

Lectures on instantons

Stefan Vandoren¹ and Peter van Nieuwenhuizen²

¹ *Institute for Theoretical Physics and Spinoza Institute*
Utrecht University, 3508 TD Utrecht, The Netherlands
 s.vandoren@phys.uu.nl

² *C.N. Yang Institute for Theoretical Physics*
State University of New York at Stony Brook, NY 11790, USA
 vannieu@insti.physics.sunysb.edu

Abstract

This is a selfcontained set of lecture notes on instantons in (super) Yang-Mills theory in four dimensions and in quantum mechanics. First the basics are derived from scratch: the regular and singular one-instanton solutions for Yang-Mills theories with gauge groups $SU(2)$ and $SU(N)$, their bosonic and fermionic zero modes, the path integral instanton measure, and supersymmetric Yang-Mills theories in Euclidean space. Then we discuss applications: the θ -angle of QCD, the solution of the $U(1)$ problem, the way Higgs fields solve the large-instanton problem, and tunneling and phase transitions in quantum mechanics and in nonabelian gauge theories. These lecture notes are an extension of a review on Yang-Mills and D-instantons written in 2000 by both authors and A.Belitsky [1].

Contents

1	Introduction	3
2	Winding number and embeddings	7
2.1	Some remarks on nonselfdual instanton solutions	16
3	Regular and singular instanton solutions	17
3.1	Lorentz and spinor algebra	18
3.2	Solving the selfduality equations	21

4	Collective coordinates, the index theorem and fermionic zero modes	25
4.1	Bosonic collective coordinates and the Dirac operator	26
4.2	Fermionic moduli and the index theorem	28
5	Construction of zero modes	34
5.1	Bosonic zero modes and their normalization	35
5.2	Construction of the fermionic zero modes	40
6	The measure for zero modes	42
6.1	The measure for the bosonic collective coordinates	42
6.2	The measure for the fermionic collective coordinates	45
7	One loop determinants	48
7.1	The exact β function for SYM theories	52
8	$\mathcal{N} = 4$ supersymmetric Yang-Mills theory	57
8.1	Minkowskian $\mathcal{N} = 4$ SYM	57
8.2	Euclidean $\mathcal{N} = 4$ SYM	58
8.3	Involution in Euclidean space	61
9	Large instantons and the Higgs effect	62
10	Instantons as most probable tunnelling paths	66
11	False vacua and phase transitions	76
12	The strong CP problem	86
13	The $U(1)$ problem	89
14	Baryon decay	91
15	Discussion	94
A	Winding number	94
B	't Hooft symbols and Euclidean spinors	97
C	The volume of the gauge orientation moduli space	100
D	Zero modes and conformal symmetries	106

1 Introduction

In the last decades enormous progress has been made in understanding nonperturbative effects, both in supersymmetric field theories and in superstring theories. By non-perturbative effects we mean effects due to solitons and instantons, whose masses and actions, respectively, are inversely proportional to the square of the coupling constant [2]. Typical examples of solitons are the kink, the vortex, and the magnetic monopole in field theory, and some D-branes in supergravity or superstring theories. In supersymmetric field theories these solutions preserve half of the supersymmetry and saturate BPS bounds. As for instantons, we have the Yang-Mills (YM) instantons in four dimensions [3, 4, 5], or tunnelling phenomena in quantum mechanics with a double-well potential as described by the kink, see e.g. [6], and there are various kinds of instantons in string theory, for example the D-instantons [7]. Also instantons preserve half the number of supersymmetries in supersymmetric field theories. Instantons can also be defined in field theories in dimensions higher than four [8], but we discuss in this chapter mainly the case of four dimensions.

Instantons in ordinary (i.e., nongravitational) quantum field theories are by definition solutions of the classical field equations in Euclidean space with finite action.¹ Only for a finite classical action S_{cl} is the factor $\exp[-\frac{1}{\hbar}S_{\text{cl}}]$ in the path integral nonvanishing. We shall consider instantons in nonabelian gauge theories in flat spacetime (there are no instantons in abelian gauge theories in flat space), both regular instantons (which actually have a singularity at $|x|^2 = \infty$) and singular instantons (which have a singularity at a point $x = x_0$ but not at $|x|^2 = \infty$). A singular gauge transformation maps the first into the second, and vice-versa². Around a given instanton solution, there are the quantum fluctuations. The action contains terms with 2, 3, 4 ... quantum fields, and one can perform perturbation theory around the instanton. The terms quadratic in quantum fields

¹In gravity there are various definitions of instantons: Einstein spaces with selfdual Weyl tensors, selfdual Riemann tensors, solutions of the Einstein equations with/without finite action etc. Since in gravity spacetime is part of the solution, one usually considers spacetime topologies which are different from that of $\mathbf{R}^4$. A selfdual Riemann tensor leads to an Einstein space ($R_{\mu\nu} = \Lambda g_{\mu\nu}$) whose Einstein-Hilbert action is either infinite (if the cosmological constant Λ is nonvanishing), or it only gets contributions from the Gibbons-Hawking boundary term [9]. In general, the semiclassical approximation of the Einstein-Hilbert action is not well defined due to the unboundedness of the action inside the path integral. To cure this, one probably has to discuss gravitational instantons inside a full theory for quantum gravity. For instanton solutions in flat space but using curvilinear coordinates (for example S^4 , or cylindrical coordinates) see [10].

²For the "regular solution", A_μ^a is finite on $\mathbf{R}^4 \cup \infty = \mathbf{S}^4$ everywhere, but this does not mean that it is regular. It is finite only because one can use two different patches to cover S^4 , and A_μ^a is regular in each patch. If one maps infinity to the origin by a space-inversion transformation ($x^\mu = y^\mu/y^2$), then one finds a singularity at the origin. In this sense the "regular solution" is singular. We further clarify this issue in the next section.

yield the propagators, which are complicated background-dependent expressions, and the terms cubic and higher in quantum fields yield the vertices. However, there is a subtlety with an instanton background: there are zero modes. A zero mode is by definition a solution of the linearized field equations for the fluctuations which is **normalizable**. (It is an eigenfunction of the quantum field operator with eigenvalue zero). In a trivial vacuum there are no zero modes: there are, of course, solutions of the linearized field equations, but they are not normalizable. We must treat the zero modes in instanton physics separately from the nonzero modes; for example, they have their own measure in the path integral. The nonzero modes live in the space orthogonal to the zero modes and in this space one can invert the linearized field equations for the fluctuations and construct propagators, and do perturbation theory.

Instantons describe tunnelling processes in Minkowski space-time from one vacuum at time t_1 to another vacuum at time t_2 . The simplest model which exhibits this phenomenon is a quantum mechanical point particle with a double-well potential having two vacua, or a periodic potential with infinitely many vacua. Classically there is no trajectory for a particle to travel from one vacuum to the other, but quantum mechanically tunnelling occurs. The tunnelling amplitude can be computed in the WKB approximation, and is typically exponentially suppressed. In the Euclidean picture, after performing a Wick rotation, the potential is turned upside down, and it is possible for a particle to propagate between the two vacua, as described by the classical solution to the Euclidean equations of motion. The claim is then that the contributions from instantons in Euclidean space yield a good approximation of the path integral in Minkowski space. We shall prove this for the case of quantum mechanics.

Also in YM theories, instantons are known to describe tunnelling processes between different vacua, labeled by an integer winding number, and lead to the introduction of the CP-violating θ -term in the action [11, 12]. It was hoped that instantons could shed some light on the mechanism of quark confinement. Although this was successfully shown in three-dimensional gauge theories (based on the Georgi-Glashow model) [13], the role of instantons in relation to confinement in four dimensions is less clear. Together with the non-perturbative chiral $U(1)$ anomaly in an instanton background, which leads to baryon number violation and a solution of the $U(1)$ problem [4, 5], instantons are used in phenomenological applications to QCD and the Standard Model. To avoid confusion, note that the triangle chiral anomalies in perturbative field theories in Minkowski space-time are canceled by choosing suitable multiplets of fermions. There remain, however, chiral anomalies at the non-perturbative level. It is hard to compute the non-perturbative terms in the effective action which lead to a breakdown of the chiral symmetry by using methods in Minkowski space-time. However, by using instantons in Euclidean space, one

can relatively easily determine these terms. The nonperturbative chiral anomalies are due to fermionic zero modes which appear in the path integral measure (in addition to bosonic zero modes). One must saturate the Grassmann integrals over these zero modes, and this leads to correlation functions of composite operators with fermion fields which do violate the chiral $U(1)$ symmetry. The new non-perturbative terms are first computed in Euclidean space, but then continued to Minkowski space where they give rise to new physical effects [5]. They have the following generic form in the effective action (we suppress here possible flavor or adjoint indices that the fermions can carry)

$$S_{\text{eff}} \propto e^{\left\{-\frac{8\pi^2}{g^2}(1+\mathcal{O}(g^2))+i\theta\right\}}(\bar{\lambda}\lambda)^n, \quad (1.1)$$

where $2n$ is the number of fermionic zero modes (n depends on the representation of the fermions and the gauge group). The prefactor is due to the classical instanton action and is clearly non-perturbative. The terms indicated by $\mathcal{O}(g^2)$ are due to standard radiative corrections computed by using Feynman graphs in an instanton background. The term $(\bar{\lambda}\lambda)^n$ involving antichiral spinors $\bar{\lambda}$ is produced if one saturates the integration in the path integral over the fermionic collective coordinates and violates in general the chiral symmetry. On top of (1.1) we have to add the contributions from anti-instantons, generating $(\lambda\lambda)^n$ terms in the effective action, where λ denotes chiral spinors. As we shall discuss, for Majorana spinors in Euclidean space the chiral and anti-chiral spinors are independent, but in Minkowski space-time they are related by complex conjugation, and one needs the sum of instanton and anti-instanton contributions to obtain a hermitean effective action.

We shall also apply the results of the general formalism to supersymmetric gauge theories, especially to the $\mathcal{N} = 4$ $SU(N)$ supersymmetric Yang-Mills (SYM) theory. Here $\mathcal{N}$ stands for the number of supersymmetries. Instantons in $\mathcal{N} = 1, 2$ models have been extensively studied in the past, see e.g.[14] for an early reference, and still are a topic of current research. For the $\mathcal{N} = 1$ models, one is mainly interested in the calculation of the superpotential and the gluino condensate [15, 16]. In some specific models, instantons also provide a mechanism for supersymmetry breaking [16], see [17] for a review on these issues. In the case of $\mathcal{N} = 2$, there are exact results for the prepotential [18] based only on general symmetry principles and electric-magnetic duality; the prepotential acquires contributions from all multi-instanton sectors. These predictions were successfully tested against direct field theoretical calculations in the one-instanton sector in [19], and for a two-instanton background in [20]. More recently, new techniques were developed to perform multi-instantons calculations in [21]. Finally, the nonperturbative structure of $\mathcal{N} = 4$ SYM has been studied thoroughly in the context of the AdS/CFT correspondence [22]. SYM instantons in the limit of large number of colors were successfully shown to reproduce the D-instanton contributions to certain correlation functions, both for single instantons

[23, 24] and for multi-instantons [25]. Other correlation functions were studied in [26, 27]. For a recent review of instantons in supersymmetric gauge theories, see [28].

The material is organized as follows. In section 2, we discuss the winding number of gauge fields, and we present the standard one-instanton solution in $SU(2)$ and in $SU(N)$. This already raises the question how to embed $SU(2)$ into $SU(N)$, and we discuss the various embeddings. In section 3 we discuss instanton solutions in general: we solve the duality condition and find multi-instanton solutions which depend on their position and their scale. We concentrate on the one-instanton solutions, and first determine the singular solutions, but then we make a (singular) gauge transformation and obtain the regular solutions. In section 4 we start the study of “collective coordinates”, the parameters on which the most general instanton solutions depend. We show that the number of collective coordinates is given by an index theorem for the Dirac operator in an instanton background. We then give a derivation of this index theorem, and conclude that a k -instanton solution in $SU(N)$ has $4Nk$ bosonic collective coordinates, $2Nk$ fermionic collective coordinates for fermions in the adjoint representation, and k fermionic collective coordinates for fermions in the defining (fundamental, vector) representation. In section 5 we explicitly construct the zero modes for gauge group $SU(N)$ in a one-instanton background. First we construct the bosonic zero modes; these are associated to the collective coordinates for translations, dilatations and gauge orientations. Next we derive the explicit formula for the general solution of the fermionic zero modes of the Dirac equation in a one-instanton background, first for $SU(2)$ and then for $SU(N)$.

In section 6 we construct the one-instanton measure for the bosonic and fermionic collective coordinates. We explain in detail the normalization of the zero modes since it is crucial for the construction of the measure. We convert the integration over the coefficients of the bosonic zero modes to an integration over the corresponding bosonic collective coordinates by the Faddeev-Popov trick, but for fermionic zero modes we do not need this procedure because in this case the coefficients of fermionic zero modes are already the fermionic collective coordinates. In section 7 we discuss the one-loop determinants in the background of an instanton, arising from integrating out the quantum fluctuations. We then apply this to supersymmetric theories, and we use an index theorem to prove that the determinants for all supersymmetric YM theories cancel each other. Furthermore, we compute the complete nonperturbative β function for supersymmetric Yang-Mills theories by assuming that the measure for the zero modes does not depend on the renormalization scale μ . However, since it is not known to which regularization scheme this procedure corresponds, this result cannot be checked by standard perturbative calculations. In section 8 we discuss the $\mathcal{N} = 4$ supersymmetric Yang-Mills theory in Euclidean space and its instantons.

The remaining sections contain applications. Section 9 discusses the problem of large

instantons and its solution in terms of Higgs fields and spontaneous symmetry breaking. Section 10 gives a detailed discussion how instantons can describe tunnelling. In section 11 we use a quantum mechanical model with a double-well potential to discuss the phase transition from a false vacuum to the true vacuum by bubble formation. Section 12 contains the strong CP problem, the mystery that the θ angle is so small. Section 13 discusses that instantons solve the $U(1)$ problem and in section 14 we finally discuss how instantons lead to baryon decay.

In a few appendices we set up our conventions and give a detailed derivation of some technical results in order to make the material self-contained. In appendix A we provide details of the calculation of the winding number. In appendix B we discuss the 't Hooft tensors and the spinor formalism in Euclidean space. In appendix C we calculate the volume of the moduli space of gauge orientations. Finally, in appendix D we show that conformal boosts and Lorentz rotations do not lead to additional zero modes.

2 Winding number and embeddings

We start with some elementary facts about instantons in $SU(N)$ Yang-Mills theories. The action, continued to Euclidean space, is

$$S = -\frac{1}{2g^2} \int d^4x \operatorname{tr} F_{\mu\nu} F_{\mu\nu} ; \quad F_{\mu\nu} = F_{\mu\nu}^a T_a . \quad (2.1)$$

The generators T_a are traceless anti-hermitean N by N matrices satisfying $[T_a, T_b] = f_{ab}^c T_c$ with real structure constants and $\operatorname{tr}(T_a T_b) = -\frac{1}{2} \delta_{ab}$. For instance, for $SU(2)$ one has $T_a = -\frac{i}{2} \tau_a$, where τ_a are the Pauli matrices and $f_{bc}^a = \epsilon_{bc}^a$. Notice that with these conventions the action is positive. Further conventions are $D_\mu Y = \partial_\mu Y + [A_\mu, Y]$ for any Lie algebra valued field Y , and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$, so that $F_{\mu\nu} = [D_\mu, D_\nu]$. The Euclidean metric is $\delta_{\mu\nu} = \operatorname{diag}(+, +, +, +)$. In (2.1), the only appearance of the coupling constant is in front of the action. The group metric $g_{ab} = \delta_{ab}$ is an invariant tensor³, so we may raise and lower indices with δ^{ab} and δ_{ab} . Thus we may also write $[T_a, T_b] = f_{abc} T_c$, and from now on we shall write group and Lorentz indices either as covariant indices or as contravariant indices, depending on which way is most convenient.

By definition, a Yang-Mills instanton is a solution of the classical Euclidean equations of motion with finite action. The classical equations of motion read

$$D_\mu F_{\mu\nu} = 0 . \quad (2.2)$$

³From $\operatorname{tr}[T_c, T_a T_b] = \operatorname{tr}([T_c, T_a] T_b + \operatorname{tr} T_a [T_c, T_b])$ it follows that g_{ab} is an invariant tensor: transforming its indices by an adjoint transformation with parameter λ^c yields zero: $\delta g_{ab} = \lambda^c f_{ca}^d g_{db} + \lambda^c f_{cb}^d g_{ac} = 0$.

To find solutions with finite action, we require that the field strength tends to zero at infinity faster than $|x|^{-2} \equiv r^{-2}$, hence the gauge fields asymptotically approach a pure gauge⁴

$$A_\mu \stackrel{|x|^2 \rightarrow \infty}{=} U^{-1} \partial_\mu U, \quad (2.3)$$

for some $U \in SU(N)$. To prove that gauge fields are pure gauge if the curvature $F_{\mu\nu}$ vanishes, is easy. Using $U \partial_\mu U^{-1} = -\partial_\mu U U^{-1}$ we must solve for U from $\partial_\mu U = -A_\mu U$, whose solution is the path-ordered integral $U = \exp[-\int^x A_\mu(y) dy^\mu]$. This expression does not depend on the path chosen because $F_{\mu\nu} = 0$. (Note, however, that if two gauge field configurations, say A_μ^I and A_μ^{II} , yield the same curvature, $F_{\mu\nu}(A^I) = F_{\mu\nu}(A^{II})$, they need not be gauge equivalent. A simple example proves this. Consider

$$A_\mu^I = \left\{ -\frac{1}{2} B y T_3, \frac{1}{2} B x T_3, 0, 0 \right\}; \quad A_\mu^{II} = \left\{ A_1^{II} = \sqrt{B} T_1, A_2^{II} = \sqrt{B} T_2, 0, 0 \right\}, \quad (2.4)$$

where B is a constant and T_a are the generators of $SU(2)$ with structure constants $f_{ab}^c = \epsilon_{abc}$. Clearly $F_{12}(A^I) = B T_3$ and also $F_{12}(A^{II}) = B T_3$ while all other components of $F_{\mu\nu}$ vanish. To prove that A_μ^I cannot be written as $U^{-1}(\partial_\mu + A_\mu^{II})U$ we note that if there was such a group element U , it should satisfy $U F_{\mu\nu} U^{-1} = F_{\mu\nu}$, hence U should commute with T_3 . This implies that U would be given by $\exp(f(x) T_3)$ for some real function $f(x)$. Then $-\frac{1}{2} B y T_3 = \partial_x f T_3 + e^{-f T_3} \sqrt{B} T_1 e^{-f T_3}$ which has no solution. One can also calculate a Wilson loop $W = \text{tr} P \exp \oint A dl$. This expression is gauge invariant, and if one chooses as loop a square in the $x - y$ plane with sides L_1 and L_2 , one finds

$$W^I = B L_1 L_2 T_3; \quad W^{II} = 2\sqrt{B}(L_1 T_1 + L_2 T_2) \quad (2.5)$$

If A_μ^I and A_μ^{II} were gauge equivalent, W^I should have been equal to W^{II} .

