinates, becomes

$$S_{\alpha\bar\beta} = i F_{\alpha\bar\beta} .$$

Using the above results, give another proof of the symmetry of the Ricci tensor: $S_{ab} = S_{ba}$. (Compare with Exercise 3.22.)

3.3.4 Hyperkähler geometry

Finally we define hyperkähler manifolds. In a hyperkähler manifold we have not one but three parallel almost complex structures I, J, and K which satisfy the quaternion algebra:

$$IJ = K = -JI, \qquad JK = I = -KJ, \qquad KI = J = -IK$$
$$I^2 = J^2 = K^2 = -\mathbb{1} ,$$

and such that the metric is hermitian relative to all three. Notice that we don't demand that the complex structures be integrable. This is a consequence of the definition.

Exercise 3.25 (Hyperkähler implies integrability)
Let (M, g, I, J, K) be a hyperkähler manifold. Prove that I, J and K are integrable complex structures.
(Hint: Associated to each of the almost complex structures there is a 2-form: ω_I, ω_J and ω_K. Because $\nabla I = 0$, Exercise 3.20 implies that $d\omega_I = 0$, and similarly for J and K. Notice that $\omega_J(X,Y) = g(X, JY) = g(X, KIY) = \omega_K(X, IY)$, whence[2] $\imath_X \omega_J = \imath_{IX}\omega_K$. A complex vector field X is of type $(1,0)$ with respect to I, if and only if $\imath_X\omega_J = i\,\imath_X\omega_K$. By the Newlander–Nirenberg theorem, it is sufficient to prove that the Lie bracket of two such complex vector fields also obeys the same relation. But this is a simple computation, where the fact that the forms ω_J and ω_K are closed is used heavily. The same proof holds mutatis mutandis *for J and K.)*

Just like an almost complex structure I on a manifold allows us to multiply vector fields by complex numbers and hence turn each tangent space into a complex vector space, the three complex structure in a hyperkähler manifold allow us to multiply by quaternions. Concretely, if $q = x + i\,y + j\,z + k\,w \in \mathbb{H}$ is a quaternion, and X is a vector field on M, then we define

$$q \cdot X \equiv x\,X + y\,I(X) + z\,J(X) + w\,K(X)\ .$$

This turns each tangent space into a quaternionic vector space (a left $\mathbb{H}$-module, to be precise) and, in particular, this means that hyperkähler manifolds are $4k$-dimensional.

One can prove, just as we did with complex manifolds, that the holonomy group of a hyperkähler manifold lies in[3] $USp(2k) \subseteq SU(2k) \subset GL(4k)$. In particular, hyperkähler manifolds are Ricci-flat.

Exercise 3.26 (The holonomy group of a hyperkähler manifold)
Prove that the holonomy group of a hyperkähler manifold is a subgroup of $USp(2k)$.
(Hint: Depending on how one looks at this, there may be nothing that needs proving. If we take as definition of $USp(2k) \subset GL(2k,\mathbb{C})$ those matrices which commute with the natural action of the quaternions on $\mathbb{C}^{2k} \cong \mathbb{H}^k$, then the result is immediate since the fact that I, J, and K are parallel means that ∇ and hence the curvature operators commute with multiplication by $\mathbb{H}$. If you have another definition of $USp(2k)$ in mind, then the exercise is to reconcile both definitions.)

Conversely, if the holonomy group of a manifold M is a subgroup of $USp(2k) \subset SO(4k)$, then decomposing tensor powers of the fundamental $4k$-dimensional representation T of $SO(2k)$ into $USp(2k)$-irreducibles, we find that $\bigwedge^2 T^*$ possesses three singlets. The resulting covariantly constant 2-forms are of course the three Kähler forms of M. A little bit closer inspection

[2]The conventions for the interior product $\imath_X$ are summarised in Exercise 3.27.
[3]Mathematicians call $Sp(k)$ what we call $USp(2k)$.

shows that the associated complex structures obey the quaternion algebra, so that M is hyperkähler.

3.4 $\mathcal{M}_k$ is hyperkähler

In this section we prove that the metric on $\mathcal{M}_k$ defined by the kinetic term in the Yang–Mills–Higgs functional is hyperkähler. We will prove this in two ways. First we will prove that the configuration space $\mathcal{A}$ is hyperkähler and that $\mathcal{M}_k$ is its hyperkähler quotient. In order to do this we will review the notion of a Kähler quotient which should be very familiar as a special case of the symplectic quotient of Marsden–Weinstein which appears in physics whenever we want to reduce a phase space with first-class constraints. In a nutshell, the group $\mathcal{G}$ of gauge transformations acts on $\mathcal{A}$ preserving the hyperkähler structure and the resulting moment mapping is nothing but the Bogomol'nyi equation. We will also give a computationally more involved proof of the hyperkähler nature of $\mathcal{M}_k$, which is independent of the hyperkähler quotient, at least on the face of it.

3.4.1 Symplectic quotients

Let (M, ω) be a symplectic manifold—that is, ω is a nondegenerate closed 2-form—and let G be a Lie group acting on M in a way that preserves ω. Let $\mathfrak{g}$ denote the Lie algebra of G and let $\{e_a\}$ be a basis for $\mathfrak{g}$ which we fix once and for all. To each e_a there is an associated vector field X_a on M. The fact that G preserves ω means that $\mathcal{L}_{X_a}\omega = 0$ for every X_a, where $\mathcal{L}_{X_a}$ is the Lie derivative along X_a. We will often abbreviate $\mathcal{L}_{X_a}$ by $\mathcal{L}_a$.

Exercise 3.27 (The Lie derivative acting on forms)
Prove that if ω is a differential form on M and X is any vector field, then the Lie derivative $\mathcal{L}_X\omega$ is given by

$$\mathcal{L}_X\omega = (d\imath_X + \imath_X d)\,\omega$$

where d is the exterior derivative and $\imath_X$ is the contraction operator characterised uniquely by the following properties:

(a) $\imath_X f = 0$ *for all functions* f*;*

(b) $\imath_X \alpha = \alpha(X)$ *for all one-forms* α*; and*

(c) $\imath_X (\alpha \wedge \beta) = (\imath_X \alpha) \wedge \beta + (-1)^p \alpha \wedge \imath_X \beta$, *for α a p-form and β any form.*

Because $d\omega = 0$, $\mathfrak{L}_a\omega = 0$ is equivalent to the one-form $\imath_a\omega$ being closed, where $\imath_a \equiv \imath_{X_a}$. Let us assume that this form is also exact, so that there is a function μ_a such that $\imath_a\omega = d\mu_a$. This would be guaranteed, for example, if M were simply-connected, or if $\mathfrak{g}$ were semi-simple. More precise conditions on the absence of this obstruction can be written down but we won't need them in what follows. The functions μ_a allow us to define a *moment mapping* $\mu : M \to \mathfrak{g}^*$ by $\mu(p) = \mu_a(p)e^a$ for every $p \in M$, where $\{e^a\}$ is the canonically dual basis for $\mathfrak{g}^*$. In other words, $\mu(p)(e_a) = \mu_a(p)$.

Exercise 3.28 (The Poisson bracket)
Prove that the Poisson bracket *defined by:*

$$\{f, g\} = \omega^{ij}\partial_i f \partial_j g \ ,$$

where $\omega^{ij}\omega_{jk} = \delta_k{}^i$, is antisymmetric and obeys the Jacobi identity. Using the above definition of the Poisson bracket (or otherwise) prove the following identity:

$$d\{\mu_a, \mu_b\} = f_{ab}{}^c d\mu_c \ , \tag{3.35}$$

where $f_{ab}{}^c$ are the structure constants of $\mathfrak{g}$ in the chosen basis.
(Hint: Prove first that $X_a\mu_b = \{\mu_a, \mu_b\}$ and take d of this expression. You may wish you use the following properties of the contraction: $[\mathfrak{L}_a, \imath_b] = f_{ab}{}^c\imath_c$.)

Now notice that since μ_a are defined by their gradients, they are defined up to a constant. If these constants can be chosen so that equation (3.35) can be integrated to

$$\{\mu_a, \mu_b\} = f_{ab}{}^c \mu_c \ , \tag{3.36}$$

then the moment mapping μ is *equivariant* and the action is called *Poisson*. In other words, the moment mapping intertwines between the action of G on M and the coadjoint action of G on $\mathfrak{g}^*$. Again one can write down precise conditions under which this is the case—conditions which would be met, for example, if $\mathfrak{g}$ were semisimple. We will assume henceforth that the necessary conditions are met and that the moment mapping is equivariant.

The components μ_a of an equivariant moment mapping can be understood as first-class constraints. It is well-known that if the constraints are irreducible, so that their gradients are linearly independent almost everywhere on the constraint submanifold, one can reduce the original symplectic manifold to a smaller symplectic manifold (or, more generally, orbifold). More precisely, the irreducibility condition on the constraints means that their zero locus $\mu^{-1}(0)$ is an embedded submanifold of M. The fact that the constraints are first class means that the vector fields X_a, when restricted to $\mu^{-1}(0)$, are tangent to $\mu^{-1}(0)$.

Equivalently, one may deduce from the equivariance of the moment mapping that G acts on $\mu^{-1}(0)$. Provided that it does so "nicely" (that is, freely and properly discontinuous) the space $\mu^{-1}(0)/G$ of G-orbits is a manifold, and a celebrated theorem of Marsden and Weinstein tells us that it is symplectic. Indeed, if we let $\pi : \mu^{-1}(0) \to \mu^{-1}(0)/G$ denote the natural projection, then the Marsden–Weinstein theorem says that there is a unique symplectic form $\bar{\omega}$ on $\mu^{-1}(0)/G$ such that its pullback $\pi^*\bar{\omega}$ to $\mu^{-1}(0)$ coincides with the restriction to $\mu^{-1}(0)$ of the symplectic form ω on M. The symplectic manifold $(\mu^{-1}(0)/G, \bar{\omega})$ is known as the *symplectic quotient* of (M, ω) by the action of G. It wouldn't be too difficult to sketch a proof of this theorem, but since we will only need the special case of a Kähler quotient, we will omit it.

3.4.2 Kähler quotients

Now suppose that (M, g, I) is Kähler with Kähler form ω. Then in particular (M, ω) is symplectic. Assume that the action of G on M is not just Poisson, but that G also acts by isometries, that is, preserving g. Because G preserves both g and ω, it also preserves I. On $\mu^{-1}(0)$ we have the induced metric: the restriction to $\mu^{-1}(0)$ of the metric on M. This gives rise to a a metric on $\mu^{-1}(0)/G$ which we will discuss below.

For every $p \in \mu^{-1}(0)$, the tangent space T_pM breaks up as

$$T_pM = T_p\mu^{-1}(0) \oplus N_p\mu^{-1}(0) \ ,$$

where $T_p\mu^{-1}(0)$ is the tangent space to $\mu^{-1}(0)$ and the *normal space* $N_p\mu^{-1}(0)$ is defined as its orthogonal complement $(T_p\mu^{-1}(0))^\perp$. Globally this means that the restriction to $\mu^{-1}(0)$ of the tangent bundle of M decomposes as:

$$TM = T\mu^{-1}(0) \oplus N\mu^{-1}(0) \ , \tag{3.37}$$

where the *normal bundle* $N\mu^{-1}(0)$ is defined as $(T\mu^{-1}(0))^\perp$. As the next exercise shows, the normal bundle is trivial because $\mu^{-1}(0)$ is defined globally by irreducible constraints.

Exercise 3.29 (Triviality of the normal bundle)
Prove that the normal space $N_p\mu^{-1}(0)$ is spanned by the gradients $\operatorname{grad}_p \mu_a$ of the constraints; or globally, that the gradients of the constraints $\{\operatorname{grad} \mu_a\}$ trivialise the normal bundle.
In fact, the converse is also true. If you feel up to it, prove that the normal bundle to a submanifold is trivial if and only if the submanifold can be described globally as the zero locus of some irreducible "constraints."
(Hint: A vector field X is tangent to $\mu^{-1}(0)$ if and only if it preserves the constraints: $d\mu_a(X) = 0$, but this is precisely $g(\operatorname{grad} \mu_a, X) = 0$, by definition of $\operatorname{grad} \mu_a$.)

Both the metric and the symplectic form restrict to $\mu^{-1}(0)$, but whereas g is nondegenerate on $\mu^{-1}(0)$, the symplectic form isn't. Thus in order to obtain a Kähler manifold it is necessary to perform a quotient. We will describe this quotient locally. To this effect, let us split the tangent space $T_p\mu^{-1}(0)$ further as:

$$T_p\mu^{-1}(0) = H_p \oplus V_p \ ,$$

where the *vertical* vectors V_p are those vectors tangent to the G-orbits and the *horizontal* vectors $H_p = V_p^\perp$ are their orthogonal complement. The vertical subspace is spanned by the Killing vectors X_a, whereas the horizontal space H_p can be identified with the tangent space to $\mu^{-1}(0)/G$ at the G-orbit of p. Indeed, given any vector field X on $\mu^{-1}(0)/G$ we define its *horizontal lift* to be the unique horizontal vector field $\tilde{X}$ on $\mu^{-1}(0)$ which projects down to X: $\pi_*\tilde{X} = X$.

Now given any two vector fields on $\mu^{-1}(0)/G$, we define their inner product to be the inner product of their horizontal lifts. This is independent on the point in the orbit to where we lift, because the metric is constant on the orbits. Hence it is a well-defined metric on $\mu^{-1}(0)/G$. In fancier language, this is the unique metric on $\mu^{-1}(0)/G$ which makes the projection π a *riemannian submersion*. (The reader will surely recognise this construction as the one which in section 3.1.2 yielded the metric on the physical configuration space $\mathcal{C}$ of the Yang–Mills–Higgs system.)

We claim that there is also a symplectic form on $\mu^{-1}(0)/G$ which makes this metric Kähler. We prefer to work with the complex structure.

By definition, if Y is *any* vector field tangent to M, its inner product with $\operatorname{grad}\mu_a$ is given by

$$g(\operatorname{grad}\mu_a, Y) = d\mu_a(Y) = \omega(X_a, Y) = g(IX_a, Y) \ ,$$

whence $\operatorname{grad}\mu_a = IX_a$. Hence if we decompose the restriction of TM to $\mu^{-1}(0)$ as

$$TM = T\mu^{-1}(0) \oplus N\mu^{-1}(0) = H \oplus V \oplus N\mu^{-1}(0) \ ,$$

and we choose as bases for V and $N\mu^{-1}(0)$, $\{X_a\}$ and $\{\operatorname{grad}\mu_a\}$ respectively, the complex structure I has the following form:

$$I = \begin{pmatrix} \bar{I} & 0 & 0 \\ 0 & 0 & \mathbb{1} \\ 0 & -\mathbb{1} & 0 \end{pmatrix} \ ,$$

whence H is a complex subspace relative to the restriction $\bar{I}$ of I. In other words, the complex structure commutes with the horizontal projection, or a little bit more precisely, if Y is a vector field on $\mu^{-1}(0)/G$ and $\tilde{Y}$ its horizontal lift, then $I\tilde{Y} = \widetilde{\bar{I}Y}$. The next exercise asks you to prove that this complex structure is integrable, whence $\mu^{-1}(0)/G$ is a complex manifold.

Exercise 3.30 (Integrability of the restricted complex structure)
Use the Newlander–Nirenberg theorem to deduce that $\bar{I}$ is integrable.
(Hint: Relate the Nijenhuis tensor $N_{\bar{I}}$ of $\bar{I}$ to that of I, which vanishes since I is integrable.)

To prove that $\bar{I}$ is parallel, we need to know how the Levi-Civita connection of $\mu^{-1}(0)/G$ is related to the one on M. The next exercise asks you to prove the relevant relation.

Exercise 3.31 (O'Neill's formula)
Let X and Y be vector fields on $\mu^{-1}(0)/G$, and let $\tilde{X}$ and $\tilde{Y}$ be their horizontal lifts. Prove the following formula

$$\nabla_{\tilde{X}}\tilde{Y} = \widetilde{\bar{\nabla}_X Y} + \tfrac{1}{2}[\tilde{X},\tilde{Y}]^v \ , \tag{3.38}$$

where ∇ is the Levi-Civita connection on $\mu^{-1}(0)/G$, and ^{v} denotes the projection onto the vertical subspace. In other words, the horizontal projection of $\nabla_{\tilde{X}}\tilde{Y}$ is precisely the horizontal lift of $\bar{\nabla}_X Y$.
(Hint: Use expressions (3.31) *and* (3.29) *to evaluate the horizontal and vertical components of $\nabla_{\tilde{X}}\tilde{Y}$.)*

In other words, formula (3.38) says that if we identify H with the tangent space to $\mu^{-1}(0)/G$, then the Levi-Civita connection on $\mu^{-1}(0)/G$ is given simply by the horizontal projection of the Levi-Civita connection on M. Or, said differently, that the covariant derivative commutes with the horizontal projection. Since the complex structures also commute with the projection, we see that $\nabla I = 0$ on M implies that $\bar{\nabla}\bar{I} = 0$ on $\mu^{-1}(0)/G$. Therefore using Exercise 3.20, $\mu^{-1}(0)/G$ is Kähler.

Notice that the above decomposition (3.37) can be thought of as

$$TM \cong T(\mu^{-1}(0)/G) \oplus \mathfrak{g}^{\mathbb{C}} \ ,$$

where $\mathfrak{g}^{\mathbb{C}}$ is the complexification of the Lie algebra of G. Therefore, morally speaking, it would seem that $\mu^{-1}(0)/G$ is the quotient of M by the action of $G^{\mathbb{C}}$. In some circumstances this is actually an accurate description of the Kähler quotient; for instance, the construction of complex projective space $\mathbb{CP}^n$ as a Kähler quotient of $\mathbb{C}^{n+1}$.

3.4.3 Hyperkähler quotients

Now let (M, g, I, J, K) be a hyperkähler manifold. We have three Kähler forms: $\omega^{(I)}$, $\omega^{(J)}$, and $\omega^{(K)}$. Suppose that G acts on M via isometries and preserving the three complex structures, hence the three Kähler forms. Assume moreover that the action of G gives rise to three equivariant moment mappings: $\mu^{(I)}$, $\mu^{(J)}$ and $\mu^{(K)}$; which we can combine into a single map

$$\mu : M \to \mathfrak{g}^* \otimes \mathbb{R}^3 \ .$$

Equivariance implies that $\mu^{-1}(0)$ is acted on by G. Assuming that $\mu^{-1}(0)/G$ is a manifold, we claim that it is actually hyperkähler.

Fix one of the complex structures, I, say; and consider the function

$$\nu = \mu^{(J)} + i\mu^{(K)} : M \to \mathfrak{g}^* \otimes \mathbb{C} \ .$$

For each Killing vector field X_a and any vector field Y,

$$\begin{aligned} d\nu_a(Y) &= \omega^{(J)}(X_a, Y) + i\omega^{(K)}(X_a, Y) = g(JX_a, Y) + ig(KX_a, Y) \\ d\nu_a(IY) &= g(JX_a, IY) + ig(KX_a, IY) = -g(KX, Y) + ig(JX, Y) \ , \end{aligned}$$

whence

$$d\nu_a(IY) = id\nu_a(Y) \ ;$$

or in other words, $\bar{\partial}\nu_a = 0$, so that ν is a holomorphic function (relative to I). This means that $\nu^{-1}(0)$ is a complex submanifold of a Kähler manifold and hence its induced metric is Kähler. Now G acts on $\nu^{-1}(0)$ in such a way that it preserves the Kähler structure, and the resulting moment mapping is clearly the restriction of $\mu^{(I)}$ to $\nu^{-1}(0)$. We may therefore perform the Kähler quotient of $\nu^{-1}(0)$ by the action of G, and obtain a manifold:

$$\nu^{-1}(0) \cap (\mu^{(I)})^{-1}(0)/G = \mu^{-1}(0)/G \ ,$$

whose metric is Kähler relative to (the complex structure induced by) I. To finish the proof that $\mu^{-1}(0)/G$ is hyperkähler, we repeat the above for J and K. This construction is called the *hyperkähler quotient*, and was described for the first time in [HKLR87].

3.4.4 $\mathcal{M}_k$ as a hyperkähler quotient

Now we will prove that $\mathcal{M}_k$ is a hyperkähler quotient of the configuration space $\mathcal{A}_k$ of fields $\mathsf{W}_{\underline{i}}$ corresponding to finite-energy configurations with monopole number k. We can think of $\mathsf{W}_{\underline{i}}$ as maps $\mathbb{R}^3 \to \mathbb{R}^4 \otimes so(3)$, and $\mathbb{R}^4$

can be thought of as a quaternionic vector space in two inequivalent ways: we first identify $\mathbb{R}^4 = \mathbb{H}$, but then we have to choose whether $\mathbb{H}$ acts by left or right multiplication. Since quaternionic multiplication is not commutative, the two actions are different. Since we will be dealing with monopoles, we choose the right action—left multiplication would correspond to anti-monopoles. Let I, J, and K denote the linear maps $\mathbb{R}^4 \to \mathbb{R}^4$ representing right multiplication on $\mathbb{H} \cong \mathbb{R}^4$ by the conjugate quaternion units $-i$, $-j$, and $-k$ respectively. The next exercise asks you to work out the explicit expressions for I, J, and K relative to a chosen basis.

Exercise 3.32 (Hyperkähler structure of $\mathbb{R}^4$)
Choose a basis $\{1, i, j, k\}$ for $\mathbb{H}$. Then relative to this basis, prove that the linear maps I, J, and K are given by the matrices:

$$I = \begin{pmatrix} i\sigma^2 & 0 \\ 0 & -i\sigma^2 \end{pmatrix} \quad J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \quad K = IJ = \begin{pmatrix} 0 & i\sigma^2 \\ i\sigma^2 & 0 \end{pmatrix} \ .$$

Notice that together with the euclidean metric on $\mathbb{R}^4$, I, J, K make $\mathbb{R}^4$ into a (linear) hyperkähler manifold.
(Hint: Remember that the matrix associated to a linear transformation is defined by $Ie_{\underline{i}} = e_{\underline{j}} I_{\underline{ji}}$. This choice makes composition of linear transformations correspond with matrix multiplication.)

We may now define a hyperkähler structure on $\mathcal{A}_k$ as follows. If $\dot{\mathsf{W}}_{\underline{i}}$ is a vector field on $\mathcal{A}_k$, then we define

$$(\hat{I}\,\dot{\mathsf{W}})_{\underline{i}}(x) = I_{\underline{ij}}\,\dot{\mathsf{W}}_{\underline{j}}(x) \ ,$$

and similarly for $\hat{J}$ and $\hat{K}$. Clearly they obey the quaternion algebra $\hat{I}\hat{J} = \hat{K}$, etc because I, J and K do. Moreover since they are constant (and so is the metric) they are certainly parallel relative to the Levi-Civita connection on $\mathcal{A}_k$ with the metric given by the Yang–Mills–Higgs functional. Hence $\mathcal{A}_k$ is an infinite-dimensional (affine) hyperkähler manifold.

Let $\mathcal{G}$ denote the group of finite-range time- and x^4-independent gauge transformations. Since the metric is gauge invariant, $\mathcal{G}$ acts on $\mathcal{A}_k$ via isometries. We also claim that $\mathcal{G}$ preserves the three complex structures and gives rise to three equivariant moment mappings. In fact, we will prove this in one go by constructing the moment mappings from the start.

The Killing vectors of the $\mathcal{G}$ action are just the infinitesimal gauge transformations and they are parametrised by square-integrable functions $\boldsymbol{\epsilon} : \mathbb{R}^3 \to so(3)$. The resulting Killing vector field is $X_{\boldsymbol{\epsilon}} \equiv \delta_{\boldsymbol{\epsilon}} \mathsf{W}_{\underline{i}} = D_{\underline{i}}\boldsymbol{\epsilon}$. For every such $\boldsymbol{\epsilon}$, define the following function:

$$\mu_{\boldsymbol{\epsilon}}^{(\hat{I})} = \tfrac{1}{2} \int d^3x\, I_{\underline{ij}} (G_{\underline{ij}} \cdot \boldsymbol{\epsilon}) \ ,$$

and the same for $\hat{J}$ and $\hat{K}$.

Now let $\dot{\mathsf{W}}$ be any tangent vector field on $\mathcal{A}_k$. (Here and in the sequel we will suppress the indices $\underline{i}$ whenever they don't play a role in an expression.) Then

$$\begin{aligned}\left(\imath_{\boldsymbol{\epsilon}}\omega^{(\hat{I})}\right)(\dot{\mathsf{W}}) &= \omega^{(\hat{I})}(D\boldsymbol{\epsilon},\dot{\mathsf{W}})\\ &= g(\hat{I}D\epsilon,\dot{\mathsf{W}})\\ &= \int d^3x\, I_{\underline{ij}}D_{\underline{j}}\boldsymbol{\epsilon}\cdot\dot{\mathsf{W}}_{\underline{i}}\\ &= \int d^3x\, I_{\underline{ij}}D_{\underline{i}}\dot{\mathsf{W}}_{\underline{j}}\cdot\boldsymbol{\epsilon} \qquad \text{(integrating by parts)}\\ &= d\mu^{(\hat{I})}_{\boldsymbol{\epsilon}}(\dot{\mathsf{W}})\ .\end{aligned}$$

Hence,

$$\imath_{\boldsymbol{\epsilon}}\omega^{(\hat{I})} = d\mu^{(\hat{I})}_{\boldsymbol{\epsilon}}\ .$$

Naturally, the same holds also for $\hat{J}$ and $\hat{K}$. Hence we can construct a moment mapping μ such that $\mu_{\boldsymbol{\epsilon}} = (\mu^{(\hat{I})}_{\boldsymbol{\epsilon}},\mu^{(\hat{J})}_{\boldsymbol{\epsilon}},\mu^{(\hat{K})}_{\boldsymbol{\epsilon}})$. The next exercise asks you to prove that it is equivariant.

Exercise 3.33 (Equivariance of the moment mapping)
Prove that the moment mapping $\mu = (\mu^{(I)},\mu^{(J)},\mu^{(K)})$ is equivariant. In other words, if $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}$ are gauge parameters, prove that

$$X_{\boldsymbol{\epsilon}}\,\mu^{(I)}_{\boldsymbol{\eta}} = d\mu^{(I)}_{\boldsymbol{\eta}}(X_{\boldsymbol{\epsilon}}) = \mu^{(I)}_{\boldsymbol{\epsilon}\times\boldsymbol{\eta}}\ ,$$

and the same for J and K.

Therefore we can apply the preceding discussion about the hyperkähler quotient to deduce that $\mu^{-1}(0)/\mathcal{G}$ is a hyperkähler manifold. But *what is* $\mu^{-1}(0)$? Configurations $\mathsf{W}_{\underline{i}}$ belonging in $\mu^{-1}(0)$ are those for which $\mu^{(\hat{I})}_{\boldsymbol{\epsilon}} = 0$ for all $\boldsymbol{\epsilon}$, and the same for $\hat{J}$ and $\hat{K}$. Since $\boldsymbol{\epsilon}$ is arbitrary, this is equivalent to demanding that $I_{\underline{ij}}G_{\underline{ij}} = 0$, and the same for $\hat{J}$ and $\hat{K}$. From the explicit expressions for the matrices I, J, and K found in Exercise 3.32, we find that

$$\begin{aligned}I_{\underline{ij}}G_{\underline{ij}} = 0 &\Rightarrow G_{12} = G_{34}\ ,\\ J_{\underline{ij}}G_{\underline{ij}} = 0 &\Rightarrow G_{13} = -G_{24}\ ,\\ K_{\underline{ij}}G_{\underline{ij}} = 0 &\Rightarrow G_{14} = G_{23}\ .\end{aligned}$$

But these make up precisely the self-duality condition on $G_{\underline{ij}}$, that is, the Bogomol'nyi equation!

Therefore $\mu^{-1}(0)$ is the submanifold of static solutions of the Bogomol'nyi equation with monopole number k (the BPS-k-monopoles) and $\mu^{-1}(0)/\mathcal{G}$ is their moduli space $\mathcal{M}_k$. In summary, $\mathcal{M}_k$ is a $4k$-dimensional hyperkähler manifold, obtained as an infinite-dimensional hyperkähler quotient of $\mathcal{A}_k$ by the action of the gauge group $\mathcal{G}$.

This "proof", although conceptually clear and offering a natural explanation of *why* $\mathcal{M}_k$ should be a hyperkähler manifold in the first place, relies rather heavily on differential geometry. Therefore a more pedestrian proof might be helpful, and we now turn to one such proof.

3.4.5 Another proof that $\mathcal{M}_k$ is hyperkähler

We start by expanding the Yang–Mills–Higgs action in terms of collective coordinates in order to obtain an expression for the metric. Let X^a, $a = 1, \ldots, 4k$, denote the collective coordinates on the moduli space $\mathcal{M}_k$ of BPS-k-monopoles. Let $\mathsf{W}_{\underline{i}}(x, X(t))$ be a family of BPS-monopoles whose t-dependence is only through the t-dependence of the collective coordinates; that is,

$$\dot{\mathsf{W}}_{\underline{i}} = \partial_a \mathsf{W}_{\underline{i}} \dot{X}^a \ . \tag{3.39}$$

Notice that $\partial_a \mathsf{W}_{\underline{i}}$ need *not* be perpendicular to the gauge orbits. Indeed, generically, we will have a decomposition

$$\partial_a \mathsf{W}_{\underline{i}} = \delta_a \mathsf{W}_{\underline{i}} + D_{\underline{i}} \boldsymbol{\epsilon}_a \ , \tag{3.40}$$

where $\delta_a \mathsf{W}_{\underline{i}}$ is the component perpendicular to the gauge orbits and $D_{\underline{i}} \boldsymbol{\epsilon}_a$ is the component tangent to the gauge orbits, hence an infinitesimal gauge-transformation. The gauge parameters $\boldsymbol{\epsilon}_a$ are determined uniquely by $\partial_a \mathsf{W}_{\underline{i}}$. Indeed, simply apply $D_{\underline{i}}$ and use the fact that $D^2 \equiv D_{\underline{i}} D_{\underline{i}}$ is negative-definite (hence invertible) to solve for $\boldsymbol{\epsilon}_a$:

$$\boldsymbol{\epsilon}_a = D^{-2} D_{\underline{i}} \partial_a \mathsf{W}_{\underline{i}} \ .$$

Exercise 3.34 (The Yang–Mills–Higgs effective action)
Compute the effective action for such a configuration of BPS-monopoles, and show that provided one sets $\mathsf{W}_0 = \dot{X}^a \boldsymbol{\epsilon}_a$, *it is given by*

$$L_{\text{eff}} = \tfrac{1}{2} g_{ab} \dot{X}^a \dot{X}^b - \frac{4\pi a}{e} |k| \ ,$$

where the metric on $\mathcal{M}_k$ *is given by*

$$g_{ab} = \int d^3x \, \delta_a \mathsf{W}_{\underline{i}} \cdot \delta_b \mathsf{W}_{\underline{i}} \ , \tag{3.41}$$

where, by construction, $\delta_a \mathsf{W}_{\underline{i}}$ *are perpendicular to the gauge orbits and satisfy the linearised Bogomol'nyi equation.*

One way to systematise the above expansion is in terms of t-derivatives. The zeroth order term is given by the potential, which is a constant since the motion is purely along the flat directions. The first order term vanishes due to our choice for W_0,[4] while the quadratic term, which describes the motion of such monopoles in the limit of zero velocity, corresponds precisely to geodesic motion on $\mathcal{M}_k$ relative to the induced metric—that is, as a particle moving freely on $\mathcal{M}_k$ or, somewhat pedantically, as a $(1+0)$-dimensional σ-model with $\mathcal{M}_k$ as its target space.

It is convenient to think of $\boldsymbol{\epsilon}_a$ as the components of a connection. We define $D_a \equiv \partial_a - e(\boldsymbol{\epsilon}_a \times -)$, whence we can think of $(\mathsf{W}_{\underline{i}}, \boldsymbol{\epsilon}_a)$ as the components of a connection on $\mathbb{R}^4 \times \mathcal{M}_k$. This allows us to interpret $\delta_a \mathsf{W}_{\underline{i}}$ as the mixed components of the curvature:

$$G_{a\underline{i}} = \partial_a \mathsf{W}_{\underline{i}} - \partial_{\underline{i}} \boldsymbol{\epsilon}_a - e\boldsymbol{\epsilon}_a \times \mathsf{W}_{\underline{i}} = \partial_a \mathsf{W}_{\underline{i}} - D_{\underline{i}} \boldsymbol{\epsilon}_a = \delta_a \mathsf{W}_{\underline{i}} \ .$$

The other components G_{ab} of the curvature may be formally computed from the Bianchi identity:

$$D_{\underline{i}} G_{ab} = -2 D_{[a} \delta_{b]} \mathsf{W}_{\underline{i}} = -D_a \delta_b \mathsf{W}_{\underline{i}} + D_b \delta_a \mathsf{W}_{\underline{i}} \ , \tag{3.42}$$

by applying $D_{\underline{i}}$ and inverting D^2 as before.

Exercise 3.35 (A somewhat more explicit formula for G_{ab})
Prove that

$$G_{ab} = -2eD^{-2}(\delta_a \mathsf{W}_{\underline{i}} \times \delta_b \mathsf{W}_{\underline{i}}) \ .$$

(Hint: Apply $D_{\underline{i}}$ to (3.42)*, and use that $\delta_a \mathsf{W}_{\underline{i}}$ is perpendicular to the gauge orbits.)*

Using these formulae it is possible to write a formal expression for the Christoffel symbols of the Levi-Civita connection. Naturally this is left as an exercise.

Exercise 3.36 (The Christoffel symbols)
Prove that

$$\Gamma_{abc} = g_{cd} \Gamma_{ab}{}^d = \int d^3x \, D_a \delta_b \mathsf{W}_{\underline{i}} \cdot \delta_c \mathsf{W}_{\underline{i}} \ . \tag{3.43}$$

Notice that $\Gamma_{abc} = \Gamma_{bac}$, since $D_{[b}\delta_{c]} \mathsf{W}_{\underline{i}} = -\frac{1}{2} D_{\underline{i}} G_{ab}$ which is orthogonal to $\delta_c \mathsf{W}_{\underline{i}}$.
(Hint: Use the explicit expressions (3.30) *and* (3.41) *and compute.)*

Using the explicit expressions found in Exercise 3.32 for the hyperkähler structure in $\mathbb{R}^4$ we now define the following two-forms on $\mathcal{M}_k$:

$$\omega_{ab}^{(I)} = \int d^3x \, I_{\underline{i}\underline{j}} \delta_a \mathsf{W}_{\underline{i}} \cdot \delta_b \mathsf{W}_{\underline{j}} \ ,$$

[4]Notice that W_0 is not zero for generic choices of $\mathsf{W}_{\underline{i}}(x, X(t))$, but it can be made to vanish after a t-dependent gauge transformation.

and similarly for J and K, and their corresponding almost complex structures

$$I_a{}^b = g^{bc}\omega^{(I)}_{ac} \quad J_a{}^b = g^{bc}\omega^{(J)}_{ac} \quad K_a{}^b = g^{bc}\omega^{(K)}_{ac} \ .$$

Exercise 3.37 (Explicit expressions for the complex structures)

(a) Prove that $I_{\underline{ij}}\delta_a \mathsf{W}_{\underline{j}}$ is orthogonal to the gauge orbits, and the same for J and K.

(b) Then derive the following formula:

$$I_a{}^b \delta_b \mathsf{W}_{\underline{i}} = -I_{\underline{ij}}\delta_a \mathsf{W}_{\underline{j}} \ , \tag{3.44}$$

and the same for $J_a{}^b$ and $K_a{}^b$.

(c) Using these expressions, prove that $I_a{}^b$, $J_a{}^b$ and $K_a{}^b$ obey the quaternion algebra.

(Hints: (a) This is equivalent to the linearised Bogomol'nyi equation, in the form $I_{\underline{ij}}G_{\underline{ij}} = 0$, etc.
(b) Argue that since $I_a{}^b\delta_b\mathsf{W}_{\underline{i}}$ is orthogonal to the gauge orbits, $\int d^3x\, I_a{}^b\delta_b\mathsf{W}_{\underline{i}}\cdot\delta_c\mathsf{W}_{\underline{i}}$ defines it uniquely. Then just compute the integral and use (a).)

We claim that the forms $\omega^{(I)}$, $\omega^{(J)}$ and $\omega^{(K)}$ are parallel. Let's see this for one of them $\omega \equiv \omega^{(I)}$, the other cases being identical. By definition,

$$\nabla_a \omega_{bc} = \partial_a \omega_{bc} - \Gamma_{ab}{}^d \omega_{dc} - \Gamma_{ac}{}^d \omega_{bd} \ .$$

We now compute this in steps. First of all we have:

$$\partial_a \omega_{bc} = \int d^3x\, I_{\underline{ij}} \left(D_a \delta_b \mathsf{W}_{\underline{i}} \cdot \delta_c \mathsf{W}_{\underline{j}} + \delta_b \mathsf{W}_{\underline{i}} \cdot D_a \delta_c \mathsf{W}_{\underline{j}} \right) \ . \tag{3.45}$$

Now we notice that $\Gamma_{ab}{}^d\omega_{dc} = -I_c{}^e\Gamma_{abe}$. Using the explicit expression (3.43), we arrive at

$$\Gamma_{ab}{}^d \omega_{dc} = -I_c{}^e \int d^3x\, D_a \delta_b \mathsf{W}_{\underline{i}} \cdot \delta_e \mathsf{W}_{\underline{i}} \ , \tag{3.46}$$

and using (3.44) we can rewrite this as

$$\Gamma_{ab}{}^d \omega_{dc} = \int d^3x\, I_{\underline{ij}} D_a \delta_b \mathsf{W}_{\underline{i}} \cdot \delta_c \mathsf{W}_{\underline{j}} \ , \tag{3.47}$$

with a similar expression for $\Gamma_{ac}{}^d\omega_{bd} = -\Gamma_{ac}{}^d\omega_{db}$. Adding it all together we find that $\nabla_a\omega_{bc} = 0$. But this means that $I_a{}^b$, $J_a{}^b$, and $K_a{}^b$ are also parallel, whence $\mathcal{M}_k$ is hyperkähler.

Chapter 4

The Effective Action for N=2 Supersymmetric Yang–Mills

In this chapter we will perform the collective coordinate expansion of the N=2 supersymmetric $SO(3)$ Yang–Mills theory defined by equation (2.13). We will also discuss the quantisation of the corresponding effective action. As we saw in the discussion in section 3.1.5 on the effective theory for the 1-monopole sector, the effective theory offers a qualitatively faithful description of the dyonic spectrum, even though quantitatively it is only an approximation. Of course, in the non-supersymmetric theory there is no reason to expect that the true quantum spectrum should resemble the classical spectrum given by the Bogomol'nyi formula, but as we saw in Chapter 2, supersymmetry protects both the formula for the bound from quantum corrections and also the saturation of the bound. Hence it makes sense to expect that in the supersymmetric theory, the quantisation of the effective action should teach us something about the full quantum theory. As we shall soon discuss, this will have a chance of holding true only in the N=4 theory, but we can already learn something from the N=2 theory we have just studied. We will therefore first discuss the fermionic collective coordinates and then the effective quantum theory in the k-monopole sector. We will see that there are $2k$ fermionic collective coordinates and that the resulting effective theory is to lowest order a $(0+1)$ supersymmetric σ-model admitting N=4 supersymmetry consistent with the fact that $\mathcal{M}_k$ is hyperkähler. The quantisation of this theory then leads to a geometric interpretation of the Hilbert space as the square-integrable $(0,q)$-forms on $\mathcal{M}_k$, and of the hamiltonian as the laplacian. This chapter is based on the work of Gauntlett [Gau94].

4.1 Fermionic collective coordinates

As we saw in section 3.2, there are $4k$ bosonic collective coordinates in the k-monopole sector. The purpose of this section is to compute the number of fermionic collective coordinates: we will see that there are $2k$ of them. We will prove this in two ways. First we can set up this problem as the computation of the index of an operator, as we did for the bosonic collective coordinates; then essentially the same calculation that was done in section 3.2 yields the answer. Alternatively, and following Zumino [Zum77], we will exhibit a supersymmetry between the bosonic and fermionic collective coordinates which will also allow us to count them.

Suppose that we start with an N=2 BPS-monopole obtained, say, in the manner of Exercise 2.17. Fermionic collective coordinates are simply fermionic flat directions of the potential; that is, fermionic configurations which do not change the potential. In order to see what this means, let us first write down the potential for a general field configuration. To this effect let us break up the lagrangian density (2.13) into kinetic minus potential terms $\mathcal{L} = \mathcal{T} - \mathcal{V}$, where

$$\mathcal{T} = \tfrac{1}{2}\|\mathsf{G}_{0i}\|^2 + \tfrac{1}{2}\|D_0\mathsf{P}\|^2 + \tfrac{1}{2}\|D_0\mathsf{S}\|^2 + i\bar{\boldsymbol{\psi}}\cdot\gamma^0 D_0\boldsymbol{\psi} \tag{4.1}$$

and

$$\begin{aligned}\mathcal{V} = \tfrac{1}{2}\|D_i\mathsf{P}\|^2 + \tfrac{1}{2}\|D_i\mathsf{S}\|^2 + \tfrac{1}{4}\mathsf{G}_{ij}\cdot\mathsf{G}_{ij} + \tfrac{1}{2}e^2\|\mathsf{P}\times\mathsf{S}\|^2 + i\bar{\boldsymbol{\psi}}\cdot\gamma_i D_i\boldsymbol{\psi}\\ - ie\bar{\psi}\cdot\gamma_5\mathsf{P}\times\boldsymbol{\psi} - ie\bar{\psi}\cdot\mathsf{S}\times\boldsymbol{\psi}\ .\end{aligned} \tag{4.2}$$

The potential is then the integral $V = \int_{\mathbb{R}^3}\mathcal{V}$. For an N=2 BPS-monopole, with $\mathsf{W}_0 = 0$, $\mathsf{S} = \alpha\boldsymbol{\phi}$, $\mathsf{P} = \beta\boldsymbol{\phi}$ with $\alpha^2+\beta^2 = 1$, the potential is given by

$$V = \tfrac{1}{2}\int_{\mathbb{R}^3}\|D_i\boldsymbol{\phi}\|^2 + \tfrac{1}{4}\int_{\mathbb{R}^3}\mathsf{G}_{ij}\cdot\mathsf{G}_{ij} + i\int_{\mathbb{R}^3}\bar{\boldsymbol{\psi}}\cdot\left(\gamma_i D_i - e\boldsymbol{\phi}(\alpha+\beta\gamma_5)\right)\boldsymbol{\psi}\ ,$$

where in the last term $\boldsymbol{\phi}$ is in the adjoint representation; that is, $\boldsymbol{\phi}\boldsymbol{\psi} = \boldsymbol{\phi}\times\boldsymbol{\psi}$. The first two terms in the potential already reproduce the potential energy of a nonsupersymmetric BPS-monopole: $\frac{4\pi a}{e}|k|$, for a k-monopole. Therefore turning on the fermions will not change the potential provided that the third term vanishes; in other words, provided that $\boldsymbol{\psi}$ satisfies the Dirac equation in the presence of the BPS-monopole. In other words, fermionic collective coordinates are in one-to-one correspondence with zero modes of the Dirac operator. We will now count the number of zero modes in two ways.

4.1.1 Computing the index

In order to count the zero modes it is again convenient to use the reformulation of the BPS-monopole as an instanton, in terms of $\mathsf{W}_{\underline{i}} = (\mathsf{W}_i, \boldsymbol{\phi})$, and to define the following four-dimensional euclidean Dirac matrices: $\bar\gamma_i = \gamma_0\gamma_i$ and $\bar\gamma_4 = \gamma_0(\alpha + \beta\gamma_5)$. In terms of these, the fermion term in the potential is given by $i\int_{\mathbb{R}^3} \boldsymbol{\psi}^\dagger \cdot \bar\gamma_{\underline{i}} D_{\underline{i}} \boldsymbol{\psi}$, keeping in mind that $\partial_4 \equiv 0$.

We want to compute the number of normalisable solutions to the equation $\bar\gamma_{\underline{i}} D_{\underline{i}} \boldsymbol{\psi} = 0$. Let us choose a Weyl basis in which $\bar\gamma_5 = \bar\gamma_1\bar\gamma_2\bar\gamma_3\bar\gamma_4$ is diagonal. In such a basis, a convenient representation of the euclidean Dirac matrices is the one given in equation (3.19). In that representation the euclidean Dirac equation becomes:

$$\begin{pmatrix} 0 & -i\sigma_i D_i - e\boldsymbol{\phi}\,\mathbb{1} \\ i\sigma_i D_i - e\boldsymbol{\phi}\,\mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\psi}_+ \\ \boldsymbol{\psi}_- \end{pmatrix} = 0 \ .$$

But notice that we have seen these very operators before, in the computation of the number of bosonic collective coordinates in section 3.2! In fact, in terms of the operator $\mathcal{D}$ defined in equation (3.17), the above Dirac equation breaks up into two equations, one for each chirality:

$$\mathcal{D}\boldsymbol{\psi}_- = 0 \qquad \text{and} \qquad \mathcal{D}^\dagger\boldsymbol{\psi}_+ = 0 \ .$$

But now in Exercise 3.5 you showed that the operator $\mathcal{D}\mathcal{D}^\dagger$ is positive, whence it has no normalisable zero modes, hence neither does $\mathcal{D}^\dagger$. Therefore we notice that fermionic zero modes in the presence of a BPS-monopole necessarily have negative chirality. (For antimonopoles, it would have been $\mathcal{D}$ which has no normalisable zero modes, and fermion zero modes would have positive chirality.)

We can arrive at the same result in a different way which doesn't use the explicit realisation of the $\bar\gamma_i$-matrices. In fact, it is an intrinsic property of fermions coupled to instantons (in four-dimensions). The next exercise takes you through it.

Exercise 4.1 (Fermion zero modes are chiral)
Consider solutions of the four-dimensional euclidean Dirac equation $\bar\gamma_{\underline{i}} D_{\underline{i}} \boldsymbol{\psi} = 0$ in the presence of an (anti)self-dual gauge field. Prove that if the gauge field is self-dual (respectively, antiselfdual), then fermion zero modes have negative (respectively, positive) chirality.
(Hint: Compute the Dirac laplacian $\bar\gamma_{\underline{i}}\bar\gamma_{\underline{j}} D_{\underline{i}} D_{\underline{j}} \boldsymbol{\psi}$ and use the fact that $D^2 = D_{\underline{i}} D_{\underline{i}}$ is negative-definite and has not normalisable zero modes.)

Finally, just as in section 3.2.1, the number of normalisable zero modes of $\mathcal{D}$ is given by its index, which was computed in section 3.2.3 to be $2k$, where k is the monopole number.

4.1.2 Using supersymmetry

We can reproduce this result in a different, but more useful way by exhibiting a supersymmetry between the bosonic and fermionic zero modes. This is based on work by Zumino [Zum77].

Let $\delta\mathsf{W}_{\underline{i}}$ be a bosonic zero mode; that is, $\delta\mathsf{W}_{\underline{i}}$ satisfies the linearised Bogomol'nyi equation (3.15) and Gauss's law (3.1). Let η_+ be a constant, *commuting* spinor of positive chirality, normalised to $\eta_+^\dagger\eta_+ = 1$. Define

$$\boldsymbol{\psi} \equiv \delta\mathsf{W}_{\underline{i}}\bar{\gamma}_{\underline{i}}\eta_+ \ . \tag{4.3}$$

It is clear that $\boldsymbol{\psi}$ has negative chirality and, as the next exercise asks you to show, $\boldsymbol{\psi}$ satisfies the Dirac equation.

Exercise 4.2 (From bosonic to fermionic zero modes)
Let $\delta\mathsf{W}_{\underline{i}}$ be a bosonic zero mode as above. With $\boldsymbol{\psi}$ defined as above, prove that $\bar{\gamma}\cdot D\boldsymbol{\psi} = 0$.
(Hint: Use Exercise 3.4.)

Conversely, suppose that $\boldsymbol{\psi}$ is a fermionic zero mode with negative chirality; that is, $\bar{\gamma}_5\boldsymbol{\psi} = -\boldsymbol{\psi}$ and $\bar{\gamma}\cdot D\boldsymbol{\psi} = 0$. Then define

$$\delta\mathsf{W}_{\underline{i}} \equiv i\eta_+^\dagger\bar{\gamma}_{\underline{i}}\boldsymbol{\psi} - i\boldsymbol{\psi}^\dagger\bar{\gamma}_{\underline{i}}\eta_+ \ .$$

The next exercise asks you to prove that $\delta\mathsf{W}_{\underline{i}}$ is a bosonic zero mode.

Exercise 4.3 (... and back)
With $\delta\mathsf{W}_{\underline{i}}$ defined as above, prove that it satisfies the linearised Bogomol'nyi equation (3.15) *and Gauss's law* (3.1).

The above result seems to suggest that there is a one-to-one correspondence between the bosonic and fermionic zero modes, but this is fictitious, since not all the fermionic zero modes obtained in this fashion are independent. Indeed, as we will now see, they are related by the complex structure.

Let $\delta_a\mathsf{W}_{\underline{i}}$ for $a = 1,\ldots,4k$ denote the bosonic zero modes, and let $\boldsymbol{\psi}_a = \delta_a\mathsf{W}_{\underline{i}}\bar{\gamma}_{\underline{i}}\eta_+$ be the corresponding fermionic zero modes. We will prove that $I_a{}^b\boldsymbol{\psi}_b = i\boldsymbol{\psi}_a$, where I is one of the complex structures of $\mathcal{M}_k$. Hence comparing with the discussion at the end of section 3.2.1, we see that unlike bosonic zero modes, $\boldsymbol{\psi}_a$ and $i\boldsymbol{\psi}_a$ are not linearly independent.

Let η_+ be a commuting spinor of positive chirality normalised to $\eta_+^\dagger\eta_+ = 1$. Define a 4×4 matrix A with entries

$$A_{\underline{ij}} \equiv \eta_+^\dagger\bar{\gamma}_{\underline{ij}}\eta_+ \ . \tag{4.4}$$

We start by listing some properties of this matrix.

Exercise 4.4 (A complex structure)
Let A be the 4×4 matrix with entries $A_{\underline{ij}}$ given by (4.4). Prove that A satisfies the following properties:

(1) A is antisymmetric;

(2) iA is real;

(3) A is antiselfdual: $A_{\underline{ij}} = -\frac{1}{2}\epsilon_{\underline{ijk\ell}}A_{\underline{k\ell}}$;

(4) $A^2 = \mathbb{1}$, so that iA is a complex structure; and

(5) $A_{\underline{ij}}\bar\gamma_{\underline{j}}\eta_+ = -\bar\gamma_{\underline{i}}\eta_+$.

(Hint: This requires the Fierz identity:

$$\eta_+\eta_+^\dagger = \tfrac{1}{4}(\mathbb{1}+\bar\gamma_5) - \tfrac{1}{8}A_{\underline{ij}}\bar\gamma_{\underline{ij}}\ ,$$

which you should prove.)

We will now prove that we can choose η_+ in such a way that that iA agrees with any one of the complex structures on $\mathbb{R}^4$ defined in Exercise 3.32, and that such an η_+ is unique up to a phase. We start by noticing that the 4×4 matrix iA defined above is real and antisymmetric, hence it belongs to $so(4)$. As a Lie algebra, $so(4)$ is isomorphic to $so(3)\times so(3)$ (see Exercise 2.32). The fact that iA is antiselfdual, means that iA belongs to one of these $so(3)$'s. In fact, it is the $so(3)$ spanned by the complex structures I, J, and K of Exercise 3.32. (Check that they are antiselfdual!) In fact, $\mathbb{R}^4$ has a two-sphere worth of complex structures: $\{aI+bJ+cK | a^2+b^2+c^2=1\}$, and from the above exercise, we see that iA defines a point in this two-sphere. In the next exercise we see this explicitly.

Exercise 4.5 (iA lives on the sphere)
Compute the matrix iA explicitly in the representation of the Dirac matrices given by equation (3.19), and show that it is given by

$$iA = \begin{pmatrix} 0 & q_3 & -q_2 & -q_1 \\ -q_3 & 0 & q_1 & -q_2 \\ q_2 & -q_1 & 0 & -q_3 \\ q_1 & q_2 & q_3 & 0 \end{pmatrix} = -q_1 K - q_2 J + q_3 I\ , \tag{4.5}$$

where I, J, and K are the complex structures in $\mathbb{R}^4$ defined in Exercise 3.32 and $q_i = \eta^\dagger\sigma_i\eta$, where η is a complex Weyl spinor normalised to $\eta^\dagger\eta = 1$. (In the Weyl basis above $\eta_+ = \begin{pmatrix}\eta\\0\end{pmatrix}$.) Prove that q_i are real and that they satisfy $\sum_i q_i^2 = 1$, hence iA defines a point in the unit two-sphere in $\mathbb{R}^3$.

Now, in the Weyl basis introduced above, $\eta_+ = \begin{pmatrix}\eta\\0\end{pmatrix}$. Any other normalised positive chirality spinor η'_+ will have the same form with η' replacing η. This new Weyl spinor will be related to η by an element of $U(2)$: $\eta' = U\eta$. The matrix iA' obtained from η' has the form given by (4.5) but with q_i replaced by $q'_i \equiv \eta'^\dagger\sigma_i\eta' = \eta^\dagger U^{-1}\sigma_i U\eta$.

Exercise 4.6 (Adjoint transformation)
In the notation above, prove that $q'_i = U_{ij}q_j$, where U_{ij} is the three-dimensional adjoint representation of $U(2)$. (Notice that because the $U(1)$ subgroup corresponding to the scalar matrices act trivially, only the $SU(2)$ subgroup acts effectively in this representation.)

Therefore the action of $U(2)$ on η induces the adjoint action on the (q_i). This action is transitive on the unit sphere $\sum_i q_i^2 = 1$, hence any two points (q_i) and (q'_i) are related by an element of $U(2)$. Notice that $U(2) = SU(2)\times U(1)$ and that the $SU(2)$ subgroup acts freely, whereas the $U(1)$ acts trivially. Hence once a complex structure has been chosen, η is unique up to the action of $U(1)$; that is, a phase.

Let us then exercise our right to choose η_+. We do so in such a way that

$$\eta_+^\dagger\bar\gamma_{\underline{ij}}\eta_+ = iI_{\underline{ij}}\ ,$$

where $I_{\underline{ij}}$ is given by Exercise 3.32. Then using equation (3.44), we have

$$\begin{aligned} {I_a}^b\,\boldsymbol{\psi}_b &= {I_a}^b\,\delta_b \mathsf{W}_{\underline{i}}\bar\gamma_{\underline{i}}\eta_+ = -I_{ij}\delta_a\mathsf{W}_{\underline{j}}\bar\gamma_{\underline{i}}\eta_+\ ,\\ &= i\delta_a\mathsf{W}_{\underline{j}}\bar\gamma_{\underline{j}}\eta_+\ ,\\ &= i\boldsymbol{\psi}_a\ , \end{aligned} \tag{4.6}$$

where in the next to last line we have used (5) in Exercise 4.5. Therefore there are only half as many linearly independent fermionic zero modes as there are bosonic ones, in agreement with the index calculation in the previous section.

4.2 The effective action

In this section we will write down the effective action governing the dynamics of the collective coordinates to lowest order. Let us introduce bosonic collective coordinates X^a, for $a = 1,\ldots,4k$. These coordinates parametrise the moduli space $\mathcal{M}_k$ of BPS-monopoles with topological number k. In addition there will be fermionic collective coordinates λ^a, for $a = 1,\ldots,4k$ satisfying the condition $\lambda^a {I_a}^b = i\lambda^b$. We now expand the supersymmetric Yang–Mills action (2.13) in terms of the collective coordinates $\{X^a,\lambda^a\}$ and keep only the lowest nontrivial order. In order to count the order of an expression we take the conventions that λ^a has order $\frac{1}{2}$, X^a has order 0, but time derivatives have order 1. These conventions are such that a free theory of bosons X^a and fermions λ^a is of quadratic order.

We start by performing an $SO(2)$ transformation which puts $\mathsf{S} = \boldsymbol{\phi}$ and $\mathsf{P} = 0$, and choosing an appropriate parametrisation for the fields $\mathsf{W}_{\underline{i}}$, P, W_0, and $\boldsymbol{\psi}$ in terms of the collective coordinates. As in the nonsupersymmetric theory, we leave W_0 and, in this case also P, undetermined for the moment. We choose to parametrise $\mathsf{W}_{\underline{i}}$ as $\mathsf{W}_{\underline{i}}(x, X(t))$, where all the time dependence comes from the collective coordinates. For this reason, equation (3.39) still holds where, as before, $\partial_a \mathsf{W}_{\underline{i}}$ need not be perpendicular to the gauge orbits. Nevertheless we can project out a part which does: $\delta_a \mathsf{W}_{\underline{i}}$ as in (3.40). Because $\delta_a \mathsf{W}_{\underline{i}}$ is perpendicular to the gauge orbits, we know that $\boldsymbol{\psi}_a$ given by (4.3) is a fermion zero mode. We therefore parametrise $\psi = \psi_a \lambda^a$. Notice that since ψ_a is commuting (because so is η_+), λ^a is anticommuting as expected.

There is no reason, in principle, to expect $\mathsf{W}_{\underline{i}}$ not to depend also on the fermionic collective coordinates. In fact, using that these are odd, we can expand $\mathsf{W}_{\underline{i}}$ as follows:

$$\mathsf{W}_{\underline{i}}(x, X, \lambda) = \mathsf{W}_{\underline{i}}(x, X) + \lambda^a \mathsf{W}_{\underline{i},a}(x, X) + \lambda^a \lambda^b \mathsf{W}_{\underline{i},ab}(x, X) + \cdots ;$$

but it is not hard to see that all terms but the first in the above expansion contribute only higher order terms to the effective action.

Because our choice of ψ is a zero mode of the Dirac equation in the presence of a BPS-monopole, the discussion of section 4.1 applies and provided that $\mathsf{P} = 0$, the potential remains at its minimum value: $\frac{4\pi a}{e} k$. However, P need not remain zero for this to be case. It can evolve along a flat direction as we now show. The P-dependent terms in the potential (4.2) can be written as follows:

$$\tfrac{1}{2} \int_{\mathbb{R}^3} \|D_{\underline{i}} \mathsf{P}\|^2 + ie \int_{\mathbb{R}^3} \boldsymbol{\psi}^\dagger \cdot (\mathsf{P} \times \boldsymbol{\psi}) ,$$

where we have used that $\bar{\gamma}_5 \boldsymbol{\psi} = \gamma_0 \gamma_5 \boldsymbol{\psi} = -\boldsymbol{\psi}$. Integrating the first term by parts, and using the invariance of the inner product in the second term, the above expression becomes:

$$-\tfrac{1}{2} \int_{\mathbb{R}^3} \mathsf{P} \cdot \left(D^2 \mathsf{P} + 2ie \boldsymbol{\psi}^\dagger \times \boldsymbol{\psi} \right) ,$$

where $D^2 = D_{\underline{i}} D_{\underline{i}}$.

Exercise 4.7 (Computing $\boldsymbol{\psi}^\dagger \times \boldsymbol{\psi}$)
Prove the following identity:

$$\boldsymbol{\psi}^\dagger \times \boldsymbol{\psi} = -\frac{1}{e} \bar{\lambda}^a \lambda^b D^2 G_{ab} ,$$

where G_{ab} are some of the components of the curvature of the connection $(\mathsf{W}_{\underline{i}}, \boldsymbol{\epsilon}_a)$ on $\mathbb{R}^4 \times \mathcal{M}_k$, which were computed in Exercise 3.35.
(Hint: You might want to use the identity:

$$\eta_+^\dagger \bar\gamma_{\underline{i}} \bar\gamma_{\underline{j}} \eta_+ = \delta_{\underline{ij}} + i I_{\underline{ij}} \ , \tag{4.7}$$

which is (up to a factor) the projector onto the I-antiholomorphic subspace of the complexification $\mathbb{C}^4$ of $\mathbb{R}^4$.)

Therefore we see that the condition that the potential remains constant demands that we either set P to zero or else

$$\mathsf{P} = 2i\bar\lambda^a \lambda^b G_{ab} \ .$$

Next we tackle the kinetic terms. Notice that either of the choices for P allow us to discard P from the kinetic terms. Indeed, if P is nonzero, then the above expression shows that it is already of order 1 and hence its contribution to the kinetic term (4.1) will be of order higher than quadratic. Having discarded P from the kinetic terms, we remain with

$$\tfrac{1}{2}\int_{\mathbb{R}^3} \|\mathsf{G}_{0\underline{i}}\|^2 + i\int_{\mathbb{R}^3} \boldsymbol{\psi}^\dagger \cdot D_0 \boldsymbol{\psi} \ .$$

The first term is computed in the following exercise.

Exercise 4.8 (The first kinetic term)
Prove that the first term in the kinetic energy above is given by

$$\tfrac{1}{2}\int_{\mathbb{R}^3} \|\mathsf{G}_{0\underline{i}}\|^2 = \tfrac{1}{2} g_{ab} \dot X^a \dot X^b + \tfrac{1}{2}\int_{\mathbb{R}^3} \|D_{\underline{i}}(\boldsymbol{\epsilon}_a \dot X^a - \mathsf{W}_0)\|^2 \ ,$$

where g_{ab} was defined in (3.41).
(Hint: Use the fact that $\delta_a \mathsf{W}_{\underline{i}}$ is perpendicular to the gauge orbits!)

Finally we come to the second kinetic term. Plugging in the expression for $\boldsymbol{\psi}$ and using equations (4.7), (3.44), and (4.6) we can rewrite the second term as

$$\begin{aligned} i\int_{\mathbb{R}^3} \boldsymbol{\psi}^\dagger \cdot D_0 \boldsymbol{\psi} = 2i g_{ab} \bar\lambda^a \dot\lambda^b + 2i\bar\lambda^a \lambda^b \dot X^c \int_{\mathbb{R}^3} \delta_a \mathsf{W}_{\underline{i}} \cdot \partial_c \delta_b \mathsf{W}_{\underline{i}} \\ + 2ie\bar\lambda^a \lambda^b \int_{\mathbb{R}^3} \mathsf{W}_0 \cdot \left(\delta_a \mathsf{W}_{\underline{i}} \times \delta_b \mathsf{W}_{\underline{j}}\right) \ . \end{aligned}$$

The next exercise finishes off the calculation.

Exercise 4.9 (... and the second kinetic term)
Prove that the second kinetic term can be written as

$$i \int_{\mathbb{R}^3} \boldsymbol{\psi}^\dagger \cdot D_0 \boldsymbol{\psi} = 2 i g_{ab} \bar{\lambda}^a \left(\dot{\lambda}^b + \Gamma_{cd}{}^b \dot{X}^c \lambda^d \right) - i \bar{\lambda}^a \lambda^b \int_{\mathbb{R}^3} \left(\mathsf{W}_0 - \boldsymbol{\epsilon}_c \dot{X}^c \right) \cdot D^2 G_{ab} \ ,$$

where the Christoffel symbols $\Gamma_{cd}{}^b$ were defined in equation (3.43).

Putting the results of Exercises 4.8 and 4.9, we find that the kinetic terms of the action are (to lowest order) given by:

$$\tfrac{1}{2} g_{ab} \dot{X}^a \dot{X}^b + 2 i g_{ab} \bar{\lambda}^a \dot{\lambda}^b + 2 i g_{ab} \Gamma_{cd}{}^b \bar{\lambda}^a \dot{X}^c \lambda^d \\ - i \bar{\lambda}^a \lambda^b \int_{\mathbb{R}^3} \left(\mathsf{W}_0 - \boldsymbol{\epsilon}_c \dot{X}^c \right) \cdot D^2 G_{ab} + \tfrac{1}{2} \int_{\mathbb{R}^3} \| D_{\underline{i}} (\boldsymbol{\epsilon}_a \dot{X}^a - \mathsf{W}_0) \|^2 \ .$$

We see that we can cancel the last two terms provided that we set

$$\mathsf{W}_0 = \boldsymbol{\epsilon}_a \dot{X}^a - 2 i \bar{\lambda}^a \lambda^b G_{ab} \ .$$

With this choice, and to lowest order, the effective action then becomes:

$$L_{\text{eff}} = \tfrac{1}{2} g_{ab} \dot{X}^a \dot{X}^b + 2 i g_{ab} \bar{\lambda}^a \left(\dot{\lambda}^b + \Gamma_{cd}{}^b \dot{X}^c \lambda^d \right) - \frac{4\pi a}{e} k \ . \tag{4.8}$$

Ignoring the constant term, this action describes a (0+1)-dimensional supersymmetric (as we shall see shortly) σ-model with target $\mathcal{M}_k$.

4.3 N=4 supersymmetry of the effective action

In general, symmetries of the theory under consideration play important roles in the effective action. Broken symmetries give rise to collective coordinates, whereas unbroken symmetries remain symmetries of the effective action. As we have seen in section 2.3.4, N=2 BPS-monopoles preserve one half of the four-dimensional N=2 supersymmetry. This supersymmetry must be present in the effective action. In 0+1 dimensions, supersymmetry charges are one-component Majorana spinors, hence one supersymmetry charge in four dimensions gives rise to four supersymmetry charges in 0+1. In this section we will prove that the effective action given by (4.8) admits N=4 supersymmetry. From the proof it follows that the same is true for any supersymmetric σ-model with hyperkähler target manifold—which is, of course, a well-known fact.

We start by discarding the constant term in the action (4.8) and rewriting the remaining terms in terms of complex coordinates adapted to the complex structure I of $\mathcal{M}_k$. To this end we define complex coordinates $(Z^\alpha, \bar{Z}^{\bar{\alpha}})$ which diagonalise the complex structure; that is, such that $I_\alpha{}^\beta = i\mathbb{1}_\alpha{}^\beta$ and $I_{\bar{\alpha}}{}^{\bar{\beta}} = -i\mathbb{1}_{\bar{\alpha}}{}^{\bar{\beta}}$. As for the fermions, equation (4.6) implies that $\lambda^a I_a{}^b = i\lambda^b$, whence $\lambda^{\bar{\alpha}} = 0$. Similarly, $\bar{\lambda}^\alpha = 0$. This prompts us to define new fermions ζ such that $\zeta^\alpha = \sqrt{2}\lambda^\alpha$ and $\zeta^{\bar{\alpha}} = \sqrt{2}\bar{\lambda}^{\bar{\alpha}}$. In terms of these new variables, the effective action remains

$$L_{\text{eff}} = g_{\alpha\bar{\beta}}\dot{Z}^\alpha \dot{\bar{Z}}^{\bar{\beta}} + i g_{\bar{\alpha}\beta}\zeta^{\bar{\alpha}}\left(\dot{\zeta}^\beta + \Gamma_{\gamma\delta}{}^\beta \dot{Z}^\gamma \zeta^\delta\right) , \tag{4.9}$$

where we have used that for a Kähler metric in complex coordinates, the only nonzero components of the metric and the Christoffel symbols are $g_{\alpha\bar{\beta}} = g_{\bar{\alpha}\beta}$ and $\Gamma_{\alpha\beta}{}^\gamma$ and $\Gamma_{\bar{\alpha}\bar{\beta}}{}^{\bar{\gamma}}$, as was proven in Exercise 3.23.

4.3.1 N=4 supersymmetry in $\mathbb{R}^4$: a toy model

In order to understand the supersymmetry of the action (4.9), let us first discuss the case of $\mathbb{R}^4$ with the standard euclidean flat metric. We can think of $\mathbb{R}^4$ as $\mathbb{C}^2$ and introduce complex coordinates $Z^\alpha, \bar{Z}^{\bar{\alpha}}$ where $\alpha = 1, 2$. The analogous action to (4.9) in this case is given simply by

$$L_{\text{eff}} = \sum_\alpha \left(\dot{Z}^\alpha \dot{\bar{Z}}^{\bar{\alpha}} + i\zeta^{\bar{\alpha}}\dot{\zeta}^\alpha\right) , \tag{4.10}$$

where ζ^α and $\zeta^{\bar{\alpha}}$ are the accompanying fermions. This toy action has four real supersymmetries. Two of them are manifest, as the next exercise asks you to show.