There is actually a way of classifying fields which satisfy the boundary condition in (2.3). It is known from homotopy theory that all gauge fields with vanishing field strength at infinity can be classified into sectors characterized by an integer number called the Pontryagin class, or the winding number, or instanton number, or the topological charge

$$k = -\frac{1}{16\pi^2} \int d^4x \text{tr} F_{\mu\nu} {}^* F_{\mu\nu}, \quad (2.6)$$

where ${}^* F_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F_{\rho\sigma}$ is the dual field strength, and $\epsilon_{1234} = 1$. Note that it is not necessary that these gauge fields satisfy the field equations, only that their field strength

⁴Another way of satisfying the finite action requirement is to first formulate the theory on a compactified $\mathbf{R}^4$, by adding and identifying points at infinity. Then the topology is that of the four-sphere, since $\mathbf{R}^4 \cup \infty \simeq S^4$. The stereographic map from $\mathbf{R}^4 \cup \infty$ to S^4 preserves the angles, and is therefore conformal. Also the YM action is conformally invariant, implying that the action and the field equations on $\mathbf{R}^4 \cup \infty$ are the same as on S^4 (the metric on the sphere is $g_{\mu\nu} = \delta_{\mu\nu}(1 + x^2)^{-2}$). The finiteness requirement is satisfied when the gauge potentials can be smoothly extended from $\mathbf{R}^4$ to S^4 . The action is then finite because S^4 is compact and A_μ is well-defined on the whole of the four-sphere.

vanishes sufficiently fast at $r = \infty$. The derivation of this result can be found in Appendix A. As part of the proof, one shows that the integrand in (2.6) is the divergence of a current

$$K_\mu = -\frac{1}{8\pi^2} \epsilon_{\mu\nu\rho\sigma} \text{tr} A_\nu (\partial_\rho A_\sigma + \frac{2}{3} A_\rho A_\sigma) . \quad (2.7)$$

The four-dimensional integral in (2.6) then reduces to an integral over a three-sphere at spatial infinity, and one can use (2.3) to show that the integer k counts how many times this spatial three-sphere covers the gauge group three-sphere $S^3 \approx SU(2) \subset SU(N)$. In more mathematical terms, the integer k corresponds to the third homotopy group $\pi_3(SU(2)) = \mathbf{Z}$. So k as defined in (2.6) does not depend on the values of the fields in the interior, but only on the fields at large $|x|^2$. This can also directly be seen: under a small variation $A_\mu \rightarrow A_\mu + \delta A_\mu$ one has $F_{\mu\nu} \rightarrow F_{\mu\nu} + D_\mu \delta A_\nu - D_\nu \delta A_\mu$, and partial integration (allowed when δA_μ is only nonzero in a region in the interior) yields $\delta A_\nu D_\mu {}^*F_{\mu\nu}$ which vanishes due to the Bianchi identity $D_{[\mu} F_{\nu\rho]} = 0$. (To prove this Bianchi identity one may use $F_{\nu\rho} = [D_\nu, D_\rho]$. In $[D_\mu, [D_\nu, D_\rho]] + [D_\nu, [D_\rho, D_\mu]] + [D_\rho, [D_\mu, D_\nu]]$ there are then 12 terms which cancel pairwise.)

Since we require instantons to have finite action, they satisfy the above boundary conditions at infinity, and hence they are classified by k , which we call the instanton number. Gauge potentials leading to field strengths with different instanton number can not be related by continuous gauge transformations. This follows from the fact that the instanton number is a gauge invariant quantity. In a given topological sector, the field configuration which minimizes the action is a solution of the field equations. (It is a priori not obvious that there exist field configurations that minimize the action, but we shall construct such solutions, thereby explicitly proving that they exist). We now show that, in a given topological sector, the solution to the field equations that minimizes the action has either a selfdual or anti-selfdual field strength

$$F_{\mu\nu} = \pm {}^*F_{\mu\nu} = \pm \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F_{\rho\sigma} . \quad (2.8)$$

This equation is understood in Euclidean space, where $(*)^2 = 1$. In Minkowski space there are no real solutions to the selfduality equations since $(*)^2 = -1$. As seen from (2.6), **instantons (with selfdual field strength) have $k > 0$ whereas anti-instantons (with anti-selfdual field strength) have $k < 0$** . (Recall that $\text{tr} T_a T_b$ is negative). To see that minimum action solutions are indeed selfdual or anti-selfdual, we perform a trick similar to the one used for deriving the BPS bound for solitons: we write the action as the square of a sum plus a total derivative term

$$\begin{aligned} S &= -\frac{1}{2g^2} \int d^4x \text{tr} F^2 = -\frac{1}{4g^2} \int d^4x \text{tr} (F \mp {}^*F)^2 \mp \frac{1}{2g^2} \int d^4x \text{tr} F {}^*F \\ &\geq \mp \frac{1}{2g^2} \int d^4x \text{tr} F {}^*F = \frac{8\pi^2}{g^2} (\pm k) . \end{aligned} \quad (2.9)$$

We used that $\text{tr}^*F^*F = \text{tr}FF$ and omitted Lorentz indices to simplify the notation. The equality is satisfied if and only if the field strength is (anti-) selfdual. The value of the action is then $S_{\text{cl}} = (8\pi^2/g^2)|k|$, and has the same value for the instanton as for the anti-instanton. However, we can also add a theta-angle term to the action, which reads

$$S_\theta = -i\frac{\theta}{16\pi^2} \int d^4x \text{tr} F_{\mu\nu}^* F^{\mu\nu} = i\theta k = \pm i\theta|k| . \quad (2.10)$$

The plus or minus sign corresponds to the instanton and anti-instanton respectively, so the theta-angle distinguishes between them. In Minkowski spacetime this term is the same because both d^4x and $F_{\mu\nu}^* F^{\mu\nu}$ produce a factor i under a Wick rotation. We give a more detailed treatment of the theta-angle term and its applications in Section 12.

It is interesting to note that the energy-momentum tensor for a selfdual (or anti-selfdual) field strength always vanishes⁵

$$T_{\mu\nu} = -\frac{2}{g^2} \text{tr} \left\{ F_{\mu\rho} F_{\nu\rho} - \frac{1}{4} \delta_{\mu\nu} F_{\rho\sigma} F_{\rho\sigma} \right\} = 0 . \quad (2.11)$$

(Because in Euclidean space $T_{44} = -\frac{1}{g^2} \text{tr} (\vec{E}^2 - \vec{B}^2)$, the Euclidean “energy” T_{44} need not be positive definite). This agrees with the observation that the instanton action $\int d^4x \text{tr} F^2 = \int d^4x \text{tr}^* FF$ is metric independent in curved space. The vanishing of the energy-momentum tensor is consistent with the fact that instantons are topological in nature. It implies that instantons do not curve Euclidean space, as follows from the Einstein equations.

An explicit construction of finite action solutions of the Euclidean classical equations of motion was given by Belavin et al. [3]. We shall derive this solution, and others, in section 3, but to get oriented we present it here, and discuss some of its properties. The gauge configuration for one instanton ($k = 1$) in $SU(2)$ contains the matrices $\sigma_{\mu\nu}$ or $\bar{\sigma}_{\mu\nu}$. One often writes it in terms of the ’t Hooft η tensors, related to $\bar{\sigma}_{\mu\nu}$ by $\bar{\sigma}_{\mu\nu} = i\eta_{\mu\nu}^a \tau_a$ where τ_a are the generators of $SU(2)$. We discuss these tensors in Appendix B. The regular one-instanton solution reads then

$$\begin{aligned} A_\mu^a(x; x_0, \rho) &= 2 \frac{\eta_{\mu\nu}^a (x - x_0)^\nu}{(x - x_0)^2 + \rho^2} , \\ A_\mu &\equiv A_\mu^a \left(\frac{\tau_a}{2i} \right) = -\frac{\bar{\sigma}_{\mu\nu} (x - x_0)^\nu}{(x - x_0)^2 + \rho^2} , \end{aligned} \quad (2.12)$$

where x_0 and ρ are arbitrary parameters called collective coordinates. They correspond to the position and the size of the instanton. The above expression solves the selfduality

⁵Note that T_{12} is proportional to $\text{tr}(F_{13}F_{23} + F_{14}F_{24})$, which is equal to minus itself due to the selfduality relations $F_{12} = F_{34}$, $F_{13} = -F_{24}$ and $F_{14} = F_{23}$. Similarly T_{11} vanishes because it is proportional to the trace of $(F_{12}^2 + F_{13}^2 + F_{14}^2) - (F_{23}^2 + F_{24}^2 + F_{34}^2)$.

equations for any value of the collective coordinates. Notice that it is regular at $x = x_0$, as long as $\rho \neq 0$. The real antisymmetric eta-symbols are defined as follows

$$\begin{aligned}\eta^a_{\mu\nu} &= \epsilon^a_{\mu\nu} & \mu, \nu &= 1, 2, 3, & \eta^a_{\mu 4} &= -\eta^a_{4\mu} = \delta^a_\mu, \\ \bar{\eta}^a_{\mu\nu} &= \epsilon^a_{\mu\nu} & \mu, \nu &= 1, 2, 3, & \bar{\eta}^a_{\mu 4} &= -\bar{\eta}^a_{4\mu} = -\delta^a_\mu.\end{aligned}\quad (2.13)$$

The η and $\bar{\eta}$ -tensors are selfdual and anti-selfdual respectively, for fixed index a . They form a basis for antisymmetric four by four matrices, and we have listed their properties in Appendix B. They are linear combinations of the Euclidean Lorentz generators $L_{\mu\nu}$, namely $\eta^a_{\mu\nu} = (J^a + K^a)_{\mu\nu}$ and $\bar{\eta}^a_{\mu\nu} = (J^a - K^a)_{\mu\nu}$, where $J^a = \epsilon^{abc}L_{bc}$ and $K^a = L_{a4}$, and $(L_{mn})_{\mu\nu} = \delta_{m\mu}\delta_{n\nu} - \delta_{m\nu}\delta_{n\mu}$ with $m, n = 1, 4$. In this subsection we use η tensors, but in later sections we shall use the matrices $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$.

The field strength corresponding to this gauge potential is (use (B.5))

$$F^a_{\mu\nu} = -4\eta^a_{\mu\nu} \frac{\rho^2}{[(x - x_0)^2 + \rho^2]^2}, \quad (2.14)$$

and it is selfdual. Thus (2.12) is a solution of the classical field equations. Far away, A^a_μ becomes proportional to the inverse radius $\frac{1}{r}$ so that it contributes a finite amount to the integral for the winding number which is of the form $\int A^3(r^3 d\Omega)$, while $F_{\mu\nu}$ becomes proportional to $\frac{1}{r^4}$, yielding a finite action. However, A^a_μ itself vanishes at $r \rightarrow \infty$, hence we have a smooth configuration on S^4 . Notice that the special point $\rho = 0$, corresponding to zero size instantons, leads to zero field strength and corresponds to pure gauge. Strictly speaking, this point must therefore be excluded from the instanton moduli space of collective coordinates. Finally one can compute the value of the action by integrating the density

$$\text{tr } F_{\mu\nu} F^{\mu\nu} = -96 \frac{\rho^4}{[(x - x_0)^2 + \rho^2]^4}. \quad (2.15)$$

Using the integral given at the end of Appendix B, one finds that this solution corresponds to $k = 1$.

One may show by direct calculation that the regular one-anti-instanton solution is also given by (2.12) but with $\bar{\eta}^a_{\mu\nu}$. (In the proof one uses that the first formula in (B.5) also holds for $\bar{\eta}^a_{\mu\nu}$).

We shall also derive the one-instanton solution in the singular gauge. In terms of η symbols it reads

$$A^a_\mu = 2 \frac{\rho^2 \bar{\eta}^a_{\mu\nu} (x - x_0)_\nu}{(x - x_0)^2 [(x - x_0)^2 + \rho^2]} = -\bar{\eta}^a_{\mu\nu} \partial_\nu \ln \left\{ 1 + \frac{\rho^2}{(x - x_0)^2} \right\}. \quad (2.16)$$

This gauge potential is singular for $x = x_0$, where it approaches a pure gauge configuration as we shall show in the next section, $A_\mu \stackrel{x \rightarrow x_0}{=} U \partial_\mu U^{-1}$. The gauge transformation U is

singular and relates the regular gauge instanton (2.12) to the singular one (2.16) at all points. The field strength in singular gauge is then (taking the instanton at the origin, $x_0 = 0$, otherwise replace $x \rightarrow x - x_0$)

$$F_{\mu\nu}^a = -\frac{4\rho^2}{(x^2 + \rho^2)^2} \left\{ \bar{\eta}_{\mu\nu}^a - 2\bar{\eta}_{\mu\rho}^a \frac{x_\rho x_\nu}{x^2} + 2\bar{\eta}_{\nu\rho}^a \frac{x_\rho x_\mu}{x^2} \right\} . \quad (2.17)$$

Notice that despite the presence of the anti-selfdual eta-tensors $\bar{\eta}$, this field strength is still selfdual, as can be seen by using the properties of the eta-tensors given in (B.5). The singular gauge is frequently used, because, as we will see later, zero modes fall off more rapidly at large x in the singular gauge. One can compute the winding number again in singular gauge. Then one finds that there is no contribution coming from infinity. Instead, all the winding is coming from the singularity at the origin. The singular solution is singular at x_0 , so one would expect that the regular solution is singular at infinity. This may seem puzzling since we saw that the regular solution was smooth on S^4 . However, to decide whether a configuration is smooth at $r \rightarrow \infty$, one should first transform the point at infinity to the origin and then study how the transformed configuration behaves near the origin. Making the coordinate transformation $x^\mu = y^\mu/y^2$ or $x^\mu = -y^\mu/y^2$, not forgetting that a vector field transforms as $A'_\mu(y) = (\partial x^\nu/\partial y^\mu) A_\nu(x)$, one finds that the transformed regular $k = 1$ solution is indeed singular at the origin⁶. In fact, it is equal to the singular $k = -1$ solution with ρ replaced by $\frac{1}{\rho}$.

At first sight it seems that there are five collective coordinates for the $k = 1$ solution. There are however extra collective coordinates corresponding to the gauge orientation. One can act with an $SU(2)$ matrix on the solution (2.12) to obtain another solution,

$$A_\mu(x; x_0, \rho, \vec{\theta}) = U^{-1}(\vec{\theta}) A_\mu(x; x_0, \rho) U(\vec{\theta}) , \quad U \in SU(2) . \quad (2.18)$$

with constant $\vec{\theta}$. One might think that these configurations should not be considered as a new solution since they are gauge equivalent to the expression given above. This is not true, however, the reason being that, after we fix the gauge, we still have left a rigid $SU(2)$ symmetry which acts as in (2.18). So in total there are eight collective coordinates, also called moduli. In principle, one could also act with the (space-time) rotation matrices $SO(4)$ on the instanton solution, and construct new solutions. However, these rotations can be undone by suitably chosen gauge transformations [29]. Actually, the Yang-Mills action is not only invariant under the Poincaré algebra (and the gauge algebra), but it is also

⁶This coordinate transformation in $\mathbf{R}^4$ can be viewed as a product of two conformal projections, one from the plane to the coordinate patch on the sphere S^4 containing the south pole, and the other from the other coordinate patch on S^4 with the north pole back to the plane. The transformed metric is $g'_{\mu\nu}(y) = \delta_{\mu\nu}/y^4$, so conformally flat. Then the action for the $A'_\mu(y)$ in y -coordinates is again the usual flat space action in (2.1), and the transformed instanton solution is an anti-instanton solution.

invariant under the conformal algebra which contains the Poincaré algebra and further the generators for dilatations (D) and conformal boosts (K_μ). As shown in Appendix D, for the Euclidean conformal group $SO(5, 1)$, the subgroup $SO(5)$ consisting of $SO(4)$ rotations and a combination of conformal boosts and translations ($R^\mu \equiv K^\mu + \rho^2 P^\mu$), leaves the instanton invariant up to gauge transformations. This leads to a 5 parameter instanton moduli space $SO(5, 1)/SO(5)$, which is the Euclidean version of the five-dimensional anti-de Sitter space AdS_5 . The coordinates on this manifold correspond to the four positions and the size ρ of the instanton. On top of that, there are still three gauge orientation collective coordinates, yielding a total of eight moduli for the $k = 1$ instanton in $SU(2)$.

Instantons in $SU(N)$ can be obtained by embedding $SU(2)$ instantons into $SU(N)$. For instance, a particular embedding is given by the following N by N matrix

$$A_\mu^{SU(N)} = \begin{pmatrix} 0 & 0 \\ 0 & A_\mu^{SU(2)} \end{pmatrix}. \quad (2.19)$$

where the instanton resides in the 2×2 matrix on the lower right. Of course this is not the most general solution, as one can choose different embeddings, see below.

One can act with a general $SU(N)$ element on the solution (2.19) and obtain a new one. Not all elements of $SU(N)$ generate a new solution. There is a stability group that leaves (2.19) invariant, acting only on the zeros, or commuting trivially with the $SU(2)$ embedding. Such group elements should be divided out, so we consider, for $N > 2$,

$$A_\mu^{SU(N)} = U \begin{pmatrix} 0 & 0 \\ 0 & A_\mu^{SU(2)} \end{pmatrix} U^\dagger, \quad U \in \frac{SU(N)}{SU(N-2) \times U(1)}. \quad (2.20)$$

One can now count the number of collective coordinates. From counting the dimension of the coset space in (2.20), one finds there are $4N - 5$ parameters. Together with the position and the scale of the $SU(2)$ solution, we find in total $4N$ collective coordinates for a one-instanton solution in $SU(N)$. It is instructive to work out the example of $SU(3)$. Here we use the eight Gell-Mann matrices $\{\lambda_\alpha\}$, $\alpha = 1, \dots, 8$. The first three λ_a , $a = 1, 2, 3$, form an $SU(2)$ algebra and are used to define the $k = 1$ instanton by contracting (2.12) or (2.16) with λ_a . The generators $\lambda_4, \dots, \lambda_7$ form two doublets under this $SU(2)$, so they act on the instanton and can be used to generate new solutions. This yields four more collective coordinates. Then there is λ_8 , corresponding to the $U(1)$ factor in (2.20). It commutes with the $SU(2)$ subgroup spanned by λ_a , and so it belongs to the stability group leaving the instanton invariant. So for $SU(3)$ and $k = 1$, there are seven gauge orientation zero modes, which agrees with $4N - 5$ for $N = 3$.

The embedding of instanton solutions as a 2×2 block inside the $N \times N$ matrix representation of $SU(N)$ is not the only embedding possible. For example, one can also use

the 3×3 matrix representation T_a of $SU(2)$, and put the instanton inside a 3×3 block of the $\mathbf{N}$ of $SU(N)$. This $\mathbf{3}$ of $SU(2)$ is sometimes called “the other $SU(2)$ in $SU(3)$ ”, but it is simply the adjoint representation of $SU(2)$, which is also the defining representation of $SO(3)$, and is given by $(T_a)_{ij} = \epsilon_{iaj}$,

$$T_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}; \quad T_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}; \quad T_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (2.21)$$

This representation has the same structure constants $f_{abc} = \epsilon_{abc}$ as the representation $T_a = \tau_a/(2i)$, but now $\text{tr}\{T_a T_b\} = -2\delta_{ab}$, four times larger.

In fact, going back to the construction of the instanton, we note that **any** representation T_a of $SU(2)$ yields an instanton solution for $SU(N)$ as long as it fits inside the $N \times N$ matrices of $SU(N)$ [30]

$$A_\mu = 2\eta_{\mu\nu}^a T_a \frac{x^\nu}{x^2 + \rho^2}. \quad (2.22)$$

The $\mathbf{2}$ of $SU(2)$ with $T_a = \frac{\tau_a}{2i}$ yields (2.19), but any other representation yields another embedding.