Exercise 4.10 (N=2 supersymmetry in flat space)
Let $\delta_{\mathbb{1}}$ and δ_I be the supersymmetries defined as follows:

$$\delta_{\mathbb{1}} Z^\alpha = \zeta^\alpha \qquad \delta_{\mathbb{1}}\zeta^\alpha = i\dot{Z}^\alpha \qquad\qquad \delta_I Z^\alpha = i\zeta^\alpha \qquad \delta_I \zeta^\alpha = \dot{Z}^\alpha$$
$$\delta_{\mathbb{1}} \bar{Z}^{\bar{\alpha}} = \zeta^{\bar{\alpha}} \qquad \delta_{\mathbb{1}}\zeta^{\bar{\alpha}} = i\dot{\bar{Z}}^{\bar{\alpha}} \qquad\qquad \delta_I \bar{Z}^{\bar{\alpha}} = -i\zeta^{\bar{\alpha}} \qquad \delta_I \zeta^{\bar{\alpha}} = -\dot{\bar{Z}}^{\bar{\alpha}} .$$

(The names chosen for these transformations will appear more natural below.) Prove that they are invariances of the toy action (4.10)*, and that they satisfy the following algebra:*

$$\delta_{\mathbb{1}}^2 = {\delta_I}^2 = i\frac{d}{dt} \qquad \text{and} \qquad \delta_{\mathbb{1}}\delta_I = -\delta_I\delta_{\mathbb{1}} .$$

We can rewrite the second of these supersymmetries in a way that makes its generalisation obvious. If we let I denote the complex structure in $\mathbb{R}^4 = \mathbb{C}^2$ which is diagonalised by our choice of complex coordinates, then the second supersymmetry δ_I can be rewritten as follows:

$$\delta_I Z^\alpha = I^\alpha{}_\beta \zeta^\beta \qquad \delta_I \zeta^\alpha = -i I^\alpha{}_\beta \dot{Z}^\beta$$
$$\delta_I \bar{Z}^{\bar{\alpha}} = I^{\bar{\alpha}}{}_{\bar{\beta}} \zeta^{\bar{\beta}} \qquad \delta_I \zeta^{\bar{\alpha}} = -i I^{\bar{\alpha}}{}_{\bar{\beta}} \dot{\bar{Z}}^{\bar{\beta}} \ ,$$

which explains the notation. It now doesn't take much imagination to write down the remaining two supersymmetries. We simply replace I in turn by each of the other two complex structures J and K of Exercise 3.32. The fact that I, J, and K satisfy the quaternion algebra is instrumental in showing that these transformations obey the right supersymmetry algebra.

Exercise 4.11 (The N=4 supersymmetry algebra)
Let $\delta_{\mathbb{1}}$ and δ_I be the supersymmetries given in Exercise 4.10. Define δ_J and δ_K in the obvious way. Let δ be any of these supersymmetries and $\delta' \neq \delta$ be a second of these supersymmetries. Prove that the following algebra is obeyed:

$$\delta^2 = i\frac{d}{dt} \qquad \text{and} \qquad \delta\delta' + \delta'\delta = 0 \ .$$

This is the N=4 supersymmetry algebra.

4.3.2 N=4 supersymmetry in hyperkähler manifolds

We now abandon our toy model and return to the action L_{eff} given by (4.9). We expect that the supersymmetry $\delta_{\mathbb{1}}$ defined in Exercise 4.10 should also be an invariance of L_{eff} and, given our choice of coordinates, that so should δI. This is because I is diagonal and constant on the chosen basis. In fact, leaving aside for the moment the issue of the invariance of L_{eff} under these transformations, Exercise 4.10 shows that they obey the right supersymmetry algebra. On the other hand, the other two complex structures J and K will not be constant in this basis, and hence the transformations δ_J and δ_K defined above will not obey the supersymmetry algebra. We will have to modify them appropriately.

To see this we will investigate the supersymmetry transformations associated to a a covariantly constant complex structure I. Let us *not* work on a complex basis adapted to I, but rather on some arbitrary basis (X^a, ζ^a). We will attempt to write down a supersymmetry transformation δ using I. Because the δ has order $\frac{1}{2}$ (being essentially a "square root" of d/dt), δX^a is determined up to an inconsequential overall constant:

$$\delta X^a = I_b{}^a \zeta^b \ . \tag{4.11}$$

Computing δ^2 we find

$$\delta^2 X^a = \partial_c I_b{}^a \, I_d{}^c \zeta^d \zeta^b + I_b{}^a \delta\zeta^b \; .$$

If we now use the fact that I is covariantly constant, so that

$$\partial_c I_b{}^a = \Gamma_{cb}{}^d I_d{}^a - \Gamma_{cd}{}^a I_b{}^d \; ,$$

we can solve for $\delta\zeta^b$ by demanding that $\delta^2 X^a = i\dot{X}^a$:

$$\delta\zeta^a = -i I_b{}^a \dot{X}^b - \Gamma_{bc}{}^a I_d{}^b \zeta^d \zeta^c \; , \tag{4.12}$$

where we have discarded a term $-\frac{1}{2} I_b{}^a T_{cd}{}^b I_e{}^c I_f{}^d \zeta^e \zeta^f$ where $T_{cd}{}^b \equiv \Gamma_{cd}{}^b - \Gamma_{dc}{}^b$ is the torsion of the connection, which in our case is zero. In order to show that $\delta^2 = id/dt$ on ζ^a, two approaches present themselves. One can use the fact that $\zeta^a = -I_b{}^a \delta X^b$ and use the fact that on (any function of) X, $\delta^2 = id/dt$:

$$\begin{aligned}
\delta^2\zeta^a &= -\delta^2 (I_b{}^a \delta X^b) \\
&= -i\partial_c I_b{}^a \dot{X}^c \delta X^b - I_b{}^a \delta^2 \delta X^b \\
&= -i\partial_c I_b{}^a \dot{X}^c \delta X^b - I_b{}^a \delta\delta^2 X^b \qquad\qquad \text{since } \delta^2\delta = \delta^3 = \delta\delta^2 \\
&= -i\partial_c I_b{}^a \dot{X}^c \delta X^b - i I_b{}^a \delta\dot{X}^b \\
&= -i\partial_c I_b{}^a \dot{X}^c \delta X^b - i I_b{}^a \dot{(I_c{}^b \zeta^c)} \\
&= -i\partial_c I_b{}^a I_d{}^b \dot{X}^c \zeta^d - i I_b{}^a \partial_d I_c{}^b \dot{X}^d \zeta^c + i\dot{\zeta}^a \\
&= i\dot{\zeta}^a \; ,
\end{aligned}$$

where in the last line we have used an identity resulting from taking the derivative of $I^2 = -\mathbb{1}$. Alternatively, one can compute $\delta^2\zeta^a$ directly. This is naturally left as an exercise.

Exercise 4.12 (Another proof that $\delta^2\zeta^a = i\dot{\zeta}^a$)
By taking δ of $\delta\zeta^a$, prove that $\delta^2\zeta^a = i\dot{\zeta}^a$.
(Hint: You might find it necessary to use two properties of the Riemann curvature tensor:

- $R_{abc}{}^d + R_{bca}{}^d + R_{cab}{}^d = 0$; *and*
- $R_{abc}{}^d I_d{}^e = R_{abd}{}^e I_c{}^d$.

You are encouraged to prove these identities. The first one is the (first) Bianchi identity, the other one follows from the fact that I is covariantly constant, and hence commutes with the curvature operator.)

Let δ_I denote the supersymmetry transformation associated to the complex structure I. If we define $\delta_{\mathbb{1}}$ as above: $\delta_{\mathbb{1}} X^a = \zeta^a$ and $\delta_{\mathbb{1}}\zeta^a = i\dot{X}^a$, then just as before $\delta_{\mathbb{1}}\delta_I = -\delta_I\delta_{\mathbb{1}}$. In other words, $\delta_{\mathbb{1}}$ and δ_I generate an N=2 supersymmetry algebra. Therefore this result holds for any Kähler manifold, and not just for $\mathbb{R}^4$ as in the previous section.

Now let J be a second covariantly constant complex structure. It will give rise to its own supersymmetry transformation given by equations (4.11) and (4.12), but with J replacing I. Let us call this supersymmetry transformation δ_J. *When will δ_I and δ_J (anti)commute?* The next exercise provides the answer.

Exercise 4.13 (Commuting supersymmetries)
Let δ_I and δ_J denote the supersymmetries generated by covariantly constant complex structure I and J. Prove that

$$\delta_I\delta_J X^a = -iI_c{}^b J_b{}^a \dot{X}^c - \Gamma_{bc}{}^a I_d{}^b J_e{}^c \zeta^d\zeta^e \ .$$

Conclude that $(\delta_I\delta_J + \delta_J\delta_I)X^a = 0$ if and only if $IJ = -JI$. Prove that if this is the case, then $(\delta_I\delta_J + \delta_J\delta_I)\zeta^a = 0$ as well, so that the two supersymmetries (anti)commute.
(Hint: To compute $\delta_I\delta_J + \delta_J\delta_I$ on ζ^a, you might find it easier to exhibit $\zeta^a = -I_b{}^a\delta_I X^b$, say, and then use that $\delta_I\delta_J + \delta_J\delta_I$ is zero on (functions of) X^a.)

This means that if $\{X^a\}$ denote the coordinates of a hyperkähler manifold and $\{\zeta^a\}$ are the accompanying fermions, then the four supersymmetries $\delta_{\mathbb{1}}$, δ_I, δ_J and δ_K satisfy an N=4 supersymmetry algebra.

4.3.3 N=4 supersymmetry of L_{eff}

It remains to show that the four supersymmetries defined above, are indeed symmetries of the effective action L_{eff}. This will be easier to ascertain if we first rewrite the supersymmetry transformations (4.11) and (4.12) in complex coordinates adapted to one of the complex structures: I, say.

Let us therefore choose coordinates $(Z^\alpha, \bar{Z}^{\bar{\alpha}}, \zeta^\alpha, \zeta^{\bar{\alpha}})$ adapted to the complex structure I. Because the the metric is hermitian relative to this complex structure (in fact, relative to all three), we can rewrite the equations (4.11) and (4.12) using the results of Exercise 3.23. Because I is constant relative to these basis, δ_I precisely agrees with δ_I in Exercise 4.10.

Now consider the second complex structure J. Because $IJ = -JI$, J maps vectors of type $(0,1)$ relative to I to vectors of type $(1,0)$ and viceversa. In other words, relative to the above basis adapted to I, J has components $J_\alpha{}^{\bar{\beta}}$ and $J_{\bar{\alpha}}{}^{\beta}$. Therefore J generates the following supersymmetry transformation:

$$\delta_J Z^\alpha = i J_{\bar\beta}{}^\alpha \zeta^{\bar\beta} \qquad \delta_J \zeta^\alpha = -i J_{\bar\beta}{}^\alpha \dot{\bar Z}^{\bar\beta} - \Gamma_{\beta\gamma}{}^\alpha J_{\bar\delta}{}^\beta \zeta^{\bar\delta}\zeta^\gamma$$
$$\delta_J \bar Z^{\bar\alpha} = J_\beta{}^{\bar\alpha} \zeta^\beta \qquad \delta_J \zeta^{\bar\alpha} = -i J_\beta{}^{\bar\alpha} \dot Z^\beta - \Gamma_{\bar\beta\bar\gamma}{}^{\bar\alpha} J_\delta{}^{\bar\beta} \zeta^\delta\zeta^{\bar\gamma} ,$$

where we have used (see Exercise 3.23) that $\Gamma_{\alpha\beta}{}^\gamma$ and $\Gamma_{\bar\alpha\bar\beta}{}^{\bar\gamma}$ are the only nonzero components of the Christoffel symbols. Similar formulas hold for δ_K.

As they stand, these supersymmetries are fermionic transformations. We can make them bosonic by introducing an anticommuting parameter ε and defining the following transformations:

$$\begin{aligned}
&\delta^\varepsilon_{\mathbb{1}} Z^\alpha = \varepsilon\zeta^\alpha && \delta^\varepsilon_{\mathbb{1}} \zeta^\alpha = i\varepsilon \dot Z^\alpha \\
&\delta^\varepsilon_{\mathbb{1}} \bar Z^{\bar\alpha} = \varepsilon\zeta^{\bar\alpha} && \delta^\varepsilon_{\mathbb{1}} \zeta^{\bar\alpha} = i\varepsilon \dot{\bar Z}^{\bar\alpha} && (4.13)\\
&\delta^\varepsilon_I Z^\alpha = i\varepsilon\zeta^\alpha && \delta^\varepsilon_I \zeta^\alpha = \dot Z^\alpha \\
&\delta^\varepsilon_I \bar Z^{\bar\alpha} = -i\varepsilon\zeta^{\bar\alpha} && \delta^\varepsilon_I \zeta^{\bar\alpha} = -\varepsilon \dot{\bar Z}^{\bar\alpha} && (4.14)\\
&\delta^\varepsilon_J Z^\alpha = i\varepsilon J_{\bar\beta}{}^\alpha \zeta^{\bar\beta} && \delta^\varepsilon_J \zeta^\alpha = -i\varepsilon J_{\bar\beta}{}^\alpha \dot{\bar Z}^{\bar\beta} - \varepsilon\Gamma_{\beta\gamma}{}^\alpha J_{\bar\delta}{}^\beta \zeta^{\bar\delta}\zeta^\gamma \\
&\delta^\varepsilon_J \bar Z^{\bar\alpha} = \varepsilon J_\beta{}^{\bar\alpha} \zeta^\beta && \delta^\varepsilon_J \zeta^{\bar\alpha} = -i\varepsilon J_\beta{}^{\bar\alpha} \dot Z^\beta - \varepsilon\Gamma_{\bar\beta\bar\gamma}{}^{\bar\alpha} J_\delta{}^{\bar\beta} \zeta^\delta\zeta^{\bar\gamma} && (4.15)\\
&\delta^\varepsilon_K Z^\alpha = i\varepsilon K_{\bar\beta}{}^\alpha \zeta^{\bar\beta} && \delta^\varepsilon_K \zeta^\alpha = -i\varepsilon K_{\bar\beta}{}^\alpha \dot{\bar Z}^{\bar\beta} - \varepsilon\Gamma_{\beta\gamma}{}^\alpha K_{\bar\delta}{}^\beta \zeta^{\bar\delta}\zeta^\gamma \\
&\delta^\varepsilon_K \bar Z^{\bar\alpha} = \varepsilon K_\beta{}^{\bar\alpha} \zeta^\beta && \delta^\varepsilon_K \zeta^{\bar\alpha} = -i\varepsilon K_\beta{}^{\bar\alpha} \dot Z^\beta - \varepsilon\Gamma_{\bar\beta\bar\gamma}{}^{\bar\alpha} K_\delta{}^{\bar\beta} \zeta^\delta\zeta^{\bar\gamma} . && (4.16)
\end{aligned}$$

The task ahead is now straightforward—albeit a little time consuming. Taking each of these supersymmetries in turn, and letting ϵ depend on time, we vary the action L_{eff}. Invariance of the action implies that

$$\delta^\varepsilon L_{\text{eff}} = \dot\varepsilon Q + \dot X ,$$

where X is arbitrary, and Q is the charge generating the supersymmetry. The next exercise summarises the results of this calculation.

Exercise 4.14 (The supersymmetry charges)
Prove that L_{eff} is invariant under the supersymmetries given by equations (4.13)–(4.16)*, with the following associated supersymmetry charges:*

$$\begin{aligned}
Q_{\mathbb{1}} &= g_{\alpha\bar\beta}\zeta^\alpha \dot{\bar Z}^{\bar\beta} + g_{\alpha\bar\beta}\zeta^{\bar\beta}\dot Z^\alpha \\
Q_I &= i g_{\alpha\bar\beta}\zeta^\alpha \dot{\bar Z}^{\bar\beta} - i g_{\alpha\bar\beta}\zeta^{\bar\beta}\dot Z^\alpha \\
Q_J &= J_{\alpha\beta}\zeta^\alpha \dot Z^\beta + J_{\bar\alpha\bar\beta}\zeta^{\bar\alpha}\dot{\bar Z}^{\bar\beta} \\
Q_K &= K_{\alpha\beta}\zeta^\alpha \dot Z^\beta + K_{\bar\alpha\bar\beta}\zeta^{\bar\alpha}\dot{\bar Z}^{\bar\beta} ,
\end{aligned}$$

where $J_{\alpha\beta} = J_\alpha{}^{\bar\gamma} g_{\beta\bar\gamma}$ *and* $J_{\bar\alpha\bar\beta} = J_{\bar\alpha}{}^{\gamma} g_{\bar\beta\gamma}$*, and similarly for* K*.*
(Hint: The calculation uses the two fundamental identities described in the hint to Exercise 4.12. In complex coordinates, and using the results of Exercise 3.23, they now look as follows:

$$\begin{aligned} R_{\alpha\bar\beta\gamma}{}^{\delta} &= R_{\gamma\bar\beta\alpha}{}^{\delta} & R_{\alpha\bar\beta\bar\gamma}{}^{\bar\delta} &= R_{\alpha\bar\gamma\bar\beta}{}^{\bar\delta} \\ R_{\alpha\bar\beta\gamma}{}^{\epsilon} J_{\epsilon}{}^{\bar\delta} &= R_{\alpha\bar\beta\bar\epsilon}{}^{\bar\delta} J_{\gamma}{}^{\bar\epsilon} & R_{\alpha\bar\beta\bar\gamma}{}^{\bar\epsilon} J_{\bar\epsilon}{}^{\delta} &= R_{\alpha\bar\beta\epsilon}{}^{\delta} J_{\bar\gamma}{}^{\epsilon} \ , \end{aligned}$$

and similarly for K*.)*

Having established the $N{=}4$ supersymmetry of the effective action L_{eff} it is now time to quantise the theory. It turns out that supersymmetry will provide a geometric description of the Hilbert space and of the hamiltonian. This will require some basic concepts of harmonic theory on Kähler manifolds. The purpose of the next section is to provide a brief review for those who are not familiar with this topic.

4.4 A brief review of harmonic theory

This section contains a brief scholium on the harmonic theory of orientable riemannian manifolds and in particular of Kähler manifolds. The reader familiar with these results can easily skip this section. Other readers are encouraged to read on. We will of necessity be brief: details can be found in many fine books on the subject [Gol62, GH78, War83, Wel80].

4.4.1 Harmonic theory for riemannian manifolds

Let M be a smooth manifold. We will let $\mathcal{E} = \bigoplus_p \mathcal{E}^p$ denote the algebra of differential forms on M. The de Rham operator $d : \mathcal{E}^p \to \mathcal{E}^{p+1}$ obeys $d^2 = 0$ and hence one can define its cohomology (*the de Rham cohomology of* M) as follows:

$$H^p_{\text{dR}}(M) = \frac{\ker d : \mathcal{E}^p \to \mathcal{E}^{p+1}}{\operatorname{im} d : \mathcal{E}^{p-1} \to \mathcal{E}^p} \ .$$

In other words, the p-th de Rham cohomology is a vector space whose elements are equivalence classes of *closed* p-forms ($d\omega = 0$)—two closed p-forms ω_1 and ω_2 being equivalent if their difference is *exact*: $\omega_1 - \omega_2 = d\theta$, for some $(p-1)$-form θ. The crown jewel of harmonic theory is the decomposition theorem of Hodge, which states that if M is a compact orientable manifold there exists a privileged representative for each de Rham cohomology class. This representative is obtained by introducing more structure on M—namely a riemannian metric. From the above definition, it is clear that the de Rham cohomology does not depend on any geometric properties of the manifold. It is precisely this reason why the Hodge theorem is of fundamental importance: because it establishes a link between the topological and the geometric properties of riemannian manifolds. (Actually, the fact that the de Rham cohomology is a topological invariant of M is not obvious. It is called the de Rham theorem and it is proven in [War83, BT81].)

We therefore let (M, g) be an m-dimensional orientable riemannian manifold. Let $\{e^i\}$ for $i = 1, \ldots, m$ be a local orthonormal basis for the 1-forms. Orthonormality means that the line element is locally $ds^2 = \sum_i e^i \otimes e^i$. In general such a basis will of course not exist globally, but will transform under a local $O(m)$ transformation when we change coordinate charts. In this basis the volume form is given by $\mathrm{vol} = e^1 \wedge e^2 \wedge \cdots \wedge e^m$. This volume form defines a local orientation in M. Orientability simply means that, unlike the 1-forms $\{e^i\}$, the volume form—and hence the orientation—does exist globally. It also means that upon changing charts, the $\{e^i\}$ will change by a local $SO(m)$ transformation. (Prove this!)

A local basis for the differential forms $\mathcal{E}$ on M is given by wedge products of these 1-forms. It is convenient to introduce *multi-indices* $I = (i_1, i_2, \ldots, i_p)$ where $1 \leq i_1 < i_2 < \cdots < i_p \leq m$. We say that I has length p or that $|I| = p$. We then define $e^I \equiv e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p}$. In this notation, $\{e^I | I, |I| = p\}$ is a local basis for $\mathcal{E}^p$; that is, any p-form ω on M can be written locally like $\sum_{|I|=p} \omega_I e^I$, where the coefficients ω_I are smooth functions. If $I = (i_1, i_2, \ldots, i_p)$ is a multi-index of length p, we let $I^c = (i_{p+1}, i_{p+2}, \ldots, i_m)$ denote the multi-index of length $m - p$ uniquely defined by the fact that $\{i_1, \ldots, i_p\} \cup \{i_{p+1}, \ldots, i_m\} = \{1, 2, \ldots, m\}$.

We can now define the Hodge $\star$-operation. This is a linear map $\star : \mathcal{E}^p \to \mathcal{E}^{m-p}$ defined by

$$\star e^I = \operatorname{sign}\sigma \; e^{I^c} \;,$$

where if $I = (i_1, \ldots, i_p)$ and $I^c = (i_{p+1}, \ldots, i_m)$, then $\operatorname{sign}\sigma$ is the sign of the permutation

$$\sigma = \begin{pmatrix} i_1 & i_2 & \cdots & i_m \\ 1 & 2 & \cdots & m \end{pmatrix} \;.$$

In particular $\star 1 = \mathrm{vol}$. The following result is important for calculations.

Exercise 4.15 (The square of the Hodge $\star$)
Prove that acting on $\mathcal{E}^p$, $\star^2 = (-)^{(m-p)p}$.

The Hodge $\star$-operator allows us to define a pointwise metric $\langle -, - \rangle$ on forms as follows:

$$\alpha \wedge \star\beta = \langle \alpha, \beta \rangle \,\mathrm{vol} \;. \tag{4.17}$$

The properties of this pointwise metric are summarised in the following exercise.

Exercise 4.16 (The pointwise metric on forms)
Prove that the basis $\{e^I\}$ is orthonormal relative to the pointwise metric defined in (4.17) and therefore that it agrees on 1-forms with the one induced by the riemannian metric on M. Conclude that the pointwise metric is positive-definite.

If, in addition, M is compact we can define an honest metric (called the *Hodge metric*) on forms by integrating the pointwise metric over the manifold relative to the volume form:

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle \operatorname{vol} = \int_M \alpha \wedge \star\beta \ .$$

If M is not compact, then we can restrict ourselves to compactly supported forms or to forms α for which the *Hodge norm* $\|\alpha\|^2 \equiv (\alpha, \alpha)$ is finite. Such forms are often called *square-integrable*.

The Hodge metric allows us to define the adjoint d^* to the de Rham operator, with which the following exercise concerns itself.

Exercise 4.17 (The adjoint de Rham operator)
Define the adjoint d^* of the de Rham operator by

$$(d\alpha, \beta) = (\alpha, d^*\beta) \ ,$$

for all forms $\alpha, \beta \in \mathcal{E}$. Prove that d^* satisfies the following properties:

(1) $d^* : \mathcal{E}^p \to \mathcal{E}^{p+1}$;

(2) $(d^*)^2 = 0$; *and*

(3) $d^* = (-)^{mp+m+1} \star d \star$ *acting on* $\mathcal{E}^p$.

Now let us define the *Hodge laplacian* $\triangle : \mathcal{E}^p \to \mathcal{E}^p$ by $\triangle \equiv dd^* + d^*d$. We say that a p-form is *harmonic* if $\triangle\alpha = 0$.

Exercise 4.18 (Harmonic forms)
Prove that a form α is harmonic if and only if $d\alpha = d^*\alpha = 0$. Prove that harmonic forms have minimal Hodge norm in their cohomology class. That is, if α is harmonic, then prove that the Hodge norm of $\alpha + d\beta$ is strictly greater than that of α.

The Hodge decomposition theorem states that in a compact orientable manifold each de Rham cohomology class has a unique harmonic representative; that is, that there is a vector space isomorphism

$$H^p_{\mathrm{dR}}(M) \cong \text{harmonic } p\text{-forms} \ .$$

The proof of this theorem is rather technical. The idea is to use the norm-minimising property to define the harmonic representative; but one then has to prove that this form is smooth. This calls for the use of regularity theorems which are beyond the scope of these notes. A proof can be found, for example, in [War83].

The Hodge decomposition theorem has a very important corollary, which the following exercise asks you to prove.

Exercise 4.19 (Poincaré duality)
Prove that the Hodge $\star$-operator commutes with the Hodge laplacian. Use the Hodge decomposition theorem to conclude that for M an m-dimensional compact orientable manifold, there is an isomorphism

$$H^p_{\mathrm{dR}}(M) \cong H^{m-p}_{\mathrm{dR}}(M) \ .$$

This isomorphism is known as Poincaré duality.

4.4.2 Harmonic theory for Kähler manifolds

Now suppose that M is a complex manifold of complex dimension n. As explained in section 3.3.2, on a complex manifold one has local coordinates $(z^\alpha, \bar{z}^{\bar{\beta}})$, where $\alpha, \beta = 1, 2, \ldots, n$. This allows us to refine the grading of the complex differential forms. We say that a complex differential form ω is of type (p,q) if it can be written in local complex coordinates as

$$\omega = \omega_{\alpha_1 \cdots \alpha_p \bar{\beta}_1 \cdots \bar{\beta}_q}(z,\bar{z}) dz^{\alpha_1} \wedge \cdots \wedge dz^{\alpha_p} \wedge d\bar{z}^{\bar{\beta}_1} \wedge \cdots \wedge d\bar{z}^{\bar{\beta}_q} \ ,$$

where $\omega_{\alpha_1 \cdots \alpha_p \bar{\beta}_1 \cdots \bar{\beta}_q}(z,\bar{z})$ are smooth functions. The algebra of complex differential forms is then bigraded as follows:

$$\mathcal{E} = \bigoplus_{0 \leq p,q \leq n} \mathcal{E}^{p,q} \ , \tag{4.18}$$

where $\mathcal{E}^{p,q}$ is the space of (p,q)-forms.

The de Rham operator d also breaks up into a type (1,0) piece and a type (0,1) piece:

$$d = \partial + \bar{\partial} \qquad \text{where} \qquad \begin{array}{l} \partial : \mathcal{E}^{p,q} \to \mathcal{E}^{p+1,q} \\ \bar{\partial} : \mathcal{E}^{p,q} \to \mathcal{E}^{p,q+1} \end{array} .$$

Breaking $d^2 = 0$ into types we find that $\partial^2 = \bar{\partial}^2 = \partial\bar{\partial} + \bar{\partial}\partial = 0$. We call $\bar{\partial}$ the *Dolbeault* operator, and its cohomology

$$H^{p,q}_{\bar{\partial}}(M) = \frac{\ker \bar{\partial} : \mathcal{E}^{p,q} \to \mathcal{E}^{p,q+1}}{\mathrm{im}\, \bar{\partial} : \mathcal{E}^{p,q-1} \to \mathcal{E}^{p,q}}$$

the *Dolbeault cohomology*.

Now suppose that we give M a hermitian metric h; that is, a riemannian metric compatible with the complex structure: $h(IX, IY) = h(X, Y)$. Such metrics always exist: one simply takes any riemannian metric g, say, and averages it over the finite group generated by I: $h(X, Y) \equiv \frac{1}{2}g(X, Y) + \frac{1}{2}g(IX, IY)$. If we forget the complex structure for a moment, M is an orientable riemannian manifold of (real) dimension $2n$. Therefore we have a Hodge $\star$-operator defined as in the previous section. The next exercise asks you to show how $\star$ interacts with the complex structure.

Exercise 4.20 (The Hodge $\star$ and the complex structure)
Prove that the Hodge $\star$-operator maps (p, q)-forms to $(n - q, n - p)$-forms:

$$\star : \mathcal{E}^{p,q} \to \mathcal{E}^{n-q,n-p} \ ,$$

and that acting on (p, q)-forms, $\star^2 = (-)^{p+q}$.
(Hint: The first part is computationally quite involved, but the idea is easy. We can always find a local basis $\{\theta^i\}$ for the (1,0)-forms on M such that the line element (relative to the hermitian metric) has the form

$$ds^2 = \sum_{i=1}^{n} \left(\theta^i \otimes \bar\theta^i + \bar\theta^i \otimes \theta^i\right) \ ,$$

where $\{\bar\theta^i\}$ are the complex conjugate (0,1)-forms. We can decompose these forms into their real and imaginary parts as follows: $\theta^j = \frac{1}{\sqrt{2}}\left(e^{2j-1} + ie^{2j}\right)$ and $\bar\theta^j = \frac{1}{\sqrt{2}}\left(e^{2j-1} - ie^{2j}\right)$. In terms of these real 1-forms, the line element becomes $ds^2 = \sum_{j=1}^{2n} e^j \otimes e^j$; in other words, they form an orthonormal basis. Therefore we know the action of the Hodge $\star$-operator on the $\{e^I\}$. Your mission, should you decide to accept it, is to find out what it is in terms of the $\theta^I \wedge \bar\theta^J$. The second part simply uses Exercise 4.15.)

Another operation that we have on a complex manifold is complex conjugation, which exchanges (p, q)-forms with (q, p)-forms. Using the Hodge $\star$-operator and complex conjugation we can define a pointwise hermitian metric for the complex forms, also denoted $\langle -, -\rangle$ as in the real case treated in the previous section. This metric is defined by

$$\alpha \wedge \star\bar\beta = \langle \alpha, \beta\rangle \,\mathrm{vol} \ .$$

Notice that relative to this metric, the decomposition in equation (4.18) is orthogonal: if β is a (p, q)-form, then $\bar\beta$ is a (q, p)-form, and $\star\bar\beta$ is a $(n - p, n - q)$-form. The only way one can obtain the volume form, which is an (n, n)-form, is to wedge with another (p, q)-form.

Exercise 4.21 (The pointwise hermitian metric)
Prove that the basis $\{\theta^I \wedge \bar\theta^J\}$ is orthonormal relative to the pointwise hermitian metric, and conclude that it is positive-definite.

If M is compact, we can then integrate this pointwise metric relative to the volume form and define an honest hermitian metric on complex forms:

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle \,\mathrm{vol} = \int_M \alpha \wedge \star\bar\beta \ . \tag{4.19}$$

This metric is again called the *Hodge metric*. As in the real case, if M is not compact, then we can still make sense of this provided we restrict our attention to square-integrable forms.

It follows from Exercise 4.17 that on a complex manifold, $d^* = -\star d \star$ regardless on which forms it is acting. Breaking d^* into types we find

$$d^* = \partial^* + \bar\partial^* \qquad \text{where} \qquad \begin{aligned} \partial^* &: \mathcal{E}^{p,q} \to \mathcal{E}^{p-1,q} \\ \bar\partial^* &: \mathcal{E}^{p,q} \to \mathcal{E}^{p,q-1} \end{aligned} \ .$$

On the other hand, breaking $-\star d \star$ into types, and comparing we see that

$$\partial^* = -\star \bar\partial \star \qquad \text{and} \qquad \bar\partial^* = -\star \partial \star \ .$$

We can therefore define two laplacian operators:

$$\square = \partial\partial^* + \partial^*\partial \qquad \text{and} \qquad \bar\square = \bar\partial\bar\partial^* + \bar\partial^*\bar\partial \ ,$$

both of which map $\mathcal{E}^{p,q} \to \mathcal{E}^{p,q}$.

Just as for de Rham cohomology, there is a Hodge decomposition theorem for Dolbeault cohomology, which says that the $\bar\partial$-cohomology on (p,q)-forms is isomorphic to the space of $\bar\square$-harmonic (p,q)-forms:

$$H^{p,q}_{\bar\partial} \cong \bar\square\text{-harmonic } (p,q)\text{-forms} \ .$$

In a generic complex manifold there is no reason to expect any relation between the Dolbeault laplacians and the Hodge laplacian $\triangle = d^*d + dd^*$; but the magic of Kähler geometry is that if M is Kähler, then

$$\boxed{\triangle = 2\square = 2\bar\square} \ . \tag{4.20}$$

This is not a hard result to obtain, but it requires quite a bit of formalism that we will not need in the remainder of this course, hence we leave it unproven and refer the interested reader to the literature [Gol62, GH78, Wel80].

As an immediate corollary of equation (4.20) and of the Hodge decomposition theorems for de Rham and Dolbeault cohomologies, we have

$$H^r_{\mathrm{dR}}(M) \cong \bigoplus_{p+q=r} H^{p,q}_{\bar\partial}(M) \;;$$

and the following exercise describes another immediate corollary of equation (4.20).

Exercise 4.22 (Serre duality)
Prove that both the Hodge $\star$-operator and complex conjugation commute with the laplacian. Use this to conclude that for M a compact Kähler manifold of complex dimension n, there exist isomorphisms:

$$H^{p,q}_{\bar\partial}(M) \cong H^{n-q,n-p}_{\bar\partial}(M) \cong H^{n-p,n-q}_{\bar\partial}(M) \;.$$

These isomorphisms are known collectively as Serre duality.

Finally, a curiosity. If we define the r-th Betti number b_r of a manifold as the real dimension of the r-th de Rham cohomology, we have as a consequence of Serre duality that for a compact Kähler manifold *all the odd Betti numbers are even.*

4.4.3 Explicit formulas for $\bar\partial$ and $\bar\partial^*$

The purpose of this section is simply to derive some explicit expressions for the differential operators $\bar\partial$ and $\bar\partial^*$. These are the expressions by which we will be able to recognise them when we quantise the effective action. Throughout this section M shall be a Kähler manifold of complex dimension n.

Let us first start by deriving formulas for d and d^*. For this we can forget momentarily the complex structure and think of M simply as an orientable riemannian manifold of dimension $2n$. Let $\{e_i\}$ denote a local orthonormal basis for the vector fields, and let $\{e^i\}$ denote the canonical dual basis for the 1-forms. They are also orthonormal relative to the induced metric. Let ∇ denote the Levi-Civita connection. We claim that d can be written as

$$d = \sum_{i=1}^{2n} e^i \wedge \nabla_{e_i} \;. \tag{4.21}$$

Proving this will be the purpose of the following exercise.

Exercise 4.23 (An explicit expression for d)
Let $d_?$ denote the right-hand side of equation (4.21).

1. *Prove that $d_?$ is independent of the orthonormal basis chosen so that it is well-defined.*

2. *Let $e^i = e^i_a dx^a$ and $e_i = e^a_i \partial_a$. Prove that $d_? = \sum_a dx^a \wedge \partial_a$, and conclude that $d_? = d$.*

(Hint: Use the fact that the Levi-Civita connection is torsionless.)

With this result we can now describe a similar formula for d^*. Letting $\{e_i\}$ and $\{e^i\}$ be as above, we will prove that

$$d^* = -\sum_{i=1}^{2n} \iota(e^i)\, \nabla_{e_i} \ , \tag{4.22}$$

where $\iota(e^i)$ is the contraction operation, defined by:

1. $\iota(e^i)\, f = 0$ for f a function;

2. $\iota(e^i)\, e^j = \delta_{ij}$; and

3. $\iota(e^i)\, (e^j \wedge \omega) = \delta_{ij}\omega - e^j \wedge \iota(e^i)\, \omega$.

The next exercise asks you to prove equation (4.22).