For $SU(3)$ there are only two possibilities. We can embed the instanton using the $\mathbf{2}$ of $SU(2)$; this yields (2.19). But we can also use the matrices T_a given in (2.21) as the first 3 generators of $SU(3)$. For $SU(N)$ we can use any spin j representation of $SU(2)$ provided it fits inside the $N \times N$ matrices. Since the action and winding number are proportional to the trace $\text{tr } T_a T_b$, which is proportional to the quadratic Casimir operator $j(j+1)$ times the dimension $2j+1$ of the spin j representation⁷, we see that we get instanton solutions with winding number $k = \pm \frac{2}{3}j(j+1)(2j+1)$. For $j = 1/2$ this reduces to $k = \pm 1$. For the first few $SU(N)$ the results are as follows

$$\begin{aligned} SU(3) : \quad & k = \pm 1; k = \pm 4 \quad (j = 1/2 \text{ and } j = 1) \\ SU(4) : \quad & k = \pm 1; k = \pm 4; k = \pm 10 \quad (j = 1/2, 1, 3/2) \\ & k = \pm 2 \quad (\text{two } j = 1/2 \text{ in block form}) \\ SU(5) : \quad & k = \pm 1, \pm 4, \pm 10, \pm 20 \quad (j = \frac{1}{2}, 1, \frac{3}{2}, 2) \\ & k = \pm 2, \pm 5 \quad (j = \frac{1}{2} \oplus \frac{1}{2} \text{ and } j = \frac{1}{2} \oplus 1). \end{aligned} \quad (2.23)$$

All these instanton solutions with winding number $|k| > 1$ still are (anti-) selfdual, so they still have minimal action, determined by the winding number, so the same as k instantons embedded as 2×2 matrices but far apart. Two instantons far apart and each of the form

⁷Use $\delta^{ab} \text{tr } T_a T_b = -\text{tr } C_2(R) = -(2j+1)C_2(R)$ where the quadratic Casimir operator for the representation R with spin j is given by $C_2(R) = -\delta^{ab} T(R)_a T(R)_b = j(j+1)$.

(2.12) repel each other (as opposed to an instanton and anti-instanton) with an interaction energy proportional to $1/r$. Bringing k instantons together such that they sit all at the same point, gives solutions of the kind above. So far apart there is a small positive interaction, but when they are brought together the interaction energy vanishes. Hence, there must be domains of attraction in between. This already shows that the interaction of instantons is a complicated problem [30]. In fact, one can deform these single-instanton solutions such that a multi-instanton solution is obtained in which the single-instantons do not attract or repel each other. In other words, in such a multi-instanton solution the positions, sizes and gauge orientations of the single instantons are collective coordinates.

For the general multi-instanton solution, the dependence on all collective coordinates is in implicit form given by the ADHM construction [31]. For a recent review, see [32]. In the next section we will obtain explicit formulas for the dependence on $5k$ collective coordinates. Explicit formulas for the dependence on all collective coordinates only exist for the $k = 2$ instanton solution [31, 33, 34, 32] and the $k = 3$ instanton solution [35].

We end this section with some remarks on embeddings into other gauge groups [36]. For $k = 1$ and gauge group $SO(N)$, it is known that there are $4N - 8$ collective coordinates. This can be understood as follows. The one-instanton solution is constructed by choosing an embedding of $SO(4) = SU(2) \times SU(2)$ generated by $\eta_{\mu\nu}^a$ and $\bar{\eta}_{\mu\nu}^a$, and putting the instanton in one of the $SU(2)$ groups. The stability group of this instanton is $SO(N - 4) \times SU(2)$, so we obtain (for $N > 4$)

$$A_{\mu}^{SO(N)} = U \begin{pmatrix} 0 & 0 \\ 0 & A_{\mu}^{SU(2)} \end{pmatrix} U^{\dagger}, \quad U \in \frac{SO(N)}{SO(N-4) \times SU(2)}. \quad (2.24)$$

The number of collective coordinates of such solutions follows from the dimension of the coset (which is $4N - 13$). Including the positions and size of the $SU(2)$ instanton, we arrive at $4N - 8$ for the total number of collective coordinates. Notice that for $N = 6$, we can use the isomorphism between $SO(6)$ and $SU(4)$. For both countings, we arrive at 16 moduli.

Similarly, we can analyze the symplectic gauge groups $USp(2N)$. Here we can simply choose the lower diagonal $SU(2) = USp(2)$ embedding inside $USp(2N)$ for a $k = 1$ instanton. The stability group of this embedding is now $USp(2N - 2)$, so for we have the following instanton solution:

$$A_{\mu}^{Sp(N)} = U \begin{pmatrix} 0 & 0 \\ 0 & A_{\mu}^{SU(2)} \end{pmatrix} U^{\dagger}, \quad U \in \frac{USp(2N)}{USp(2N-2)}. \quad (2.25)$$

The dimension of $USp(2N)$ is $N(2N+1)$,⁸ and so the total number of collective coordinates that follows from this construction is $5 + (4N - 1) = 4(N + 1)$, which is the correct number

⁸The dimension of $U(2N)$ is $4N^2$ and the generators have the form $\begin{pmatrix} a_1 + is_1 & b \\ -b^{\dagger} & a_2 + is_2 \end{pmatrix}$ where a_i is

[36]. For $N = 2$, we have the isomorphism $USp(2) = SO(5)$, which in both countings leads to 12 collective coordinates.

For higher instanton number, not all instantons can be constructed from a properly chosen embedding. There the ADHM formalism must be used. We just mention here that the total number of collective coordinates is $4kN$, $4k(N - 2)$ and $4k(N + 1)$ for the gauge groups $SU(N)$, $SO(N)$ and $USp(2N)$ respectively. The geometric relation between instanton moduli spaces and quaternionic manifolds (whose dimension is always a multiple of four) can e.g. be found in [37].

2.1 Some remarks on nonselfdual instanton solutions

Note that we have not shown that all solutions of (2.2) with finite action are given by selfdual (or anti-selfdual) field strengths. In principle there could be configurations which are extrema of the action, but are neither selfdual nor anti-selfdual⁹. For the gauge group $SU(2)$ this has been a long standing question. The first result was established in [38, 39, 40] where it was shown that for gauge groups $SU(2)$ and $SU(3)$, nonselfdual solutions cannot be local minima, hence if they exist, they should correspond to saddle points. The existence of nonselfdual solutions with finite action and gauge group $SU(2)$ was first established in [41], for $k = 0$, and later for $k \neq 0$ in [42]. For gauge group $SU(3)$ some results have been obtained in [43, 44]. The situation seems to be quite complicated, and no elegant and simple framework to address these issues has been found so far. For bigger gauge groups, it is easier to construct non-selfdual (or anti-selfdual) solutions. This becomes clear in the example of $SO(4) = SU(2) \times SU(2)$. If we associate a selfdual instanton to the first factor, and an anti-selfdual instanton to the second factor, the total field strength satisfies the equations of motion (2.2) but is neither selfdual nor anti-selfdual. Even simpler is the example of $SU(4)$. By choosing two commuting $SU(2)$ subgroups, we can embed both an

antisymmetric and s_i is symmetric. Complex symplectic matrices $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ satisfy $M^T \Omega + \Omega M = 0$ where $\Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$. The restriction that the unitary generators be also symplectic leads to $N^2 + N(N - 1)$ constraints ($D + A^T = 0$ and $C - C^T = B - B^T = 0$).

⁹It is possible to construct solutions for $SU(2)$ that are not selfdual, but not with finite action. An example is $A_\mu = -\frac{1}{2}\sigma_{\mu\nu}\frac{x_\nu}{r^2}$. Its field strength is $F_{\mu\nu} = \frac{1}{2}\sigma_{\mu\nu}/r^2 + \frac{1}{2}(x^\mu\sigma_{\mu\rho} - x^\nu\sigma_{\nu\rho})x^\rho/r^4$. One can check that it satisfies the second order equation of motion (2.2) (both $\partial_\mu F_{\mu\nu}$ and $[A_\mu, F_{\mu\nu}]$ vanish), but this configuration is not selfdual since $F_{\mu\nu} - *F_{\mu\nu} = \frac{1}{2}\sigma_{\mu\nu}\frac{1}{r^2}$. Because this field strength does not tend to zero fast enough at infinity, the action evaluated on this solution diverges logarithmically.

$SU(2)$ instanton and an anti-instanton inside $SU(4)$,

$$A_\mu^{SU(4)} = \begin{pmatrix} A_\mu^+ & 0 \\ 0 & A_\mu^- \end{pmatrix}, \quad (2.26)$$

where $A_\mu^\pm$ denotes the (anti-) selfdual $SU(2)$ gauge potentials with topological charges $k^\pm$. Clearly the total field strength is neither selfdual nor anti-selfdual, but satisfies the second order equations of motion. The instanton action is finite and the total topological charge is $k^+ - k^-$.

From the embedding (2.26) one can generate more solutions by acting on the gauge potential with a global gauge transformation $U \in SU(4)$. In this way, one generates new exact and nonselfdual solutions which are not of the form (2.26).

For $SU(N)$ gauge groups, one has even more possibilities. One can embed k_+ instantons and k_- anti-instantons on the (block)-diagonal of $SU(N)$, as long as $2(k_+ + k_-) \leq N$. If we take both $k_+ > 0$ and $k_- > 0$, the solution is clearly not selfdual or anti-selfdual and the instanton action, including the theta-angle, is given by

$$S = \frac{8\pi^2}{g^2}(k_+ + k_-) + i\theta(k_+ - k_-). \quad (2.27)$$

In a supersymmetric theory, these solutions will not preserve any supersymmetry. This is interesting in the context of the AdS/CFT correspondence that relates $\mathcal{N} = 4$ SYM theory to type IIB superstrings. In [45], it is shown that these non-selfdual Yang-Mills instantons are related to non-extremal (non BPS) D-instantons in IIB supergravity.

3 Regular and singular instanton solutions

To find explicit instanton solutions, we solve the selfduality (or anti-selfduality) equations $F_{\mu\nu} = {}^*F_{\mu\nu}$ where ${}^*F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}$ with $\mu, \nu = 1, 4$ and $\epsilon_{1234} = \epsilon^{1234} = 1$. Since $D_\mu {}^*F_{\mu\nu}$ vanishes identically due to the Bianchi identity, we then have a solution of the field equations, $D_\mu F_{\mu\nu} = 0$. The main idea is to make a suitable ansatz, and then to check that it yields solutions. The ansatz is (we restrict ourselves for the moment to the gauge group $SU(2)$)

$$A_\mu(x) = \alpha \sigma_{\mu\nu} \partial_\nu \ln \phi(x^2), \quad (3.1)$$

where α is a real constant to be fixed and $\sigma_{\mu\nu}$ is the 2×2 matrix representation of the Lorentz generators in Euclidean space. Since we shall be using these matrices $\sigma_{\mu\nu}$ a lot, we first discuss their properties in some detail, and then we shall come back below (3.22) to the construction of instanton solutions.

3.1 Lorentz and spinor algebra

In Euclidean space, a suitable 4×4 matrix representation of the Dirac matrices is given by

$$\gamma^\mu = \begin{pmatrix} 0 & -i(\sigma^\mu)^{\alpha\beta'} \\ i(\bar{\sigma}^\mu)_{\alpha'\beta} & 0 \end{pmatrix}, \quad \begin{aligned} \sigma^\mu &= (\vec{\tau}, iI) \\ \bar{\sigma}^\mu &= (\vec{\tau}, -iI) \end{aligned}, \quad (3.2)$$

where $\vec{\tau}$ are the Pauli matrices. We use slashes instead of dots on the spinor indices to indicate that we are in Euclidean space. All four Dirac matrices are hermitian, and satisfy $\{\gamma^\mu, \gamma^\nu\} = 2\delta^{\mu\nu}$. The matrix γ^5 is diagonal

$$\gamma^5 \equiv \gamma^1\gamma^2\gamma^3\gamma^4 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad (3.3)$$

and chiral spinors correspond to projections with $\frac{1}{2}(1 \pm \gamma_5)$ which yield the upper or lower two components of a nonchiral four-component spinor.

Since we are in Euclidean space, it does not matter whether we write the index μ as a contravariant or covariant index. In Minkowski space this representation (with γ^4 replaced by γ^0 where $\gamma^4 = i\gamma^0$, so that $(\gamma^k)^2 = +1$ but $(\gamma^0)^2 = -1$) is used for two-component spinor formalism. Four-component spinors are then decomposed into two-component spinors as $\psi = \begin{pmatrix} \lambda^\alpha \\ \bar{\chi}_{\dot{\alpha}} \end{pmatrix}$, and this explains the position of the spinor indices on σ^μ and $\bar{\sigma}^\mu$ in (3.2). The Euclidean Lorentz generators ($SO(4)$ generators) acting on 4-component spinors are $M_{\mu\nu} = \frac{1}{4}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$ and satisfy the Euclidean Lorentz algebra

$$[M_{\mu\nu}, M_{\rho\sigma}] = \delta_{\nu\rho}M_{\mu\sigma} - \delta_{\nu\sigma}M_{\mu\rho} - \delta_{\mu\rho}M_{\nu\sigma} + \delta_{\mu\sigma}M_{\nu\rho}. \quad (3.4)$$

However, this representation is reducible: the upper and lower components of ψ form separate representations

$$M_{\mu\nu} = \frac{1}{2} \begin{pmatrix} (\sigma^{\mu\nu})^\alpha{}_\beta & 0 \\ 0 & (\bar{\sigma}^{\mu\nu})_{\alpha'\beta'} \end{pmatrix}. \quad (3.5)$$

In terms of σ^μ and $\bar{\sigma}^\mu$ we then find the following two inequivalent spinor representations of $SO(4)$: $M_{\mu\nu} = \frac{1}{2}\sigma_{\mu\nu}$ and $M_{\mu\nu} = \frac{1}{2}\bar{\sigma}_{\mu\nu}$, where

$$\sigma^{\mu\nu} = \frac{1}{2}(\sigma^\mu\bar{\sigma}^\nu - \sigma^\nu\bar{\sigma}^\mu); \quad \bar{\sigma}^{\mu\nu} = \frac{1}{2}(\bar{\sigma}^\mu\sigma^\nu - \bar{\sigma}^\nu\sigma^\mu). \quad (3.6)$$

(It is customary not to include the factor $\frac{1}{2}$ in $M_{\mu\nu} = \frac{1}{2}\sigma_{\mu\nu}$ into the definition of $\sigma^{\mu\nu}$).

The matrices $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ satisfy some properties which we shall need repeatedly. First of all, they are anti-selfdual and selfdual, respectively

$$\sigma_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\sigma_{\rho\sigma}; \quad \bar{\sigma}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\bar{\sigma}_{\rho\sigma}. \quad (3.7)$$

This follows most easily by noting that the matrices γ_μ satisfy $\gamma_{[\mu}\gamma_{\nu]} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\gamma_\rho\gamma_\sigma\gamma_5$ where $\gamma_{\mu\nu} \equiv \gamma_{[\mu}\gamma_{\nu]} = \frac{1}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$. For example, $\gamma_1\gamma_2 = -\gamma_3\gamma_4\gamma_5$ because $\epsilon_{1234} = +1$. From this (anti)-selfduality one derives another useful property

$$\epsilon_{\mu\nu\rho\sigma}\sigma_{\sigma\tau} = \delta_{\mu\tau}\sigma_{\nu\rho} - \delta_{\nu\tau}\sigma_{\mu\rho} + \delta_{\rho\tau}\sigma_{\mu\nu} . \quad (3.8)$$

It is easiest to prove (3.8) by substituting (3.7) into the left-hand side, and decomposing the product of two ϵ -tensors into a sum of products of Kronecker tensors. Another proof is based on the ‘‘Schouten identity’’ which is the observation that a totally antisymmetric tensor with 5 indices vanishes in 4 dimensions (because there are always at least two indices equal). Writing the left-hand side of (3.8) as $\epsilon_{\mu\nu\rho\alpha}\delta_\beta^\tau\sigma_{\alpha\beta}$ and using the Schouten identity

$$\epsilon_{\mu\nu\rho\alpha}\delta_\beta^\tau = \epsilon_{\beta\nu\rho\alpha}\delta_\mu^\tau + \epsilon_{\mu\beta\rho\alpha}\delta_\nu^\tau + \epsilon_{\mu\nu\beta\alpha}\delta_\rho^\tau + \epsilon_{\mu\nu\rho\beta}\delta_\alpha^\tau , \quad (3.9)$$

the identity (3.7) can be used to prove the property (3.8) (the last term in (3.9) yields minus the contribution of the term on the left-hand side). In a similar way one may prove

$$\epsilon_{\mu\nu\rho\sigma}\bar{\sigma}_{\sigma\tau} = -\delta_{\mu\tau}\bar{\sigma}_{\nu\rho} + \delta_{\nu\tau}\bar{\sigma}_{\mu\rho} - \delta_{\rho\tau}\bar{\sigma}_{\mu\nu} . \quad (3.10)$$

The extra overall minus sign is due to the extra minus sign in the selfduality relation in (3.7).

Further identities are the commutator of two Lorentz generators, and the anticommutator which is proportional to the unit matrix in spinor space

$$\begin{aligned} [\sigma_{\mu\nu}, \sigma_{\rho\sigma}] &= 2\delta_{\nu\rho}\sigma_{\mu\sigma} + \text{three more terms} , \\ \{\sigma_{\mu\nu}, \sigma_{\rho\sigma}\} &= 2(\delta_{\mu\sigma}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\sigma}) + 2\epsilon_{\mu\nu\rho\sigma} . \end{aligned} \quad (3.11)$$

One easy way to prove or check these identities is to use 4×4 Dirac matrices; for example $\{\gamma_1\gamma_2, \gamma_3\gamma_4\} = 2\gamma_5$ and $\{\gamma_{12}, \gamma_{13}\} = 0$ but $\{\gamma_{12}, \gamma_{12}\} = -2$ and $[\gamma_{12}, \gamma_{13}] = -2\gamma_{23}$. Because $\gamma_5 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$, it is clear that the $\bar{\sigma}_{\mu\nu}$ satisfy the same commutation and anticommutation relations, but with a different sign for the ϵ symbol. In particular,

$$\{\bar{\sigma}_{\mu\nu}, \bar{\sigma}_{\rho\sigma}\} = 2(\delta_{\mu\sigma}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\sigma}) - 2\epsilon_{\mu\nu\rho\sigma} . \quad (3.12)$$

All these identities can also be derived using two-component spinor formalism for vectors. For example, a vector v^μ is written as $v^{\alpha\alpha'} \equiv v^\mu\sigma_\mu^{\alpha\alpha'}$, and then one may use such identities as

$$\delta_\mu^\nu \sim \delta_{\alpha\alpha'}^{\beta\beta'} \sim \delta_\alpha^\beta \delta_{\alpha'}^{\beta'} ; \quad \delta_{\mu\nu} \sim \epsilon_{\alpha\beta}\epsilon_{\alpha'\beta'} . \quad (3.13)$$

If one never introduces any vector indices at all but only uses spinor indices, this spinor formalism turns about all identities into trivialities, but we prefer to also keep vector indices

around. The other extreme is to expand $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ into Pauli matrices τ_a as $\sigma_{\mu\nu} = i\bar{\eta}_{\mu\nu}^a \tau_a$ and $\bar{\sigma}_{\mu\nu} = i\eta_{\mu\nu}^a \tau_a$ where $\eta_{\mu\nu}^a$ and $\bar{\eta}_{\mu\nu}^a$ are constructed from ϵ_{aij} and δ_{ai} tensors, as in (2.13). A whole calculus of these “t Hooft-tensors” can be set-up, and is often used. We discuss it in appendix B. We shall not limit ourselves to one of these extremes; proofs are given either by using 2-component spinors or 4×4 Dirac matrices, depending on which approach is simplest for a given problem.