Exercise 4.24 (An explicit expression for d^*)
Let $d^*_?$ stand for the right-hand side of equation (4.22). Prove that $d^*_? = - \star d \star$, whence it agrees with d^*. (We are using (3) in Exercise 4.17, with $m = 2n$.)
(Hint: Prove first that $d^*_?$ is well-defined; that is, it is independent of the choice of orthonormal frame. Because of this and by linearity, conclude that it is sufficient to compare $d^*_?$ and $- \star d \star$ on a p-form of the form $f e^1 \wedge e^2 \wedge \cdots \wedge e^p$. Moreover argue that it is sufficient to compute this at a point where $\nabla_{e_i} e^j = 0$. Then do it.)

As a corollary of the previous exercise, it follows that relative to a coordinate basis, we can write

$$d^* = -\sum_a \iota(dx^a)\, \partial_a \ , \tag{4.23}$$

where now $\iota(dx^a)\, dx^b = g^{ab}$.

We can now re-introduce the complex structure. Let $\{\theta_i, \bar\theta_i\}$ be a complex basis for the complex vector fields, and let $\{\theta^i, \bar\theta^i\}$ be the canonical dual basis for the complex 1-forms. In terms of the above basis $\{e^i\}$, θ^i is given as in Exercise 4.20. The canonical dual basis for the vector fields are related by

$$\theta_i = \frac{1}{\sqrt{2}}\left(e_{2i-1} - i e_{2i}\right) \qquad\qquad \bar\theta_i = \frac{1}{\sqrt{2}}\left(e_{2i-1} + i e_{2i}\right) \ .$$

Inverting this change of basis, and using equations (4.21) and (4.22), we find

$$d = \sum_{i=1}^{n} \left(\theta^i \wedge \nabla_{\theta_i} + \bar\theta^i \wedge \nabla_{\bar\theta_i} \right) ,$$

and

$$d^* = - \sum_{i=1}^{n} \left(\iota(\theta^i)\, \nabla_{\theta_i} + \iota(\bar\theta^i)\, \nabla_{\bar\theta_i} \right) .$$

Breaking up into types, one concludes that

$$\bar\partial = \sum_{i=1}^{n} \bar\theta^i \wedge \nabla_{\bar\theta_i} \qquad \text{and} \qquad \bar\partial^* = - \sum_{i=1}^{n} \iota(\theta^i)\, \nabla_{\theta_i} .$$

Or in a coordinate basis,

$$\bar\partial = \sum_{\bar\alpha=1}^{n} d\bar z^{\bar\alpha} \wedge \partial_{\bar\alpha} \qquad \text{and} \qquad \bar\partial^* = - \sum_{\alpha=1}^{n} \iota(dz^\alpha)\, \partial_\alpha . \tag{4.24}$$

These equations will be important in the sequel.

4.5 Quantisation of the effective action

In this section we discuss the canonical quantisation of the effective action (4.8). We will be able to identify the Hilbert space with the square-integrable $(0,q)$-forms on the moduli space $\mathcal{M}_k$. We will exploit the supersymmetry to write the hamiltonian as the anticommutator of supersymmetry charges which, under the aforementioned isomorphism, will be identified as the Dolbeault operator $\bar\partial$ and its adjoint under the Hodge metric. This will then allow us to identify the ground states of the effective quantum theory as the harmonic $(0,q)$-forms on the moduli space.

4.5.1 Canonical analysis

The first step in this direction is to find the expression for the canonical momenta. Then we write the hamiltonian and the supersymmetry charges in terms of the momenta. We write down the Poisson brackets and make sure that the classical algebra is indeed the $N{=}4$ supersymmetry algebra. Most of these calculations are routine, and are therefore left as exercises.

The first exercise starts you in this path by asking you to compute the canonical momenta.

Exercise 4.25 (The canonical momenta)
Prove that the canonical momenta defined by L_{eff} *take the following form:*

$$\begin{aligned}
P_\alpha &= \frac{\partial L_{\text{eff}}}{\partial \dot{Z}^\alpha} = g_{\alpha\bar\beta}\dot{\bar{Z}}^{\bar\beta} + i\Gamma_{\alpha\beta\bar\gamma}\zeta^{\bar\gamma}\zeta^\beta\\
\bar{P}_{\bar\alpha} &= \frac{\partial L_{\text{eff}}}{\partial \dot{\bar{Z}}^{\bar\alpha}} = g_{\bar\alpha\beta}\dot{Z}^\beta\\
\pi_\alpha &= \frac{\partial L_{\text{eff}}}{\partial \dot\zeta^\alpha} = -ig_{\alpha\bar\beta}\zeta^{\bar\beta}\\
\pi_{\bar\alpha} &= \frac{\partial L_{\text{eff}}}{\partial \dot\zeta^{\bar\alpha}} = 0 \ ,
\end{aligned}$$

where $\Gamma_{\alpha\beta\bar\gamma} = \Gamma_{\alpha\beta}{}^\delta g_{\delta\bar\gamma}$.

The fact that $\pi_{\bar\alpha} = 0$ is not very important. It is simply a consequence of the fact that the fermionic part of the effective lagrangian is already in first order form, so that morally speaking $\{\zeta^{\bar\alpha}\}$ play the role of momenta while $\{\zeta^\alpha\}$ are coordinates.

The effective hamiltonian H_{eff} is defined as usual by:

$$H_{\text{eff}} = \dot{Z}^\alpha P_\alpha + \dot{\bar{Z}}^{\bar\alpha}\bar{P}_{\bar\alpha} + \dot\zeta^\alpha\pi_\alpha - L_{\text{eff}} \ .$$

The next exercise asks you to compute it.

Exercise 4.26 (The effective hamiltonian)
Prove that the effective hamiltonian is given by

$$H_{\text{eff}} = g^{\alpha\bar\beta}P_\alpha\bar{P}_{\bar\beta} + g^{\alpha\bar\beta}\Gamma_{\alpha\gamma}{}^\delta\bar{P}_{\bar\beta}\pi_\delta\zeta^\gamma \ .$$

Next we write the supersymmetry charges obtained in Exercise 4.14 in terms of momenta. This is another easy exercise.

Exercise 4.27 (The supersymmetry charges revisited)
Prove that the supersymmetry charges obtained in Exercise 4.14 have the following form:

$$\begin{aligned}
Q_{\mathbb{1}} &= \zeta^\alpha P_\alpha + \zeta^{\bar\alpha}\bar{P}_{\bar\alpha} = \zeta^\alpha P_\alpha + ig^{\alpha\bar\beta}\pi_\alpha\bar{P}_{\bar\beta}\\
Q_I &= i\zeta^\alpha P_\alpha - i\zeta^{\bar\alpha}\bar{P}_{\bar\alpha} = i\zeta^\alpha P_\alpha + g^{\alpha\bar\beta}\pi_\alpha\bar{P}_{\bar\beta}\\
Q_J &= J_\alpha{}^{\bar\alpha}\zeta^\alpha\bar{P}_{\bar\alpha} + J_{\bar\alpha}{}^\alpha\zeta^{\bar\alpha}P_\alpha - iJ_{\bar\alpha}{}^\alpha\Gamma_{\alpha\beta\bar\gamma}\zeta^{\bar\alpha}\zeta^{\bar\gamma}\zeta^\beta\\
Q_K &= K_\alpha{}^{\bar\alpha}\zeta^\alpha\bar{P}_{\bar\alpha} + K_{\bar\alpha}{}^\alpha\zeta^{\bar\alpha}P_\alpha - iK_{\bar\alpha}{}^\alpha\Gamma_{\alpha\beta\bar\gamma}\zeta^{\bar\alpha}\zeta^{\bar\gamma}\zeta^\beta \ .
\end{aligned}$$

The canonical Poisson brackets are defined to be the following:

$$\{P_\alpha, Z^\beta\} = \delta_\alpha^\beta \qquad \{\bar{P}_{\bar\alpha}, \bar{Z}^{\bar\beta}\} = \delta_{\bar\alpha}^{\bar\beta} \qquad \{\pi_\alpha, \zeta^\beta\} = \delta_\alpha^\beta \ .$$

Exercise 4.28 (Some checks)
As a check on our calculations, show that the supersymmetry transformations given in equations (4.13)–(4.16) *are indeed generated via Poisson brackets by the supersymmetry charges computed in the previous exercise.*

Finally, we are ready to verify that we have a classical realisation of the N=4 supersymmetry algebra. Let $q = q_1\, i + q_2\, j + q_3\, k + q_4 \in \mathbb{H}$ be a quaternion. Let $Q_q \equiv q_1\, Q_I + q_2\, Q_J + q_3\, Q_K + q_4\, Q_{\mathbb{1}}$. The next exercise asks you to prove that the supersymmetry charges obey the N=4 supersymmetry algebra.

Exercise 4.29 (Classical N=4 supersymmetry algebra)
Let $q, q' \in \mathbb{H}$ be quaternions. Then prove that under Poisson bracket:

$$\{Q_q, Q_{q'}\} = i(\bar{q}q')H_{\text{eff}} \ ,$$

where $\bar{q} = -q_1\, i - q_2\, j - q_3\, k + q_4$ is the conjugate quaternion and $\bar{q}q' = \sum_i q_i q_i'$ is the quaternionic product.

Because the supersymmetry charges generate under Poisson bracket the supersymmetry transformations, the above exercise implies that the effective hamiltonian indeed generates time translation. If you feel up to it you can check this directly from the expression of the hamiltonian.

4.5.2 The quantisation of the effective hamiltonian

To quantise the effective hamiltonian we first need to identify the Hilbert space. Let us quickly quantise the bosons. We choose to realise Z^α and $\bar{Z}^{\bar{\alpha}}$ as multiplication operators and hence P_α and $\bar{P}_{\bar{\alpha}}$ will be realised as derivatives:

$$P_\alpha \mapsto -i\frac{\partial}{\partial Z^\alpha} \qquad \text{and} \qquad \bar{P}_{\bar{\alpha}} \mapsto -i\frac{\partial}{\bar{Z}^{\bar{\alpha}}} \ .$$

For the fermions, we notice that the canonical Poisson brackets can be rewritten in terms of ζ^α and $\zeta^{\bar{\alpha}}$ as follows:

$$\{\zeta^\alpha, \zeta^{\bar{\beta}}\} = i g^{\alpha\bar{\beta}} \ .$$

Upon quantisation this gives rise to the following anticommutation relations

$$\zeta^\alpha \zeta^{\bar{\beta}} + \zeta^{\bar{\beta}} \zeta^\alpha = g^{\alpha\bar{\beta}} \ ,$$

with all other anticommutators vanishing. Of course, $g^{\alpha\bar{\beta}}$ is a function of $Z^\alpha, \bar{Z}^{\bar{\alpha}}$; but for each point $(Z^\alpha, \bar{Z}^{\bar{\alpha}})$ in $\mathcal{M}_k$, the above anticommutation relations define a Clifford algebra. In other words, this defines a *Clifford bundle* on $\mathcal{M}_k$. Fixing a point in $\mathcal{M}_k$, we have a standard Clifford algebra of the type studied in section 2.2.2. It has a unique irreducible representation constructed as follows. We choose a Clifford vacuum $|\Omega\rangle$, defined by the condition

$$\zeta^\alpha|\Omega\rangle = 0 \qquad \text{for all } \alpha \ .$$

The representation is then built on $|\Omega\rangle$ by acting with the $\zeta^{\bar\alpha}$.

We now tensor together the representations of the bosons and the fermions and what we have is linear combinations of objects of the form

$$f(Z,\bar Z)\zeta^{\bar\alpha}\zeta^{\bar\beta}\cdots\zeta^{\bar\gamma}|\Omega\rangle \ .$$

If we take the $f(Z,\bar Z)$ smooth, this space is clearly isomorphic to the space $\oplus_{0\leq p\leq 2k}\mathcal{E}^{0,p}$ of differential forms of type $(0,p)$ on $\mathcal{M}_k$:

$$f(Z,\bar Z)\zeta^{\bar\alpha}\zeta^{\bar\beta}\cdots\zeta^{\bar\gamma}|\Omega\rangle \leftrightarrow f(Z,\bar Z)d\bar Z^{\bar\alpha}\wedge d\bar Z^{\bar\beta}\wedge\cdots\wedge d\bar Z^{\bar\gamma} \ .$$

Of course the Hilbert space will consist of (the completion of) the subspace formed by those forms which are square integrable relative to a suitable inner product. As we saw in section 4.4.2, the natural inner product to consider is the Hodge metric given by (4.19). Therefore we have the following geometric interpretation of the Hilbert space $\mathcal{H}$ of the quantum effective theory:

$$\boxed{\mathcal{H} \cong \bigoplus_{0\leq p\leq 2k} \mathcal{E}^{0,p}_{L^2}} \ , \tag{4.25}$$

where $\mathcal{E}^{0,p}_{L^2}$ denotes the space of $(0,p)$-forms on $\mathcal{M}_k$ with finite Hodge norm; that is, square-integrable.

In order to identify the hamiltonian we will use supersymmetry. The expressions for the supersymmetry charges and the hamiltonian, being polynomial, suffer from ordering ambiguities. One way to get around this problem is to define the quantisation in a way that the $N{=}4$ supersymmetry algebra is realised quantum-mechanically, and in such a way that we can identify the resulting operators geometrically. The hamiltonian can be defined as the square of any of the supersymmetry charges, but we find it more convenient to take complex linear combinations of $Q_{\mathbb{1}}$ and Q_I. Indeed, let us define

$$Q = i\tfrac{1}{2}\left(Q_{\mathbb{1}} + iQ_I\right) \qquad \text{and} \qquad Q^* = -i\tfrac{1}{2}\left(Q_{\mathbb{1}} - iQ_I\right) \ .$$

The classical expressions for these charges are very simple

$$Q = i\zeta^{\bar\alpha}\bar P_{\bar\alpha} \qquad \text{and} \qquad Q^* = -i\zeta^\alpha P_\alpha \ ,$$

and they obey the following algebra

$$\{Q, Q^*\} = i\tfrac{1}{2} H_{\text{eff}} \ . \tag{4.26}$$

The quantisation is now clear. Quantise the charges Q and Q^* as follows:

$$Q \mapsto \zeta^{\bar\alpha} \frac{\partial}{\partial \bar Z^{\bar\alpha}} \qquad \text{and} \qquad Q^* \mapsto -\zeta^{\alpha} \frac{\partial}{\partial Z^{\alpha}} \ . \tag{4.27}$$

But notice that we have seen these operators before. Indeed, acting on $f \equiv f_{\bar\alpha\bar\beta\cdots\bar\gamma} \zeta^{\bar\alpha}\zeta^{\bar\beta}\cdots\zeta^{\bar\gamma}|\Omega\rangle$, we find that

$$Qf = \partial_{\bar\delta} f_{\bar\alpha\bar\beta\cdots\bar\gamma} \zeta^{\bar\delta}\zeta^{\bar\alpha}\zeta^{\bar\beta}\cdots\zeta^{\bar\gamma}|\Omega\rangle \ .$$

Under the isomorphism (4.25), this corresponds to the form $\bar\partial f$. In other words, $Q \mapsto \bar\partial$.

How about Q^*? Acting on a $(0,0)$-form, Q^* is zero, since ζ^α annihilates the Clifford vacuum. Acting on a $(0,1)$-form $f_{\bar\alpha}\zeta^{\bar\alpha}|\Omega\rangle$, we find

$$Q^* f_{\bar\alpha}\zeta^{\bar\alpha}|\Omega\rangle = -\partial_\beta f_{\bar\alpha} g^{\bar\alpha\beta}|\Omega\rangle \ .$$

In other words, up to a sign, it is given by the divergence. This fact persists to higher $(0,p)$-forms. Indeed, the next exercise asks you to show that $Q^* = \bar\partial^*$, the adjoint of $\bar\partial$ under the Hodge metric.

Exercise 4.30 (Q^* is $\bar\partial^*$)
Show that under the isomorphism (4.25)*, the quantisation of Q^* given by* (4.27) *agrees with $\bar\partial^* = - \star \partial \star$, the adjoint of $\bar\partial$ under the Hodge metric.*
(Hint: Compare with equation (4.24)*.)*

Finally, we quantise the hamiltonian by demanding that the N=4 supersymmetry algebra be preserved quantum-mechanically. In other words, and taking into account equation (4.26), we quantise the hamiltonian as follows:

$$H_{\text{eff}} \mapsto 2\left(QQ^* + Q^*Q\right) \ .$$

Under the identification $Q \leftrightarrow \bar\partial$ and $Q^* \leftrightarrow \bar\partial^*$, the quantum effective hamiltonian agrees with twice the Dolbeault laplacian $\bar\Box$ or—since $\mathcal{M}_k$ is (hyper)Kähler—with the Hodge laplacian $\triangle$.

This result doesn't just provide a beautiful geometric interpretation of the effective quantum theory, but also allows us to use geometric information to derive physical results. For example, the ground states of the theory will be in one-to-one correspondence with (square-integrable) harmonic $(0,p)$-forms. This sort of reasoning will play a crucial role in the test of Montonen–Olive duality in N=4 supersymmetric Yang–Mills theory, which shall be the focus of the next chapter.

Chapter 5

The Effective Action for N=4 Supersymmetric Yang–Mills

In the previous chapter we found that the low energy effective action for the collective coordinates of N=2 supersymmetric Yang–Mills was given by supersymmetric quantum mechanics on the moduli space of BPS-monopoles. In this chapter we will do the same for N=4 super Yang–Mills. As we saw when we discussed that theory in Chapter 2, N=4 super Yang–Mills is a prime candidate to exhibit Montonen–Olive duality: not just are the masses and the structure of the multiplets protected by supersymmetry, but the massive vector bosons and the BPS-monopole belong to isomorphic multiplets. Therefore it would be possible for this theory to afford two inequivalent descriptions: one the standard one and a dual description where the perturbative fields are those in the multiplet containing the BPS-monopole. The structure of this chapter is therefore very similar to that of the previous chapter. We will first count the number of fermionic collective coordinates and will perform the collective coordinate expansion of the action up to second order. The resulting theory is again a $(0 + 1)$ supersymmetric σ-model, this time admitting N=8 supersymmetry due to the fact that there are twice as many fermionic collective coordinates as in the N=2 case. The quantisation of the effective action will proceed along lines similar to the previous chapter: this time the Hilbert space will be isomorphic to square integrable forms on the monopole moduli space, and the hamiltonian will once again be given by the laplacian. This chapter is based on the work of Blum [Blu94].

5.1 Fermionic collective coordinates

We saw in section 3.2 that there are $4k$ bosonic coordinates in the k-monopole sector; and, as we saw in section 4.1, N=2 supersymmetry contributed $2k$ fermionic collective coordinates. In this section we will show that for N=4 supersymmetric Yang–Mills the number of fermionic collective coordinates will double. We can understand this heuristically in a very simple matter. It follows from the discussion in section 4.1, that fermionic collective coordinates are in one-to-one correspondence with zero modes of the Dirac equation in the monopole background. The Dirac operator is the same in both the N=2 and the N=4 theories, but it acts on different types of fermions. In the N=2 theory, ψ was an unconstrained Dirac spinor (it came from a Weyl spinor in six dimensions); whereas in the N=4 theory, it acts on a quartet of Majorana fermions (the dimensional reduction from ten-dimensions of a Majorana–Weyl fermion). But now four Majorana spinors have twice the number of degrees of freedom that an unconstrained Dirac spinor does: $4 \times 4 = 16$ real components to only 4 complex.

To make this argument precise, we need to look in detail at how the Dirac operator breaks up. We start with a monopole background like the one in 2.4.3. Namely, we choose $\mathsf{W}_0 = 0$, $\mathsf{S}_I = a_I \boldsymbol{\phi}$, and $\mathsf{P}_J = b_J \boldsymbol{\phi}$, where a_I and b_J are real numbers satisfying $\sum_I (a_I^2 + b_I^2) = 1$, and where $(\mathsf{W}_i, \boldsymbol{\phi})$ define a k-monopole. Because the scalar fields are collinear, the potential remains at the minimum provided that the fermions satisfy the Dirac equation:

$$\bar{\gamma}_{\underline{i}} D_{\underline{i}} \boldsymbol{\psi} = 0 \ ,$$

where $\bar{\gamma}_i = \gamma_0 \gamma_i$, and $\bar{\gamma}_4 = -i\gamma_0 \left(a_I \alpha^I + b_J \beta^J \gamma_5 \right)$.

Exercise 5.1 (Euclidean Clifford algebra)
Prove that the matrices $\bar{\gamma}_{\underline{i}}$ defined above satisfy a euclidean Clifford algebra in four-dimensions.

From Exercise 4.1 we know that the normalisable zero modes of the Dirac operator $\bar{\gamma}_{\underline{i}} D_{\underline{i}}$ will have negative chirality with respect to $\bar{\gamma}_5$. But remember that $\boldsymbol{\psi}$ is also Majorana. We now check what chirality with respect to $\bar{\gamma}_5$ and the Majorana condition imply on a spinor.

Recall that our choice (2.32) of ten-dimensional Γ-matrices is such that the Majorana condition in ten-dimensions translates directly into the Majorana condition in four-dimensions. In four-dimensional Minkowski spacetime, there cannot be Majorana–Weyl spinors, but the euclidean $\bar{\gamma}$-matrices preserve the Majorana condition as the next exercise asks you to show.

Exercise 5.2 ($\bar\gamma_{\underline{i}}$ and the Majorana condition)
Let $\boldsymbol{\psi}$ be a quartet of Majorana spinors. Prove that $\bar\gamma_{\underline{i}}\boldsymbol{\psi}$ is again Majorana. Deduce that one can simultaneously impose the Majorana and $\bar\gamma_5$-chirality conditions.

Let us now start by choosing a explicit realisation for the γ-matrices:

$$\gamma_0 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \qquad \gamma_i = \begin{pmatrix} i\sigma_i & 0 \\ 0 & -i\sigma_i \end{pmatrix} \qquad \gamma_5 = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \tag{5.1}$$

The next exercise asks you to compute the charge conjugation matrix in this realisation.

Exercise 5.3 (The charge conjugation matrix explicitly)
Prove that the charge conjugation matrix C in the above realisation can be chosen to be

$$C = \begin{pmatrix} i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} .$$

(Hint: Using that $C^t = -C$ and that $C\gamma_\mu = -\gamma_\mu^t C$ determine C up to a constant multiple. A possible choice for this multiple is then one for which $C^\dagger C = \mathbb{1}$. That is the choice exhibited above.)

Now let $\boldsymbol{\psi}$ denote a quartet of Majorana spinors which in addition obey $\bar\gamma_5\boldsymbol{\psi} = -\boldsymbol{\psi}$. Because $\bar\gamma_5 = -\gamma_5\bar\gamma_4$ (prove it!), the chirality condition on $\boldsymbol{\psi}$ means that $\bar\gamma_4\boldsymbol{\psi} = -\gamma_5\boldsymbol{\psi}$. This means that the euclidean Dirac equation $\bar\gamma_{\underline{i}}D_{\underline{i}}\boldsymbol{\psi} = 0$ becomes $(\bar\gamma_i D_i - \gamma_5 D_4)\boldsymbol{\psi} = 0$. For the explicit realisation (5.1), this has the virtue that the Dirac operator doesn't see the internal $SU(4)$ indices. Indeed, the Dirac operator is given by:

$$\bar\gamma_{\underline{i}}D_{\underline{i}} = \begin{pmatrix} 0 & -\mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} \otimes \mathbb{1}_4 = (-i\sigma_2 \otimes \mathcal{D}) \otimes \mathbb{1}_4) , \tag{5.2}$$

where $\mathbb{1}_4$ is the identity matrix in the internal $SU(4)$ space, and $\mathcal{D} = iD_i\sigma_i + e\boldsymbol{\phi}\mathbb{1}$ is the operator introduced in (3.17).

Notice that if $\boldsymbol{\psi}$ had the opposite chirality with respect to $\bar\gamma_5$, then it would have been $\mathcal{D}^\dagger$ which would have appeared. This is as expected from the results of Exercise 3.5 and Exercise 4.1.

We are now ready to count the zero modes of the euclidean Dirac operator, by relating them to zero modes of $\mathcal{D}$, which we have already calculated to be $2k$. We first choose the explicit realisation for the α^I and β^J matrices found in Exercise 2.33: $\alpha^I = e_I^+$ and $\beta^J = e_J^-$. Next we exploit the internal $SU(4)$ invariance to fix $a_1 = 1$ and all the other a_I and b_J to zero. This means that $\bar\gamma_5 = -i\gamma_0\gamma_5 \otimes \alpha^1 = (\sigma_3 \otimes \mathbb{1}) \otimes (\sigma_2 \otimes \mathbb{1})$. From Exercise 5.3 we know that the charge conjugation matrix is given by $C = (\sigma_3 \otimes i\sigma_2) \otimes \mathbb{1}_4$. The next exercise asks you to write down the typical quartet of Majorana spinors $\boldsymbol{\psi}$ which in addition are chiral with respected to $\bar\gamma_5$.

Exercise 5.4 ("Majorana–Weyl" spinors)
Prove that every quartet of Majorana spinors $\boldsymbol{\psi}$ *obeying* $\bar{\gamma}_5\boldsymbol{\psi} = \pm\boldsymbol{\psi}$ *is of the form:*

$$\left(\begin{pmatrix}\eta\\ -i\sigma_2\eta^*\end{pmatrix}\begin{pmatrix}\zeta\\ -i\sigma_2\zeta^*\end{pmatrix}\begin{pmatrix}\mp i\eta\\ \pm\sigma_2\eta^*\end{pmatrix}\begin{pmatrix}\mp i\zeta\\ \pm\sigma_2\eta^*\end{pmatrix}\right) ,$$

where η *and* ζ *are complex two-component spinors.*

Finally we count the zero mode of the euclidean Dirac operator $\bar{\gamma}_{\underline{i}}D_{\underline{i}}$.

Exercise 5.5 (Counting zero modes)
Show that $\boldsymbol{\psi}$ *is a "Majorana–Weyl" zero mode of the euclidean Dirac operator, in the sense of the previous exercise, if and only if* η *and* ζ *are zero modes of* $\mathcal{D}$*. Therefore if* η_a *for* $a = 1, \ldots, 2k$ *is a basis for the normalisable zero modes of* $\mathcal{D}$*, then the* $4k$ *spinors*

$$\left(\begin{pmatrix}\eta_a\\ -i\sigma_2\eta_a^*\end{pmatrix}\begin{pmatrix}0\\ 0\end{pmatrix}\begin{pmatrix}\mp i\eta_a\\ \pm\sigma_2\eta_a^*\end{pmatrix}\begin{pmatrix}0\\ 0\end{pmatrix}\right) \qquad \left(\begin{pmatrix}0\\ 0\end{pmatrix}\begin{pmatrix}\eta_a\\ -i\sigma_2\eta_a^*\end{pmatrix}\begin{pmatrix}0\\ 0\end{pmatrix}\begin{pmatrix}\mp i\eta_a\\ \pm\sigma_2\eta_a^*\end{pmatrix}\right)$$

form a basis for the normalisable zero modes of $\bar{\gamma}_{\underline{i}}D_{\underline{i}}$*.*

In summary, there are $4k$ fermionic collective coordinates for $N{=}4$ supersymmetric Yang–Mills with gauge group $SO(3)$.

Part II

Arbitrary Gauge Groups

Chapter 6

Monopoles for Arbitrary Gauge Groups

In this chapter we start the study of electromagnetic duality in supersymmetric gauge theories with an arbitrary gauge group. We will be interested in this part of the notes only on $N{=}4$ super Yang–Mills. Our principal aim is to frame an analogue of the Montonen–Olive duality conjecture for these theories, to develop testable predictions and then to test them. This will occupy several chapters, but in this one we will start with the analysis of the kind of monopole solutions that can exist in a Yang–Mills–Higgs theory with gauge group G, taken to be a compact, connected Lie group, and a Higgs field with values in the adjoint representation. We will cover the homotopy classification of topologically stable solutions and the generalised Dirac quantisation condition. This chapters borrows quite a lot from the magnificent lectures of Coleman [Col77], and from the paper of Goddard, Nuyts and Olive [GNO77].

6.1 Topologically stable solutions

Let G be a compact connected Lie group, and Φ a scalar field taking values in some finite-dimensional representation $\mathbb{V}$ of G. We will assume that there is a G-invariant potential $V(\Phi)$ which is positive semi-definite and also that $\mathbb{V}$ admits a G-invariant metric. This is necessary in order to write down the kinetic term for Φ in the action. We will let $\mathfrak{g}$ denote the Lie algebra of G. We will fix once and for all an invariant metric on $\mathfrak{g}$. As the next exercise shows, such a metrics always exists.

Exercise 6.1 (Invariant metrics exist)
Prove that there exists a G-invariant metric in the Lie algebra of a compact Lie

group.
(Hint: Start with any metric and average over the group with respect to the Haar measure. Does this argument work for any *representation?)*

We will denote both metrics on $\mathfrak{g}$ and on $\mathbb{V}$ by $(-,-)$, hoping that no confusion will arise. The lagrangian density of the Yang–Mills–Higgs system is given by

$$\mathcal{L} = -\tfrac{1}{4}(\mathsf{G}_{\mu\nu}, \mathsf{G}^{\mu\nu}) + \tfrac{1}{2}(D_\mu\Phi, D^\mu\Phi) - V(\Phi) \ , \tag{6.1}$$

where

- $D_\mu = \partial_\mu\Phi - e\mathsf{W}_\mu \cdot \Phi$,
- $\mathsf{G}_{\mu\nu} = \partial_\mu\mathsf{W}_\nu - \partial_\nu\mathsf{W}_\mu - e[\mathsf{W}_\mu, \mathsf{W}_\nu]$,

where W_μ are the $\mathfrak{g}$-valued gauge potentials, and by $\cdot$ we mean the action of $\mathfrak{g}$ on the representation $\mathbb{V}$.

Let M_0 denote the *manifold of vacua*: those values of Φ for which $V(\Phi) = 0$. Because V is G-invariant, G will map M_0 to M_0; in other words, G stabilises M_0.

We now choose the temporal gauge $\mathsf{W}_0 = 0$. As the next exercise shows, this can always be done and leaves intact the freedom of performing time-independent gauge transformations.

Exercise 6.2 (The temporal gauge)
Prove that the temporal gauge exists by exhibiting a gauge transformation which makes $\mathsf{W}_0 = 0$. Prove that this gauge is preserved by time-independent gauge transformations.
(Hint: Use path-ordered exponentials.)

In this gauge, the energy density corresponding to the lagrangian density (6.1) is given by

$$\mathcal{H} = \tfrac{1}{2}(\dot{\mathsf{W}}_i, \dot{\mathsf{W}}_i) + \tfrac{1}{2}(\dot{\Phi}, \dot{\Phi}) + \tfrac{1}{4}(\mathsf{G}_{ij}, \mathsf{G}_{ij}) + \tfrac{1}{2}(D_i\Phi, D_i\Phi) + V(\Phi) \ ,$$

where a dot indicates the time derivative, and where repeated indices are summed. The energy is of course the integral over space $\mathbb{R}^3$ of the energy density $\mathcal{H}$, and hence finite-energy configurations must obey the following asymptotic conditions as $|\vec{\boldsymbol{r}}| \to \infty$:

- $\dot{\mathsf{W}}_i = 0$ and $\dot{\Phi} = 0$, whence fields are asymptotically static;
- $\mathsf{G}_{ij} = 0$ faster than $O(1/r)$;
- $D_i\Phi = 0$ faster than $O(1/r)$; and

- $V(\Phi) = 0$.

In particular this last condition says that Φ defines a map from the asymptotic 2-sphere $S^2_\infty \subset \mathbb{R}^3$ to the manifold of vacua M_0. Because M_0 is stabilised by G, it will be foliated by orbits of G. For example, in $G = SO(3)$ and the Higgs is in the adjoint, the leaves of the foliation of $SO(3)$ in $\mathbb{R}^3$ are the round spheres centred at the origin, with a "singular" orbit corresponding to the sphere of zero size. A priori there is no reason to expect that the mapping $S^2_\infty \to M_0$ defined by the asymptotics of the Higgs field should lie on only one of the orbits, but because $D_i\Phi = 0$ in this limit, this is actually the case. The proof is left as an exercise.

Exercise 6.3 ($\Phi(S^2_\infty) \subset M_0$ lies in a single orbit)
Prove that the image of S^2_∞ lies in a single orbit in M_0.
(Hint: Integrate the equation $D_i\Phi = 0$ on S^2_∞.)

A sufficient—but as shown by Coleman [Col77]) not necessary—condition for a finite-energy configuration to be non-dissipative is that it should be topologically stable. As explained in section 1.2.2, a way to guarantee the topological stability of a field configuration is for the map $\Phi : S^2_\infty \to M_0$ to belong to a nontrivial homotopy class. It therefore behoves us to study the homotopy classes of maps from the asymptotic two-sphere S^2_∞ to the manifold of vacua M_0 or, more precisely, to the orbit to which $\Phi(S^2_\infty)$ belongs. To this effect we find it useful to set down some basic notions about homotopy groups. Readers familiar with this material can easily skip the next section.

6.1.1 Some elements of homotopy

This section contains a brief review of homotopy theory. Homotopy theory is the study of continuous change, and is particularly concerned with the determination of quantities which are impervious to such changes. Roughly speaking a homotopy is a continuous deformation parametrised by the unit interval $I = [0, 1]$. We will have to be a little bit more precise than this in what follows, but we will avoid getting too technical. In particular, we will not give many proofs. Luckily for us, the aspects of homotopy theory that we will need in these notes can be understood quite intuitively. Proofs can be given but they rely quite a bit on point-set topology. Since that is not the main point of these notes, we simply point the reader who wishes to look at the proofs of the statements made in this section to the old but still excellent book by Steenrod [Ste51].

The useful objects in homotopy theory are not just topological spaces, but spaces with a privileged point called the *basepoint*. A map between two such spaces is understood to be a continuous function which sends basepoint to basepoint. Let X and Y be two topological spaces with basepoints x_0 and y_0 respectively, and let f_0 and f_1 be two continuous functions $X \to Y$ taking x_0 to y_0. We say that these two functions are *homotopic* if there exists a family of functions parametrised by the interval which interpolates continuously between them. More precisely, f_0 is homotopic to f_1 (written $f_0 \simeq f_1$) if there exists a continuous function $F : X \times I \to Y$, such that for all $x \in X$, $F(x,0) = f_0(x)$ and $F(x,1) = f_1(x)$ and such that for all $t \in I$, $F(x_0,t) = y_0$. This last condition says that the homotopy is relative to the basepoint.

The fundamental group

A good example with which to visualise these definitions is to take X to be the circle. We can think of the circle as the unit interval with endpoints identified. Then a map from the circle to Y as a map $f : I \to Y$ with $f(0) = f(1) = y_0$. That is, a continuous loop based at y_0. Then two such loops are homotopic if they can be continuously deformed to each other through loops which are based at y_0.

The set of homotopy equivalence classes of maps $f : X \to Y$ with $f(x_0) = y_0$ is written $[X, x_0; Y, y_0]$. In the special case above where X is the circle, the set of homotopy equivalence classes is written $\pi_1(Y, y_0)$. But $\pi_1(Y, y_0)$ is more than just a set: indeed, based loops can be composed. Given two loops f_1 and f_2 based at y_0, we can form a third loop $f_1 * f_2$ by simply going first along f_1 and then along f_2 at twice the speed. In other words,

$$(f_1 * f_2)(t) = \begin{cases} f_1(2t) & \text{for } t \in [0, \frac{1}{2}], \\ f_2(2t-1) & \text{for } t \in [\frac{1}{2}, 1]. \end{cases}$$

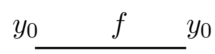
Notice that composition is not just defined for loops but also for paths, provided that the first path ends where the second begins. As the following exercise shows, composition of based loops induces a well-defined operation on homotopy classes, which makes $\pi_1(Y, y_0)$ into a group.

Exercise 6.4 ($\pi_1(Y, y_0)$ is a group)
In this exercise we prove that $\pi_1(Y, y_0)$ is a group, with group multiplication given by composition of loops. The proof consists of several steps which are all very easy. The idea is to first prove that $*$ makes sense in $\pi_1(Y, y_0)$ and then that $*$ on loops satisfies all the properties of a group up to homotopy. This means that in $\pi_1(Y, y_0)$ they are satisfied exactly.

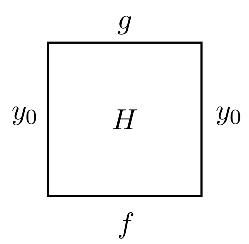
- $*$ is well-defined in homotopy. *Prove that if $f_0 \simeq f_1$ and $g_0 \simeq g_1$ are loops, then $f_0 * g_0 \simeq f_1 * g_1$. (This allows us to work with loops, knowing that up* *to homotopy it doesn't really matter which loop we choose to represent its homotopy class.)*

- $*$ is associative in $\pi_1(Y, y_0)$. *Prove that if f_1, f_2, and f_3 are loops then $(f_1 * f_2) * f_3 \simeq f_1 * (f_2 * f_3)$. In other words, $*$ is associative up to homotopy.*
- $\pi_1(Y, y_0)$ has an identity. *Prove that the constant loop sending all the circle to y_0 is an identity for $*$ up to homotopy; that is, if k denotes the constant loop, then $k * f \simeq f * k \simeq f$ for any loop f.*
- Inverses exist. *Let f be a loop, and let $\bar{f}$ denote the loop obtained by following f backwards in time: $\bar{f}(t) = f(1-t)$. Prove that $f * \bar{f} \simeq \bar{f} * f \simeq k$, where k is the constant loop.*

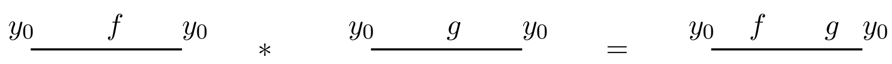
(Hint: It may be convenient to devise a pictorial way to denote loops and homotopies. For instance a loop f based at y_0 can be depicted as a unit interval whose endpoints are marked y_0:



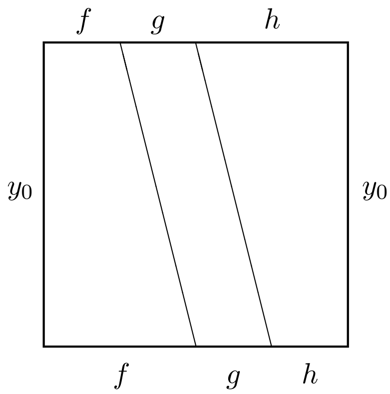
Similarly if f and g are two such loops, a homotopy H between them can be depicted as a square whose left and right edges are marked y_0 and whose top and bottom edges correspond to f and g:



Composition of loops can then be depicted simply as pasting the intervals together by their endpoints and contracting (reparametrising time) so that the resulting interval has again unit length. The following picture illustrates this:



In this language, the group properties become almost self-evident. For example, the associativity property of $*$ simply becomes



and similarly for the other axioms.)

The group $\pi_1(Y, y_0)$ is known as the *first homotopy group* of the pointed space (Y, y_0). If Y is path-connected, so that any two points in Y can be joined by a continuous path, then the first homotopy group does not depend (up to isomorphism) on the basepoint. This fact has a simple proof which we leave to the next exercise. Incidentally, the condition of connectedness and path-connectedness are not equivalent, but they do agree for manifolds, and hence for all the spaces we will be considering in these notes.

Exercise 6.5 ($\pi_1(Y, y_0) \cong \pi_1(Y, y_1)$)
Let Y be path-connected and y_0 and y_1 be two points in Y. Fix a path $\gamma : I \to Y$ with $\gamma(0) = y_0$ and $\gamma(1) = y_1$. Because Y is path-connected, γ exists. We can use this path to turn any loop f_1 based at y_1 into a loop based on y_0: one simply composes $\bar{\gamma} * f_1 * \gamma$, where $\bar{\gamma}(t) = \gamma(1-t)$. Prove that this defines a group isomorphism $\pi_1(Y, y_1) \cong \pi_1(Y, y_0)$.

Therefore when Y is connected, it makes sense to talk about $\pi_1(Y)$ without reference to a basepoint. This group is called the *fundamental group* of Y. If this group is trivial, so that all loops are homotopic to the constant map, then Y is said to be *simply-connected*.

Notice that the isomorphism in Exercise 6.5 depends on the choice of path γ joining the two basepoints. How does the isomorphism depend on γ? It is easy to show (do it!) that if γ' is any other path which is homotopic to γ with endpoints fixed, then the isomorphisms induced by γ and γ' agree. On the other hand, paths in different homotopy classes generally define different isomorphisms. This can be formulated in a way that shows an action of the fundamental group of the space on itself by conjugation. This will play a role later and we will discuss this further we study the addition of topological charges for largely separated monopoles.

The fundamental groups of some manifolds are well known. Here are some examples.

- $\mathbb{R}^n$ is simply-connected for any n.
- The punctured plane $\mathbb{R}^2\backslash\{0\}$ is no longer simply-connected; in fact, $\pi_1(\mathbb{R}^2\backslash\{0\}) \cong \mathbb{Z}$. The isomorphism is given by the following well-known integral formula from complex analysis. To see notice that $\mathbb{R}^2\backslash\{0\} = \mathbb{C}^\times$ is the punctured complex plane. Let γ be a loop in $\mathbb{C}^\times$, and compute the contour integral
$$\oint_\gamma \frac{1}{2\pi i}\frac{dz}{z} .$$
It is well known that this is an integer and is a homotopy invariant of the loop.
- $\pi_1(S^1) \cong \mathbb{Z}$. This is just the above example in disguise. We can think of S^1 as the unit circle in the complex plane $S^1 \subset \mathbb{C}^\times$. Any loop (or homotopy for that matter) in $\mathbb{C}^\times$ can be projected onto the unit circle by $t \mapsto \gamma(t) \mapsto \gamma(t)/|\gamma(t)|$. The isomorphism $\pi_1(S^1) \cong \mathbb{Z}$ is known as the degree of the map. It basically counts the number of times one circle winds around another.
- Puncturing $\mathbb{R}^n$, for $n > 2$, does not alter the fundamental group. In fact, any loop in $\mathbb{R}^n\backslash\{0\}$ is homotopic to a loop on its unit sphere $S^{n-1} \subset \mathbb{R}^n$, again by projecting. But for $n > 2$, it is intuitively clear that any loop on S^{n-1} is homotopic to a constant. For $n = 3$ it is the well-known principle that "you cannot lasso an orange." You cannot lasso higher-dimensional oranges either.
- You can make a non-simply-connected space out of $\mathbb{R}^3$ by removing a circle (or a knot), say, or an infinite line.

However it is not only by making holes in a space that we can generate nontrivial loops. We can also identify points. For example, if we take the sphere S^n, for $n > 1$, and identify antipodal points, we describe a space which is not simply-connected. The space in question is the space of lines through the origin in $\mathbb{R}^{n+1}$: since every line through the origin will intersect the unit sphere in two antipodal points. We call this space the *real projective space* $\mathbb{RP}^n$. We can lift a loop in $\mathbb{RP}^n$ up to S^n. This procedure is locally well-defined once we choose a starting point in S^n. There is no further choice and when we are done with the lift, we are either at the starting point or at its antipodal point, since both points map down to the same point in $\mathbb{RP}^n$.

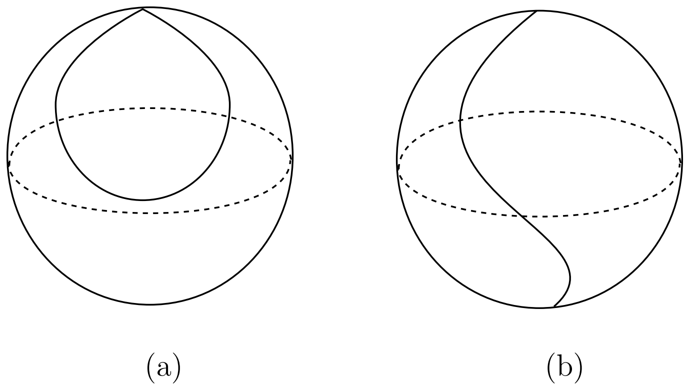


Figure 6.1: The two possible lifts to S^n of a loop in $\mathbb{RP}^n$.

In the former case, the loop has lifted to an honest loop in S^n, which is depicted by (a) in Figure 6.1. Since S^n is simply-connected, we can project the homotopy to $\mathbb{RP}^n$ and this gives a homotopy for the original loop. On the other hand, if the loop ends at the antipodal point, as shown in (b) in Figure 6.1, there is clearly no way to deform it to the constant map while keeping endpoints fixed, so it defines a nontrivial loop in $\mathbb{RP}^n$. However notice that the loop obtained by going twice around the loop lifts to an honest loop in the sphere, and is hence trivial. This shows that $\pi_1(\mathbb{RP}^n) \cong \mathbb{Z}_2$. The two-to-one map $\rho : S^n \to \mathbb{RP}^n$ is a *covering map*, in that it is a local homeomorphism and every point p in $\mathbb{RP}^n$ has a neighbourhood U such that its inverse image by the covering map $\rho^{-1}(U) \subset S^n$ consists of two disconnected neighbourhoods. Because S^n is simply-connected, we say that S^n is the *universal covering space* of $\mathbb{RP}^n$. All reasonable spaces X (certainly all manifolds and hence all spaces considered in these notes) possess a universal covering space $\tilde{X}$. This space is simply-connected and is such that it admits a free action of the fundamental group of X. In the case of S^n, it admits an action of $\mathbb{Z}_2$, sending a point on the sphere to its antipodal point. The case $n = 3$ is particularly interesting, because it is intimately related with two of our favourite Lie groups: $SU(2)$ and $SO(3)$.

Exercise 6.6 ($SU(2)$ and $SO(3)$)
Prove that the Lie group $SU(2)$ of 2×2 special unitary matrices is parametrised by a three-sphere S^3 and that the group $SO(3)$ of 3×3 special orthogonal matrices is parametrised by the real projective space $\mathbb{RP}^3$. Prove that there is a group homomorphism $SU(2) \to SO(3)$ which sends both $\mathbb{1}$ and $-\mathbb{1}$ in $SU(2)$ to $\mathbb{1}$ in $SO(3)$. Notice that $\mathbb{1}$ and $-\mathbb{1}$ generate the centre of $SU(2)$, which is isomorphic to Z_2. Hence $SU(2)$ is the universal covering group *of $SO(3)$.*

This situation persists for other Lie groups. Every semisimple compact

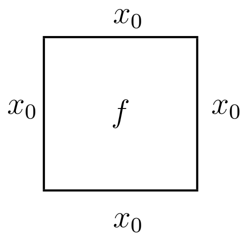
Lie group G has a universal covering group $\tilde{G}$, sharing the same Lie algebra $\mathfrak{g}$. The fundamental group $\pi_1(G)$ is naturally identified with a subgroup of the centre of $\tilde{G}$. We will be able to compute $\pi_1(G)$ by comparing the finite-dimensional irreducible representations of G with those of $\tilde{G}$, which are those of $\mathfrak{g}$. For example, not every irreducible representation of $SU(2)$ is a representation of $SO(3)$: only those with integer spin. Since representations of $SU(2)$ can have integer or half-integer spin, this means that $SU(2)$ has twice as many irreducible representations as $SO(3)$—which is precisely the order of $\pi_1(SO(3)) \cong \mathbb{Z}_2$. This is no accident, as we will see later on.

Before we abandon the subject of the fundamental group, we mention one last fact. Notice that all the fundamental groups that we have discussed so far are abelian. This is not always the case. In fact, the fundamental group of any compact Riemann surface of genus $g > 1$ is non-abelian. However, there is an important class of manifolds for which the fundamental group is abelian.

Exercise 6.7 ($\pi_1(G)$ is abelian)
Let G be a connected Lie group. Prove that $\pi_1(G)$ is abelian.
(Hint: In a Lie group there are two ways to compose loops. We can use the loop composition $*$ defined above, or we can use pointwise group multiplication, provided that the loops are based at the identity. Indeed, if f and g are loops in G based at the identity, one can define $(f \bullet g)(t) = f(t)\, g(t)$. Prove that $f * g \simeq f \bullet g$, so that we can use group multiplication to define the multiplication in the fundamental group. Use this to write down a homotopy between $f * g$ and $g * f$, for any two loops f and g in G.)

Higher homotopy groups

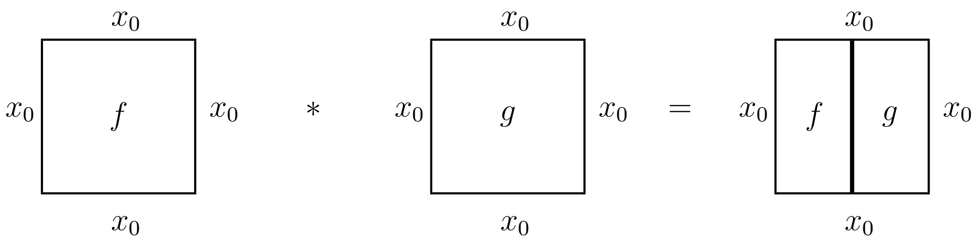
The fundamental group has higher dimensional analogues obtained by substituting the circle by a sphere. Just like we could think of the circle as the interval I with edges identified, we can think of the n-sphere as the multi-interval I^n with its boundary ∂I^n identified to one point. There is a rich theory for all n, but we will only need $n = 2$ in these notes, so we will concentrate mainly on this case. Any map $S^2 \to X$ can be thought of as a map $I^2 \to X$ which sends the boundary ∂I^2 to the basepoint $x_0 \in X$. Just as in Exercise 6.4, we choose to depict such a map f as a rectangle I^2 with the basepoint x_0 along the edges to remind us that x_0 is where ∂I^2 gets mapped to:



We will denote the homotopy classes of such maps by $\pi_2(X, x_0)$. Just like for π_1, we can turn this space into a group. We first discuss composition. Two maps $S^2 \to X$ can be composed by adjoining the squares, just like we did for loops. However in this case there seems to be an ambiguity: we can adjoin the squares horizontally or vertically. We will see that there is indeed no such ambiguity, but for the present we choose to resolve it by composing them horizontally. In other words, if f and g are two maps $I^2 \to X$, we define the composition $f * g$ by

$$(f * g)(t_1, t_2) = \begin{cases} f(2t_1, t_2) & \text{for } t_1 \in [0, \frac{1}{2}], \\ g(2t_1 - 1, t_2) & \text{for } t_1 \in [\frac{1}{2}, 1]. \end{cases}$$

Clearly the resulting map is continuous, since the boundary conditions agree: $f(1, t_2) = g(0, t_2) = x_0$ for all t_2. Pictorially this composition corresponds to the following diagram:



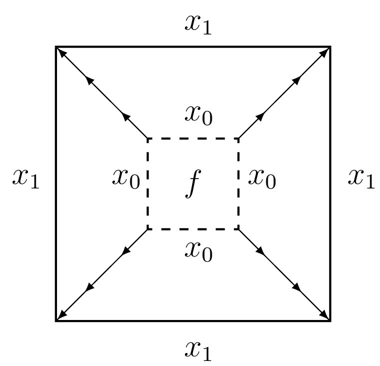
Just like we did for $\pi_1(X, x_0)$, we can show that $\pi_2(X, x_0)$ is a group.

Exercise 6.8 ($\pi_2(X, x_0)$ is a group)
This exercise follows similar steps to Exercise 6.4. Prove the following:

- $*$ is well-defined in homotopy.
- $*$ is associative up to homotopy.

- $\pi_2(X, x_0)$ has an identity. *Prove that the constant map sending all of I^2 to x_0 is an identity for $*$ up to homotopy.*
- Inverses exist. *Let f be a map $I^2 \to X$, let $\bar{f}$ denote the map obtained by following f backwards in the first of the two times: $\bar{f}(t_1, t_2) = f(1 - t_1, t_2)$. Prove that $\bar{f}$ is the inverse of f up to homotopy.*

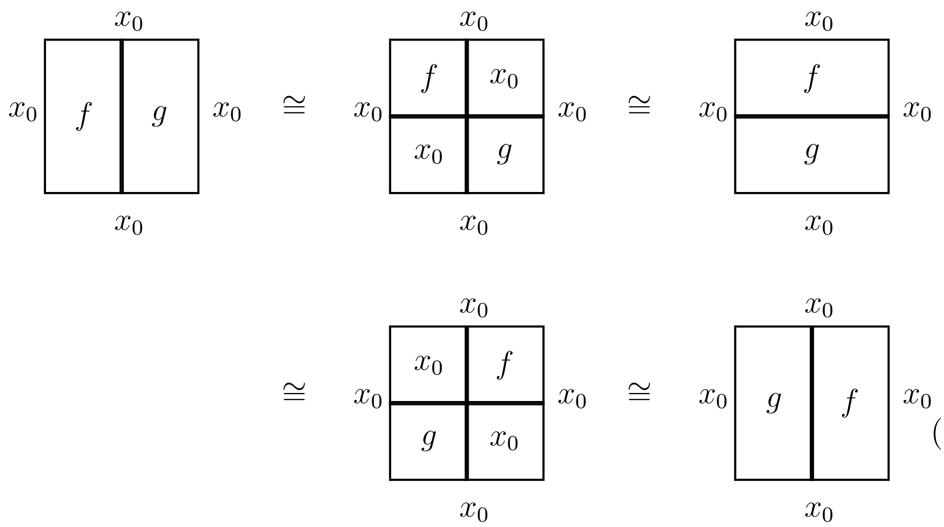
If X is path-connected, it follows that $\pi_2(X, x_0)$ doesn't depend on the basepoint (up to isomorphism). Indeed, let f represent a homotopy class in $\pi_2(X, x_0)$. Given any other basepoint $x_1 \in X$, let γ be a path from x_0 to x_1. The following diagram represents a homotopy class in $\pi_2(X, x_1)$:



where the arrows represent the path γ going from x_0 to x_1. The next exercise asks you to show that this map is an isomorphism.

Exercise 6.9 **($\pi_2(X, x_0) \cong \pi_2(X, x_1)$)**
Prove that the above map is an isomorphism $\pi_2(X, x_0) \cong \pi_2(X, x_1)$; and that the isomorphism only depends on the homotopy class of the path γ used to define it.

Hence for path-connected X it makes sense to talk about $\pi_2(X)$ without reference to the basepoint, provided that we are only interested in its isomorphism class. This group is a higher-dimensional analogue of the fundamental group $\pi_1(X)$. Unlike the fundamental group, $\pi_2(X)$ is always abelian. The following sequence of homotopies proves this assertion:



. (6.2)

This proof also shows that there is no ambiguity in the composition after all. Similarly one proves that also $\pi_k(X)$ for $k > 2$ are abelian.

Finally we discuss some examples of higher homotopy groups. Unlike the fundamental group, for which there are theorems (for instance, the Van Kampen theorem) allowing us to compute $\pi_1(X)$ starting from a decomposition of X into simpler spaces, the computation of the higher homotopy groups is a very difficult problem. For example, not all the homotopy groups of the 2-sphere S^2 are known! Nevertheless, here are some examples of higher homotopy groups:

- $\pi_k(S^n) = 0$ for $k < n$, and $\pi_n(S^n) \cong \mathbb{Z}$. The first equality is the fact that one cannot k-lasso an n-orange, for $k < n$; whereas the last isomorphism is given by the *degree* of the map, as was the case for n=1.

- $\pi_k(S^1) = 0$ for $k > 1$. This follows because S^1 can be covered by a contractible space: $S^1 \cong \mathbb{R}/\mathbb{Z}$.

- $\pi_k(T^n) = 0$ for $k > 1$, where T^n is an n-torus. This follows for the same reason as the above example: $T^n \cong \mathbb{R}^n/\Lambda$, where Λ is some lattice.

- $\pi_k(\Sigma_g) = 0$ for $k > 1$, where Σ_g is a compact Riemann surface of genus g. Again, the proof is as above, since Σ_g can be written as the quotient of the Poincaré upper half plane by a Fuchsian group: $\Sigma_g \cong H/\Gamma$.

- $\pi_2(G) = 0$, for any topological group G. This result of É. Cartan will play a very important role in the next section.
- The homotopy groups of the spaces determining a fibration $F \to E \to B$, where F is the typical fibre, E is the total space, and B is the base, are related by a useful gadget known as the *exact homotopy sequence* of the fibration:

$$\begin{aligned}\cdots \to \pi_n(B) \to \pi_{n-1}(F) \to \pi_{n-1}(E) \to \pi_{n-1}(B) \to \cdots \\ \cdots \to \pi_2(E) \to \pi_2(B) \to \pi_1(F) \to \pi_1(E) \to \pi_1(B)\ .\end{aligned}$$

  The "exactness" of this sequence simply means that every arrow is a group homomorphism such that its kernel (the normal subgroup sent to the identity) precisely agrees with the image of the preceding arrow. If the fibration is principal, so that F is a Lie group, then the sequence extends one more term to include a map $\pi_1(B) \to \pi_0(F)$. Where $\pi_0(F)$ is the set of connected components of the typical fibre. One can define $\pi_0(X)$ in this way for any space X, but for a general X, $\pi_0(X)$ is only a set. It is when $X = G$ is a group, that $\pi_0(G)$ also inherits a group operation. Indeed, $\pi_0(G) \cong G/G_0$, where G_0 is the connected component of the identity. It is in this case that it makes sense to speak of a group homomorphism $\pi_1(B) \to \pi_0(G)$.

A lot more could be said about higher homotopy groups, but this about covers all that we will need in the sequel.

6.1.2 Homotopy classification of finite-energy configurations

After this brief review of homotopy theory, we return to the problem at hand. Let us fix a basepoint $\vec{\boldsymbol{r}}_0$ in the two-sphere at infinity; for example, we could choose the north pole. Let $\Phi(\vec{\boldsymbol{r}}_0) = \phi_0 \in M_0$. The G-orbit of ϕ_0 will be the set $G \cdot \phi_0 = \{g \cdot \phi_0 | g \in G\} \subset M_0$. If we let $H = H_{\phi_0} \subset G$ denote the stability subgroup of ϕ_0: $H_{\phi_0} = \{h \in G | h \cdot \phi_0 = \phi_0\}$, then $G \cdot \phi_0 \cong G/H$. The asymptotics of the Higgs field define a map from the two-sphere to G/H taking the basepoint to ϕ_0. In other words it defines an element in the second homotopy group $\pi_2(G/H, \phi_0)$.

Because G is connected, this class is gauge-invariant, as the next exercise asks you to show. This shows that the homotopy class is physical.

Exercise 6.10 (Gauge invariance of the homotopy class)
Prove that gauge related Higgs field configurations define asymptotics which are homotopic.

(Hint: Use the fact that G is connected to write down an explicit homotopy between the two configurations.)

As shown in the previous section, $\pi_2(G/H, \phi_0)$ is an abelian group, and because G/H is connected it does not depend (up to isomorphism) on ϕ_0. Because of this fact we will drop the reference to the basepoint when unnecessary. We will now prove that $\pi_2(G/H)$ is isomorphic to the subgroup of $\pi_1(H)$ given by those homotopy classes of loops in H which are null-homotopic in G. From what was said in the previous section, you will immediately recognise this statement as part of the exact homotopy sequence associated to the fibration $H \to G \to G/H$. But rather than appealing to such heavy machinery, we will prove most of this statement here using more pedestrian methods.

We first associate a loop in H with each Φ. Let $\Sigma^\pm$ denote an open cover for the asymptotic two-sphere S^2_∞; more concretely, we take their union to be S^2_∞ and their intersection to be a small band around the equator. Since $\Sigma^\pm$ are homeomorphic to disks (hence contractible), we can find local gauge transformations $g_\pm : \Sigma^\pm \to G$ such that

$$\Phi(x) = g_\pm(x) \cdot \phi_0 \qquad \text{for } x \in \Sigma^\pm \ .$$

Then on the intersection $\Sigma^+ \cap \Sigma^-$, we have

$$g_+(x) \cdot \phi_0 = g_-(x) \cdot \phi_0 \ ,$$

whence $g_+(x)^{-1} g_-(x) \cdot \phi_0 = \phi_0$, whence $g_+(x)^{-1} g_-(x)$ defines an element of H. Restricting to the equator, we have a continuous map $x \mapsto h(x) = g_+(x)^{-1} g_-(x) \in H$; that is, a loop in H. Because $g_\pm(x)$ are defined only up to right multiplication by H, we can always arrange so that $h(x)$ is the identity for some x in the equator. This way h defines an element in $\pi_1(H, \mathbb{1})$. Furthermore, this loop is trivial in G. Indeed, since G is path-connected, there exist paths $t \mapsto g_\pm(t,x)$ from the identity to $g_\pm(x)$. Defining $h(t,x) = g_+(t,x)^{-1} g_-(t,x)$ provides a homotopy in G from the identity to the loop $h(x)$.

Alternatively we can understand this in a more explicit way. A map $\Phi : S^2_\infty \to G/H$ can be thought of as a loop of loops: In the above picture, there is a family parametrised by the interval $s \in [0,1]$ of loops based at ϕ_0 and each loop in the family is in turn parametrised by the interval $t \in [0,1]$, with the condition that the initial and final loops are trivial. Therefore we can redraw Figure 6.2 as a map from the square to G/H where the edges are mapped to ϕ_0:

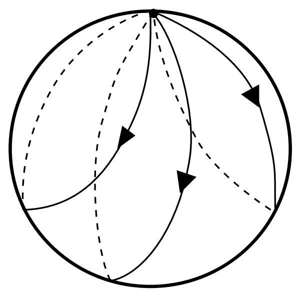
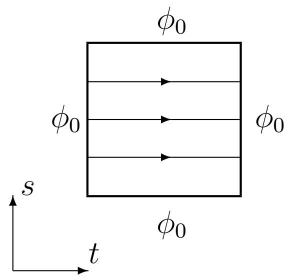


Figure 6.2: A map $S^2 \to G/H$ as a loop of loops.



where the three horizontal lines are precisely the three loops depicted above.

Consider now a fixed loop, that is, a fixed value of s. Because $D_i\Phi = 0$ on $S^2\infty$, we can solve for Φ.

Exercise 6.11 (Solving for Φ)
Prove that for a fixed s, $\Phi(s,t)$ given by

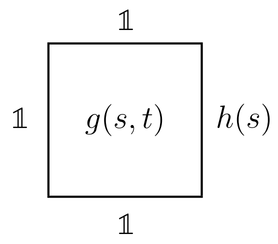
$$\Phi(s,t) = P\exp\left(e\int_0^t dt'\,\mathsf{W}_i(s,t')\frac{\partial x^i}{\partial t'}\right)\cdot\phi_0 \tag{6.3}$$

is a solution of $D_i\Phi$ with the boundary conditions $\Phi(s,0) = \phi_0$. Here s,t are coordinates for S^2_∞ and $x(s,t)$ are the coordinates in $\mathbb{R}^3$. Similarly $\mathsf{W}_i(s,t)$ is short for $\mathsf{W}_i(x(s,t))$.

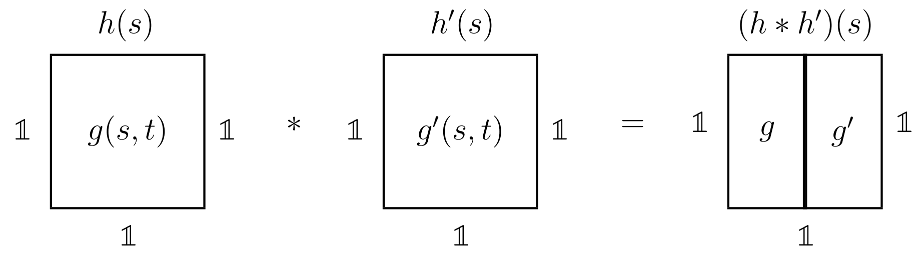
Let $g(s,t) \in G$ be the group element defined by $\Phi(s,t) = g(s,t)\cdot\phi_0$ in (6.3). From its definition it follows that $g(s,0) = 1$ and since $\frac{\partial x^i}{\partial t} = 0$ at $s = 0$ and $s = 1$, it follows that $g(0,t) = g(1,t) = 1$. How about $g(s,1)$? Because $\Phi(s,1) = \phi_0$, $g(s,1)\cdot\phi_0 = \phi_0$, whence $g(s,1) = h(s)$ is in H. Since $h(0) = h(1) = 1$, as s varies, $h(s)$ defines a loop in H:

$$h(s) \equiv P\exp\left(e\int_0^1 dt\,\mathsf{W}_i(s,t)\frac{\partial x^i}{\partial t}\right) \tag{6.4}$$

Moreover, this loop in H is trivial in G, the homotopy being given by $g(s,t)$ itself, as the following figure shows:



This approach has the added benefit that it is very easy to see that the map $\pi_2(G/H) \to \pi_1(H)$ sending the class of $\Phi(s,t)$ to the class of $h(s)$ is a group homomorphism. Indeed rotating the squares, we have the composition:

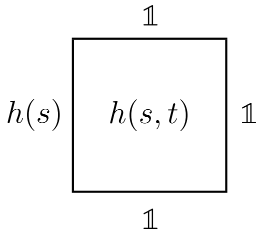


from where we see that if $\Phi \mapsto h$ and $\Phi' \mapsto h'$, then $\Phi * \Phi' \mapsto h * h'$. In summary we have a map $\pi_2(G/H) \to \pi_1(H)$, sending $\Phi(s,t)$ to $h(s)$, which is a homomorphism of groups. Notice that both groups are abelian by Exercise 6.7 and equation (6.2). Also, because any loop in H can be thought of as a loop in G, we have a natural map $\pi_1(H) \to \pi_1(G)$, which is also a homomorphism of abelian groups. We can thus compose these two maps, and what we have shown is that the composition:

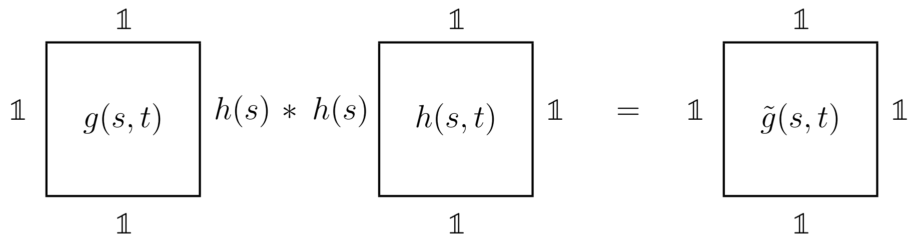
$$\pi_2(G/H) \to \pi_1(H) \to \pi_1(G) \qquad \text{is zero.} \tag{6.5}$$

Conversely it follows readily from what we said above that any loop in H which is trivial in G of necessity comes from a map Φ in $\pi_2(G/H)$. Indeed, if $h(s)$ is a loop in H which is null-homotopic in G, let $g(s,t)$ denote the homotopy in G. We can then *define* $\Phi'(s,t) = g(s,t) \cdot \phi_0$. It is not hard to convince oneself that this $\Phi'(s,t)$ gives rise to a loop in H which is homotopic to the one from which we obtained it.

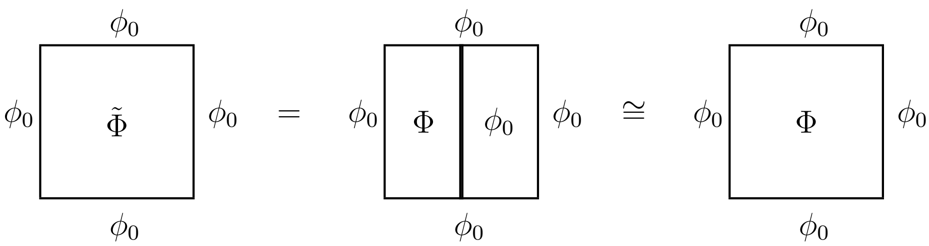
In other words, we have proven that the above sequence (6.5) is *exact* at $\pi_1(H)$, since the kernel of the arrow leaving $\pi_1(H)$ coincides with the image of the arrow entering $\pi_1(H)$. We are still not done, though: for we still have to show that the map $\pi_2(G/H) \to \pi_1(H)$ is one-to-one; that is, that if the loop $h(s)$ in H defined by $\Phi(s,t)$ is null-homotopic in H, then the map $\Phi(s,t)$ was already null-homotopic in G/H. Indeed, suppose that there exists a homotopy in H interpolating between $h(s)$ and the constant loop based at the identity; that is, that there exists a map $h(s,t)$:



Then we can compose this with $g(s,t)$ as follows:



Now define $\tilde{\Phi}(s,t) = \tilde{g}(s,t) \cdot \phi_0$. Because $h(s,t) \in H$, this map is homotopic to $\Phi(s,t)$. In fact, acting on ϕ_0 with the above maps we find:



But now notice that $\tilde{g}(s,t)$ defines an element in $\pi_2(G)$. It is now that we must invoke the result of É. Cartan mentioned in the previous section, that

$\pi_2(G) = 0$.[1] This means that there exists a homotopy $H(s,t,u)$ interpolating continuously between $\tilde{g}(s,t)$ and $\mathbb{1}$. Acting on ϕ_0, we see that $H(s,t,u)\cdot\phi_0$ provides the desired homotopy between $\Phi(s,t)$ and the constant map ϕ_0.

In summary we have proven that

$$\boxed{\pi_2(G/H) \cong \ker\left(\pi_1(H) \to \pi_1(G)\right)}\ , \tag{6.6}$$

or equivalently the exactness of the sequence:

$$0 \to \pi_2(G/H) \to \pi_1(H) \to \pi_1(G)\ .$$

In particular, if G is simply-connected, as is often the case, then every loop in G is null-homotopic, and we find that $\pi_2(G/H) \cong \pi_1(H)$.