The index structure of the ansatz for A_μ in (3.1) merits a short discussion. A Lie-algebra valued gauge field A_μ has indices i, j for a representation R of an $SU(N)$ group. For $SU(2)$ the generators in the defining representation are the Pauli matrices τ^a divided by $2i$, hence $A_\mu = (A_\mu)^i_j = A_\mu^a \left(\frac{\tau_a}{2i}\right)^i_j$. The ansatz for the instanton can then be written as

$$(A_\mu)^i_j = (\sigma_{\mu\nu})^i_j x^\nu f(x^2) . \quad (3.14)$$

The indices μ, ν are Lorentz indices, but the indices i, j are $SU(2)$ indices. Hence the matrix $(\sigma_{\mu\nu})^i_j$ carries simultaneously spacetime indices and internal $SU(2)$ indices. The matrices $\sigma_{\mu\nu}$ are indeed proportional to τ_a , $\sigma_{\mu\nu} = i\bar{\eta}_{\mu\nu}^a \tau_a$, as one may check for specific values of μ and ν , using

$$\begin{aligned} (\sigma_{\mu\nu})^i_j &= \frac{1}{2} \left\{ (\sigma_\mu)^{i\beta'} (\bar{\sigma}_\nu)_{\beta'j} - (\sigma_\nu)^{i\beta'} (\bar{\sigma}_\mu)_{\beta'j} \right\} \\ (\sigma_\mu)^{i\beta'} &= \{ \vec{\tau}, i \} \quad , \quad (\bar{\sigma}_\mu)_{\beta'j} = \{ \vec{\tau}, -i \} . \end{aligned} \quad (3.15)$$

The matrices $\eta_{\mu\nu}^a$ and $\bar{\eta}_{\mu\nu}^a$ are actually invariant tensors of a particular $SU(2)$ group. There are three groups $SU(2)$: the gauge group $SU(2)_g$ and the rotation group $SO(4) = SU(2)_L \times SU(2)_R$ generated by $\eta_{\mu\nu}^a$ and $\bar{\eta}_{\mu\nu}^a$. The tensor $\eta_{\mu\nu}^a$ is invariant under the combined $SU(2)_g$ gauge transformations acting on the index a generated by ϵ_{abc} , and the $SU(2)_L$ Lorentz transformations generated by $\eta_{\rho\sigma}^b$. Indeed, under infinitesimal variations with parameter λ^a we find, using (B.5),

$$\delta\eta^a = \epsilon^{abc} \eta^b \lambda_g^c + \frac{1}{2} \lambda_L^c [\eta^c, \eta^a] = 0 \quad \text{if } \lambda_g^a = \lambda_L^a . \quad (3.16)$$

Furthermore, $\eta_{\mu\nu}^a$ is separately invariant under the $SU(2)_R$ subgroup of the Lorentz group generated by $\bar{\eta}_{\rho\sigma}^b$; this follows from $[\eta^a, \bar{\eta}^b] = 0$. In fact, $\eta^a = L^a_4 + \frac{1}{2} \epsilon^{abc} L_{bc}$ and $\bar{\eta}^a = -L^a_4 + \frac{1}{2} \epsilon^{abc} L_{bc}$ from which $[\eta^a, \bar{\eta}^b] = 0$ easily follows.¹⁰

Spinor indices are raised and lowered by ϵ -tensors following the northwest-southeast convention: $v^{\alpha'} = \epsilon^{\alpha'\beta'} v_{\beta'}$ and $v^\alpha = \epsilon^{\alpha\beta} v_\beta$. So $(\bar{\sigma}^\mu)^{\beta'\alpha} = \epsilon^{\beta'\delta'} \epsilon^{\alpha\gamma} (\bar{\sigma}^\mu)_{\delta'\gamma}$. There are various definitions of these ϵ tensors in the literature; we define

$$\epsilon^{\alpha\beta} = -\epsilon^{\alpha'\beta'} . \quad (3.17)$$

¹⁰The 4×4 matrices L^a_4 and L_{bc} have entries $(L^a_4)_{\mu\nu} = \delta^a_\mu \delta_{4\nu}$ and $(L_{bc})_{\mu\nu} = \delta_{b\mu} \delta_{c\nu} - \delta_{c\mu} \delta_{b\nu}$. They form the defining representation of the Euclidean Lorentz algebra.

Note that numerically $\epsilon^{\alpha\beta} = \epsilon_{\alpha\beta}$ but also $\epsilon^{\alpha'\beta'} = \epsilon_{\alpha'\beta'}$ because one needs two ϵ tensors to raise or lower both indices of an ϵ tensor. We fix the overall sign by $\epsilon^{\alpha\beta} = \epsilon^{ij}$ where $\epsilon^{12} = 1$. A crucial relation in the spinor formalism which we shall frequently use is

$$\bar{\sigma}_{\mu,\alpha' i} = \sigma_{\mu,i \alpha'} , \quad (3.18)$$

where we recall that $\sigma_{\mu,i \alpha'} = \sigma_{\mu}^{j \beta'} \epsilon_{ji} \epsilon_{\beta' \alpha'}$.

Using 2-component spinor indices for vectors,

$$(\bar{\sigma}^{\mu})_{\alpha' \alpha} A_{\mu} \equiv A_{\alpha' \alpha} \text{ and } (\bar{\sigma}_{\nu})_{\beta' j} x^{\nu} \equiv x_{\beta' j} , \quad (3.19)$$

the ansatz for the instanton solution in (3.1) with spinor indices for A_{μ} becomes

$$\begin{aligned} (\bar{\sigma}^{\mu})_{\alpha' \alpha} (A_{\mu})^i_j &\equiv A_{\alpha' \alpha}^i_j = (\bar{\sigma}^{\mu})_{\alpha' \alpha} (\sigma_{\mu\nu})^i_j x^{\nu} f(x^2) \\ &= \left\{ \delta_{\alpha'}^{\beta'} \delta_{\alpha}^i x_{\beta' j} - \epsilon_{\alpha' \beta'} \epsilon_{\alpha j} x^{i \beta'} \right\} f(x^2) = \left\{ \delta_{\alpha}^i x_{\alpha' j} + \epsilon_{\alpha j} x^i_{\alpha'} \right\} f(x^2) . \end{aligned} \quad (3.20)$$

The trace over (ij) clearly vanishes, and this fixes the relative sign. We worked out the matrix $(\bar{\sigma}^{\mu})_{\alpha' \alpha} (\sigma_{\mu\nu})^i_j$ using

$$\bar{\sigma}_{\alpha' \alpha}^{\mu} \sigma_{\mu}^{i \beta'} = 2 \delta_{\alpha'}^{\beta'} \delta_{\alpha}^i , \quad (3.21)$$

and

$$\bar{\sigma}_{\alpha' \alpha}^{\mu} (\bar{\sigma}_{\mu})_{\beta' j} = \bar{\sigma}_{\alpha' \alpha}^{\mu} (\sigma_{\mu})_{j \beta'} = \bar{\sigma}_{\alpha' \alpha}^{\mu} \sigma_{\mu}^{k \gamma'} \epsilon_{kj} \epsilon_{\gamma' \beta'} = 2 \epsilon_{\alpha' \beta'} \epsilon_{\alpha j} . \quad (3.22)$$

3.2 Solving the selfduality equations

Let us now come back to the construction of instanton solutions. Substituting the ansatz for A_{μ} in (3.1) into the definition of $F_{\mu\nu}$ yields with (3.11)

$$\begin{aligned} F_{\mu\nu} &= \alpha \sigma_{\nu\rho} \partial_{\mu} \partial_{\rho} \ln \phi - \alpha \sigma_{\mu\rho} \partial_{\nu} \partial_{\rho} \ln \phi + \alpha^2 [\sigma_{\mu\rho}, \sigma_{\nu\sigma}] (\partial_{\rho} \ln \phi) (\partial_{\sigma} \ln \phi) \\ &= (\alpha \sigma_{\nu\rho} \partial_{\mu} \partial_{\rho} \ln \phi - \mu \leftrightarrow \nu) + 2\alpha^2 (\sigma_{\mu\sigma} \partial_{\nu} \ln \phi \partial_{\sigma} \ln \phi - \mu \leftrightarrow \nu) \\ &\quad - 2\alpha^2 \sigma_{\mu\nu} (\partial \ln \phi)^2 . \end{aligned} \quad (3.23)$$

We want to solve the equation $F_{\mu\nu} = {}^*F_{\mu\nu}$. The dual of $F_{\mu\nu}$ can be written as an expression without any ϵ tensor by using the identities for $\epsilon_{\mu\nu\rho\sigma} \sigma_{\sigma\tau}$ and $\epsilon_{\mu\nu\rho\sigma} \sigma_{\rho\sigma}$ in (3.7) and (3.8). One finds

$$\begin{aligned} {}^*F_{\mu\nu} &= \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F_{\rho\sigma} = \alpha \epsilon_{\mu\nu\rho\sigma} \sigma_{\sigma\alpha} \partial_{\rho} \partial_{\alpha} \ln \phi \\ &\quad + 2\alpha^2 \epsilon_{\mu\nu\rho\sigma} \sigma_{\rho\beta} \partial_{\sigma} \ln \phi \partial_{\beta} \ln \phi - \alpha^2 \epsilon_{\mu\nu\rho\sigma} \sigma_{\rho\sigma} (\partial \ln \phi)^2 \\ &= \sigma_{\nu\rho} (\alpha \partial_{\rho} \partial_{\mu} \ln \phi - 2\alpha^2 \partial_{\rho} \ln \phi \partial_{\mu} \ln \phi) - \mu \leftrightarrow \nu \\ &\quad + \sigma_{\mu\nu} (\alpha \partial^2 \ln \phi) . \end{aligned} \quad (3.24)$$

Equating $F_{\mu\nu}$ to ${}^*F_{\mu\nu}$ yields two equations for ϕ , namely one for the terms with $\sigma_{\nu\rho}$ and the other for the terms with $\sigma_{\mu\nu}$

$$\begin{aligned}\alpha\partial_\mu\partial_\rho\ln\phi - 2\alpha^2\partial_\mu\ln\phi\partial_\rho\ln\phi &= \alpha\partial_\mu\partial_\rho\ln\phi - 2\alpha^2\partial_\mu\ln\phi\partial_\rho\ln\phi, \\ -2\alpha^2(\partial\ln\phi)^2 &= \alpha\partial^2\ln\phi.\end{aligned}\tag{3.25}$$

The first equation is identically satisfied (for that reason we equated $F_{\mu\nu}$ to ${}^*F_{\mu\nu}$), while the second equation can be rewritten as $\partial^2\ln\phi + 2\alpha(\partial\ln\phi)^2 = 0$. For $\alpha = \frac{1}{2}$ it simplifies to $\partial^2\phi/\phi = 0$. (Setting $\alpha = \frac{1}{2}$ is not a restriction because rescaling $\phi \rightarrow \phi^{1/2\alpha}$ achieves the same result).

Setting $\phi = \frac{1}{x^2}$ yields for $x \neq 0$ a solution: $\partial^2\phi/\phi = x^2\partial_\mu(-2x^\mu/x^4) = 0$. However, this is also a solution at $x = 0$ because ∂^2x^{-2} is proportional to a delta function (note that the dimensions match) and $x^2\delta^4(x) = 0$

$$\partial^2\frac{1}{x^2} = -4\pi^2\delta^4(x).\tag{3.26}$$

To check the coefficient, we integrate over a small ball, which includes the point $x = 0$; we obtain then $\int \partial^2\frac{1}{x^2}d^4x = \int r^3drd\Omega_\mu\partial_\mu\frac{1}{x^2} = \int r^3d\Omega_\mu(-2x^\mu/x^4) = -4\pi^2$. (The surface of a sphere in 4 dimensions is $2\pi^2$).

We have thus found a selfdual solution

$$A_\mu(x) = \frac{1}{2}\sigma_{\mu\nu}\partial_\nu\ln\left[1 + \frac{\rho^2}{(x-a)^2}\right].\tag{3.27}$$

We have added unity to ϕ in order that $A_\mu(x)$ vanishes for large $|x|$. A more general solution is given by $A_\mu(x) = \frac{1}{2}\sigma_{\mu\nu}\partial_\nu\ln\phi$ with

$$\phi = 1 + \sum_{i=1}^k \frac{\rho_i^2}{(x-a_i)^2},\tag{3.28}$$

which also solves $\partial^2\phi/\phi = 0$. These are a class of k -instanton solutions, parameterized by $5k$ collective coordinates. In particular for $k = 1$ we find the one-instanton solution

$$\begin{aligned}A_\mu^{\text{sing}}(x) &= \frac{1}{2}\sigma_{\mu\nu}\partial_\nu\ln\left[1 + \frac{\rho^2}{(x-a)^2}\right], \\ &= -\sigma_{\mu\nu}\frac{\rho^2(x-a)_\nu}{(x-a)^2((x-a)^2 + \rho^2)} \quad (k=1, \text{ singular}).\end{aligned}\tag{3.29}$$

This solution is clearly singular at $x = a$, but one can remove the singularity at $x = a$ by a singular gauge transformation (which maps the singularity to $x^2 = \infty$). To determine this gauge transformation we first study the structure of the singularity. Near $x = a$ the singular solution becomes

$$A_\mu^{\text{sing}}(x) \approx -\sigma_{\mu\nu}\frac{(x-a)_\nu}{(x-a)^2},\tag{3.30}$$

which is a pure gauge field with $U(x-a)$ in (A.9)

$$U^{-1}\partial_\mu U = -\sigma_{\mu\nu}\frac{(x-a)_\nu}{(x-a)^2}; \quad U(x) = \frac{x_4 + ix_k\sigma_k}{\sqrt{x^2}} = i\bar{\sigma}_\mu x_\mu/\sqrt{x^2}. \quad (3.31)$$

Note that U is unitary, and U^{-1} equals $-i\sigma_\mu x_\mu/\sqrt{x^2}$, which follows from the property $\sigma_\rho\bar{\sigma}_\mu + \sigma_\mu\bar{\sigma}_\rho = 2\delta_{\rho\mu}$.

From (3.29) and (3.30) it follows that we can write A_μ^{sing} as

$$A_\mu^{\text{sing}}(x) = \frac{\rho^2}{(x-a)^2 + \rho^2} U^{-1}\partial_\mu U. \quad (3.32)$$

It is now clear that an opposite gauge tranformation removes the singularity at $x=0$

$$\begin{aligned} A_\mu^{\text{reg}}(x) &= U(\partial_\mu + A_\mu^{\text{sing}})U^{-1} = \partial_\mu U U^{-1} \left(-1 + \frac{\rho^2}{(x-a)^2 + \rho^2} \right) \\ &= (U\partial_\mu U^{-1}) \frac{(x-a)^2}{(x-a)^2 + \rho^2}. \end{aligned} \quad (3.33)$$

The expressions $U^{-1}\partial_\mu U$ and $U\partial_\mu U^{-1}$ are closely related; in fact, one finds by direct evaluation

$$U\partial_\mu U^{-1} = -\bar{\sigma}_{\mu\nu}\frac{(x-a)_\nu}{(x-a)^2}. \quad (3.34)$$

Thus the regular one-instanton solution is given by

$$A_\mu^{\text{reg}} = -\bar{\sigma}_{\mu\nu}\frac{(x-a)_\nu}{(x-a)^2 + \rho^2} \quad (k=1, \text{regular}). \quad (3.35)$$

Of course, the singular and the regular solution are both selfdual, because self-duality is a gauge-invariant property, but the field strengths differ by a gauge transformation. Setting $a=0$ for simplicity, one finds for the field strengths in the regular and singular gauge

$$\begin{aligned} F_{\mu\nu}^{\text{reg}} &= 2\bar{\sigma}_{\mu\nu}\frac{\rho^2}{[x^2 + \rho^2]^2} \quad (k=1, \text{regular}), \\ F_{\mu\nu}^{\text{sing}} &= U^{-1}F_{\mu\nu}^{\text{reg}}U = -\frac{ix^\rho\sigma_\rho}{\sqrt{x^2}}\frac{2\bar{\sigma}_{\mu\nu}\rho^2}{(x^2 + \rho^2)^2}\frac{ix^\sigma\bar{\sigma}_\sigma}{\sqrt{x^2}} \quad (k=1, \text{singular}). \end{aligned} \quad (3.36)$$

It is clear that $F_{\mu\nu}^{\text{reg}}$ is selfdual because $\bar{\sigma}_{\mu\nu}$ is selfdual, but also $F_{\mu\nu}^{\text{sing}}$ is selfdual¹¹ as is clear from acting with $\epsilon^{\mu\nu\rho\sigma}$ on $\bar{\sigma}_{\mu\nu}$.

¹¹Using some further identities which follow from the results for $[\gamma_{\mu\nu}, \gamma_\rho]$ and $\{\gamma_{\mu\nu}, \gamma_\rho\}$

$$\begin{aligned} \bar{\sigma}_\mu\sigma_\nu\rho &= \delta_{\mu\nu}\bar{\sigma}_\rho - \delta_{\mu\rho}\bar{\sigma}_\nu - \epsilon_{\mu\nu\rho\sigma}\bar{\sigma}_\sigma; & \bar{\sigma}_{\mu\nu}\bar{\sigma}_\rho &= \delta_{\nu\rho}\bar{\sigma}_\mu - \delta_{\mu\rho}\bar{\sigma}_\nu - \epsilon_{\mu\nu\rho\sigma}\bar{\sigma}_\sigma, \\ \sigma_{\mu\nu}\sigma_\rho &= \delta_{\nu\rho}\sigma_\mu - \delta_{\mu\rho}\sigma_\nu + \epsilon_{\mu\nu\rho\sigma}\sigma_\sigma; & \sigma_\rho\bar{\sigma}_{\mu\nu} &= \delta_{\rho\mu}\sigma_\nu - \delta_{\rho\nu}\sigma_\mu + \epsilon_{\rho\mu\nu\sigma}\sigma_\sigma, \end{aligned}$$

one finds for the $k=1$ singular solution

$$F_{\mu\nu}^{\text{sing}} = \frac{2\rho^2}{(x^2 + \rho^2)^2} \left(-2\frac{x_\mu x_\rho}{x^2}\sigma_{\rho\nu} + 2\frac{x_\nu x_\rho}{x^2}\sigma_{\rho\mu} + \sigma_{\mu\nu} \right)$$

In this form the selfduality is no longer manifest.

The action for the one-instanton solution is, of course, proportional to the winding number

$$S = -\frac{1}{2g^2} \int \text{tr } F_{\mu\nu} F_{\mu\nu} d^4x = -\frac{1}{2g^2} \int \text{tr } F_{\mu\nu} {}^*F_{\mu\nu} d^4x = \frac{8\pi^2}{g^2} . \quad (3.37)$$

The same result is obtained by direct evaluation of this integral.

The anti-instanton (the solution with $k = -1$) is closely related to the instanton solution. Recall that we derived the instanton solution by making the ansatz $A_\mu = \alpha \sigma_{\mu\nu} \partial_\nu \ln \phi$, evaluating $F_{\mu\nu}$ and ${}^*F_{\mu\nu}$ in terms of $\sigma_{\mu\nu}$ matrices, and then setting $F_{\mu\nu} = {}^*F_{\mu\nu}$. For the anti-instanton solution we make the ansatz $A_\mu = \beta \bar{\sigma}_{\mu\nu} \partial_\nu \ln \phi$. The expression for $F_{\mu\nu}$ is unchanged (except that A_μ contains $\bar{\sigma}_{\mu\nu}$ instead of $\sigma_{\mu\nu}$), but the $\bar{\sigma}_{\mu\nu}$ are selfdual instead of anti-selfdual, hence the expression for $\epsilon_{\mu\nu\rho\sigma} \bar{\sigma}_{\sigma\tau}$ has opposite signs from $\epsilon_{\mu\nu\rho\sigma} \sigma_{\sigma\tau}$. The equation with $\partial_\mu \partial_\rho \ln \phi$ again cancels if $F_{\mu\nu} = -{}^*F_{\mu\nu}$, which leads to opposite winding number ($k = -1$). The other equation is again $\partial^2 \ln \phi + 2\beta(\partial \ln \phi)^2 = 0$, hence $\beta = \frac{1}{2}$ and again $\phi = 1 + \sum_{i=1}^N \frac{\rho_i^2}{(x-a_i)^2}$. This yields for the singular-gauge anti-instanton solution

$$A_\mu^{\text{sing}} = -\bar{\sigma}_{\mu\nu} \frac{\rho^2(x-a)_\nu}{(x-a)^2[(x-a)^2 + \rho^2]} , \quad (k = -1, \text{ singular}) \quad (3.38)$$

Setting again temporarily $a = 0$, we find near $x = 0$

$$A_\mu^{\text{sing}} \approx -\bar{\sigma}_{\mu\nu} x_\nu / x^2 = U \partial_\mu U^{-1} , \quad (3.39)$$

with the same $U = i\bar{\sigma}^\mu x_\mu / \sqrt{x^2}$ as before. Similarly as for the instanton, we have

$$\begin{aligned} A_\mu^{\text{sing}} &= U \partial_\mu U^{-1} \frac{\rho^2}{x^2 + \rho^2} \\ A_\mu^{\text{reg}} &= U^{-1} (\partial_\mu + A_\mu^{\text{sing}}) U \\ &= \partial_\mu U^{-1} U \left(\frac{\rho^2}{x^2 + \rho^2} - 1 \right) \\ &= U^{-1} \partial_\mu U \left(\frac{x^2}{x^2 + \rho^2} \right) . \end{aligned} \quad (3.40)$$

Using the expression for $U^{-1} \partial_\mu U$ in (3.31) one finds

$$A_\mu^{\text{reg}} = -\sigma_{\mu\nu} \frac{(x-a)_\nu}{(x-a)^2 + \rho^2} \quad (k = -1, \text{ regular}) \quad (3.41)$$

The curvatures for the anti-instanton solution are obtained by interchanging $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ in the instanton solution

$$F_{\mu\nu}^{\text{reg}} = 2\sigma_{\mu\nu} \frac{\rho^2}{[(x-a)^2 + \rho^2]^2} \quad (k = -1, \text{ regular}) . \quad (3.42)$$

So, the only difference between the instanton and anti-instanton solutions is the exchange between $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ in $F_{\mu\nu}$ and A_μ . For the instanton solution, $F_{\mu\nu}^{\text{reg}}$ and A_μ^{reg} depend

on $\bar{\sigma}_{\mu\nu}$, but A_μ^{sing} depends on $\sigma_{\mu\nu}$, and $F_{\mu\nu}^{\text{sing}}$ also depends on $\sigma_{\mu\nu}$ (setting $a = 0$ again for notational simplicity),

$$F_{\mu\nu}^{\text{sing}} = U F_{\mu\nu}^{\text{reg}} U^{-1} = \frac{ix^\rho \bar{\sigma}_\rho}{\sqrt{x^2}} \frac{2\sigma_{\mu\nu} \rho^2}{(x^2 + \rho^2)^2} \frac{-ix^\sigma \sigma_\sigma}{\sqrt{x^2}} \quad (k = -1, \text{ singular}) . \quad (3.43)$$

If one evaluates the product of the σ matrices as in footnote 11, one finds an expression for $F_{\mu\nu}^{\text{sing}}$ in which the anti-selfduality is no longer manifest.

4 Collective coordinates, the index theorem and fermionic zero modes

We found in section 2 one-instanton solutions ($k = 1$) in $SU(N)$ with $4N$ parameters. The question arises whether these are all the solutions. To find this out, one can consider small deformations of the solution, $A_\mu + \delta A_\mu$, and study when they preserve selfduality. Expanding to first order in the deformation, and using that the variation of a curvature is the covariant derivative of the variation of the gauge field, this leads to the condition

$$D_\mu \delta A_\nu - D_\nu \delta A_\mu = *(D_\mu \delta A_\nu - D_\nu \delta A_\mu) , \quad (4.1)$$

where the covariant derivative depends only on the classical solution but not on δA_μ . In addition we require that the new solution is not related to the old one by a gauge transformation. This can be achieved by requiring that the small deformations are orthogonal to any small gauge transformation $D_\mu \Lambda$, for any function Λ , i.e.

$$\int d^4x \text{tr} \{ (D_\mu \Lambda) \delta A_\mu \} = 0 . \quad (4.2)$$

This certainly rules out deformations of the form $\delta A_\mu = D_\mu \Lambda$. After partial integration the orthogonality requirement leads to the usual gauge condition in the background field formalism

$$D_\mu \delta A^\mu = 0 . \quad (4.3)$$

At this point the reader may start feeling uneasy because the conditions (4.1) and (4.2) may seem too strong. First of all, the deformation should be a solution but need not be (anti-) selfdual. Furthermore, the field equation for the fluctuations consists of the sum of a classical piece and a piece from the gauge fixing term, so that, requiring each part to vanish separately may seem too restrictive. However, one can prove the following general result [46]. Arbitrary solutions of the fluctuations around an (anti-) instanton which are square-integrable so that they do not change the winding number, are themselves also (anti-) selfdual and transversal. To prove this property, note that the field equations

for the fluctuations read $D_\mu F^{\mu\nu}(A + \delta A) + D_\nu(D^\mu \delta A_\mu) = 0$. The second term comes from the gauge-fixing term. Taking the D^ν derivative, the first term vanishes while the second term yields $D^2(D^\mu \delta A_\mu) = 0$, hence $D^\mu \delta A_\mu$ on-shell. The terms in the classical action which are quadratic in the fluctuations can be written as $-\frac{1}{8}(f_{\mu\nu} - {}^*f_{\mu\nu})^2$ where $f_{\mu\nu} = D_\mu \delta A_\nu - D_\nu \delta A_\mu$. The minimum of the action yields a solution, hence $f_{\mu\nu} = {}^*f_{\mu\nu}$ on-shell. Thus imposing (4.1) and (4.3) is not too restrictive.

The requirement that δA_μ be square integrable is due to the fact that the inner product of zero modes δA_μ will later give us the metric or moduli space, which in turn will give us the integration measure of the moduli space. Also, for the index theorem which will be used to determine the number of zero modes, one needs the L^2 norm for fluctuations. It is remarkable that the zero modes which satisfy the differential equations in (4.1) and (4.3) are all square integrable.

In references [46, 36] the solutions of (4.1) subject to the condition (4.3) were studied using the Atiyah-Singer index theorem. Index theory turns out to be a useful tool when counting the number of solutions to a certain linear differential equation of the form $\hat{D}T = 0$, where $\hat{D}$ is some differential operator and T is a tensor. We will elaborate on this in the next subsection and also when studying fermionic collective coordinates. The ultimate result of [36] is that there are $4Nk$ solutions, leading indeed to $4N$ collective coordinates for $k = 1$ [46]. An assumption required to apply index theorems is that the space has to be compact. One must therefore compactify Euclidean space to a four-sphere S^4 , as was already discussed in footnote 4.

4.1 Bosonic collective coordinates and the Dirac operator

In this section we will make more precise statements about the number of solutions to the selfduality equations by relating it to the index of the Dirac operator. The problem is to determine the number of solutions to the (anti-)selfduality equations with topological charge k . For definiteness we consider anti-instantons, so we look for deformations which satisfy an anti-selfduality equation.

As explained in the last subsection, we study deformations of a given classical solution $A_\mu^{\text{cl}} + \delta A_\mu$. Let us define $\phi_\mu \equiv \delta A_\mu$ and $f_{\mu\nu} \equiv D_\mu \phi_\nu - D_\nu \phi_\mu$. The covariant derivative here contains only A_μ^{cl} . The constraints can then be written as

$$\bar{\sigma}_{\mu\nu} D_\mu \phi_\nu = 0 ; \quad D_\mu \phi_\mu = 0 , \quad (4.4)$$

which are $3 + 1$ relations. Indeed, more explicitly, $(\bar{\sigma}_{\mu\nu})_\alpha^{\alpha'} D_\mu \phi_\nu$ are 3 Lie-algebra valued expressions because $\alpha, \alpha' = 1, 2$ and $\text{tr} \bar{\sigma}_{\mu\nu} = 0$. To prove the first relation, multiply by $\bar{\sigma}_{\rho\sigma}$ and take the trace. Since the trace of $[\bar{\sigma}_{\rho\sigma}, \bar{\sigma}_{\mu\nu}]$ vanishes, while $\{\bar{\sigma}_{\rho\sigma}, \bar{\sigma}_{\mu\nu}\} = 2(\delta_{\mu\sigma} \delta_{\rho\nu} -$

$\delta_{\nu\sigma}\delta_{\rho\mu}) - 2\epsilon_{\rho\sigma\mu\nu}$, one finds $D_\sigma\phi_\rho - D_\rho\phi_\sigma - \epsilon_{\mu\nu\rho\sigma}D_\mu\phi_\nu = 0$, which is the anti-selfduality condition (3 relations). Both equations can be written as one simple equation as follows:

$$\bar{\sigma}_\mu\sigma_\nu D_\mu\phi_\nu = 0 , \quad (4.5)$$

because $\bar{\sigma}_\mu\sigma_\nu = \delta_{\mu\nu} + \bar{\sigma}_{\mu\nu}$, and the spinor structures of $\delta_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ are independent.

Introducing two-component spinor notation with

$$\bar{\mathcal{D}} = \bar{\sigma}_{\mu,\alpha'\beta}D_\mu = \bar{D}_{\alpha'\beta} ; \quad \sigma_\nu^{\alpha\beta'}\phi_\nu = \Phi^{\alpha\beta'} , \quad (4.6)$$

the deformations of an anti-instanton can be written as follows

$$\bar{\mathcal{D}}\Phi = \bar{D}_{\alpha'\beta}\Phi^{\beta\gamma'} = 0 . \quad (4.7)$$

Note that $\Phi^{\beta\gamma'}$ is in the adjoint representation, so (4.7) stands for $\partial_{\alpha'\beta}\Phi^{\beta\gamma'} + [A_{\alpha'\beta}, \Phi^{\beta\gamma'}] = 0$. Using the explicit representation of the matrices σ_μ in (B.9), we can represent the quaternion Φ by

$$\Phi = \begin{pmatrix} a & b^* \\ b & -a^* \end{pmatrix} , \quad (4.8)$$

with a and b complex adjoint-valued functions. Then (4.7) reduces to two spinor equations, one for

$$\lambda = \begin{pmatrix} a \\ b \end{pmatrix} ; \quad \bar{\mathcal{D}}\lambda = 0 , \quad (4.9)$$

and one for $i\sigma^2\lambda^* = \begin{pmatrix} b^* \\ -a^* \end{pmatrix}$. Conversely, for each spinor solution λ to the Dirac equation, one may show that also $i\sigma^2\lambda^*$ is a solution. (Use $(\bar{\sigma}^\mu)^* = -\sigma^2\bar{\sigma}^\mu\sigma^2$). Indeed, if λ yields a deformation $(\delta A_1, \delta A_2, \delta A_3, \delta A_4)$, then $i\sigma^2\lambda^*$ corresponds to the deformation $(\delta A'_1, \delta A'_2, \delta A'_3, \delta A'_4)$ with $\delta A'_1 = -\delta A_3, \delta A'_3 = \delta A_1, \delta A'_2 = \delta A_4$ and $\delta A'_4 = -\delta A_2$. They are not related by a Lorentz transformation because the coordinates x^μ are not transformed. Thus given λ , we obtain two linearly independent deformations of the (anti-) instanton. As we already stressed, the spinors λ are in the adjoint representation. We shall discuss other representations later.

Given a solution λ of the spinor equation, one can still construct **two** other solutions of the deformation of the anti-instanton, which differ by a factor i

$$\Phi^{(1)} = \begin{pmatrix} a & b^* \\ b & -a^* \end{pmatrix} , \quad \Phi^{(2)} = \begin{pmatrix} ia & -ib^* \\ ib & ia^* \end{pmatrix} . \quad (4.10)$$

The reason we do not count $i\lambda$ as a different solution for the spinors but treat $\Phi^{(1)}$ and $\Phi^{(2)}$ as independent has to do with reality properties: δA_μ^a should be real, and $\Phi^{(1)}$ and $\Phi^{(2)}$ yield different variations δA_μ . Namely, $a = \phi_3 + i\phi_4$ and $b = \phi_1 + i\phi_2$, so

$$\Phi^{(1)} : \delta A_4 = \phi_4 , \delta A_3 = \phi_3 , \delta A_1 = \phi_1 , \delta A_2 = \phi_2 ,$$

$$\Phi^{(2)} : \delta A_4 = \phi_3, \delta A_3 = -\phi_4, \delta A_1 = -\phi_2, \delta A_2 = \phi_1. \quad (4.11)$$

It may seem miraculous that we find a second solution without any hard work, but closer inspection reveals that no miracle is at work: under the substitutions $\delta A_1 \rightarrow \delta A_2, \delta A_2 \rightarrow -\delta A_1, \delta A_3 \rightarrow \delta A_4, \delta A_4 \rightarrow -\delta A_3$, one of the anti-selfduality equations is exchanged with the gauge condition, and the other two duality equations get interchanged. Also for solitons this way of counting zero modes is encountered: for example for vortices one complex fermion zero mode corresponds to two real bosonic zero modes [47].

In fact, because $\Phi^{(2)} = \Phi^{(1)}i\sigma_3$, one might wonder whether $\Phi^{(3)} = \Phi^{(1)}(-i\sigma_1)$ and $\Phi^{(4)} = \Phi^{(1)}(-i\sigma_2)$ yield further solutions. One obtains

$$\Phi^{(3)} = \begin{pmatrix} -ib^* & -ia \\ ia^* & -ib \end{pmatrix}; \quad \Phi^{(4)} = \begin{pmatrix} b^* & -a \\ -a^* & -b \end{pmatrix} \quad (4.12)$$

which are just the $\Phi^{(1)}$ constructed from $\sigma_2\lambda^*$ and $i\sigma_2\lambda^*$. So there are no further independent solutions [46]. Therefore, the number of solutions for Φ is twice the number of solutions for a single two-component adjoint spinor. So, the problem of counting the number of bosonic collective coordinates is now translated to the computation of the Dirac index, which we discuss next.

4.2 Fermionic moduli and the index theorem

Both motivated by the counting of bosonic collective coordinates, as discussed in the last subsection, and by the interest of coupling Yang-Mills theory to fermions, we study the Dirac equation in the background of an anti-instanton. We start with a massless four-component complex (Dirac) fermion ψ , in an arbitrary representation (adjoint, fundamental, etc) of an arbitrary gauge group

$$\gamma_\mu D_\mu \psi = \not{D}\psi = 0. \quad (4.13)$$

We recall that a Dirac spinor can be decomposed into its chiral and anti-chiral components

$$\psi = \begin{pmatrix} \lambda^\alpha \\ \bar{\chi}_{\alpha'} \end{pmatrix}; \quad \lambda \equiv \frac{1}{2}(1 + \gamma^5)\psi, \quad \bar{\chi} \equiv \frac{1}{2}(1 - \gamma^5)\psi. \quad (4.14)$$

We use the Euclidean representation for the Clifford algebra discussed before

$$\gamma^\mu = \begin{pmatrix} 0 & -i\sigma^{\mu\alpha\beta'} \\ i\bar{\sigma}^{\mu}_{\alpha'\beta} & 0 \end{pmatrix}, \quad \gamma^5 = \gamma^1\gamma^2\gamma^3\gamma^4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (4.15)$$

In Euclidean space the Lorentz group decomposes according to $SO(4) = SU(2) \times SU(2)$. The spinor indices α and α' correspond to the doublet representations of these two $SU(2)$

factors. As opposed to the case of Minkowski space, λ^α and $\bar{\chi}'_\alpha$ are not in complex-conjugate representations. The Dirac equation then becomes

$$\bar{\mathcal{D}}\lambda = 0, \quad \mathcal{D}\bar{\chi} = 0, \quad (4.16)$$

where $\mathcal{D}$ and $\bar{\mathcal{D}}$ are two-by-two matrixes, see (4.6), and λ and $\bar{\chi}$ are independent complex two-component spinors. We now show that in the presence of an anti-instanton, (4.16) has zero modes for λ , but not for $\bar{\chi}$. Conversely, in the background of an instanton, $\mathcal{D}$ has zero modes, but $\bar{\mathcal{D}}$ has not. A zero mode is by definition a solution of the linearized field equations for the quantum fluctuations **which is normalizable**. The fermionic fields are treated as quantum fields (there are no background fermionic fields), so normalizable solutions of (4.16) are zero modes.

The argument goes as follows. Given a zero mode $\bar{\chi}$ for $\mathcal{D}$, it also satisfies $\bar{\mathcal{D}}\mathcal{D}\bar{\chi} = 0$. In other words, $\ker \mathcal{D} \subset \ker \{\bar{\mathcal{D}}\mathcal{D}\}$ where $\ker$ denotes the kernel. Next we evaluate

$$\bar{\mathcal{D}}\mathcal{D} = \bar{\sigma}_\mu\sigma_\nu D_\mu D_\nu = D^2 + \frac{1}{2}\bar{\sigma}_{\mu\nu}F_{\mu\nu}, \quad (4.17)$$

where we have used $\bar{\sigma}_\mu\sigma_\nu + \bar{\sigma}_\nu\sigma_\mu = 2\delta_{\mu\nu}$, and $\bar{\sigma}_{\mu\nu}$ was defined in (3.6). But notice that the anti-instanton field strength is anti-selfdual whereas the tensor $\bar{\sigma}_{\mu\nu}$ is selfdual, so the second term vanishes. From this it follows that $\bar{\chi}$ satisfies $D^2\bar{\chi} = 0$. Now we can multiply $D^2\bar{\chi}$ with its conjugate $\bar{\chi}^*$ and integrate to get, after partial integration and assuming that the fields go to zero at infinity¹², $\int d^4x |D_\mu\bar{\chi}|^2 = 0$. From this it follows that $\bar{\chi}$ is covariantly constant, $D_\mu\bar{\chi} = 0$, and so $F_{\mu\nu}\bar{\chi} = 0$. Since $F_{\mu\nu}\bar{\chi} = F_{\mu\nu}^a T_a\bar{\chi}$, with T_a the generators of the gauge group $SU(2)$ in a representation R , we conclude that $F_{\mu\nu}^a(x)T_a\bar{\chi}(x)$ must vanish at all points x . Since $F_{\mu\nu}^a$ is proportional to $\eta_{a\mu\nu}$ (or $\bar{\eta}_{a\mu\nu}$), and $\eta_{a\mu\nu}\eta_{b\mu\nu}$ is proportional to δ_{ab} , we find that $T_a\bar{\chi}(x)$ vanishes for all a and all x . Then $D_\mu\bar{\chi} = 0$ reduces to $\partial_\mu\bar{\chi} = 0$, and this implies that $\bar{\chi} = 0$. We conclude that $\mathcal{D}\bar{\chi}$ has no square-integrable solutions. Stated differently, $-D^2$ is a positive definite operator and has no zero modes. Note that this result is independent of the representation of the fermion.