We can illustrate this theorem with a simple example. Suppose that $G = SU(2)$ and $H = U(1)$, then we have that $SU(2)/U(1) \simeq S^2$ and $U(1) \simeq S^1$, and indeed $\pi_2(S^2) \cong \pi_1(S^1) \cong \mathbb{Z}$. As an abelian group, $\mathbb{Z}$ is freely generated by 1, hence it suffices to determine where 1 gets sent to under the map. In the above case the map $\pi_2(SU(2)/U(1)) \to \pi_1(U(1))$ sends the generator to the generator since the two groups are isomorphic, $SU(2)$ being simply-connected. On the other hand now consider $G = SO(3)$ and $H = SO(2)$. Again we have that $SO(3)/SO(2) \simeq S^2$ and $SO(2) \simeq S^1$, hence as abstract abelian groups $\pi_2(SO(3)/SO(2))$ and $\pi_1(SO(2))$ are both isomorphic to $\mathbb{Z}$, but the theorem says more. It says that the generator of $\pi_2(SO(3)/SO(2))$ cannot be sent to the generator of $\pi_1(SO(2))$, since the image of $\pi_2(SO(3)/SO(2))$ is not all of $\pi_1(SO(2))$ but only the kernel of the map $\pi_1(SO(2)) \cong \mathbb{Z} \to \mathbb{Z}_2 \cong \pi_1(SO(3))$. This map is simply reduction modulo 2 and its kernel consists of the even integers, that is, the subgroup generated by 2. Hence the generator of $\pi_2(SO(3)/SO(2))$ must get sent to twice the generator of $\pi_1(SO(2))$. This can also be understood more pictorially from the discussion surrounding Figure 6.1 and from Exercise 6.6, and it is a good exercise to do so.

Adding topological charges

We now briefly discuss to what extent we can add the topological charges of distant monopoles. It is physically intuitive, despite the fact that the equations describing monopoles are nonlinear, that one should be able to patch distant monopoles together to form a multi-monopole solution. It also seems physically intuitive that the charge of this monopole solution should

[1] There is to my knowledge no "simple" proof of this fact, but the interested reader is encouraged to go through the one in the book of Bröcker and tom Dieck [BtD85].

be given purely in terms of the charges of the constituents and not depend on the details on how the solutions were patched together. We will now see, however that this is not quite right. We will see that when the unbroken gauge group H is *disconnected* there is an ambiguity in the addition of the monopole charges.

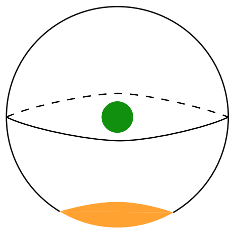


Figure 6.3: A monopole.

Consider first a monopole configuration. For convenience we will draw monopole configurations in such a way that the asymptotic sphere at spatial infinity is brought forth to a finite distance from the origin. Physically, we are assuming that the fields reach their asymptotic values to a good approximation in a finite distance. More formally, we are performing a "conformal compactification" of our space. This will be perfectly valid for our purposes, since many physical quantities (e.g., the monopole charge) will turn out to be conformally invariant. The Higgs configuration is gauge related to one which is constant almost everywhere on the asymptotic sphere. This is the so-called *unitary gauge*. More precisely, the unitary gauge is one where the Higgs field is constant *throughout* the sphere. It follows that the unitary gauge is *singular* whenever the Higgs configuration has nontrivial topological charge, since we have seen that regular gauge transformations are homotopies. For our purposes it will be sufficient to consider gauges in which the Higgs is constant almost everywhere on the sphere. Such configurations are depicted in Figure 6.3, where the Higgs is constant everywhere but in the shaded region at the south pole.

Now suppose that we have two monopoles. They are assumed to be so separated that their asymptotic spheres do not intersect. In other words in the space on and outside their two asymptotic spheres, the fields have already attained their asymptotic values. It is as if the monopoles were non-intersecting bubbles in the Higgs vacuum.

We can make a 2-monopole solution by patching together two monopoles in the following way. Let us denote by Φ_1 and Φ_2 the Higgs fields for each of the monopoles. It is of course necessary that the image of the Higgs fields lie in the same G-orbit of the manifold of vacua. In a sense different G-orbits are like different superselection sectors. We start by gauge transforming the Higgs fields in such a way that they are equal to ϕ_0 almost everywhere on their asymptotic spheres. We can further orient the monopoles in such a way that the regions in which Φ_1 and Φ_2 are allowed to fluctuate do not intersect. We can do this independently for each monopole because we can always perform gauge transformations which are "compactly supported" in the sense that they are the identity far away from the centre of the monopole. After these gauge transformations, we have a configuration where on and outside the asymptotic spheres (except for the shaded regions in Figure 6.4) the Higgs field is constant and equal to ϕ_0. In particular we have continuity in the Higgs field along the dotted line in Figure 6.4.

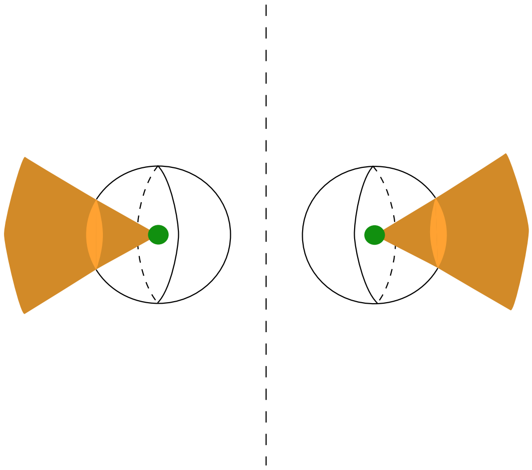


Figure 6.4: Patching two monopoles.

It is clear that the resulting field configuration is continuous, but is this procedure unambiguous? Suppose that we had used a different set of gauge transformations in order to make Φ_1 and Φ_2 equal almost everywhere asymptotically. Once the Higgs are set to ϕ_0 we are still allowed to make an gauge transformation in the stability subgroup H. Would the resulting two-monopole configuration be homotopic (i.e, gauge-equivalent) to the one resulting from our first attempt at patching? Clearly if H is connected, then we can always make a homotopy so that any discontinuity in the H-gauge transformation can be undone. But how about if H is disconnected? In this case there is a potential ambiguity in the patching prescription.

Another way to understand this is as follows. The question boils down to whether $\pi_2(G/H, \phi_1)$ and $\pi_2(G/H, \phi_2)$ can be composed meaningfully. Because G is path-connected, we know that G/H is path-connected, hence $\pi_2(G/H, \phi_1) \cong \pi_2(G/H, \phi_1)$. But remember that this isomorphism depends on the path used to connect ϕ_1 and ϕ_2. If all such paths were homotopic—that is, if G/H were simply-connected—then all such isomorphisms would be one and the same and we could unambiguously compose elements in $\pi_2(G/H, \phi_1)$ and $\pi_2(G/H, \phi_2)$. In other words, if G/H were simply-connected, then we could add without ambiguity the topological charges of each of the monopoles constituting a given two-monopole solution to derive its charge. It turns out, thanks to Theorem 16.11 in [Ste51], that it is enough to check that H be connected. If H is connected, the theorem states, that the isomorphism $\pi_2(G/H, \phi_1) \cong \pi_2(G/H, \phi_2)$ is independent of the path used to go between ϕ_1 and ϕ_2. If H is not connected, however, there is a potential ambiguity. We *can* patch separated monopoles together, but the topological charge of the resulting two-monopole configuration will not be given simply in terms of the topological charges of its constituents. We need more information: namely the details on how the solutions were put together.

6.2 The Dirac quantisation condition

In this section we start the analysis of the generalised Dirac quantisation condition obeyed by these monopole solutions. The results in this section are based on the seminal paper of Goddard, Nuyts and Olive [GNO77].

We start by considering a monopole in the unitary gauge, where the Higgs field Φ is constant and equal to ϕ_0, say, almost everywhere on the asymptotic sphere. Looking back at Figure 6.3, we have $\Phi = \phi_0$ everywhere on the sphere but on the shaded region around the south pole. Because $D_i\Phi = 0$ everywhere on the sphere, on the part where Φ is constant, this condition becomes $\mathsf{W}_i \cdot \Phi = 0$, hence W_i takes values in the Lie algebra $\mathfrak{h}$ of the stability subgroup $H \subset G$, and so hence so does the path-ordered exponential in equation (6.3).

We will now assume that the field strength G_{ij} has the following asymptotic form:

$$\mathsf{G}_{ij} = \epsilon_{ijk} \frac{x^k}{|x|^3} \frac{\mathsf{Q}(x)}{4\pi} \ , \tag{6.7}$$

where the *magnetic charge* $\mathsf{Q}(x)$ is Lie algebra valued, and hence takes values in $\mathfrak{h}$ almost everywhere on S^2_∞. It may seem surprising at first that the magnetic charge is *not* constant; but in the presence of a non-abelian gauge symmetry, constancy is not a gauge-invariant statement. The correct non-abelian generalisation of constancy is covariantly constant; and, as the next exercise asks you to show, this is indeed the case.

Exercise 6.12 (The magnetic charge is covariantly constant)
Prove that $D_i\mathsf{Q}(x) = 0$ on the sphere.
(Hint: Analyse the Bianchi identity and the equations of motion on the asymptotic sphere in the Ansatz (6.7) and show that

$$\textit{Bianchi identity} \Rightarrow x^k D_k \mathsf{Q}(x) = 0$$
$$\textit{equation of motion} \Rightarrow \epsilon_{ijk} x^k D_j \mathsf{Q}(x) = 0 \ .$$

Deduce that these two equations together imply that $D_k\mathsf{Q} = 0$.)

The quantisation condition will come from demanding that the map $h(s)$ defined in (6.4) does indeed trace a loop in H, so that $h(1) = h(0)$. To make this condition into something amenable to computation we will derive another expression for $h(1)$ in the unitary gauge. This will bring to play a non-abelian version of Stoke's theorem.

Let us define the group element $g(s,t)$ by $\Phi(s,t) = g(s,t) \cdot \phi_0$ in (6.3). Because $D_i\Phi = 0$, it follows that $D_i g(s,t) = 0$, where the covariant derivative is now in the adjoint representation. Define the covariant derivative along the curves of constant s, by $D_t = \frac{\partial x^i}{\partial t} D_i$, and the covariant derivative along the curves of constant t, by $D_s = \frac{\partial x^i}{\partial s} D_i$. The next exercise asks you to prove the non-abelian Stoke's theorem.

Exercise 6.13 (The non-abelian Stoke's theorem)
The point of this exercise is to prove the following formula.

$$h(s)^{-1} \frac{dh(s)}{ds} = -e \int_0^1 dt\, g(s,t)^{-1} \mathsf{G}_{ij} g(s,t) \frac{\partial x^i}{\partial t} \frac{\partial x^j}{\partial s} \ . \tag{6.8}$$

We proceed in steps:

1. *Use the fact that $D_t \Phi(s,t) = 0$ implies $D_t g(s,t) = 0$, provided the curve $s =$ constant lies in the region where $\Phi(s,t) = \phi_0$ is constant, and prove that this is equivalent to $D_t \circ g(s,t) = g(s,t) \circ \partial_t$ as operators.*

2. *Show that*

$$\begin{aligned} \partial_t \left(g(s,t)^{-1} D_s g(s,t) \right) &= g(s,t)^{-1} [D_t, D_s] g(s,t) \\ &= -e \frac{\partial x^i}{\partial t} \frac{\partial x^j}{\partial s} g(s,t)^{-1} \mathsf{G}_{ij} g(s,t) \ . \end{aligned}$$

3. *Finally integrate the above expression over $t \in [0,1]$ and use the fact that $g(s,t)^{-1} D_s g(s,t)\big|_{t=0} = 0$ because $g(s,0) = 1$ and $x^i(s,0)$ is constant for all s, and that $g(s,t)^{-1} D_s g(s,t)\big|_{t=0} = h(s)^{-1} \frac{dh(s)}{ds}$.*

With the Ansatz (6.7) for G_{ij} we now have that

$$g(s,t)^{-1} \mathsf{G}_{ij} g(s,t) = \frac{1}{4\pi} \epsilon_{ijk} \frac{x^k}{|x|^3} g(s,t)^{-1} \mathsf{Q}(x) g(s,t) \ .$$

But now notice that because $\mathsf{Q}(x)$ is covariantly constant,

$$g(s,t)^{-1} \mathsf{Q}(x(s,t)) g(s,t) = \mathsf{Q}(x(0,0)) \equiv \mathsf{Q} \in \mathfrak{h} \ .$$

In other words,

$$h(s)^{-1} \frac{dh(s)}{ds} = -\frac{e}{4\pi} \mathsf{Q} \int_0^1 dt\, \epsilon_{ijk} \frac{x^k}{|x|^3} \frac{\partial x^i}{\partial t} \frac{\partial x^j}{\partial s} \ ,$$

which can be trivially solved for $h(s)$ to yield:

$$h(s) = \exp \left[-\frac{e}{4\pi} \mathsf{Q} \int_0^s ds' \int_0^1 dt\, \epsilon_{ijk} \frac{x^k}{|x|^3} \frac{\partial x^i}{\partial t} \frac{\partial x^j}{\partial s'} \right] \cdot h(0) \ ,$$

and in particular

$$h(1) = \exp\left[-\frac{e}{4\pi}\mathsf{Q}\int_0^1 ds \int_0^1 dt\, \epsilon_{ijk}\frac{x^k}{|x|^3}\frac{\partial x^i}{\partial t}\frac{\partial x^j}{\partial s}\right]\cdot h(0) \ , \tag{6.9}$$

where $\exp : \mathfrak{h} \to H$ is the exponential map. Alternatively, if you are more familiar with matrix groups, you can embed H inside a matrix group (every Lie group has a faithful finite-dimensional matrix representation) and then the above differential equation for the matrix $h(s)$ is solved by the above expression, but where exp of a matrix is now defined by its power series.

Let us first compute the above integral. Notice that because the integrand is invariant under rescalings of x we can evaluate it on the unit sphere in $\mathbb{R}^3$; that is, we take $|x| = 1$. We then rewrite it in a more invariant looking form. To this end, it suffices to notice that the integrand is the pull back via the embedding $S^2 \to \mathbb{R}^3$, $(s,t) \mapsto x^i(s,t)$ of the form $\omega = \frac{1}{2}\epsilon_{ijk}x^k\, dx^i \wedge dx^j$, whose exterior derivative $d\omega = \frac{1}{2}\epsilon_{ijk}\, dx^i \wedge dx^j \wedge dx^k$ is precisely 3 times the volume form in $\mathbb{R}^3$ relative to the standard euclidean metric. Therefore using Stoke's theorem and understanding the unit sphere in $\mathbb{R}^3$ as the boundary of the unit ball $S^2 = \partial B^3$, we have that

$$\int_{\partial B^3} \omega = \int_{B^3} d\omega = 3\,\mathrm{vol}(B^3) = 4\pi \ .$$

We can then rewrite equation (6.9) as follows

$$h(1) = \exp\left[-e\mathsf{Q}\right]\cdot h(0) \ ,$$

whence the the Dirac quantisation condition $h(1) = h(0)$, becomes

$$\boxed{\exp e\mathsf{Q} = \mathbb{1} \in H} \ . \tag{6.10}$$

Before undertaking a general analysis of this equation, let us make sure that it reduces to the familiar condition (1.6) for $G = SO(3)$ and a nonzero Higgs field in the adjoint representation, as was the case in Chapter 1. The adjoint representation of $SO(3)$ is three-dimensional. Choose the Higgs field to point in the z-direction. The stability subgroup if the subgroup of rotations about the z-axis. It is the $SO(2)$ subgroup consisting of matrices of the form

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \ ,$$

where θ runs from 0 to 2π. The magnetic charge Q is given by:

$$\mathsf{Q} = g \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

where g is what appears in (1.5). With these definitions, we see that the Dirac quantisation condition (6.10) becomes

$$\exp eg \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(eg) & \sin(eg) & 0 \\ -\sin(eg) & \cos(eg) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbb{1} ,$$

whence $eg \in 2\pi\mathbb{Z}$ in complete agreement with (1.6). We are clearly on the right track. In order to analyse the Dirac quantisation condition properly we will need quite a bit of technology concerning compact Lie groups. This is the purpose of the following section. Readers who already know this material are encouraged to skim through the section for notation.

6.3 Some facts about compact Lie groups and Lie algebras

In this section we collect without proof those results from the theory of compact Lie groups that are relevant for the analysis of the Dirac quantisation condition. There are many fine books on the subject. A quick and efficient introduction to the main results can be found in the second chapter of Pressley and Segal's book on loop groups [PS86]. A fuller treatment of the parts we will need can be can be found in the book by Adams [Ada69] and also in the more comprehensive book by Bröcker and tom Dieck [BtD85]. For the results on Lie algebras we have followed the book by Humphreys [Hum72].

6.3.1 Compact Lie groups

Suppose that G is a compact connected Lie group. Any connected abelian subgroup is clearly a torus. Let T be a fixed maximal *connected* abelian subgroup of G; that is, a *maximal torus*. Maximal tori obviously exist because any one-parameter subgroup is a connected abelian subgroup. One of the key theorems in the structure of compact Lie groups is the fact that all maximal tori are conjugate in G. This implies, in particular, that the dimension of all maximal tori are the same: it is an invariant of G known as the *rank* of G. Another way to rephrase this theorem is that any element in G is conjugate to an element in T, or simply that every group element in G lies in some maximal torus. Generic elements will lie in just one maximal torus: these are called *regular* elements, whereas there exist also *singular* elements which lie in more than one.

The prototypical compact connected Lie group is $U(n)$ and many of the results in the theory of compact Lie groups, when restricted to $U(n)$, reduce to well-known facts. For instance, a maximal torus in $U(n)$ can be taken to be the set of diagonal matrices; hence the rank of $U(n)$ is n, and the rank for the $SU(n)$ subgroup is $n-1$. The theorem about maximal tori being conjugate, is simply the fact that any set of commuting unitary matrices can be simultaneously diagonalised by a unitary transformation. The regular elements are those matrices which have distinct eigenvalues.

Let $\mathfrak{g}$ and $\mathfrak{t}$ denote the Lie algebras of G and T, respectively. $\mathfrak{t}$ is a *maximal toral subalgebra*. A lot can be learned about G by studying the action of T on $\mathfrak{g}$. Because T is abelian, any finite-dimensional complex representation is completely reducible into one-dimensional representations. But $\mathfrak{g}$ is a real representation, so we complexify it first: define $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes \mathbb{C}$, and extend the action of G (and hence the one of T) complex-linearly. We can now decompose $\mathfrak{g}_{\mathbb{C}}$ as representations of T as follows:

$$\mathfrak{g}_{\mathbb{C}} = \mathfrak{t}_{\mathbb{C}} \oplus \left(\bigoplus_{\alpha} \mathfrak{g}_\alpha \right) ,$$

where $\mathfrak{t}_{\mathbb{C}} = \mathfrak{t} \otimes \mathbb{C}$ is the subspace on which T acts trivially, and $\mathfrak{g}_\alpha$ is the subspace of $\mathfrak{g}_{\mathbb{C}}$ defined as follows:

$$v \in \mathfrak{g}_\alpha \Leftrightarrow \exp X \cdot v = e^{i\alpha(X)} v ,$$

where $X \in \mathfrak{t}$ and $\alpha : \mathfrak{t} \to \mathbb{R}$ is a real linear function. The α's appearing in the above decomposition are known as the *(infinitesimal) roots* of G. Notice that if α is a root, so is $-\alpha$ since if $v \in \mathfrak{g}_\alpha$ its complex conjugate $\bar{v} \in \mathfrak{g}_{-\alpha}$.

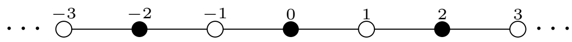
The complexified Lie algebra of $U(n)$ is the Lie algebra of all $n\times n$ complex matrices. The roots are given by α_{ij} where $1 \leq i, j \leq n$, $i \neq j$, and the root subspace corresponding to α_{ij} is spanned by the matrices E_{ij} with a 1 in the (ij) entry and zeroes everywhere else. Acting on the diagonal matrix $X = \operatorname{diag}(x_1, x_2, \ldots, x_n) \in \mathfrak{t}$, $\alpha_{ij}(X) = x_i - x_j$.

Let us think of $U(1)$ as the group of complex numbers of unit norm. A homomorphism $\chi : T \to U(1)$ is called a *character* of T. Characters can be multiplied pointwise and indeed form a group called the *character group* of T. Characters are uniquely determined by their derivatives at the identity. In other words, if χ is a character and $\exp X$ belongs to T, then

$$\chi(\exp X) = e^{iw(X)} , \tag{6.11}$$

where $w \in \mathfrak{t}^*$ is an *infinitesimal character* or a *weight*. The set of infinitesimal characters define a lattice in $\mathfrak{t}^*$ called the *weight lattice* of G and denoted $\Lambda_w(G)$. The roots are particular examples of weights, and taking integer linear combinations of the roots, we obtain a sublattice of the weight lattice known as the *root lattice* and denoted $\Lambda_r(G)$. The root lattice only depends on the Lie algebra, whence Lie groups sharing the same Lie algebra have the same root lattice. On the other hand, the weight lattice identifies the Lie group. If G is semisimple, then both the weight and root lattices span $\mathfrak{t}^*$. It means that the quotient $\Lambda_w(G)/\Lambda_r(G)$ is a finite abelian group. We will see later that it is the fundamental group of the dual group of G.

We can illustrate this with $SU(2)$ and $SO(3)$. The weights of $SU(2)$ form a one-dimensional integral lattice isomorphic to $\mathbb{Z}$, shown below. The weight $m \in \mathbb{Z}$ corresponds to twice the "magnetic quantum number," since for $SU(2)$ the magnetic quantum number, like the spin, can be half-integral. In the case of $SO(3)$ only integral spin representations can occur, hence its weight lattice (shown below with filled circles) corresponds to the sublattice consisting of even integers:



In the semisimple case, $\mathfrak{t}^*$ is called the *root space* of $\mathfrak{g}$. On the other hand, if G is not semisimple, the roots will only span a subspace of $\mathfrak{t}^*$ which is then the root space of its maximal semisimple subalgebra $[\mathfrak{g}, \mathfrak{g}] \subset \mathfrak{g}$. This shows that the Lie algebra of a compact Lie group is *reductive*; that is, the direct product of a semisimple Lie algebra and an abelian algebra—namely, its centre.

From the definition of the roots of $U(n)$ above we see that they don't span $\mathfrak{t}^*$, since they annihilate the scalar matrices. This is to be expected since the scalar matrices are in the centre of the Lie algebra $u(n)$ of $U(n)$. The traceless matrices in $u(n)$ span the complement of the scalars matrices and generate the Lie algebra $su(n)$ of $SU(n)$, which is semisimple (in fact, simple). The space spanned by the roots is the root space of $su(n)$.

The root subspaces $\mathfrak{g}_\alpha$ are one-dimensional. Choose vectors $e_\alpha \in \mathfrak{g}_\alpha$ such that $e_{-\alpha} = \bar{e}_\alpha$. Then e_α, $e_{-\alpha}$ and their bracket $h_\alpha = -i[e_\alpha, e_{-\alpha}] \in \mathfrak{t}$ define an embedding of $sl(2, \mathbb{C})$ in $\mathfrak{g}_{\mathbb{C}}$:

$$[h_\alpha, e_\alpha] = 2ie_\alpha \qquad [h_\alpha, e_{-\alpha}] = -2ie_{-\alpha} \qquad \text{and} \qquad [e_\alpha, e_{-\alpha}] = ih_\alpha \ .$$

Explicitly the embedding is given by $e \mapsto e_\alpha$, $f \mapsto e_{-\alpha}$ and $h \mapsto h_\alpha$, where

$$e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad f = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \qquad \text{and} \qquad h = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \ .$$

It therefore follows that $\exp(2\pi h_\alpha) = \mathbb{1}$. It also follows from the representation theory of $sl(2,\mathbb{C})$ that for any root $\beta \in \mathfrak{t}^*$, $\beta(h_\alpha) \in \mathbb{Z}$ and that, in particular, $\alpha(h_\alpha) = 2$. The h_α are known as *coroots* and their integer linear combinations span a lattice in $\mathfrak{t}$ called the *coroot lattice* and denoted $\Lambda_r^\vee(G)$. If Λ is a lattice, then the dual lattice is the set of linear functions $\Lambda \to \mathbb{Z}$ and is denoted Λ^*. This relation is reflexive because $\Lambda^{**} = \Lambda$. In this notation we now see that the coroot lattice is a sublattice of the dual root lattice: $\Lambda_r^\vee(G) \subseteq \Lambda_r(G)^*$. We will see later that the two lattices will agree when G is simply-connected.

Despite the fact that the coroot lattice lives naturally in $\mathfrak{t}$, one often sees in the literature where the coroot lattice is a lattice in $\mathfrak{t}^*$, just like the root and weight lattices. In my opinion this causes more confusion than it is worth, but for the sake of comparison let us see how this goes. In order to identify $\mathfrak{t}$ and $\mathfrak{t}^*$ we need a new piece of information: namely, a metric. We saw in Exercise 6.1 that the Lie algebra of every compact Lie group has an invariant metric, so we will fix one such G-invariant metric $(-,-)$ on $\mathfrak{g}$. Its restriction to T will also be denoted $(-,-)$. Because this metric is invariant and non-degenerate, we can use it to identify $\mathfrak{t}$ and $\mathfrak{t}^*$. In particular there is an element $\alpha^\vee \in \mathfrak{t}^*$ such that for all $X \in \mathfrak{t}$, $\alpha^\vee(X) = (h_\alpha, X)$. In terms of the root α, we have that $\alpha^\vee = 2\alpha/(\alpha,\alpha)$. We call $\alpha^\vee$ the *inverse root* corresponding to the root α. Taking integer linear combinations of the coroots, we span the *inverse root lattice* of G. The weight, root and inverse root lattices are all subsets of $\mathfrak{t}^*$, but notice that whereas the weight and root lattices are intrinsic, the inverse root lattice depends on the chosen metric. In particular the inverse root lattice cannot be meaningfully compared with either the root or weight lattices, since we can scale it at will by rescaling the metric. As we do not wish to advocate its use, we will not give it a symbol, but the reader should beware that sometimes the symbol we use for the coroot lattice is reserved for the inverse root lattice, relative to some "standard" metric.

Using the metric on $\mathfrak{t}^*$ one can measure lengths of roots, and it can be proven that if $\mathfrak{g}$ is simple, then there are at most two lengths of roots, called *long* and *short* roots. Simple Lie algebras for which all roots are the same length are called *simply-laced*. For these simple Lie algebras, we can choose the metric so that $(\alpha,\alpha) = 2$ for all roots. Under this metric, the roots and the inverse roots agree.

6.3.2 The Weyl group

Because the maximal torus T is abelian, conjugation by elements of T is trivial. Moreover generic elements of G will conjugate T to another maximal torus. However there are some elements of G which conjugate T back to T. The largest such subgroup of G is called the *normaliser* of T and is denoted $N(T)$; that is,

$$N(T) = \left\{h \in G \mid hTh^{-1} = T\right\} \ .$$

It follows from this definition that $N(T)$ is indeed a subgroup of G and that T is contained in $N(T)$ as a normal subgroup. Because $T \subset N(T)$ is a normal subgroup, it follows that $N(T)/T$ is a group. This group is the *Weyl group* of G relative to the maximal torus T. It is the group of symmetries of the maximal torus. Although it is defined relative to T, the Weyl group $N(T')/T'$ corresponding to any other maximal torus T' is conjugate (and hence isomorphic) to $N(T)/T$. Hence it makes sense to talk about *the* Weyl group W of G, up to isomorphism.

The Weyl group W is a finite group, generated by reflections corresponding to the roots. More precisely, if α is a root, then consider the group element

$$\exp \frac{\pi}{2} \left(e_\alpha + e_{-\alpha}\right) \in N(T) \ .$$

The adjoint action of this group element on $\mathfrak{t}$ corresponds to a reflection ρ_α on the *reflection hyperplane* $H_\alpha \subset \mathfrak{t}$ defined by $H_\alpha = \{X \in \mathfrak{t} \mid \alpha(X) = 0\}$. Indeed, one computes that for all $X \in \mathfrak{t}$,

$$\rho_\alpha(X) = X - \alpha(X) h_\alpha \ .$$

It can be proven that the ρ_α generate W.

For example, the Weyl group of $U(n)$ is $\mathfrak{S}_n$, the symmetric group in n objects, and it acts by permuting the entries of the diagonal matrices in $\mathfrak{t}$. This is also the Weyl group of $SU(n)$ since the roots of $U(n)$ are the roots of $SU(n)$.

Elements of $\mathfrak{t}$ not belonging to any hyperplane H_α are called *regular*; whereas those who are not regular are called *singular*. Regular elements fall into connected components called *Weyl chambers*. The Weyl group permutes the Weyl chambers and no two elements in the same Weyl chamber are Weyl-related. Fix a Weyl chamber C and call it *positive*. The roots can then be split into two sets, *positive* and *negative* roots, according to whether they are positive or negative on C—they cannot be zero, because C does not intersect any hyperplane H_α. A positive root is called *simple* if H_α is a wall of C. If G is a simple group of rank ℓ, then there are ℓ simple roots. Every positive root is a linear combination of the simple roots with non-negative integer coefficients, hence the simple roots generate the root lattice, and their associated reflections generate the Weyl group. The positive Weyl chamber is sometimes called the *fundamental Weyl chamber*. Its closure (that is, including the walls) is a fundamental domain for the action of the Weyl group on $\mathfrak{t}$: every point in $\mathfrak{t}$ is Weyl-related to a unique point in the closure of the fundamental Weyl chamber.

In the case of $U(n)$, a choice of fundamental Weyl chamber consists in choosing diagonal matrices whose entries are ordered in a particular way. For instance, we can choose a descending order, in which case the positive roots of $U(n)$ are the α_{ij} with $i < j$. The simple roots are then clearly the $\alpha_{i,i+1}$.

Again using the metric on $\mathfrak{t}$ there is a dual picture of this construction in $\mathfrak{t}^*$, where the hyperplanes H_α now are defined as the hyperplanes perpendicular to the roots. This picture is independent of the metric since the notion of perpendicularity does not depend on the choice of G-invariant metric on $\mathfrak{t}$. The Weyl group acts on $\mathfrak{t}^*$ and it is once again generated by reflections associated to every root. If α and β are roots, we have

$$\rho_\alpha(\beta) = \beta - (\beta, \alpha^\vee)\,\alpha = \beta - (\beta, \alpha)\,\alpha^\vee \ , \tag{6.12}$$

where $\alpha^\vee = 2\alpha/(\alpha, \alpha)$ is the inverse root. Once again the complement of the hyperplanes is divided into connected components called the *dual Weyl chambers* and one any one of them can be chosen to be the *positive* or *fundamental* dual Weyl chamber. The walls of the fundamental dual Weyl chamber are the hyperplanes perpendicular to the simple roots. Once again the closure of the fundamental dual Weyl chamber is a fundamental domain for the action of the Weyl group in $\mathfrak{t}^*$.

Those weights of G which lie in the closure of the fundamental dual Weyl chamber are called *dominant*. We write this set $\Lambda_w^+(G)$. It is a semigroup of $\Lambda_w(W)$; that is, if w_1 and w_2 are dominant, so is their sum $w_1 + w_2$, but there are no inverses. Every irreducible representation of G has a unique *highest weight* which is dominant. Therefore $\Lambda^+(W)$ is in one-to-one correspondence with the set of finite-dimensional irreducible representations of G.

6.3.3 Root systems and simple Lie algebras

We have seen that the Lie algebra of a compact Lie group is reductive. Because semisimple Lie algebras split in turn into their simple factors, we see that the Lie algebra of a compact Lie group is a direct sum of abelian and simple Lie algebras. This does not mean that any compact Lie group is the direct product of simple Lie groups and a torus, but it turns out that it is covered finitely by a compact Lie group of this type. Hence to a large extent it is enough to study simple Lie groups and abelian Lie groups separately and only at the end put the structures together. Let us therefore assume that G is a simple Lie group. It is a remarkable fact that compact simple Lie groups are essentially classified up to “finite ambiguity” by their root systems. We will begin to describe this process now.

First of all we need to axiomatise the notion of a root system. A subset Φ of a euclidean space E is a *root system* if the following conditions are obeyed:

1. Φ is finite, spans E and does not contain the origin;
2. If $\alpha \in \Phi$, then $-\alpha \in \Phi$ and no other multiples of α are in Φ;
3. The reflections ρ_α (see (6.12)) leave Φ invariant; and
4. For all $\alpha, \beta \in \Phi$, $(\alpha^\vee, \beta) \in \mathbb{Z}$, where $(-,-)$ is the metric in E.

This last condition is extremely restrictive. It essentially says that only very few angles can occur between roots. Indeed, notice that $(\alpha^\vee, \beta) = 2|\beta|/|\alpha| \cos\vartheta$, where ϑ is the angle between α and β. Now, $(\alpha^\vee, \beta)(\beta^\vee, \alpha) = 4\cos^2\vartheta$ is a non-negative integer. Taking into account that $(\alpha^\vee, \beta)$ and $(\beta^\vee, \alpha)$ have the same sign we are left with the possibilities listed in Table 6.1, where we have chosen $(\alpha, \alpha) < (\beta, \beta)$ for definiteness and have omitted the trivial case $\alpha = \pm\beta$.

$(\alpha, \beta^\vee)$	$(\alpha^\vee, \beta)$	ϑ	$(\beta, \beta)/(\alpha, \alpha)$
0	0	$\pi/2$	undetermined
1	1	$\pi/3$	1
-1	-1	$2\pi/3$	1
1	2	$\pi/4$	2
-1	-2	$3\pi/4$	2
1	3	$\pi/6$	3
-1	-3	$5\pi/6$	3

Table 6.1: Allowed angles between roots in a root system.

If a root system Φ admits a split $\Phi = \Phi_1 \cup \Phi_2$ into disjoint sets so that every element of Φ_1 is orthogonal to every element of Φ_2, we say that it is *reducible*; otherwise it is *simple*. The simple root systems have been classified. There are four infinite families: A_ℓ, for $\ell \geq 1$, B_ℓ and C_ℓ for $\ell \geq 2$, and D_ℓ for $\ell \geq 3$; and five exceptional root systems G_2, F_4, E_6, E_7 and E_8. There are two “accidental” isomorphisms in the above list: $B_2 = C_2$ and $A_3 = D_3$. In all cases, the subscript indicates the rank. The simply-laced root systems are those in the A, D and E series. They are listed in Table 6.2 in a graphical notation that will be explained shortly.