For the λ -equation, we have $\mathcal{D}\bar{\mathcal{D}}\lambda = 0$, i.e. $\ker \bar{\mathcal{D}} \subset \ker \{\mathcal{D}\bar{\mathcal{D}}\}$, and we obtain

$$\mathcal{D}\bar{\mathcal{D}} = D^2 + \frac{1}{2}\sigma_{\mu\nu}F_{\mu\nu}. \quad (4.18)$$

This time the second term does not vanish in the presence of an anti-instanton, so zero modes cannot be ruled out. In fact, there do exist fermionic zero modes, because we shall construct them. Knowing that $\mathcal{D}$ has no zero modes, one easily concludes that $\ker \bar{\mathcal{D}} = \ker \{\mathcal{D}\bar{\mathcal{D}}\}$ and $\ker \mathcal{D} = \ker \{\bar{\mathcal{D}}\mathcal{D}\} = 0$.

¹²Normalizability of zero modes requires that $\bar{\chi}$ tends to zero faster than $1/r^2$ (usually like $1/r^3$ or sometimes $1/r^4$). Then the boundary term with $\bar{\chi}^*D_\mu\bar{\chi}$ indeed vanishes.

For massive spinors no zero modes are possible. To prove this one may repeat the same steps as for massless spinors, but now one finds that $\bar{\mathcal{D}}\lambda = im\bar{\chi}$ and $\mathcal{D}\bar{\chi} = -im\lambda$, and iteration yields $\bar{\mathcal{D}}\mathcal{D}\bar{\chi} = m^2\bar{\chi}$. The crucial observation is that m^2 is positive, while $\bar{\mathcal{D}}\mathcal{D}$ is negative definite. Hence, no zero modes exist for massive spinors.

Now we can count the number of solutions using index theorems. The index of the Dirac operator is defined as

$$\text{Ind } \bar{\mathcal{D}} = \dim \ker \mathcal{D} \bar{\mathcal{D}} - \dim \ker \bar{\mathcal{D}} \mathcal{D} . \quad (4.19)$$

This index will give us the number of zero modes, since the second term is zero and since any renormalizable solution of $\mathcal{D} \bar{\mathcal{D}}\lambda = 0$ satisfies $\bar{\mathcal{D}}\lambda = 0$ as we have shown. There are several ways to compute its value. We begin by writing the index as follows

$$\text{Ind } \bar{\mathcal{D}} = \lim_{M^2 \rightarrow 0} \text{Tr} \left\{ \frac{M^2}{-\mathcal{D} \bar{\mathcal{D}} + M^2} - \frac{M^2}{-\bar{\mathcal{D}} \mathcal{D} + M^2} \right\} , \quad (4.20)$$

where M is an arbitrary parameter. The trace Tr stands for a sum over group indices and spinor indices, and includes an integration over space-time. We shall discuss that this expression (before taking the limit) is independent of M . This implies that the operators $\mathcal{D} \bar{\mathcal{D}}$ and $\bar{\mathcal{D}} \mathcal{D}$ not only have the same spectrum but also the same density of states for non-zero eigenvalues¹³. That they have the same non-zero eigenvalues is clear: if ψ is an eigenfunction of $\bar{\mathcal{D}} \mathcal{D}$, then $\mathcal{D}\psi$ is an eigenfunction of $\mathcal{D} \bar{\mathcal{D}}$ with the same nonvanishing eigenvalue and $\mathcal{D}\psi$ does not vanish. Conversely, if ψ is an eigenfunction of $\mathcal{D} \bar{\mathcal{D}}$ with nonzero eigenvalue, then $\bar{\mathcal{D}}\psi$ does not vanish and is an eigenfunction of $\bar{\mathcal{D}} \mathcal{D}$ with the same nonvanishing eigenvalue.

To show that (4.20) is independent of M^2 , we rewrite the index in terms of four-dimensional Dirac matrices,

$$I(M^2) \equiv \text{Ind } \bar{\mathcal{D}} = \text{Tr} \left\{ \frac{M^2}{-\mathcal{D}^2 + M^2} \gamma_5 \right\} , \quad (4.21)$$

where now $\mathcal{D} = D_\mu \gamma^\mu$.

$$\mathcal{D}_{4 \times 4} = \begin{pmatrix} 0 & \mathcal{D}_{2 \times 2} \\ \bar{\mathcal{D}}_{2 \times 2} & 0 \end{pmatrix} \quad (4.22)$$

We rewrote the trace of the two terms in (4.20) over a two-dimensional spinor space as the trace of one term over a four-dimensional spinor space. It has been argued that indepen-

¹³One can also (as is customary in the literature) place the system in a large box to discretize the spectrum, and let the boundary conditions for the eigenfunctions of $\mathcal{D} \bar{\mathcal{D}}$ determine the boundary conditions for the eigenfunctions of $\bar{\mathcal{D}} \mathcal{D}$, and vice-versa, such that the non-zero eigenvalues are the same. Such a treatment for the kink has been worked out in detail in [47]. However, in the limit of infinite volume, the densities of states can become different, as we shall discuss.

dence of M^2 follows by taking the M^2 -derivative (see [46]),

$$\frac{\partial}{\partial M^2} I(M^2) = -Tr \left\{ \frac{\not{D}^2}{(-\not{D}^2 + M^2)^2} \gamma_5 \right\} . \quad (4.23)$$

Using that γ_5 anticommutes with $\not{D}$ and that the trace is cyclic, we find

$$-Tr \frac{\not{D} \not{D} \gamma_5}{A} = Tr \frac{\not{D} \gamma_5 \not{D}}{A} = Tr \frac{\not{D} \not{D} \gamma_5}{A} , \quad (4.24)$$

where $A = (-\not{D}^2 + M^2)^2$. Hence $Tr(\not{D}^2 \gamma_5 / A)$ would seem to vanish and this would prove that $I(M^2)$ is independent of M^2 . The problem with this proof is that one can give a counter example: one can repeat all the steps for the supersymmetric kink, and this would then imply that the densities for chiral and anti-chiral fermion modes are equal. However, one can directly calculate these densities for the supersymmetric kink, and one then finds that they are different [48]

$$\Delta\rho(k^2) = -\frac{2m}{k^2 + m^2} , \quad (4.25)$$

where m is the mass of the fluctuating fields far away from the kink. Applied to the case of instantons, the situation was considered in [46]. In [49, 50, 51] it was noted that the proof of [46] was incomplete. Cyclicity of the trace (on which the proof in [46] that $\Delta\rho(k^2)$ vanishes is based), breaks down due to the presence of massless fluctuating fields¹⁴. One can directly compute $I(M^2)$, using a more detailed index theorem [50, 51], and then finds that the densities of chiral and antichiral fermionic modes in an instanton background are equal,

$$\Delta\rho(k^2) = 0 \quad \text{for instantons} . \quad (4.26)$$

Given that the density of states of the operator $\not{D} \bar{\not{D}}$ in (4.20) is the same as the density of states of the operator $\bar{\not{D}} \not{D}$, there is a pairwise cancellation in (4.20) coming from the sum over eigenstates with non-zero eigenvalues, both for the discrete and continuous spectrum. So the only contribution is coming from the zero modes, for which the first term in (4.20) simply gives one for each zero mode, and the second term vanishes because there are no zero modes. The result is then clearly an integer, namely $\dim \{\ker \bar{\not{D}}\}$. Since $I(M^2)$ is independent of M^2 , one can evaluate it in the large M^2 limit instead of the small M^2 limit. The calculation is then identical to the calculation of the chiral anomaly, which we now review.

¹⁴One would expect that at the regularized level the trace is cyclic but one may expect that one should also regularize infrared aspects of the problem. Consider for example quantum mechanics for a harmonic oscillator with mass term $\frac{1}{2}m^2 q^2$. Define $a = \sqrt{\frac{m}{2}}q + ip/\sqrt{2m}$ and $a^\dagger = \sqrt{\frac{m}{2}}q - ip/\sqrt{2m}$. For m tending to zero, the vacuum is annihilated by $p + \mathcal{O}(m)$ but the vacuum becomes non-normalizable when m vanishes. Still, at finite m , $tr[p, q] = 0$.

The chiral anomaly is equal to the regulated trace of the matrix γ_5 . It can be written as

$$\text{Ind } \bar{\mathcal{D}} = \text{tr} \int dx \langle x | \frac{M^2}{-\mathcal{D}\bar{\mathcal{D}} + M^2} - \frac{M^2}{-\bar{\mathcal{D}}\mathcal{D} + M^2} | x \rangle , \quad (4.27)$$

where tr denotes the trace over group indices and spinor indices. Because $\gamma_5 = \text{diag}(+1, +1, -1, -1)$ one finds in (4.27) a relative minus sign between the first and the second term. We have chosen a quantum mechanical representation for the trace in a Hilbert space spanned by the eigenfunctions $|x\rangle$ of the position operator. The operators D_μ depend on the operators $\hat{x}^\mu$ and the operators $\hat{p}_\mu$. When $\hat{x}$ reaches $|x\rangle$, it becomes a c -number x . Similarly $\hat{p}_\mu|x\rangle = -\frac{\hbar}{i}\frac{\partial}{\partial x^\mu}|x\rangle$. The latter statement follows by contracting with a complete set of momentum eigenstates, using that $\langle k|\hat{p}_\mu = \hbar k_\mu \langle k|$ because $\hat{p}_\mu$ is hermitian

$$\begin{aligned} \langle k | \hat{p}_\mu | x \rangle &= \hbar k_\mu \langle k | x \rangle = \hbar k_\mu \frac{e^{-ikx}}{(2\pi)^2} \\ &= -\frac{\hbar}{i} \frac{\partial}{\partial x^\mu} \frac{e^{-ikx}}{(2\pi)^2} = -\frac{\hbar}{i} \frac{\partial}{\partial x^\mu} \langle k | x \rangle = \langle k | \frac{-\hbar}{i} \frac{\partial}{\partial x^\mu} | x \rangle . \end{aligned} \quad (4.28)$$

So, from now on we will replace the operators $D_\mu(\hat{x}, \hat{p}_x)$ by $D_\mu(x, -\frac{\hbar}{i}\frac{\partial}{\partial x})$. These $\frac{\partial}{\partial x}$ act on the x in $|x\rangle$ and do not act on $|k\rangle$.

Let us now insert a complete set of eigenstates of $\bar{\mathcal{D}}\mathcal{D}$ and $\mathcal{D}\bar{\mathcal{D}}$, respectively. The index becomes then

$$\text{Ind } \bar{\mathcal{D}} = \text{tr} \sum_{m,n} \int dx \langle x | n_L \rangle \langle n_L | \mathcal{O}_L | m_L \rangle \langle m_L | x \rangle - \text{same with } L \leftrightarrow R , \quad (4.29)$$

where $\mathcal{O}_L$ is the first operator in (4.27) and $\mathcal{O}_R$ the second. As we already discussed the eigenfunctions $\langle x | n_L \rangle = \varphi_n^{(L)}(x)$ and $\langle x | n_R \rangle = \varphi_n^{(R)}(x)$ have the same nonvanishing eigenvalues λ_n and the same densities.

So the eigenfunctions with nonzero eigenvalues do not contribute to the index. (Note that it does not make sense to look for eigenfunctions of $\mathcal{D}$ or $\bar{\mathcal{D}}$ because these operators change the helicity of the spinors). There are in general a finite number of zero modes in the L sector but none in the R sector. Hence

$$\begin{aligned} \text{Ind } \bar{\mathcal{D}} &= \int d^4x \left(\sum_n \varphi_n^{(L)}(x) \varphi_m^{(L)}(x)^* - \sum_m \varphi_n^{(R)}(x) \varphi_m^{(R)}(x)^* \right) \frac{M^2 \delta_{mn}}{\lambda_n^2 + M^2} \\ &+ \sum_\alpha \int d^4x \varphi_\alpha^{(L)}(x) \varphi_\alpha^{(L)}(x)^* = n^{(L)} , \end{aligned} \quad (4.30)$$

where $\varphi_\alpha^{(L)}(x)$ are the (square-integrable) zero modes, and $n^{(L)}$ is the number of these. A sum over spinor indices is taken in (4.30).

To actually compute the index (namely, to compute the integer $n^{(L)}$), we use momentum eigenstates instead of eigenfunctions of $\not{D}\bar{\not{D}}$ and $\bar{\not{D}}\not{D}$

$$\text{Ind } \bar{\not{D}} = \int d^4x \int d^4k \int d^4k' \text{tr} \langle x | k' \rangle \langle k' | \frac{M^2}{-\not{D}^2 + M^2} \gamma_5 | k \rangle \langle k | x \rangle , \quad (4.31)$$

where we recall that $\gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\not{D} = \begin{pmatrix} 0 & -i\not{D} \\ i\bar{\not{D}} & 0 \end{pmatrix}$.

As we have discussed, the operator $D_\mu = \frac{\partial}{\partial x^\mu} + [A_\mu(x), \cdot]$ acts on the coordinates x in $|x\rangle$ but not on the k in $|k\rangle\langle k|$, and the trace tr sums over the group indices and the spinor indices of γ^μ in $\not{D} = \gamma^\mu D_\mu$ and γ_5 . Using $\langle k|x\rangle = e^{-ikx}/(2\pi)^2$ and pulling these plane waves to the left, the derivatives ∂_μ act on the c -numbers x in e^{-ikx} and are replaced by $\partial_\mu - ik_\mu$. The matrix element $\langle k'|M^2(-\not{D}^2 + M^2)^{-1}\gamma_5|k\rangle$ is equal to $\langle k'|k\rangle$ times the operator $[M^2/(-\not{D}^2 + M^2)]\gamma_5$ and $\langle k'|k\rangle = \delta^4(k - k')$. When the plane wave e^{-ikx} has been pulled all the way to the left, the plane waves $e^{ik'x}$ and e^{-ikx} in $\langle x|k'\rangle = e^{ik'x}/(2\pi)^2$ and $\langle k|x\rangle = e^{-ikx}/(2\pi)^2$ cancel each other, and one is left with

$$\text{Ind } \bar{\not{D}} = \int d^4x \int \frac{d^4k}{(2\pi)^4} \text{tr} \left\{ \frac{M^2}{-(-i\not{k} + \not{D})^2 + M^2} \gamma_5 \right\} . \quad (4.32)$$

The denominator can be written as

$$\frac{1}{(k^2 + M^2) - (-2ik \cdot D + D_\mu D_\mu + \frac{1}{2}\gamma_\mu \gamma_\nu F_{\mu\nu})} , \quad (4.33)$$

and we can exhibit the M^2 dependence by rescaling $k_\mu = M\kappa_\mu$, yielding

$$\text{Ind } \bar{\not{D}} = \int d^4x M^4 \int \frac{d^4\kappa}{(2\pi)^4} \text{tr} \left\{ \frac{1}{(\kappa^2 + 1) - \left(-\frac{2i\kappa_\mu D_\mu}{M} + \frac{D_\mu D_\mu}{M^2} + \frac{1}{2} \frac{\gamma_\mu \gamma_\nu F_{\mu\nu}}{M^2} \right)} \gamma_5 \right\} . \quad (4.34)$$

Expanding the denominator, only terms due to expanding two, three or four times can contribute in the limit $M \rightarrow \infty$, but only the terms with at least four Dirac matrices can contribute to the trace due to the matrix γ_5 . Thus we only need retain the square of $\frac{1}{2}\gamma_\mu \gamma_\nu F_{\mu\nu}$, and the index becomes

$$\begin{aligned} \text{Ind } \bar{\not{D}} &= \int d^4x \int \frac{d^4\kappa}{(2\pi)^4} \frac{1}{(\kappa^2 + 1)^3} \text{tr} \left(\frac{1}{2} F_{\mu\nu} \gamma_{\mu\nu} \frac{1}{2} F_{\rho\sigma} \gamma_{\rho\sigma} \gamma_5 \right) \\ &= \int d^4x \frac{2\pi^2}{(2\pi)^4} \int_0^\infty \frac{r^3 dr}{(r^2 + 1)^3} \frac{1}{4} (\text{tr } T_a T_b) (\text{tr } \gamma_{\mu\nu} \gamma_{\rho\sigma} \gamma_5) F_{\mu\nu}^a F_{\rho\sigma}^b \\ &= \int d^4x \frac{1}{32\pi^2} (\text{tr } T_a T_b) (\epsilon_{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^b) , \end{aligned} \quad (4.35)$$

where we used that $\int d\Omega_\mu = 2\pi^2$ and $\int_0^\infty \frac{r^3 dr}{(r^2+1)^3} = \frac{1}{4}$. Note that both a trace over group indices and a trace over spinor indices has been taken. The result for the index is twice the product of the winding number in (2.6) and a group theory factor

$$\text{Ind } \bar{\mathcal{D}} = 2 \left(\frac{1}{32\pi^2} \int d^4x F_{\mu\nu}^a * F_{\mu\nu}^b \right) \text{tr } T_a T_b . \quad (4.36)$$

For a representation R of $SU(N)$ for the fermions, we define $\text{tr } T_a^R T_b^R = -\delta_{ab} T(R)$. By definition one has $T(R) = \frac{1}{2}$ for the fundamental representation, and then $T(R) = N$ for the adjoint representation¹⁵. Hence, finally,

$$\begin{aligned} \text{Ind } \bar{\mathcal{D}} &= |k| && \text{for the fundamental representation,} \\ &= 2N|k| && \text{for the adjoint representation .} \end{aligned} \quad (4.37)$$

(For an anti-instanton, k is negative. The factor 2 corresponds to our earlier observation that $i\sigma_2\lambda^*$ is also a zero mode if λ is a zero mode.) Furthermore, as shown in the last subsection, an (anti-) instanton in $SU(N)$ has twice as many bosonic collective coordinates as there are fermionic zero modes in the adjoint representation. This proves that there are $4Nk$ bosonic collective coordinates for an instanton with winding number k and gauge group $SU(N)$.

5 Construction of zero modes

In two later sections we will show how to set up and do (one-loop) perturbation theory around an (anti-) instanton. This will require the reduction of the path integral measure over instanton field configurations to an integral over the moduli space of collective coordinates. In order to achieve this we need to know the explicit form of the bosonic and fermionic zero modes. This is the content of this section. We follow closely [52].

¹⁵To compute $T(R)$ for the adjoint representation, write the carrier space for the adjoint representation of $SU(N)$ as $u^i \bar{v}_j - \frac{1}{N} \delta_j^i (u^k \bar{v}_k)$. Then, for $i \neq j$, $T_a^{\text{adj}} u^i \bar{v}_j = (T_a^{(f)})^i_{i'} u^{i'} \bar{v}_j + (T_a^{f*})^j_{j'} u^i \bar{v}_{j'}$. For a diagonal generator A of the fundamental representation of $SU(N)$ with entries $(i\alpha_1, \dots, i\alpha_N)$ with real α_j one has $Au^i = i\alpha^i u^i$ and $Au^i \bar{v}_j = (i\alpha^i - i\alpha^j) u^i \bar{v}_j$, so $A(u^i \bar{v}_j - \frac{1}{N} \delta_j^i u^k \bar{v}_k) = (i\alpha^i - i\alpha^j)(u^i \bar{v}_j - \frac{1}{N} \delta_j^i u^k \bar{v}_k)$ and $\sum \alpha_i = 0$. Hence $\text{tr } A^2 = -\sum_{i=1}^N (\alpha^i)^2$ for the fundamental representation, but $\text{tr } A^2 = -\sum_{i,j} (\alpha^i - \alpha^j)^2$ for the adjoint representation. The latter sum can also be written as

$$\sum_{i,j=1}^N (\alpha_i - \alpha_j)^2 = \left(\sum \alpha_i^2 \right) N - 2 \left(\sum \alpha_i \right) \left(\sum \alpha_j \right) + \left(\sum \alpha_j^2 \right) N = 2 \left(\sum \alpha_i^2 \right) N .$$

So $T(R^{\text{adj}}) = 2NT(R^f)$.