Notice that the above definition of a root system is symmetrical with respect with the interchange $\alpha \leftrightarrow \alpha^\vee$ of a root and the inverse root. In particular this shows that the set $\Phi^\vee \subset \mathsf{E}$ consisting of the dual roots $\alpha^\vee$ is again a root system, and that it is simple if Φ is. In this case $\Phi^\vee$ must be again one of the simple root systems listed above. From the definition of $\alpha^\vee$, it follows that for simply-laced root systems (where all roots have the same length) we can choose the metric so that $\alpha^\vee = \alpha$, hence simply laced root systems are self-dual. More generally, since $(\alpha^\vee, \alpha^\vee) = 4/(\alpha, \alpha)$, long and short roots are interchanged. A quick look at Table 6.2 reveals that G_2 and F_4 are also self-dual, whereas $B_\ell^\vee = C_\ell$, and viceversa, since $\Phi^{\vee\vee} = \Phi$.

Let $\alpha_1, \alpha_2, \ldots, \alpha_\ell$ be a set of simple roots. Let $\alpha_i^\vee = 2\alpha_i/(\alpha_i, \alpha_i)$. Then the inner product $a_{ij} = (\alpha_i, \alpha_j^\vee)$ is an integer. The set $\{a_{ij}\}$ of all such integers are called the *Cartan integers* and the matrix (a_{ij}) is known as the *Cartan matrix*. They are independent of the invariant metric chosen for $\mathfrak{t}$. If two complex simple Lie algebras have the same Cartan matrix, then they are isomorphic. The Cartan matrices of the simple Lie algebras are listed in [Hum72], for example.

There is also a graphical notation for root systems. Let Φ be a root system of rank ℓ, and let $\alpha_1, \alpha_2, \ldots, \alpha_\ell$ be a set of simple roots. The *Coxeter graph* of Φ is the graph consisting of ℓ vertices and such that the ith vertex is joined to the jth vertex by $(\alpha_i, \alpha_j^\vee)(\alpha_i^\vee, \alpha_j)$ lines. From Table 6.1, we know that this number can be 0, 1, 2 or 3. The Coxeter graph can be shown to determine the Weyl group, but does not determine the root system because when two vertices are connected by more than one line, it fails to tell us which of the two vertices corresponds to the shorter root. In other words, the Coxeter graph cannot tell between Φ and $\Phi^\vee$. In order to distinguish them it is necessary to decorate the diagram further: we colour those vertices which corresponds to the short roots, if any are present. The resulting diagram is called the *Dynkin diagram*. The Dynkin diagrams corresponding to the simple root systems are listed in Table 6.2, the vertex labelled i corresponds to α_i, and the filled vertices corresponds to the short roots.

Reconstructing the group

From the above discussion about compact Lie groups it follows that the root system associated to a compact simple Lie group is simple. Hence it has to be one of the roots systems listed above. This prompts the question of the reconstruction of the group from the root system. It turns out that this is possible up to a finite ambiguity. In a nutshell, given the root system of a compact group, one can obtain a finite covering group of the group in question. In this section we will consider only simple Lie groups.

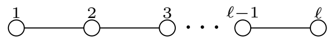
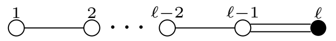
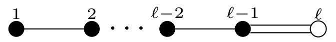
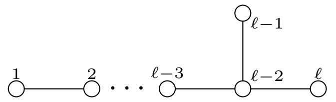
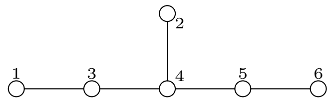
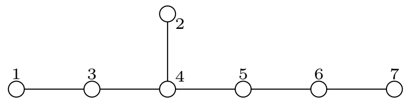
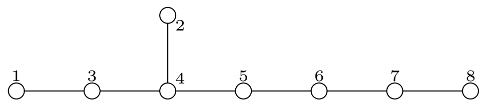
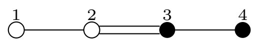
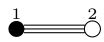
Φ	Dynkin diagram
A_ℓ	
B_ℓ	
C_ℓ	
D_ℓ	
E_6	
E_7	
E_8	
F_4	
G_2	

Table 6.2: Dynkin diagrams of the simple root systems.

Given an simple root system Φ of rank ℓ, one can construct a unique simple complex Lie algebra. Associated with each simple root α_i there exist three generators $e_i = e_{\alpha_i}$, $f_i = e_{-\alpha_i}$ and $h_i = h_{\alpha_i}$. These 3ℓ elements, subject to the so-called Serre relations (see [Hum72], for example) generate a complex simple Lie algebra $\mathfrak{g}_{\mathbb{C}}$, where $\mathfrak{t}_{\mathbb{C}}$ is spanned by the h_i and whose root system relative to this maximal torus is Φ. As a linear space, the Lie algebra will be generated by ℓ elements h_i spanning the *Cartan subalgebra* and generators e_α for each root α, whose Lie brackets can be written down as follows:

$$[e_\alpha, e_\beta] = \begin{cases} n_{\alpha\beta} e_{\alpha+\beta} & \text{if } \alpha+\beta \text{ is a root;} \\ h_\alpha & \text{if } \beta = -\alpha\text{; and} \\ 0 & \text{otherwise,} \end{cases}$$

where h_α belongs to the Cartan subalgebra. Furthermore, for every h in the Cartan subalgebra, $[h, e_\alpha] = \alpha(h) e_\alpha$. It is possible to choose a basis in which the $n_{\alpha\beta}$ are all nonzero integers. These integers obey $n_{\alpha\beta} = -n_{-\alpha,-\beta}$.

Every complex simple Lie algebra has in general several real forms; that is, real subalgebras. Among these real forms there is a unique *compact real form*; that is, one for which the Killing form is negative-definite. It is easy to write this form down explicitly. It is generated over the reals by the ih_α, and the combinations

$$\frac{i}{\sqrt{2}}(e_\alpha + e_{-\alpha}) \qquad \text{and} \qquad \frac{1}{\sqrt{2}}(e_\alpha - e_{-\alpha}) \ .$$

It is easy to check that the real linear combinations of these generators close under the Lie bracket. It is also easy to compute the Killing form and see that it is indeed negative-definite.

Next, each compact real form $\mathfrak{g}$ of a simple Lie algebra is the Lie algebra of a unique simply-connected compact simple Lie group $\tilde{G}$, whose maximal torus T is obtained by exponentiating the $\{ih_\alpha\}$, and whose root system relative to T agrees with the one of G. Therefore we have almost come full circle. I say almost, because we are left with a simply-connected compact simple Lie group, even though we started with a compact simple Lie group G which was not assumed to be simply-connected. Therefore we need more information. The information we need is of course the fundamental group $\pi_1(G)$ of G, which is a finite subgroup of the centre of $\tilde{G}$. Remarkably, the centre of $\tilde{G}$ can be read off simply from Lie algebraic data. We review this now.

The centre of $\tilde{G}$

Let $\{\alpha_i\}$ for $i = 1, \ldots, \ell$ denote the simple roots of G relative to a maximal torus T. We define ℓ *fundamental weights* $\{\lambda_i\}$ by the requirement: $(\lambda_i, \alpha_j^\vee) = \delta_{ij}$. Alternatively, $\lambda_i(h_{\alpha_j}) = \delta_{ij}$. In other words, the fundamental weights generate a lattice which is dual to the coroot lattice $\Lambda_r^\vee(G)$. The lattice generated by the fundamental weights is the weight lattice of the simply-connected Lie group $\tilde{G}$: $\Lambda_w(\tilde{G})$. This lattice contains the root lattice $\Lambda_r(\tilde{G})$ as a sublattice, and the quotient $\Lambda_w(\tilde{G})/\Lambda_r(\tilde{G})$ is a finite abelian group isomorphic to the centre $Z(\tilde{G})$ of $\tilde{G}$. It is sometimes called the *fundamental group* of the root system, since it is the fundamental group of the *adjoint group*, the Lie group whose weight lattice agrees with its root lattice.

Let us now explain why $\Lambda_w(\tilde{G})/\Lambda_r(\tilde{G})$ is isomorphic to the centre $Z(\tilde{G})$ of $\tilde{G}$. First of all, notice that $Z(\tilde{G})$ is contained in every maximal torus of $\tilde{G}$. In fact, $Z(\tilde{G})$ *is* the intersection of all the maximal tori of $\tilde{G}$.

Exercise 6.14 (The centre is the intersection of all the maximal tori) *Prove that the centre $Z(\tilde{G})$ is the intersection of all the maximal tori of $\tilde{G}$.*

(Hint: Use that any element can be conjugated to any given maximal torus and the fact that an element of the centre is invariant under conjugation.)

Fix a maximal torus $\tilde{T}$ of $\tilde{G}$ and let $\exp : \mathfrak{t} \to \tilde{T}$ be the restriction of the exponential map to $\mathfrak{t}$. Because $\tilde{T}$ is abelian, the exponential map is a homomorphism of abelian groups. We find it convenient in what follows to include a factor of 2π in the exponential map. We will introduce then the *reduced exponential map*, denoted $\underline{\exp}$, and defined by $\underline{\exp}\, X = \exp 2\pi i\, X$. Clearly $\underline{\exp}$ is also a group homomorphism and, in particular, its kernel is a lattice $\Lambda_I(\tilde{G})$, called the *integer lattice* of $\tilde{G}$. The reduced exponential map yields an isomorphism $\tilde{T} \cong \mathfrak{t}/\Lambda_I(\tilde{G})$, whence we see that the integer lattice is the lattice of periods of the maximal torus $\tilde{T}$. It follows from (6.11) that h belongs to the integer lattice if and only if for every weight w of $\tilde{G}$, $w(h) \in \mathbb{Z}$. In other words, the integer lattice and the weight lattice are dual:

$$\Lambda_I(\tilde{G}) = \Lambda_w(\tilde{G})^* \ .$$

Let $\Lambda_Z(\tilde{G}) = \underline{\exp}^{-1} Z(\tilde{G})$ denote those elements of $\mathfrak{t}$ which the reduced exponential map sends to the centre of $\tilde{G}$. $\Lambda_Z(\tilde{G})$ too is a lattice called the *central lattice* of $\tilde{G}$, which by definition contains the integer lattice. Because $\underline{\exp}$ is a group homomorphism, we have that $Z(\tilde{G})$ is canonically isomorphic to $\Lambda_Z(\tilde{G})/\Lambda_I(\tilde{G})$. We now claim that the central lattice is the dual of the root lattice.

Exercise 6.15 (The central and root lattices are dual)
Prove that $X \in \mathfrak{t}$ belongs to the central lattice if and only if for every root $\alpha \in \mathfrak{t}^*$, $\alpha(X) \in \mathbb{Z}$.
(Hint: X belongs to the central lattice if and only $\exp 2\pi i\, X$ is central in $\tilde{G}$, which in turn is equivalent to the statement that for every root α, $\exp(2\pi i\, X)\, \exp(te_\alpha) = \exp(te_\alpha)\, \exp(2\pi i\, X)$, for all t. Now use that $[X, e_\alpha] = \alpha(X) e_\alpha$.)

Therefore $Z(\tilde{G}) \cong \Lambda_Z(\tilde{G})/\Lambda_I(\tilde{G}) = \Lambda_r(\tilde{G})^*/\Lambda_w(\tilde{G})^*$, which as the next exercise asks you to show is isomorphic to $\Lambda_w(\tilde{G})/\Lambda_r(\tilde{G})$ as we had claimed.

Exercise 6.16 (Some facts about lattices)
Let $\Lambda_1 \supseteq \Lambda_2$ be lattices. Prove the following:

- *Duality reverses inclusions: $\Lambda_1^* \subseteq \Lambda_2^*$;*
- *$\Lambda_1/\Lambda_2 \cong \Lambda_2^*/\Lambda_1^*$; and*
- *if $\Lambda_3 \subseteq \Lambda_2$ is a third lattice, then*

$$\Lambda_1/\Lambda_2 \cong (\Lambda_1/\Lambda_3) \,/\, (\Lambda_2/\Lambda_3) \ .$$

Since the root lattice is contained in the fundamental weight lattice, we can write $\alpha_i = \sum_j M_{ij}\lambda_j$, for some integers M_{ij}. Now, taking the inner product with $\alpha_j^\vee$ and using the definition of the fundamental weights, we find that $(\alpha_i, \alpha_j^\vee) = M_{ij}$. In other words, (M_{ij}) is the Cartan matrix. Hence in order to write the fundamental weights in terms of the roots, it is necessary to invert the Cartan matrix. If the Cartan matrix has unit determinant then its inverse has integer entries and the fundamental weights belong to the root lattice. In this case the root lattice and the fundamental weight lattice agree, and $\tilde{G}$ has no centre. In general, the order of the group $Z(\tilde{G})$ is given by the determinant of the Cartan matrix, since this is the only denominator in which we incur in the process of expressing the fundamental weights in terms of the roots. In many cases, the order of $Z(\tilde{G})$ is enough to determine the group uniquely. For example, the Cartan matrices of G_2, F_4 and E_8 have unit determinant, the ones of B_ℓ, C_ℓ and E_7 have determinant 2, and the one of E_6 has determinant 3. Hence the fundamental groups of the roots systems are respectively 1, $\mathbb{Z}_2$ and $\mathbb{Z}_3$. In the other cases, the order does not generally determine the group, and one has to work a little harder: the Cartan matrix of A_ℓ has determinant $\ell + 1$, from where it follows that if $\ell+1$ is prime, then the fundamental group of A_ℓ is $\mathbb{Z}_{\ell+1}$, since the only finite abelian group of prime order is the cyclic group. (Proof: Take any element not equal to the identity. It generates a cyclic subgroup whose order must divide the order of the group.) In fact, this persists for all ℓ, but this requires an explicit computation. Finally the Cartan matrix of D_ℓ has determinant 4, which again does not determine the fundamental group. It turns out that for ℓ even the fundamental group is $\mathbb{Z}_2 \times \mathbb{Z}_2$, whereas for ℓ odd it is $\mathbb{Z}_4$. A useful mnemonic in this case is to remember that $D_3 = A_3$.

An example: $A_3 = D_3$

Let us in fact work out this example to see how to go about these calculations. Let us consider the root system $A_3 = D_3$ whose simply-connected compact Lie group is $SU(4) \cong Spin(6)$. We can read off the Cartan matrix from its Dynkin diagram listed in Table 6.2:

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} ,$$

which has indeed determinant 4. Inverting this matrix we can read off the expression for the fundamental weights in terms of the roots:

$$\begin{aligned}\lambda_1 &= \tfrac{3}{4}\alpha_1 + \tfrac{1}{2}\alpha_2 + \tfrac{1}{4}\alpha_3\\ \lambda_2 &= \tfrac{1}{2}\alpha_1 + \alpha_2 + \tfrac{1}{2}\alpha_3\\ \lambda_3 &= \tfrac{1}{4}\alpha_1 + \tfrac{1}{2}\alpha_2 + \tfrac{3}{4}\alpha_3\ .\end{aligned}$$

We can now compute the factor group Λ_w/Λ_r. Its elements are the cosets $0+\Lambda_r$, $\lambda_1+\Lambda_r$, $\lambda_2+\Lambda_r$ and $\lambda_3+\Lambda_r$, which possess the following multiplication table:

	0	λ_1	λ_2	λ_3
0	0	λ_1	λ_2	λ_3
λ_1	λ_1	λ_2	λ_3	0
λ_2	λ_2	λ_3	0	λ_1
λ_3	λ_3	0	λ_1	λ_2

,

where all entries are understood modulo Λ_r. It follows clearly that this is the cyclic group $\mathbb{Z}_4$.

Another example: D_4

Finally let us work a second example. We pick now one of our favourite root systems: D_4, whose simply-connected compact Lie group is $Spin(8)$. From Table 6.2 we can read off the Cartan matrix:

$$\begin{pmatrix} 2 & -1 & 0 & 0\\ -1 & 2 & -1 & -1\\ 0 & -1 & 2 & 0\\ 0 & -1 & 0 & 2\end{pmatrix}\ ,$$

and inverting it, we can read off the expression of the fundamental weights in terms of the roots:

$$\begin{aligned}\lambda_1 &= \alpha_1 + \alpha_2 + \tfrac{1}{2}\alpha_3 + \tfrac{1}{2}\alpha_4 & \lambda_2 &= \alpha_1 + 2\alpha_2 + \alpha_3 + \alpha_4\\ \lambda_3 &= \tfrac{1}{2}\alpha_1 + \alpha_2 + \alpha_3 + \tfrac{1}{2}\alpha_4 & \lambda_4 &= \tfrac{1}{2}\alpha_1 + \alpha_2 + \tfrac{1}{2}\alpha_3 + \alpha_4\ .\end{aligned}$$

We can now compute the factor group Λ_w/Λ_r. Because $\lambda_2 \in \Lambda_r$, it has as elements the cosets of 0, λ_1, λ_3 and λ_4. The multiplication table for this group can be read off easily:

	0	λ_1	λ_3	λ_4
0	0	λ_1	λ_3	λ_4
λ_1	λ_1	0	λ_4	λ_3
λ_3	λ_3	λ_4	0	λ_1
λ_4	λ_4	λ_3	λ_1	0

,

where all entries are understood modulo Λ_r. It is clear that this group is $\mathbb{Z}_2 \times \mathbb{Z}_2$. It has three proper $\mathbb{Z}_2$ subgroups, each one generated by one of the cosets $\lambda_1 + \Lambda_r$, $\lambda_3 + \Lambda_r$ and $\lambda_4 + \Lambda_r$. The representations with highest weights λ_1, λ_3 and λ_4 are all eight-dimensional. They correspond to the vector and the two spinor representations $\Delta_\pm$ of $Spin(8)$. Alternatively, they correspond to the three inequivalent embeddings $Spin(7) \subset Spin(8)$. Given any one of these eight-dimensional representations there exists an $Spin(7)$ subgroup of $Spin(8)$ under which the representation remains irreducible and can be identified with the unique spinorial representation of $Spin(7)$. The dihedral group $\mathcal{D}_3$ of automorphisms of the Dynkin diagram is the group of outer automorphisms of $Spin(8)$. It is called the *triality group* in the physics literature, and it permutes the three inequivalent $Spin(7)$ subgroups and thus the three eight-dimensional representations. In terms of the weights, it permutes λ_1, λ_3 and λ_4.

All the connected compact simple Lie groups

It is now time to summarise what we have learned so far in a table. Table 6.3 lists the simple root systems, their Weyl groups, the associated simple complex Lie algebras, their simply-connected simple compact Lie groups, and their centres. An eternal thorny issue about the notation in Table 6.3: the compact Lie group associated to the root system of type C_ℓ is called in the physics literature $USp(2\ell)$ and in the mathematics literature $Sp(\ell)$. *From here until the end of this section, all Lie groups are connected, compact and simple unless otherwise explicitly stated.*

We start by associating with every Lie group G a subgroup of the centre of its universal cover $\tilde{G}$ and viceversa. Representations of G are also representations of $\tilde{G}$, whence the weight lattice $\Lambda_w(G)$ is contained in the weight lattice $\Lambda_w(\tilde{G})$. We thus have the following inclusions of lattices in $\mathfrak{t}^*$:

$$\Lambda_r(\tilde{G}) \subseteq \Lambda_w(G) \subseteq \Lambda_w(\tilde{G}) \ .$$

Dualising and keeping in mind Exercise 6.16, we have in $\mathfrak{t}$ the following lattices:

$$\Lambda_Z(\tilde{G}) \supseteq \Lambda_w(G)^* \supseteq \Lambda_I(\tilde{G}) \ ,$$

where we have used that $\Lambda_I(\tilde{G}) = \Lambda_w(\tilde{G})^*$, and that $\Lambda_Z(\tilde{G}) = \Lambda_r(\tilde{G})^*$. Applying the reduced exponential map $\underline{\exp} : \mathfrak{t} \to \tilde{G}$ to these lattices and remembering that $\underline{\exp}$ is a group homomorphism when restricted to the maximal torus, we find that $\underline{\exp}\, \Lambda_w(G)^* = \Gamma_G \subseteq Z(\tilde{G})$ is a subgroup of the centre. The subgroup Γ_G is naturally isomorphic to $\Lambda_w(G)^*/\Lambda_I(\tilde{G})$, since the integral lattice is the kernel of the reduced exponential map. Using the fact that

Φ	W or $\vert W\vert$	$\mathfrak{g}_{\mathbb{C}}$	$\tilde{G}$	$Z(\tilde{G})$
A_ℓ	$\mathfrak{S}_{\ell+1}$	$sl(\ell+1,\mathbb{C})$	$SU(\ell+1)$	$\mathbb{Z}_{\ell+1}$
B_ℓ	$(\mathbb{Z}_2)^\ell \rtimes \mathfrak{S}_\ell$	$so(2\ell+1,\mathbb{C})$	$Spin(2\ell+1)$	$\mathbb{Z}_2$
C_ℓ	$(\mathbb{Z}_2)^\ell \rtimes \mathfrak{S}_\ell$	$sp(2\ell,\mathbb{C})$	$USp(2\ell)$	$\mathbb{Z}_2$
D_ℓ	$(\mathbb{Z}_2)^{\ell-1} \rtimes \mathfrak{S}_\ell$	$so(2\ell,\mathbb{C})$	$Spin(2\ell)$	$\begin{cases}\mathbb{Z}_4 & \ell \text{ is odd}\\ \mathbb{Z}_2\times\mathbb{Z}_2 & \ell \text{ is even}\end{cases}$
G_2	$\mathcal{D}_6$	G_2	G_2	1
F_4	$2^7\,3^2$	F_4	F_4	1
E_6	$2^7\,3^4\,5$	E_6	E_6	$\mathbb{Z}_3$
E_7	$2^{10}\,3^4\,5\,7$	E_7	E_7	$\mathbb{Z}_2$
E_8	$2^{10}\,3^5\,5^2\,7$	E_8	E_8	1

Table 6.3: Simple root systems, their Weyl groups, their complex Lie algebras, compact Lie groups and their centres.

$\Lambda_I(\tilde{G}) = \Lambda_w(\tilde{G})^*$, we have

$$\Gamma_G = \Lambda_w(G)^*/\Lambda_w(\tilde{G})^* \cong \Lambda_w(\tilde{G})/\Lambda_w(G) ,$$

where we have again used Exercise 6.16. Since G is determined by its weight lattice, this actually tells us that $G \cong \tilde{G}/\Gamma_G$. Since $\tilde{G}$ is simply-connected, this implies that $\pi_1(G) \cong \Gamma_G$.

Conversely, if $\Gamma \subseteq Z(\tilde{G})$ is a subgroup of the centre of $\tilde{G}$. The preimage of Γ via the reduced exponential map $\underline{\exp} : \mathfrak{t} \to \tilde{G}$, is a sublattice of the central lattice and contains the integer lattice:

$$\Lambda_I(\tilde{G}) \subseteq \Lambda_\Gamma \subseteq \Lambda_Z(\tilde{G}) , \tag{6.13}$$

which upon dualising gives in $\mathfrak{t}^*$ the following series of lattices:

$$\Lambda_w(\tilde{G}) \supseteq \Lambda_\Gamma^* \supseteq \Lambda_r(\tilde{G}) . \tag{6.14}$$

It is not hard to see that Λ_Γ^* is the weight lattice of the group G defined by $\tilde{G}/\Gamma$.

In summary, there is a one-to-one correspondence between Lie groups with the same universal covering group $\tilde{G}$ and subgroups of the centre $Z(\tilde{G})$; or, equivalently, between Lie groups with the same Lie algebra $\mathfrak{g}$ and lattices Λ containing the root lattice and contained in the fundamental weight lattice. Since the centre $Z(\tilde{G})$ is finite, it has a finite number of subgroups, and hence there are only a finite number of Lie groups covered by the same simply-connected Lie group. This is what we meant earlier by "finite ambiguity."

Minding Table 6.3, we can now list all the connected compact simple Lie groups.

For the root systems E_8, F_4 and G_2, the centre is trivial, so they are the only groups with that root system. Similarly, the centres of E_6, E_7, B_ℓ and C_ℓ are not trivial but have no proper nontrivial subgroups, hence there are only two groups associated with each of those root systems: the simply-connected group and the adjoint group: E_6 and $E_6/\mathbb{Z}_3$, E_7 and $E_7/\mathbb{Z}_2$, $Spin(2\ell+1)$ and $SO(2\ell+1) = Spin(2\ell+1)/\mathbb{Z}_2$, and $USp(2\ell)$ and $USp(2\ell)/\mathbb{Z}_2$. A similar story holds for A_ℓ with $\ell+1$ prime: there are only two groups with that root system, $SU(\ell+1)$ and $SU(\ell+1)/\mathbb{Z}_{\ell+1}$. For general ℓ, however, the centre of $SU(\ell+1)$ has subgroups corresponding to the divisors of $\ell+1$: $\mathbb{Z}_m \subset \mathbb{Z}_{\ell+1}$ if and only if m divides $\ell+1$. So we have a whole hierarchy of groups $SU(\ell+1)/\mathbb{Z}_m$ where m runs over the divisors or $\ell+1$, interpolating between the simply-connected $SU(\ell+1)$ and the adjoint group $SU(\ell+1)/\mathbb{Z}_{\ell+1}$. For the root system $D_{2\ell+1}$, the centre is $\mathbb{Z}_4$ which has a single nontrivial proper subgroup isomorphic to $\mathbb{Z}_2$. Hence there are three groups: $Spin(4\ell+2)$, $SO(4\ell+2) = Spin(4\ell+2)/\mathbb{Z}_2$ and $Spin(4\ell+2)/\mathbb{Z}_4$. Finally, the root systems $D_{2\ell}$ has centre $\mathbb{Z}_2 \times \mathbb{Z}_2$ which has three proper subgroups isomorphic to $\mathbb{Z}_2$. Hence there are five groups in this family: $Spin(4\ell)$, $SO(4\ell) = Spin(4\ell)/\mathbb{Z}_2$, $Spin(4\ell)/\mathbb{Z}_2'$, $Spin(4\ell)/Z_2''$, and the adjoint group $Spin(4\ell)/(\mathbb{Z}_2\times\mathbb{Z}_2)$. In the next section we will see that many of these groups are mapped to each other by a duality transformation.

6.3.4 Some simple examples

We illustrate some of the results above with some simple examples: the simple root systems of rank 2: A_2, $B_2 = C_2$ and G_2.

The simple root system A_2

The root system A_2 is defined by the Cartan matrix

$$(A_{ij}) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} .$$

Therefore the simple roots are given in terms of the fundamental weights as follows: $\alpha_1 = 2\lambda_1 - \lambda_2$ and $\alpha_2 = -\lambda_1 + 2\lambda_2$. Inverting these relations we see that $\lambda_1 = \frac{2}{3}\alpha_1 + \frac{1}{3}\alpha_2$ and $\lambda_2 = \frac{1}{3}\alpha_1 + \frac{2}{3}\alpha_2$. This clearly shows that the order of the fundamental group of A_2 is 3, and hence that $\Lambda_w/\Lambda_r \cong \mathbb{Z}_3$. Indeed, notice that this group has as elements the cosets $0 + \Lambda_r$ and $\lambda_1 + \Lambda_r$ and

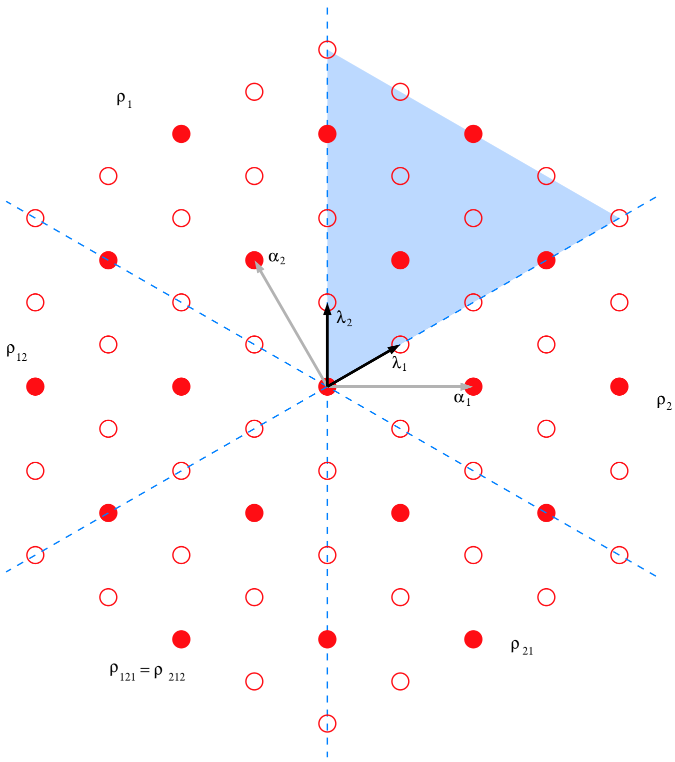


Figure 6.5: The root system A_2.

$\lambda_2 + \Lambda_r$, with multiplication table:

	0	λ_1	λ_2
0	0	λ_1	λ_2
λ_1	λ_1	λ_2	0
λ_2	λ_2	0	λ_1

where all entries are to be understood modulo Λ_r. We can choose a euclidean metric on $\mathbb{R}^2$ and represent these lattices pictorially. This is done in Figure 6.5, which also shows the hyperplanes perpendicular to the roots as dashed lines, and the positive dual Weyl chamber as shaded. The Weyl group is the dihedral group $\mathfrak{D}_3 \cong \mathfrak{S}_3$, the symmetries of an equilateral triangle, and it clearly permutes the dual Weyl chambers. Indeed, in Figure 6.5 all chambers but the fundamental are labelled with the element of the Weyl with which it is associated. Since the Weyl group is generated by reflections on the hyperplanes perpendicular to the simple roots, I have chosen to write the Weyl group elements in this way: the notation ρ_i means the reflection ρ_{α_i} and $\rho_{ij\cdots k} = \rho_i \rho_j \cdots \rho_k$. The filled circles defines the root lattice Λ_r and these together with the open circle define the weight lattice Λ_w. The fundamental weights and simple roots are also shown. Λ_w is the weight lattice of the group $SU(3)$, whereas Λ_r is the weight lattice of the adjoint group $SU(3)/\mathbb{Z}_3$.

The simple root systems $B_2 = C_2$

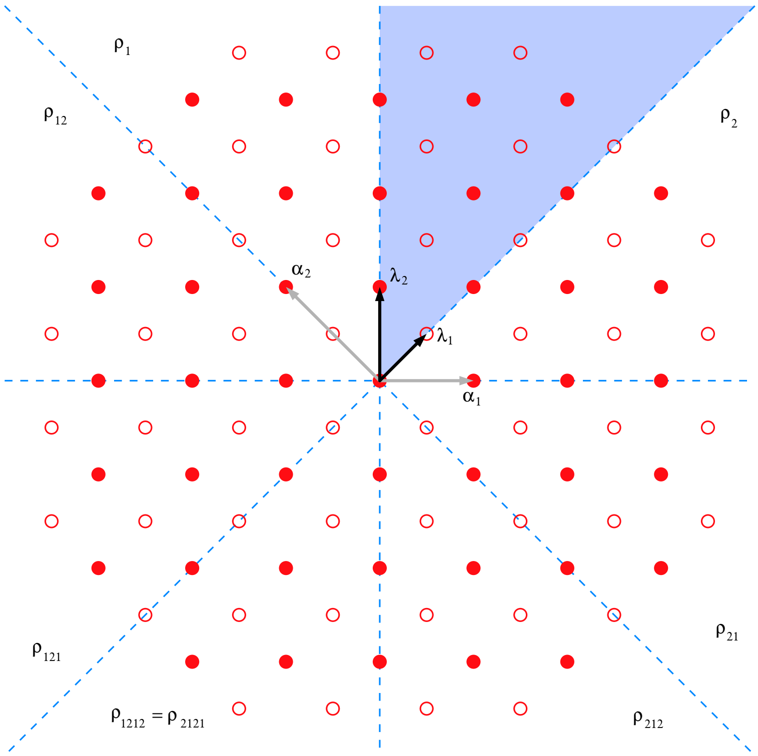


Figure 6.6: The root system B_2.

The root system B_2 is defined by the Cartan matrix

$$(A_{ij}) = \begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix} .$$

Therefore the simple roots are given in terms of the fundamental weights as follows: $\alpha_1 = 2\lambda_1 - \lambda_2$ and $\alpha_2 = -2\lambda_1 + 2\lambda_2$. Inverting these relations we see that $\lambda_1 = \alpha_1 + \frac{1}{2}\alpha_2$ and $\lambda_2 = \alpha_1 + \alpha_2$. This clearly shows that the order of the fundamental group of B_2 is 2, and hence that $\Lambda_w/\Lambda_r \cong \mathbb{Z}_2$. Again, one can see this directly: the cosets $0 + \Lambda_r$ and $\lambda_1 + \Lambda_r$ are the elements of the fundamental group with multiplication table

	0	λ_1
0	0	λ_1
λ_1	λ_1	0

where all entries are again to be understood modulo Λ_r. We can choose a euclidean metric on $\mathbb{R}^2$ and represent these lattices pictorially. This is done in Figure 6.6, which also shows the hyperplanes perpendicular to the roots as dashed lines, and the positive dual Weyl chamber as shaded. The Weyl group is isomorphic to the dihedral group $\mathcal{D}_4$ of symmetries of the square, and again the Weyl chambers have been decorated with the corresponding element of the Weyl group. Once again the filled circles define the root lattice Λ_r and these together with the open circles define the weight lattice Λ_w. The fundamental weights and simple roots are also shown. Λ_w is the weight lattice of the group $Spin(5) \cong USp(4)$, whereas Λ_r is the weight lattice of the group $SO(5)$. The weight λ_1 is the highest weight of the irreducible spinorial representation Δ of $Spin(5)$ obtained as the unique irreducible representation of the Clifford algebra in five-dimensional euclidean space.

The dual root system C_2 has as Cartan matrix the transpose of the Cartan matrix of B_2. They are of course isomorphic root systems, but the isomorphism interchanges long and short roots: $\alpha_1 \leftrightarrow \alpha_2$. This essentially rotates the root diagram by $\pi/4$, and chooses a different fundamental dual Weyl chamber.

The simple root system G_2

The root system G_2 is defined by the Cartan matrix

$$(A_{ij}) = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix} .$$

Therefore the simple roots are given in terms of the fundamental weights as follows: $\alpha_1 = 2\lambda_1 - \lambda_2$ and $\alpha_2 = -3\lambda_1 + 2\lambda_2$. Inverting these relations we see that $\lambda_1 = 2\alpha_1 + \alpha_2$ and $\lambda_2 = 3\alpha_1 + 2\alpha_2$. Hence the root and weight lattices agree. We can choose a euclidean metric on $\mathbb{R}^2$ and represent this lattice pictorially. This is done in Figure 6.7, which also shows the hyperplanes perpendicular to the roots as dashed lines, and the positive dual Weyl chamber as shaded. The Weyl group is now $\mathcal{D}_6$, the symmetries of the regular hexagon, and it permutes the Weyl chambers as shown in the figure. Now the open circles define the root/weight lattice Λ. The fundamental weights and simple roots are also shown. Notice that the long roots form a root system of type A_2, indicative of the fact that $SU(3)$ is a maximal subgroup of G_2.