5.1 Bosonic zero modes and their normalization

In order to construct the bosonic zero modes and discuss perturbation theory, we first decompose the fields into a background part and quantum fields

$$A_\mu = A_\mu^{\text{cl}}(\gamma) + A_\mu^{\text{qu}} . \quad (5.1)$$

Here γ_i denote a set of collective coordinates, and, for gauge group $SU(N)$, $i = 1, \dots, 4Nk$. Before we make the expansion of the action, we should first fix the gauge and introduce ghosts, c , and anti-ghosts, b . We choose the background gauge condition

$$D_\mu^{\text{cl}} A_\mu^{\text{qu}} = 0 . \quad (5.2)$$

The gauge-fixing term is then $\mathcal{L}_{\text{fix}} = -\frac{1}{g^2} \text{tr}(D_\mu A_\mu^{\text{qu}})^2$ and the ghost action is $\mathcal{L}_{\text{ghost}} = -b^a (D_\mu (A_\mu^{\text{cl}}) D_\mu (A_\mu^{\text{cl}} + A_\mu^{\text{qu}}) c)^a$. The action, expanded through quadratic order in the quantum fields, is of the form

$$S = \frac{8\pi^2}{g^2} |k| + \frac{1}{g^2} \text{tr} \int d^4x \{ A_\mu^{\text{qu}} M_{\mu\nu} A_\nu^{\text{qu}} + 2b M^{\text{gh}} c \} , \quad (5.3)$$

with $M^{\text{gh}} = D^2$ and

$$\begin{aligned} M_{\mu\nu} &= (D^2 \delta_{\mu\nu} - D_\nu D_\mu + F_{\mu\nu}) + D_\mu D_\nu \equiv M_{\mu\nu}^{(1)} + M_{\mu\nu}^{(2)} , \\ &= D^2 \delta_{\mu\nu} + 2F_{\mu\nu} , \end{aligned} \quad (5.4)$$

where we have dropped the subscript cl. Here, $M^{(1)}$ stands for the quadratic operator coming from the classical action, and $M^{(2)}$ is due to the gauge fixing term¹⁶. (Recall that $F_{\mu\nu}$ acts on A_ν^{qu} as $[F_{\mu\nu}, A_\nu^{\text{qu}}]$). In an expansion as in (5.3), one encounters zero modes (i.e. normalizable eigenfunctions of the operator $M_{\mu\nu}$ with zero eigenvalues). They are of the form

$$Z_\mu^{(i)} \equiv \frac{\partial A_\mu^{\text{cl}}}{\partial \gamma_i} + D_\mu^{\text{cl}} \Lambda^i , \quad (5.5)$$

where the gauge parameter Λ^i is chosen to keep Z_μ in the background gauge, so that

$$D_\mu^{\text{cl}} Z_\mu^{(i)} = 0 . \quad (5.6)$$

The first term in (5.5) is a solution of $M^{(1)}$ (i.e. an eigenfunction with zero eigenvalue), as follows from taking the derivative with respect to γ_i of the field equation. Namely, $\delta S^{\text{cl}} / \delta A_\mu^{\text{cl}} = 0$ for all γ_i , so

$$0 = \frac{\partial}{\partial \gamma_i} \frac{\delta S^{\text{cl}}}{\delta A_\mu^{\text{cl}}(x)} = \int \frac{\delta^2 S^{\text{cl}}}{\delta A_\nu^{\text{cl}}(y) \delta A_\mu^{\text{cl}}(x)} \partial_{\gamma_i} A_\nu^{\text{cl}}(y) d^4y . \quad (5.7)$$

¹⁶To arrive at this expression for $M_{\mu\nu}^{(1)}$, use that $F_{\mu\nu} = F_{\mu\nu}^{\text{cl}} + (D_\mu^{\text{cl}} A_\nu^{\text{qu}} - D_\nu^{\text{cl}} A_\mu^{\text{qu}}) + [A_\mu^{\text{qu}}, A_\nu^{\text{qu}}]$ and note that $-\frac{1}{2g^2} \text{tr} 2F_{\mu\nu}^{\text{cl}} [A_\mu^{\text{qu}}, A_\nu^{\text{qu}}] = \frac{1}{g^2} \text{tr} A_\mu^{\text{qu}} [F_{\mu\nu}^{\text{cl}}, A_\nu^{\text{qu}}]$.

The term $D_\mu \Lambda$ is also a solution of $M^{(1)}$, since it is a pure gauge transformation.¹⁷ The sum of the two terms is also a solution of $M^{(2)}$, because Λ is chosen such that Z_μ is in the background gauge. As we shall show, the solutions in (5.5) are normalizable, hence they are zero modes. Due to these zero modes, we cannot integrate over all quantum fluctuations, since the corresponding determinants would vanish and yield divergences in the path integral. They must therefore be extracted from the quantum fluctuations, in a way we will describe in a more general setting in the next subsection. It will turn out to be important to compute the matrix of inner products

$$U^{ij} \equiv \langle Z^{(i)} | Z^{(j)} \rangle \equiv -\frac{2}{g^2} \int d^4x \operatorname{tr} \{ Z_\mu^{(i)} Z^{\mu(j)} \} = \frac{1}{g^2} \int Z_\mu^{(i)a} Z_\mu^{(j)a} d^4x . \quad (5.8)$$

We put a factor $\frac{1}{g^2}$ in front of the usual L^2 inner product because the metric U^{ij} will be used to construct a measure $(\det U^{ij})^{1/2}$ for the zero modes, and this measure is also needed if one considers the quantum mechanics of zero modes $\gamma_i(t)$. The action for these time-dependent $\gamma_i(t)$ is $U^{ij} \dot{\gamma}_i \dot{\gamma}_j$ with the same prefactor $\frac{1}{g^2}$ as in the Yang-Mills gauge action.

We now evaluate this matrix for the anti-instanton. For the four translational zero modes, one can easily keep the zero mode in the background gauge by choosing $\Lambda^i = A_\nu^{\text{cl}}$. Indeed,

$$Z_\mu^{(\nu)} = \frac{\partial A_\mu^{\text{cl}}}{\partial x_\nu^\rho} + D_\mu A_\nu^{\text{cl}} = -\partial_\nu A_\mu^{\text{cl}} + D_\mu A_\nu^{\text{cl}} = F_{\mu\nu}^{\text{cl}} , \quad (5.9)$$

which satisfies the background gauge condition. The norms of these zero modes are

$$U^{\mu\nu} = \frac{8\pi^2 |k|}{g^2} \delta^{\mu\nu} = S_{\text{cl}} \delta^{\mu\nu} . \quad (5.10)$$

As indicated, this result actually holds for any k , and arbitrary gauge group.

Next we consider the dilatational zero mode corresponding to ρ and limit ourselves to $k = -1$. Taking the derivative with respect to ρ leaves the zero mode in the background gauge, so we can set $\Lambda^\rho = 0$. In the singular gauge of (3.38) we have

$$Z_\mu^{(\rho)} = -2 \frac{\rho \bar{\sigma}_{\mu\nu} x_\nu}{(x^2 + \rho^2)^2} . \quad (5.11)$$

(To show that (5.6) is satisfied, note that $(\partial/\partial x^\mu) Z_\mu^{(\rho)} = 0$ since $\bar{\sigma}_{\mu\nu}$ is antisymmetric, while $[A_\mu^{\text{cl}}, Z_\mu^{(\rho)}] = 0$ since both involve $\bar{\sigma}_{\mu\nu} x^\nu$). Using (B.18) and (B.21), one easily computes that

$$U^{\rho\rho} = \frac{16\pi^2}{g^2} = 2S_{\text{cl}} . \quad (5.12)$$

¹⁷This is also easy to prove by direct calculation: $(D^2 \delta_{\mu\nu} - D_\nu D_\mu + F_{\mu\nu}) D_\nu \Lambda$ is equal to $D_\nu [D_\nu, D_\mu] \Lambda + F_{\mu\nu} D_\nu \Lambda$, and this vanishes since $[D_\nu, D_\mu] = F_{\nu\mu}$ and $D_\nu F_{\nu\mu} = 0$. More generally, $\delta S^{\text{cl}} / \delta A_\nu^{\text{cl}} \sim D_\mu^{\text{cl}} F_{\mu\nu}^{\text{cl}}$ is gauge-covariant, hence $D_\mu^{\text{cl}} F_{\mu\nu}^{\text{cl}} (A_\rho + D_\rho \Lambda) - D_\mu^{\text{cl}} F_{\mu\nu}^{\text{cl}} (A_\rho) = [D_\mu^{\text{cl}} F_{\mu\nu}^{\text{cl}}, \Lambda]$ which vanishes on-shell (field equations transform into field equations). Hence $\frac{\delta^2 S^{\text{cl}}}{\delta A_\mu^{\text{cl}} \delta A_\rho^{\text{cl}}} (D_\rho \Lambda) = M_{\mu\nu}^{(1)} D_\nu \Lambda$ vanishes.

In regular gauge one finds $Z_\mu^{(\rho)} = \frac{2\rho\sigma_{\mu\nu}x^\nu}{(x^2+\rho^2)^2}$ which has clearly the same norm. This result can also be derived from $\frac{\partial}{\partial\rho}A_\mu^{\text{reg}}(k=-1) = \frac{\partial}{\partial\rho}U^{-1}(\partial_\mu + A_\mu^{\text{sing}}(k=-1))U$ and the identity $U^{-1}\bar{\sigma}_{\mu\nu}x^\nu U = -\sigma_{\mu\nu}x^\nu$.

The gauge-orientation zero modes can be obtained from (2.18). By expanding $U(\theta) = \exp(\theta^a T_a)$ infinitesimally in (2.18) we get to lowest order in θ (the case of general θ will be discussed shortly)

$$\frac{\partial A_\mu}{\partial\theta^a} = [A_\mu, T_a] , \quad (5.13)$$

which is not in the background gauge (the matrices T_a are in the fundamental representation). To satisfy (5.6) we have to add appropriate gauge transformations, which differ for different generators of $SU(N)$. First, for the $SU(2)$ subgroup corresponding to the instanton embedding, we add, for the singular gauge,

$$\Lambda_a = -\frac{\rho^2}{x^2 + \rho^2} T_a , \quad (5.14)$$

and find that

$$Z_{\mu(a)} = D_\mu \left[\frac{x^2}{x^2 + \rho^2} T_a \right] . \quad (5.15)$$

(using $\partial_\mu \frac{x^2}{x^2 + \rho^2} = -\partial_\mu \frac{\rho^2}{x^2 + \rho^2}$). One can now show, using (B.5), that the zero mode (5.15) is in the background gauge, and its norm reads¹⁸

$$U_{ab} = \frac{4\pi^2}{g^2} \rho^2 \delta_{ab} = \frac{1}{2} \delta_{ab} \rho^2 S_{\text{cl}} . \quad (5.16)$$

¹⁸A few details may be helpful. One finds for this zero mode in the singular gauge, using (3.38) and (B.15),

$$Z_\mu^{(a)} = 2x_\mu \rho^2 (x^2 + \rho^2)^{-2} T_a + 2\eta_{b\mu\nu} \epsilon_{bac} T_c x_\nu \rho^2 / (x^2 + \rho^2)^2 .$$

It is covariantly transversal: ∂_μ acting on the first term plus the commutator of A_μ^{sing} with the second term vanishes upon using (B.5). (The commutator of the first term with A_μ^{sing} is proportional to $(\bar{\sigma}_{\mu\nu}x^\nu)x^\mu$ and vanishes). The norm is due to integrating the sum of the square of the first term and the second term, using (B.21) with $n=1$ and $m=4$. All terms which contribute to $Z_\mu^{(a)}$, namely $\frac{\partial}{\partial\theta^a}A_\mu^{\text{sing}}$ and $\partial_\mu\Lambda^{a,\text{sing}}$ and $[A_\mu^{\text{sing}}, \Lambda^{a,\text{sing}}]$ fall off as $1/r^3$ for large $|x|$, and $Z_\mu^{(a)}$ itself is nonsingular at $x=0$.

In regular gauge one finds from (3.40)

$$\partial_\gamma A_\mu^{\text{reg}} = \partial_\gamma U^{-1}(\partial_\mu + A_\mu^{\text{sing}})U = U^{-1}\partial_\gamma A_\mu^{\text{sing}}U ,$$

and the transversality condition becomes

$$\begin{aligned} & U^{-1}D_\mu(A^{\text{sing}})U[U^{-1}\partial_\gamma A_\mu^{\text{sing}}U + U^{-1}D_\mu(A^{\text{sing}})UU^{-1}\Lambda^{a,\text{sing}}U] \\ & = D_\mu(A^{\text{reg}})[\partial_\gamma A_\mu^{\text{reg}} + D_\mu(A_\mu^{\text{reg}})U^{-1}\Lambda^{a,\text{sing}}U] = 0 . \end{aligned}$$

Hence, $\Lambda^{a,\text{reg}} = U^{-1}\Lambda^{a,\text{sing}}U$, and now all contributions to $Z_\mu^{(a)}$ in the regular gauge fall only off as $1/r$. Only their sum $Z_\mu^{(a),\text{reg}}$ falls off as $1/r^3$, just as $Z_\mu^{(a),\text{sing}}$. It is clearly simpler to work in the singular gauge, because then all integrals separately converge.

We need the gauge-orientation zero modes for arbitrary values of θ because this is needed for the group (Haar) measure. They are obtained as follows. By differentiating $U(\theta)$ and using that $U^{-1} \frac{\partial}{\partial \theta^\alpha} U$ is equal to $e_\alpha^a(\theta) T_a$, where the function $e_\alpha^a(\theta)$ is called the group vielbein (with α a curved and a a flat index according to the usual terminology¹⁹), one obtains

$$\frac{\partial}{\partial \theta^\alpha} A_\mu(\theta) = [A_\mu(\theta), e_\alpha^a(\theta) T_a] \quad (5.17)$$

For $\Lambda_{(\alpha)}$ we take now $\Lambda_{(\alpha)}(\theta) = -\frac{\rho^2}{x^2 + \rho^2} e_\alpha^a(\theta) T_a$, and then we obtain for the gauge zero modes at arbitrary θ

$$\begin{aligned} Z_{\mu(\alpha)}(\theta) &= D_\mu(A(\theta)) \left(\frac{x^2}{x^2 + \rho^2} e_\alpha^a(\theta) T_a \right) \\ &= U^{-1} \left[D_\mu(A(\theta = 0)) \left(\frac{x^2}{x^2 + \rho^2} \partial_\alpha U U^{-1} \right) \right] U. \end{aligned} \quad (5.18)$$

We define²⁰ $\partial_\alpha U U^{-1} = f_\alpha^a(\theta) T_a$. Note that $\text{tr} \partial_\alpha U U^{-1} \partial_\beta U U^{-1} = \text{tr} (U^{-1} \partial_\alpha U U^{-1} \partial_\beta U) = e_\alpha^a e_\beta^b \text{tr} T_a T_b = f_\alpha^a f_\beta^b \text{tr} T_a T_b$. Hence the left-invariant metric $e_\alpha^a e_\beta^b \delta_{ab}$ is equal to the right-invariant metric. There is a geometrical interpretation of these results [53, 54].

There are only two differences with the $\theta = 0$ case

- (i) the factors $U(\theta)$ and $U^{-1}(\theta)$ in front and at the back; these drop out in the trace
- (ii) the factors of f_α^a multiplying T_a . Taking the trace one obtains the group metric

$$U^{\alpha\beta}(\theta) = \langle Z_\mu^{(\alpha)} | Z_\mu^{(\beta)} \rangle = e_\alpha^a(\theta) e_\beta^b(\theta) U_{ab}(\theta = 0) = e_\alpha^a(\theta) e_\beta^a(\theta) (\tfrac{1}{2} \rho^2 S_{\text{cl}}). \quad (5.19)$$

Hence, in the square root of the determinant of U one finds a factor $\det e_\alpha^a$ (because $\det(e_\alpha^a \delta_{ab} e_\beta^b) = (\det e_\alpha^a)^2$), and this yields the Haar measure

$$\mu(\theta) = \det e_\alpha^a(\theta) d^3 \theta. \quad (5.20)$$

Using this measure one can calculate the group volume V of $SU(2)$, $V = \int (\det e_\alpha^a) d^3 \theta$, which is independent of the choice of coordinates θ . (We chose the parametrization $U(\theta) = \exp \theta^a T_a$, but any other parametrization yields the same result.)

¹⁹The group vielbein is given by

$$e_\alpha^a(\theta) T_a = T_\alpha + \frac{1}{2!} [T_\alpha, \theta \cdot T] + \frac{1}{3!} [[T_\alpha, \theta \cdot T], \theta \cdot T] + \dots,$$

whereas the adjoint matrix representation $M^{\text{adj}}(\theta)$ is given by

$$e^{-\theta \cdot T} T_a e^{\theta \cdot T} = M^{\text{adj}}(\theta)_a^b T_b = T_a + [T_a, \theta \cdot T] + \dots$$

One has $M^{\text{adj}}(\theta)_a^b = (\exp \theta^c f_{\cdot c}^{\cdot})_a^b$. There is a relation between the group vielbein and the adjoint matrix: $(\theta^\beta \frac{\partial}{\partial \theta^\beta} + 1) e_\alpha^b(\theta) = (M^{\text{adj}}(\theta))_\alpha^b$

²⁰The functions $e_\alpha^a(\theta)$ are sometimes called the left-invariant one-forms, while $f_\alpha^a(\theta)$ are the right-invariant one-forms.