6.4 The magnetic dual of a compact Lie group

We now start to analyse the Dirac quantisation condition (6.10) in more detail. The punch-line is that the Dirac quantisation condition says that the

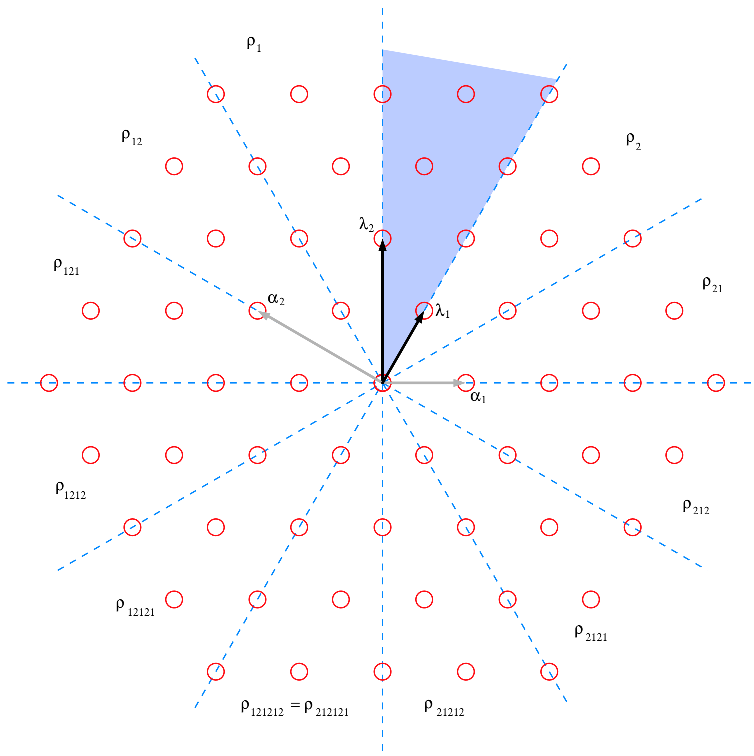


Figure 6.7: The root system G_2.

magnetic charge (suitably normalised) is a dominant weight of a connected compact Lie group $H^\vee$, the (magnetic) dual group of H.

First of all notice that it is irrelevant for these purposes that H be connected, since the image of the exponential map lies in the connected component of the identity. (*Proof*: If $g = \exp X$, then $g(t) = \exp(1-t)X$ is a path to the identity.) Therefore we will assume from now on that H is connected. It is also compact since it is a closed subgroup of a compact Lie group. So we are in the situation that we have just discussed. Because physics is gauge invariant, we have to identify different charges Q which are gauge related via the unbroken gauge group H. Q belongs to the Lie algebra $\mathfrak{h}$ of H and H acts on its Lie algebra via conjugation. One way to fix this gauge invariance is to to choose a fixed maximal torus T in H, with Lie algebra $\mathfrak{t}$ and use our gauge freedom to conjugate Q to lie in $\mathfrak{t}$. As discussed above, this does not fix the gauge completely, because there will be elements of H which stabilise T; in other words, we have to still take into account the action of the Weyl group. The action of the Weyl group is fixed by choosing Q in the *closure* of the fundamental Weyl chamber C, since this is a fundamental domain for the action of the Weyl group.

Therefore in the Dirac quantisation condition (6.10), we can take $e\mathsf{Q}$ to lie in the closure $\bar{C}$ of the fundamental Weyl chamber in $\mathfrak{t}$. The exponential map in (6.10) is then the exponential map $\mathfrak{t} \to T$, and (6.10) says that $e\mathsf{Q}/2\pi$ belongs to the integer lattice $\Lambda_I(H)$ of H. We saw above that $\Lambda_I(H) = \Lambda_w(H)^*$, whence the Dirac quantisation condition becomes:

$$e\mathsf{Q}/2\pi \in \Lambda_w(H)^*/W \cong \Lambda_w(H)^* \cap \bar{C} \ ,$$

where W is the Weyl group. On the other hand, the integer lattice can be thought of as the weight lattice of a connected compact Lie group $H^\vee$ known as the *(magnetic) dual group* of H. As we will see below, this group is a quotient of the simply-connected compact simple Lie group whose root system is dual to the root system of H. Now, dual root systems share the same Weyl group. This follows from the fact that the Weyl group is generated by those reflections ρ_α in (6.12) corresponding to simple roots. But from (6.12) it follows that $\rho_\alpha = \rho_{\alpha^\vee}$. Therefore we can fix the Weyl symmetry by going to the fundamental dual Weyl chamber of $H^\vee$. In other words, the Dirac quantisation condition can be rewritten as

$$e\mathsf{Q}/2\pi \in \Lambda_w(H^\vee)/W \cong \Lambda_w^+(H^\vee) \ ,$$

where $\Lambda_w^+(H^\vee)$ are the dominant weights of $H^\vee$, which are in one-to-one correspondence with the finite-dimensional irreducible representations of $H^\vee$. We now turn to a more detailed description of the dual group.

6.4.1 Some lattices and dual groups

In our flash review of compact Lie groups, we have already encountered several lattices. We will now review their interrelations and in particular how they can be used to describe the dual of a connected compact Lie group. For the purposes of studying the Dirac quantisation condition, we will take the unbroken gauge group H to be a compact connected Lie group. Such a group is covered finitely by a compact group $\tilde{H} = \tilde{K} \times S$, where $\tilde{K}$ is a simply-connected compact Lie group (hence semisimple, and in turn the product of simple factors $\tilde{K}_1 \times \cdots \times \tilde{K}_p$) and S is a torus. S is the connected component of the identity of the centre of $\tilde{H}$. We define the *dual group* $H^\vee$ of H to be the compact connected Lie group whose weight lattice is dual to the weight lattice of H. From this it follows that just like H is a finite quotient of $\tilde{H}$, so will $H^\vee$ be a finite quotient of $\tilde{H}^\vee = \tilde{K}_1^\vee \times \cdots \times \tilde{K}_p^\vee \times S^\vee$. It is impractical to treat the general case, so we will discuss separately the cases of H abelian and H simple. From these ingredients it should be possible to treat the case of general H should the urge arise. All Lie groups in this section are compact and connected unless stated otherwise.

H abelian

If H is abelian, then it is a torus. Let $\mathfrak{h}$ be its Lie algebra. The (reduced) exponential map is surjective and defines a diffeomorphism $H \cong \mathfrak{h}/\Lambda$, where $\Lambda \subset \mathfrak{h}$ is the lattice of periods of H. As we reviewed above, this lattice is dual to the weight lattice $\Lambda_w(H) \subset \mathfrak{h}^*$. By definition this is the weight lattice of the dual group $H^\vee$. Hence we have a diffeomorphism $H^\vee \cong \mathfrak{h}^*/\Lambda_w(H)$. Notice that $\mathfrak{h}^\vee$ is identified with $\mathfrak{h}^*$.

H simple

Let H be a simple Lie group, $\tilde{H}$ its universal covering group, and $\mathfrak{h}$ its Lie algebra. Let T be a fixed maximal torus and $\mathfrak{t} \subset \mathfrak{h}$ its Lie algebra. We let $\mathfrak{t}^*$ be the space of linear forms $\mathfrak{t} \to \mathbb{R}$. The root lattices $\Lambda_r(H)$ and $\Lambda_r(\tilde{H})$ in $\mathfrak{t}^*$ agree, since as explained above they only depend on the Lie algebra. We will therefore write it as $\Lambda_r(\mathfrak{h})$. The weight lattices $\Lambda_w(H)$ and $\Lambda_w(\tilde{H})$ are different, with $\Lambda_w(\tilde{H})$ depending only on the Lie algebra again, since it is the lattice of fundamental weights. We will then often write it as $\Lambda_w(\mathfrak{h})$. We have the following inclusions:

$$\Lambda_r(\mathfrak{h}) \subseteq \Lambda_w(H) \subseteq \Lambda_w(\mathfrak{h}) \ , \tag{6.15}$$

where the first inclusion is an equality when H is the adjoint group, and the last inclusion is an equality when $H = \tilde{H}$. From Table 6.3 we see that for E_8, F_4 and H_2, the adjoint group is simply-connected, so that in these cases, and in these cases only, are both inclusions equalities.

The dual of these lattices give rise to lattices in $\mathfrak{t}$. Dualising the lattices reverses the inclusions in (6.15), so we have

$$\Lambda_r(\mathfrak{h})^* \supseteq \Lambda_w(H)^* \supseteq \Lambda_w(\mathfrak{h})^* \ . \tag{6.16}$$

We have met some of these lattices before. $\Lambda_w(\mathfrak{h})^* = \Lambda_I(\tilde{H}) = \Lambda_r^\vee(\mathfrak{h})$, which again only depends on the Lie algebra. Similarly, $\Lambda_w(H)^* = \Lambda_I(H)$ is the integer lattice of H: those elements $h \in T$ such that $2\pi h$ lies in the kernel of the exponential map $\exp : \mathfrak{t} \to T$. It clearly depends on H, as it will be different for H and for $\tilde{H}$.

Now with the same notation as above, let us consider the inverse root system $\Phi^\vee$. As mentioned above, it is a simple root system. Therefore by the construction outlined above, there will be a complex simple Lie algebra $\mathfrak{h}_{\mathbb{C}}^\vee$ associated to $\Phi^\vee$, which has a unique compact real form $\mathfrak{h}^\vee$, which can be integrated to a unique connected compact simply-connected Lie group $\tilde{H}^\vee$. We will now exhibit the dual group $H^\vee$ of H as a quotient of $\tilde{H}^\vee$ by a finite subgroup of its centre.

The inverse root system requires for its very definition the existence of the metric: $\alpha^\vee = 2\alpha/(\alpha,\alpha)$. We can undo this dependence by using the metric to map each $\alpha^\vee \in \mathfrak{t}^*$ to a unique $\alpha^* \in \mathfrak{t}$ such that if $\beta \in \mathfrak{t}^*$, $\beta(\alpha^*) = (\beta, \alpha^\vee)$. We have met these α^* before: they are nothing but the coroots h_α. The coroots generate a root system in $\mathfrak{t}$ whose root lattice is the coroot lattice $\Lambda^\vee(\mathfrak{h})$ of $\mathfrak{h}$ and whose fundamental weight lattice is the dual lattice to the root lattice of $\mathfrak{h}$.

Exercise 6.17 (The dual fundamental weights)
Let $\Lambda_w(\mathfrak{h}^\vee) \subset \mathfrak{t}$ denote the lattice of fundamental weights of the dual root system. Prove that

$$\Lambda_w(\mathfrak{h}^\vee) = \Lambda_r(\mathfrak{h})^* \ .$$

In other words, if we let $\mathfrak{t}^\vee = \mathfrak{t}^*$, then on $\mathfrak{t}^{\vee *} = \mathfrak{t}$, we have a root lattice $\Lambda_r(\mathfrak{h}^\vee) = \Lambda_r^\vee(\mathfrak{h}) = \Lambda_w(\mathfrak{h})^*$ and a fundamental weight lattice $\Lambda_w(\mathfrak{h}^\vee) = \Lambda_r(\mathfrak{h})^*$. On $\mathfrak{t}^\vee = \mathfrak{t}^*$ there is also a notion of reduced exponential map $\underline{\exp} : \mathfrak{t}^\vee \to T^\vee$ which is given by the canonical projection $\mathfrak{t}^* \to \mathfrak{t}^*/\Lambda_r(\mathfrak{h})$. The centre of $\tilde{H}^\vee$ is given by

$$Z(\tilde{H}^\vee) \cong \Lambda_w(\mathfrak{h}^\vee)/\Lambda_r(\mathfrak{h}^\vee) \cong \Lambda_r(\mathfrak{h})^*/\Lambda_w(\mathfrak{h})^* \cong \Lambda_w(\mathfrak{h})/\Lambda_r(\mathfrak{h}) \cong Z(\tilde{H}) \ ,$$

where we have used Exercise 6.16.

Now, by definition, the weight lattice $\Lambda_w(H^\vee)$ of $H^\vee$ is dual of the weight lattice $\Lambda_w(H)$ of H:

$$\Lambda_w(H^\vee) \equiv \Lambda_w(H)^* \ ,$$

which sits between the above two lattices in $\mathfrak{t}$:

$$\Lambda_r(\mathfrak{h}^\vee) \subseteq \Lambda_w(H^\vee) \subseteq \Lambda_w(\mathfrak{h}^\vee) \ .$$

From the above discussion surrounding equations (6.13) and (6.14), we know that $H^\vee$ is given by $\tilde{H}^\vee/\Gamma^\vee$ where $\Gamma^\vee \subseteq Z(\tilde{H}^\vee)$ is the subgroup of the centre of $\tilde{H}^\vee$ defined by $\Lambda_w(H^\vee)^*/\Lambda_w(\mathfrak{h}^\vee)^* = \Lambda_w(H)/\Lambda_r(\mathfrak{h})$. But consider now the subgroup $\Gamma \equiv \Gamma_H \subseteq Z(\tilde{H})$ which defines $H = \tilde{H}/\Gamma$. Taking into account Exercise 6.16, we find that it is given by

$$\Gamma \cong \Lambda_w(\mathfrak{h})/\Lambda_w(H) \cong (\Lambda_w(\mathfrak{h})/\Lambda_r(\mathfrak{h}))\,/\,(\Lambda_w(H)/\Lambda_r(\mathfrak{h})) \cong Z(\tilde{H})/\Gamma^\vee \ ,$$

whence

$$|\Gamma||\Gamma^\vee| = |Z(\tilde{H})| \ . \tag{6.17}$$

Let us now look at examples of dual groups. Above we listed the connected compact simple Lie groups. We now do the same for their duals. This has been done in [GNO77]. We list the results in Table 6.4. Most cases

H	$H^\vee$
$SU(pq)/\mathbb{Z}_p$	$SU(pq)/\mathbb{Z}_q$
$Spin(2\ell+1)$	$USp(2\ell)/\mathbb{Z}_2$
$SO(2\ell+1)$	$USp(2\ell)$
$SO(2\ell)$	$SO(2\ell)$
$Spin(4\ell+2)$	$Spin(4\ell+2)/\mathbb{Z}_4$
$Spin(4\ell)$	$Spin(4\ell)/(\mathbb{Z}_2\times\mathbb{Z}_2)$
$Spin(8\ell)/\mathbb{Z}_2'$	$Spin(8\ell)/\mathbb{Z}_2'$
$Spin(8\ell)/\mathbb{Z}_2''$	$Spin(8\ell)/\mathbb{Z}_2''$
$Spin(8\ell+4)/\mathbb{Z}_2'$	$Spin(8\ell+4)/\mathbb{Z}_2''$
G_2	G_2
F_4	F_4
E_6	$E_6/\mathbb{Z}_3$
E_7	$E_7/\mathbb{Z}_2$
E_8	E_8

Table 6.4: The connected compact simple Lie groups and their duals.

can be determined without any computation, but some of the D_ℓ series turn out to be subtle, and require an explicit description of the weight and root lattices. They are listed, for example, in [Hum72].

Equation (6.17) tells us that the orders of Γ and $\Gamma^\vee$ are complementary in $|Z(\tilde{H})| = |Z(\tilde{H}^\vee)|$. This means that the dual of the simply-connected group $\tilde{H}$ is the adjoint group $\tilde{H}^\vee/Z(\tilde{H}^\vee)$. This already tells us the last five entries of Table 6.4 as well as the second, third, fifth and sixth entries, and the special case $p=1$, in the first entry. But, in fact, the rest of the first entry also requires no further calculation. Since any subgroup of a cyclic group is cyclic and is moreover unique, the dual of $SU(pq)/\mathbb{Z}_p$ has to be $SU(pq)/\mathbb{Z}_q$, since given $\mathbb{Z}_p \subset \mathbb{Z}_{pq}$ there is a unique subgroup of $\mathbb{Z}_{pq}$ of order q, and it is $\mathbb{Z}_q$. The same argument also applies to $D_{2\ell+1}$, since the centre is cyclic in this case: whence $SO(4\ell+2)$ is self-dual. For the groups with root system $D_{2\ell}$ one has to work harder.

An example: *Spin*(8) and its quotients

As an example we will work out the example of $Spin(8)$ and its factor groups. The root system of $Spin(8)$ is D_4 and we worked out the Cartan matrix, the centre and the fundamental weights above. The lattice of fundamental weights $\Lambda_w = \mathbb{Z}\langle\lambda_i\rangle$ is the integer span of the fundamental weights λ_i. The root lattice is the sublattice of the fundamental weights generated by the combinations:

$$\alpha_1 = 2\lambda_1 - \lambda_2 \qquad \alpha_2 = -\lambda_1 + 2\lambda_2 - \lambda_3 - \lambda_4$$
$$\alpha_3 = -\lambda_2 + 2\lambda_3 \qquad \alpha_4 = -\lambda_2 + 2\lambda_4 \ .$$

Equivalently it is the lattice consisting of elements $\sum_{i=1}^4 n_i\lambda_i$ where $n_i \in \mathbb{Z}$ such that n_1, n_3 and n_4 are either all even or all odd. There are three intermediate lattices corresponding to the weight lattices of the three subgroups $SO(8)$, $Spin(8)/\mathbb{Z}_2'$ and $Spin(8)/\mathbb{Z}_2''$: $\Lambda_1 = \Lambda_r \cup (\Lambda_r + \lambda_1)$, $\Lambda_3 = \Lambda_r \cup (\Lambda_r + \lambda_3)$, and $\Lambda_4 = \Lambda_r \cup (\Lambda_r + \lambda_4)$. Equivalently,

$$\Lambda_1 = \{\sum_{i=1}^4 n_i\lambda_i \mid n_3 \equiv n_4\,(2)\}$$
$$\Lambda_3 = \{\sum_{i=1}^4 n_i\lambda_i \mid n_1 \equiv n_4\,(2)\}$$
$$\Lambda_4 = \{\sum_{i=1}^4 n_i\lambda_i \mid n_1 \equiv n_3\,(2)\} \ ,$$

all other integers n_i unconstrained. We can easily find a $\mathbb{Z}$-basis for these lattices as follows:

$$\Lambda_1 = \mathbb{Z}\langle\lambda_1, \lambda_2, \lambda_3 \pm \lambda_4\rangle$$
$$\Lambda_3 = \mathbb{Z}\langle\lambda_2, \lambda_3, \lambda_1 \pm \lambda_4\rangle$$
$$\Lambda_4 = \mathbb{Z}\langle\lambda_2, \lambda_4, \lambda_1 \pm \lambda_3\rangle \ .$$

The dual picture is as follows. Take as a basis the canonical dual basis $\{\alpha^i\}$ to the roots: $\alpha^i(\alpha_j) = \delta^i{}_j$. Their $\mathbb{Z}$-span is the lattice Λ_r^* and all lattices of interest are contained in it, so their elements will be integer linear combinations of the α^i. Given a sublattice $\Lambda \subseteq \Lambda_w$ described as the $\mathbb{Z}$-span of some vectors v_i in the weight lattice Λ_w, the dual lattice will be the sublattice $\Lambda^* \subseteq \Lambda_r^*$ given by the $\mathbb{Z}$-span of the canonical dual basis v^i to the v_i. Let $v_i = \sum_j M_i{}^j\lambda_j$, where $M_i{}^j \in \mathbb{Z}$ since Λ is a sublattice of Λ_w. Similarly $v^i = \sum_j N^i{}_j\alpha^j$, where $N^i{}_j \in \mathbb{Z}$. We can solve for N in terms of M and the Cartan matrix C as follows. By definition, $v^i(v_j) = \delta^i{}_j$, whence

$$\begin{aligned}\delta^i{}_j &= v_j(v^i)\\ &= N^i{}_k M_j{}^\ell \lambda_\ell(\alpha^k)\\ &= N^i{}_k M_j{}^\ell (C^{-1})_\ell{}^k\end{aligned}$$

where we have used that $\lambda_\ell = (C^{-1})_\ell{}^k \alpha_k$. In other words, $N = (CM^{-1})^t$. Computing this for each of the lattices above, we find:

$$\begin{aligned}
\Lambda_1^* &= \mathbb{Z}\langle 2\alpha^1 - \alpha^2, \alpha^1 - \alpha^2 + \alpha^3 + \alpha^4, \alpha^2 - \alpha^3 - \alpha^4, \alpha^3 - \alpha^4 \rangle \\
\Lambda_3^* &= \mathbb{Z}\langle \alpha^1 - \alpha^2 + \alpha^4, \alpha^1 - \alpha^2 + \alpha^3 + \alpha^4, \alpha^2 - 2\alpha^3, \alpha^1 - \alpha^4 \rangle \\
\Lambda_4^* &= \mathbb{Z}\langle \alpha^1 - \alpha^2 + \alpha^3, \alpha^1 - \alpha^2 + \alpha^3 + \alpha^4, \alpha^1 - \alpha^3, \alpha^2 - 2\alpha^4 \rangle \ .
\end{aligned}$$

We can understand these lattices as sublattices of Λ_r^* by changing basis to the α^i and constraining the coefficients. We find

$$\begin{aligned}
\Lambda_1^* &= \{\sum_{i=1}^4 n_i \alpha^i \mid n_3 \equiv n_4\,(2)\} \\
\Lambda_3^* &= \{\sum_{i=1}^4 n_i \alpha^i \mid n_1 \equiv n_4\,(2)\} \\
\Lambda_4^* &= \{\sum_{i=1}^4 n_i \alpha^i \mid n_1 \equiv n_3\,(2)\} \ ,
\end{aligned}$$

whence we conclude that all three lattices are self-dual, in agreement with Table 6.4.

Another example: ***Spin*****(12) and its quotients**

As a final example, and to illustrate the other behaviour of the $D_{2\ell}$ series, we will work out the example of *Spin*(12) and its factor groups. The root system of *Spin*(12) is D_6, whose Cartan matrix follows from Table 6.2:

$$\begin{pmatrix}
2 & -1 & 0 & 0 & 0 & 0 \\
-1 & 2 & -1 & 0 & 0 & 0 \\
0 & -1 & 2 & -1 & 0 & 0 \\
0 & 0 & -1 & 2 & -1 & -1 \\
0 & 0 & 0 & -1 & 2 & 0 \\
0 & 0 & 0 & -1 & 0 & 2
\end{pmatrix} \ .$$

The fundamental weights are given by

$$\begin{aligned}
\lambda_1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \tfrac{1}{2}\alpha_5 + \tfrac{1}{2}\alpha_6 \\
\lambda_2 &= \alpha_1 + 2\alpha_2 + 2\alpha_3 + 2\alpha_4 + \alpha_5 + \alpha_6 \\
\lambda_3 &= \alpha_1 + 2\alpha_2 + 3\alpha_3 + 3\alpha_4 + \tfrac{3}{2}\alpha_5 + \tfrac{3}{2}\alpha_6 \\
\lambda_4 &= \alpha_1 + 2\alpha_2 + 3\alpha_3 + 4\alpha_4 + 2\alpha_5 + 2\alpha_6 \\
\lambda_5 &= \tfrac{1}{2}\alpha_1 + \alpha_2 + \tfrac{3}{2}\alpha_3 + 2\alpha_4 + \tfrac{3}{2}\alpha_5 + \alpha_6 \\
\lambda_6 &= \tfrac{1}{2}\alpha_1 + \alpha_2 + \tfrac{3}{2}\alpha_3 + 2\alpha_4 + \alpha_5 + \tfrac{3}{2}\alpha_6 \ .
\end{aligned}$$

It follows that the centre $\Lambda_w/\Lambda_r \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ consists of the following Λ_r-cosets: 0, λ_1, λ_5 and λ_6, with multiplication table:

	0	λ_1	λ_5	λ_6
0	0	λ_1	λ_5	λ_6
λ_1	λ_1	0	λ_6	λ_5
λ_5	λ_5	λ_6	0	λ_1
λ_6	λ_6	λ_5	λ_1	0

,

where as usual all entries are modulo Λ_r.

Letting $\Lambda_r = \mathbb{Z}\langle\lambda_i\rangle$, the root lattice is the sublattice $\Lambda_r = \mathbb{Z}\langle\alpha_i\rangle$ spanned by the following combinations:

$$\begin{aligned} \alpha_1 &= 2\lambda_1 - \lambda_2 & \alpha_2 &= -\lambda_1 + 2\lambda_2 - \lambda_3 \\ \alpha_3 &= -\lambda_2 + 2\lambda_3 - \lambda_4 & \alpha_4 &= -\lambda_3 + 2\lambda_4 - \lambda_5 - \lambda_6 \\ \alpha_5 &= -\lambda_4 + 2\lambda_5 & \alpha_6 &= -\lambda_4 + 2\lambda_6 \end{aligned}$$

Equivalently it is the lattice consisting of elements $\sum_{i=1}^6 n_i\lambda_i$ where $n_i \in \mathbb{Z}$ such that $n_1 + n_3$, n_5 and n_6 are either all even or all odd. There are three intermediate lattices corresponding to the weight lattices of the three subgroups $SO(12)$, $Spin(12)/\mathbb{Z}_2'$ and $Spin(12)/\mathbb{Z}_2''$: $\Lambda_1 = \Lambda_r \cup (\Lambda_r + \lambda_1)$, $\Lambda_5 = \Lambda_r \cup (\Lambda_r + \lambda_5)$, and $\Lambda_6 = \Lambda_r \cup (\Lambda_r + \lambda_6)$. Equivalently,

$$\begin{aligned} \Lambda_1 &= \{\sum_{i=1}^{6} n_i\lambda_i \mid n_5 \equiv n_6\,(2)\} \\ \Lambda_5 &= \{\sum_{i=1}^{6} n_i\lambda_i \mid n_1 + n_3 \equiv n_6\,(2)\} \\ \Lambda_6 &= \{\sum_{i=1}^{6} n_i\lambda_i \mid n_1 + n_3 \equiv n_5\,(2)\}\ , \end{aligned}$$

all other integers n_i unconstrained. We can easily find a $\mathbb{Z}$-basis for these lattices as follows:

$$\begin{aligned} \Lambda_1 &= \mathbb{Z}\langle\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \pm \lambda_6\rangle \\ \Lambda_5 &= \mathbb{Z}\langle\lambda_2, \lambda_4, \lambda_5, \lambda_1 + \lambda_3, \lambda_1 + \lambda_6, \lambda_3 + \lambda_6\rangle \\ \Lambda_6 &= \mathbb{Z}\langle\lambda_2, \lambda_4, \lambda_6, \lambda_1 + \lambda_3, \lambda_1 + \lambda_5, \lambda_3 + \lambda_5\rangle\ . \end{aligned}$$

Following the discussion given in the previous example, the dual lattices are given by

$$\begin{aligned}
\Lambda_1^* &= \mathbb{Z}\langle 2\alpha^1 - \alpha^2, \alpha^1 - 2\alpha^2 + \alpha^3, \alpha^2 - 2\alpha^3 + \alpha^4, \alpha^3 - 2\alpha^4 + \alpha^5 + \alpha^6, \\
&\qquad \alpha^5 - \alpha^6, \alpha^4 - \alpha^5 - \alpha^6\rangle \\
\Lambda_5^* &= \mathbb{Z}\langle \alpha^1 - \alpha^2 + \alpha^3 - \alpha^6, \alpha^1 - \alpha^3 + \alpha^4 - \alpha^6, \alpha^3 - 2\alpha^4 + \alpha^5 + \alpha^6, \\
&\qquad \alpha^1 - 2\alpha^2 + \alpha^3, \alpha^5 - 2\alpha^6, \alpha^1 - \alpha^3 + \alpha^6, \rangle \\
\Lambda_6^* &= \mathbb{Z}\langle \alpha^1 - \alpha^2 + \alpha^3 - \alpha^5, \alpha^1 - \alpha^3 + \alpha^4 - \alpha^5, \alpha^3 - 2\alpha^4 + \alpha^5 + \alpha^6, \\
&\qquad \alpha^1 - 2\alpha^2 + \alpha^3, \alpha^1 - \alpha^3 + \alpha^5, \alpha^4 - 2\alpha^6\rangle \ .
\end{aligned}$$

We can understand these lattices as sublattices of Λ_r^* by changing basis to the α^i and constraining the coefficients. We find

$$\begin{aligned}
\Lambda_1^* &= \{\sum_{i=1}^6 n_i\alpha^i \mid n_5 \equiv n_6\,(2)\} \\
\Lambda_5^* &= \{\sum_{i=1}^6 n_i\alpha^i \mid n_1 + n_3 \equiv n_5\,(2)\} \\
\Lambda_6^* &= \{\sum_{i=1}^6 n_i\alpha^i \mid n_1 + n_3 \equiv n_6\,(2)\} \ ,
\end{aligned}$$

whence we conclude that Λ^1 is self-dual, whereas duality interchanges the groups whose weight lattices are Λ_5 and Λ_6. It can be shown that the group whose weight lattice is Λ_1 is $SO(12)$. Again this is in agreement with Table 6.4.

Bibliography

[Ada69] J. F. Adams, *Lectures on Lie groups*, W. A. Benjamin, 1969.

[AH85] M. F. Atiyah and N. J. Hitchin, *Low energy scattering of non-abelian monopoles*, Phys. Lett. **107A** (1985), 21–25.

[AH88] ———, *The geometry and dynamics of magnetic monopoles*, Princeton University Press, 1988.

[Bes86] A. L. Besse, *Einstein manifolds*, Springer–Verlag, 1986.

[Blu94] J. D. Blum, *Supersymmetric quantum mechanics of monopoles in N=4 Yang–Mills theory*, Phys. Lett. **B333** (1994), 92–97.

[Bog76] E. B. Bogomol'nyi, *The stability of classical solutions*, Soviet J. Nuc. Phys. **24** (1976), 449–454.

[BT81] R. Bott and L. W. Tu, *Differential forms in algebraic topology*, Springer–Verlag, 1981.

[BtD85] T. Bröcker and T. tom Dieck, *Representations of compact Lie groups*, Springer–Verlag, 1985.

[Cal78] C. Callias, *Axial anomalies and index theorems on open spaces*, Comm. Math. Phys. **62** (1978), 213.

[Col77] S. Coleman, *Classical lumps and their quantum descendants*, New Phenomena in Subnuclear Physics (A. Zichichi, ed.), Plenum, 1977, Erice lectures on 't Hooft Polyakov monopoles, larger gauge groups, homotopy considerations; comparison with sine-Gordon.

[CPNS76] S. Coleman, S. Parke, A. Neveu, and C. M. Sommerfield, *Can one dent a dyon?*, Phys. Rev. **D15** (1976), 544–545.

[DHdV78] A. D'Adda, R. Horsley, and P. di Vecchia, *Supersymmetric magnetic monopoles and dyons*, Phys. Lett. **76B** (1978), 298–302.

[Dir31] P. A. M. Dirac, *Quantised singularities in the electromagnetic field*, Proc. R. Soc. **A133** (1931), 60–72.

[Gau94] J. P. Gauntlett, *Low energy dynamics of N=2 supersymmetric monopoles*, Nuc. Phys. **411B** (1994), 433–460.

[GH78] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, 1978.

[GM86] G. W. Gibbons and N. S. Manton, *Classical and quantum dynamics of BPS monopoles*, Nuc. Phys. **274B** (1986), 183–224.

[GNO77] P. Goddard, J. Nuyts, and D. Olive, *Gauge theories and magnetic charge*, Nuc. Phys. **B125** (1977), 1–28.

[Gol62] S. I. Goldberg, *Curvature and homology*, Academic Press, 1962.

[HKLR87] N. J. Hitchin, A. Karlhede, U. Lindström, and M. Roček, *Hyperkähler metrics and supersymmetry*, Comm. Math. Phys. **108** (1987), 535–589.

[Hum72] J. E. Humphreys, *Introduction to lie algebras and representation theory*, vol. 9, Springer–Verlag, 1972.

[JT80] A. Jaffe and C. Taubes, *Vortices and monopoles*, Birkhäuser, 1980.

[JZ75] B. Julia and A. Zee, *Poles with both magnetic and electric charges in non-abelian gauge theories*, Phys. Rev. **D11** (1975), 2227–2232.

[KN63] S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, vol. 1, Interscience, 1963.

[KN69] ———, *Foundations of differential geometry*, vol. 2, Interscience, 1969.

[Lic76] A. Lichnerowicz, *Global theory of connections and holonomy groups*, Noordhoff International Pub., 1976.

[Man82] N. S. Manton, *A remark on the scattering of BPS monopoles*, Phys. Lett. **110B** (1982), 54–56.

[MO77] C. Montonen and D. Olive, *Magnetic monopoles as gauge particles*, Phys. Lett. **72B** (1977), 117–120.

[Oli79] D. Olive, *The electric and magnetic charges as extra components of four-momentum*, Nuc. Phys. **B153** (1979), 1–12.

[Osb79] H. Osborn, *Topological charges for N=4 supersymmetric gauge theories and monopoles of spin 1*, Phys. Lett. **83B** (1979), 321–326.

[Pol74] A. M. Polyakov, *Particle spectrum in quantum field theory*, JETP Letters **20** (1974), 194–95.

[PS75] M. K. Prasad and C. M. Sommerfield, *Exact classical solution for the 't Hooft monopole and the Julia–Zee dyon*, Phys. Rev. Lett. **35** (1975), 760–762.

[PS86] A. Pressley and G. Segal, *Loop groups*, Clarendon, 1986.

[Sal89] S. Salamon, *Riemannian geometry and holonomy groups*, Longman Scientific and Technical, 1989.

[Ser73] J. P. Serre, *A course in arithmetic*, Springer–Verlag, 1973.

[Soh85] M. F. Sohnius, *Introducing supersymmetry*, Phys. Rep. **128** (1985), 39–204.

[Ste51] N. E. Steenrod, *The topology of fibre bundles*, Princeton University Press, 1951.

[Tau83] C. H. Taubes, *Stability in Yang–Mills theories*, Comm. Math. Phys. **91** (1983), 473–540.

[tH74] G. 't Hooft, *Magnetic monopoles in unified gauge theories*, Nuc. Phys. **B79** (1974), 276–284.

[War83] F. W. Warner, *Foundations of differentiable manifolds and lie groups*, vol. 94, Springer–Verlag, 1983.

[Wei79] E. J. Weinberg, *Parameter counting for multimonopole solutions*, Phys. Rev. **D20** (1979), 936–944.

[Wel80] R. O. Wells, Jr., *Differential analysis on complex manifolds*, Springer–Verlag, 1980.

[Wit79] E. Witten, *Dyons of charge $e\theta/2\pi$*, Phys. Lett. **86B** (1979), 283–287.

[WO78] E. Witten and D. Olive, *Supersymmetry algebras that include topological charges*, Phys. Lett. **78B** (1978), 97–101.

[Zum77] B. Zumino, *Euclidean supersymmetry and the many-instanton problem*, Phys. Lett. **69B** (1977), 369–371.