We have now calculated all norms. It is fairly easy to prove that there is no mixing between the different modes, for example $U^{\mu(\rho)} = U^\mu_a = U^{(\rho)}_a = 0$. Thus the matrix U^{ij} for $SU(2)$ is eight by eight, with non-vanishing entries along the block-diagonal

$$U^{ij} = \begin{pmatrix} \delta^{\mu\nu} S_{\text{cl}} & & \\ & 2S_{\text{cl}} & \\ & & \frac{1}{2}g_{\alpha\beta}(\theta)\rho^2 S_{\text{cl}} \end{pmatrix}_{8 \times 8}, \quad (5.21)$$

The square root of the determinant is

$$\sqrt{U} = \frac{1}{2}S_{\text{cl}}^4 \rho^3 \sqrt{\det g_{\alpha\beta}(\theta)} = \frac{2^{11}\pi^8 \rho^3}{g^8} \sqrt{\det g_{\alpha\beta}(\theta)} \quad (\text{for } SU(2)). \quad (5.22)$$

Let us now consider the remaining generators of $SU(N)$ by first analyzing the example of $SU(3)$. For simplicity, we restrict ourselves again to lowest order in θ^a . There are seven gauge orientation zero modes, three of which are given by (5.15) by taking for T_a the first three Gell-Mann matrices $\lambda_1, \lambda_2, \lambda_3$ multiplied by $-\frac{i}{2}$. For the other four zero modes, corresponding to $\lambda_4, \dots, \lambda_7$, the formula (5.13) still holds, but we have to change the gauge transformation in order to keep the zero mode in background gauge,

$$\Lambda_k = \left[\sqrt{\frac{x^2}{x^2 + \rho^2}} - 1 \right] T_k, \quad k = 4, 5, 6, 7, \quad (5.23)$$

with $T_k = (-i/2)\lambda_k$. The difference in x -dependence of the gauge transformations (5.14) and (5.23) is due to the change in commutation relations. Namely, $\sum_{a=1}^3 [\lambda_a, [\lambda_a, \lambda_\beta]] = -(3/4)\lambda_\beta$ for $\beta = 4, 5, 6, 7$, whereas it is $-2\lambda_\beta$ for $\beta = 1, 2, 3$. (These are the values of the Casimir operator of $SU(2)$ on doublets and triplets, respectively). As argued before, there is no gauge orientation zero mode associated with λ_8 , since it commutes with the $SU(2)$ embedding. The zero modes are then

$$Z_{\mu(k)} = D_\mu \left[\sqrt{\frac{x^2}{x^2 + \rho^2}} T_k \right], \quad k = 4, 5, 6, 7, \quad (5.24)$$

with norms²¹

$$U_{kl} = \frac{1}{4}\delta_{kl}\rho^2 S_{\text{cl}}, \quad (5.25)$$

and are orthogonal to (5.15), such that $U_{ka} = 0$. This construction easily generalizes to $SU(N)$. One first chooses an $SU(2)$ embedding, and this singles out 3 generators. The

²¹These zero modes are given by $Z_{\mu(k)} = \rho^2 x^\nu / (\sqrt{x^2}(x^2 + \rho^2)^{3/2}) (\delta_{\mu\nu} T_k + 2\eta_{a\mu\nu} [T_a, T_k])$ in the singular gauge, see (3.38). For the first three zero modes we found instead $Z_{\mu(a)} = \rho^2 x^\nu / (x^2 + \rho^2)^2 (2\delta_{\mu\nu} T_a + 2\eta_{b\mu\nu} [T_b, T_a])$ with $[T_b, T_a] = \epsilon_{bac} T_c$. The norm of (5.24) is proportional to $\text{tr } T_k T_l + 4\text{tr } [T_a, T_k][T_a, T_l] = 4\text{tr } T_k T_l$, where we used (B.5).

other generators can then be split into $2(N-2)$ doublets under this $SU(2)$ and the rest are singlets. There are no zero modes associated with the singlets, since they commute with the $SU(2)$ chosen. For the doublets, each associated zero mode has the form as in (5.24), with the same norm $\frac{1}{4}\rho^2 S_{\text{cl}}$. This counting indeed leads to $4N-5$ gauge orientation zero modes. Straightforward calculation for the square-root of the complete determinant then yields an extra factor $(\frac{1}{4}\rho^2 S_{\text{cl}})^{2(N-2)}$, and so

$$\sqrt{U} = \frac{2^{2N+7}}{\rho^5} \left(\frac{\pi\rho}{g} \right)^{4N} \quad (\text{for } SU(N)) . \quad (5.26)$$

This result is a factor 2^{4N-5} smaller than [52, 1], since we chose $U(\theta) = \exp \theta^a T_a$ instead of $\exp(2\theta^a T_a)$. This ends the discussion about the (bosonic) zero mode normalization.

5.2 Construction of the fermionic zero modes

In this subsection we will explicitly construct the fermionic zero modes (normalizable solutions of the Dirac equation) in the background of a single anti-instanton. For an $SU(2)$ adjoint fermion, there are 4 zero modes according to (4.37), and these can be written as follows [63]

$$\lambda^\alpha = -\frac{1}{2}\sigma_{\rho\sigma}{}^\alpha{}_\beta \left(\xi^\beta - \sigma_\nu^{\beta\gamma'} \bar{\eta}_{\gamma'}(x-x_0)^\nu \right) F_{\rho\sigma} . \quad (5.27)$$

The $SU(2)$ indices u and v are carried by $(\lambda^\alpha)^u{}_v$ and $(F_{\rho\sigma})^u{}_v$.

To prove that these spinors are solution of the Dirac equation, use $\bar{\sigma}_\mu \sigma_{\rho\sigma} = \delta_{\mu\rho} \bar{\sigma}_\sigma - \delta_{\mu\sigma} \bar{\sigma}_\rho - \epsilon_{\mu\rho\sigma\tau} \bar{\sigma}_\tau$. Then $\bar{\not{D}}\lambda$ vanishes since $D_\mu F_{\rho\sigma}$ vanishes when contracted with $\eta_{\mu\rho}, \eta_{\mu\sigma}$ or $\epsilon_{\mu\rho\sigma\tau}$. Actually, this expression also solves the Dirac equation for higher order k , but there are then additional solutions, $4|k|$ in total for $SU(2)$, see (4.37). The four fermionic collective coordinates are denoted by ξ^α and $\bar{\eta}_{\gamma'}$, where $\alpha, \gamma' = 1, 2$ are spinor indices in Euclidean space²². They are the fermionic partners of the translational and dilatational collective coordinates in the bosonic sector. These solutions take the same form in any gauge, one just takes the corresponding gauge for the field strength. The canonical dimension of ξ and $\bar{\eta}$ is $-1/2$ and $1/2$, respectively.

For $SU(N)$ (and always $k = -1$) there are a further set of $2 \times (N-2)$ zero modes in the adjoint representation, and their explicit form depends on the gauge chosen. In regular gauge, with color indices $u, v = 1, \dots, N$ explicitly written, the gauge field is given by (3.41) (setting $x_0 = 0$, otherwise replace $x \rightarrow x - x_0$)

$$A_\mu{}^u{}_v = A_\mu^a (T_a)^u{}_v = -\frac{\sigma_{\mu\nu}{}^u{}_v x_\nu}{x^2 + \rho^2} , \quad \sigma_{\mu\nu}{}^u{}_v = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{\mu\nu}{}^\alpha{}_\beta \end{pmatrix} . \quad (5.28)$$

²²To check that the expression with $\bar{\eta}$ is a solution, one may use that $\bar{\sigma}_\rho \sigma_{\mu\nu} \sigma_\rho = 0$. Note that one may change the value of x_0 in (5.27) while keeping $F_{\mu\nu}$ fixed, because the difference is a solution with ξ^β .

Then the corresponding fermionic instanton in the adjoint representation reads

$$\lambda^{\alpha u}{}_v = \frac{\rho}{\sqrt{(x^2 + \rho^2)^3}} (\mu^u \delta^{\alpha}{}_v + \epsilon^{\alpha u} \bar{\mu}_v) . \quad (5.29)$$

Here we have introduced Grassmann collective coordinates

$$\mu^u = (\mu^1, \dots, \mu^{N-2}, 0, 0) ; \epsilon^{\alpha u} = \begin{pmatrix} 0, \dots, 0, \\ \epsilon^{\alpha \beta'} \\ 0, \dots, 0, \end{pmatrix} \text{ with } N - 2 + \beta' = u , \quad (5.30)$$

and similarly for $\bar{\mu}_v$ and $\delta^{\alpha}{}_v$. Thus the $SU(N)$ structure for the fermionic instanton is as follows

$$\lambda \propto \begin{pmatrix} 0 & \mu \\ \bar{\mu} & \xi, \bar{\eta} \end{pmatrix} . \quad (5.31)$$

The canonical dimension of μ and $\bar{\mu}$ is $-1/2$. To prove that $(\lambda^{\alpha})^u{}_v$ in (5.29) satisfies the Dirac equation $\bar{\sigma}^{\mu}(\partial_{\mu}\lambda + [A_{\mu}, \lambda]) = 0$, note that the terms $(A_{\mu})^u{}_w \mu^w$ and $\bar{\mu}_w (A_{\mu})^w{}_v$ vanish due to the index structure of A_{μ} and $\mu, \bar{\mu}$. Because A_{μ} has only nonzero entries in the lower right block, there cannot be fermionic instantons in the upper left block.

In singular gauge, the gauge field is given by (3.38)

$$A_{\mu u}{}^v = -\frac{\rho^2}{x^2(x^2 + \rho^2)} \bar{\sigma}_{\mu\nu}{}^v{}_u x_{\nu} . \quad (5.32)$$

Notice that the position of the color indices is different from that in regular gauge. This is due to the natural position of indices on the sigma matrices²³. The fermionic anti-instanton in singular gauge reads [55]

$$\lambda^{\alpha u}{}_v = \frac{\rho}{\sqrt{x^2(x^2 + \rho^2)^3}} (\mu_u x^{\alpha v} + x^{\alpha}{}_u \bar{\mu}^v) , \quad (5.33)$$

where for fixed α , the N -component vectors μ_u and $x^{\alpha v}$ are given by

$$\mu_u = (\mu_1, \dots, \mu_{N-2}, 0, 0) , \quad x^{\alpha v} = (0, \dots, 0, x^{\mu} \sigma_{\mu}^{\alpha \beta'}) \quad \text{with } N - 2 + \beta' = v . \quad (5.34)$$

Further, $x^{\alpha}{}_u = x^{\alpha v} \epsilon_{vu}$ and $\bar{\mu}^v$ also has $N - 2$ nonvanishing components. The particular choice of zeros in the last two entries corresponds to the choice of embedding the $SU(2)$ instanton in the lower-right block of $SU(N)$. Notice that the adjoint field λ is indeed traceless in its color indices. This follows from the observation that μ and $\bar{\mu}$ only appear at

²³To be very precise, we could have used different Pauli matrices $(\tau^a)^u{}_v$ for the internal $SU(2)$ generators. Then we could have defined a matrix $(\bar{\sigma}_{\mu\nu})^u{}_v$ by $\bar{\sigma}_{\mu\nu} = i\eta_{a\mu\nu} \tau^a$, and the $SU(2)$ indices in (5.32) and (5.33) would have appeared in the same position as in (5.29). It is simpler to work with only one kind of Pauli matrices.

the off-diagonal blocks inside $SU(N)$. In general μ and $\bar{\mu}$ are independent, but if there is a reality condition on λ in Euclidean space, the μ and $\bar{\mu}$ are related by complex conjugation. We will discuss this in a concrete example when we discuss instantons in $\mathcal{N} = 4$ super Yang-Mills theory. We should also mention that while the bosonic collective coordinates are related to the rigid symmetries of the theory, this is not obviously true for the fermionic collective coordinates, although, as we will see later, the ξ and $\bar{\eta}$ collective coordinates can be obtained from ordinary supersymmetry and conformal supersymmetry in super Yang-Mills theories.

A similar construction holds for a fermion in the fundamental representation. Now there is only one fermionic collective coordinate, see (4.37), which we denote by $\mathcal{K}$. The explicit expression for $k = -1$ in singular gauge is²⁴

$$(\lambda^\alpha)_u = \frac{\rho}{\sqrt{x^2(x^2 + \rho^2)^3}} x^\alpha{}_u \mathcal{K} . \quad (5.35)$$

In regular gauge it is given by

$$(\lambda^\alpha)^u = \frac{\epsilon^{\alpha u}}{(x^2 + \rho^2)^{3/2}} \mathcal{K} . \quad (5.36)$$

The Dirac equation for $(\lambda^\alpha)^u$ is proportional to

$$-3x^\mu \bar{\sigma}_{\mu, \alpha' \beta} \epsilon^{\beta u} - \bar{\sigma}_{\mu, \alpha' \beta} \epsilon^{\beta v} (\sigma_{\mu\nu})^u{}_v x^\nu \quad (5.37)$$

and to show that this vanishes one may use $(\sigma_{\mu\nu})^u{}_v \epsilon^{\beta v} = (\sigma_{\mu\nu})^{u\beta}$ and the symmetry of the Lorentz generators $(\sigma_{\mu\nu})^{u\beta} = (\sigma_{\mu\nu})^{\beta u}$ and $\bar{\sigma}_\mu \sigma_{\mu\nu} = 3\bar{\sigma}_\nu$.

6 The measure for zero modes

Having determined the bosonic and fermionic zero modes for $k = \pm 1$ instantons with $SU(N)$ gauge group, we now discuss the measure for the zero mode sector of path integrals. The one-loop corrections due to the nonzero modes, will be discussed in the next section.

6.1 The measure for the bosonic collective coordinates

We now construct the measure on the moduli space of bosonic collective coordinates, and show that the matrix U plays the role of a Jacobian. We first illustrate the idea for a

²⁴The color index should again be written as $(\lambda^\alpha)_{u'}$ because $\lambda^{\text{reg}, u} = (U)^{uv'} \lambda_{v'}^{\text{sing}}$ with $U^{uv'} = \sigma_\mu^{uv'} x_\mu / \sqrt{x^2}$. However, we drop these primes. The proof that (5.35) satisfies the Dirac equation uses $\bar{\sigma}_{\mu\alpha'\beta} \sigma_{\rho u}^\beta x^\mu x^\rho = \epsilon_{\alpha' u} x^2$ and $(\bar{\sigma}_{\mu\rho})_{\alpha' v} (\bar{\sigma}_{\mu\nu})_u{}^v = -(\bar{\sigma}_{\mu\rho})_{\alpha' v} (\bar{\sigma}_{\mu\nu})_{vu} = 3\delta_{\rho\nu} \epsilon_{\alpha' u}$.

generic system without gauge invariance, with fields ϕ^A , and action $S[\phi]$ (for example, the kink in one dimension). We expand around the instanton solution

$$\phi^A(x) = \phi_{\text{cl}}^A(x, \gamma) + \phi_{\text{qu}}^A(x, \gamma) . \quad (6.1)$$

The collective coordinates are denoted by γ and, for notational simplicity, we assume there is only one. At this point the fields ϕ_{qu}^A can still depend on the collective coordinate, as they can include zero modes. The action, up to terms quadratic in the quantum fields, is

$$S = S_{\text{cl}} + \frac{1}{2} \phi_{\text{qu}}^A M_{AB}(\phi_{\text{cl}}) \phi_{\text{qu}}^B . \quad (6.2)$$

The operator M has zero modes given by

$$Z^A = \frac{\partial \phi_{\text{cl}}^A}{\partial \gamma} , \quad (6.3)$$

since, as we explained in (5.7), $M_{AB}Z^B$ is just the derivative of the field equation $\partial S_{\text{cl}}/\partial \phi_{\text{cl}}^A$ with respect to the collective coordinate. More generally, if the operator M is hermitian (or rather self-adjoint²⁵), it has a complete set of eigenfunctions F_α with eigenvalues ϵ_α ,

$$M_{AB}F_\alpha^B = \epsilon_\alpha F_\alpha^A . \quad (6.4)$$

One of the solutions is of course the zero mode $Z = F_0$ with $\epsilon_0 = 0$. Any function can be expanded into a basis of eigenfunctions, in particular the quantum fields,

$$\phi_{\text{qu}}^A = \sum_\alpha \xi_\alpha F_\alpha^A , \quad (6.5)$$

with coefficients ξ_α . The eigenfunctions have norms, determined by their inner product

$$\langle F_\alpha | F_\beta \rangle = \int d^4x F_\alpha^A(x) F_\beta^A(x) . \quad (6.6)$$

The eigenfunctions can always be chosen orthogonal, such that $\langle F_\alpha | F_\beta \rangle = \delta_{\alpha\beta} u_\alpha$. The action then becomes

$$S = S_{\text{cl}} + \frac{1}{2} \sum_\alpha \xi_\alpha \xi_\alpha \epsilon_\alpha u_\alpha . \quad (6.7)$$

If there is a coupling constant in front of the action (6.2), we rescale the inner product with the coupling, such that (6.7) still holds. This was done in (5.8). The path-integral measure is now *defined* as

$$[d\phi] \equiv \prod_{\alpha=0}^{\infty} \sqrt{\frac{u_\alpha}{2\pi}} d\xi_\alpha . \quad (6.8)$$

²⁵More precisely, if there is an inner product $(\phi_1, \phi_2) = \int \phi_1^A H_{AB} \phi_2^B d^4x$ with real ϕ_1, ϕ_2 and with metric H_{AB} , and $H_{AB}H^{BC} = \delta_A^C$, then one may define $\phi^A H_{AB} = \phi_B$ so that $(\phi_1, \phi_2) = \int \phi_1 A \phi_2^A d^4x$. If one further defines $H^{BC} M_{CD} = M^B_D$, then M^A_B is hermitian if $(\phi_1, M\phi_2) = (M\phi_1, \phi_2)$. The need for a matrix to define an inner product is familiar from spinors, but for bosons the metric is in general trivial ($H_{AB} = \delta_{AB}$).

We perform the Gaussian integration over the ξ_α and get

$$\int [d\phi] e^{-S[\phi]} = \int \sqrt{\frac{u_0}{2\pi}} d\xi_0 e^{-S_{\text{cl}}} (\det' M)^{-1/2} . \quad (6.9)$$

One sees that if there were no zero modes, the measure in (6.8) produces the correct result with the determinant of M . In the case of zero modes, the determinant of M is zero, and the path integral would be ill-defined. Instead, we must leave out the zero mode in M , take the amputated determinant (denoted by $\det'$), and integrate over the mode ξ_0 . By slightly changing some parameters in the action (for example by adding a small mass term) the zero mode turns into a non-zero mode, and then one needs $\sqrt{\frac{u_0}{2\pi}} d\xi_0$ as measure. So, continuity fixes the measure for the zero modes as in (6.8).

The next step is to convert the ξ_0 integral to an integral over the collective coordinate γ [56]. This can be done by inserting unity into the path integral. Consider the identity

$$1 = \int d\gamma \delta(f(\gamma)) \frac{\partial f}{\partial \gamma} , \quad (6.10)$$

which holds for any (invertible) function $f(\gamma)$. Taking $f(\gamma) = -\langle \phi - \phi_{\text{cl}}(\gamma) | Z \rangle$, and recalling that the original field ϕ is independent of γ , we get

$$1 = \int d\gamma \left(u_0 - \left\langle \phi_{\text{qu}} \left| \frac{\partial Z}{\partial \gamma} \right\rangle \right) \delta(\langle \phi_{\text{qu}} | Z \rangle) = \int d\gamma \left(u_0 - \left\langle \phi_{\text{qu}} \left| \frac{\partial Z}{\partial \gamma} \right\rangle \right) \delta(\xi_0 u_0) . \quad (6.11)$$

This trick is similar to the Faddeev-Popov trick for gauge fixing. In the semiclassical approximation, the term $\langle \phi_{\text{qu}} | \frac{\partial Z}{\partial \gamma} \rangle$ is subleading and we will neglect it²⁶. The integration over ξ_0 is now trivial and one obtains

$$\int [d\phi] e^{-S} = \int d\gamma \sqrt{\frac{u_0}{2\pi}} e^{-S_{\text{cl}}} (\det' M)^{-1/2} . \quad (6.12)$$

For a system with more zero modes Z^i with norms-squared U^{ij} , the result is²⁷

$$\int [d\phi] e^{-S} = \int \prod_{i=1} \frac{d\gamma_i}{\sqrt{2\pi}} (\det U)^{1/2} e^{-S_{\text{cl}}} (\det' M)^{-1/2} . \quad (6.13)$$

²⁶It will contribute however to a two-loop contribution. To see this, one first writes this term in the exponential, where it enters without $\hbar$, so it is at least a one-loop effect. Then ϕ_{qu} has a part proportional to the zero mode, which drops out by means of the delta function insertion. The other part of ϕ_{qu} is genuinely quantum and contains a power of $\hbar$ (which we have suppressed). Therefore, it contributes at two loops [57] (see also [58] for related matters).

²⁷One obtains from (6.11) $\det \langle \partial_{\gamma_i} A_\mu^{\text{cl}} | Z^{(j)} \rangle$ times $(\det U^{ab})^{-1/2}$. The matrix elements $\langle \partial_{\gamma_i} A_\mu^{\text{cl}} | Z^{(j)} \rangle$ are equal to $\langle Z^{(i)} | Z^{(j)} \rangle = U^{ij}$ minus $\langle D_\mu \Lambda^{(a)} | Z_\mu^{(b)} \rangle$. The latter term can be partially integrated, and vanishes since there are no boundary contributions, neither in the singular nor in the regular gauge. (For the regular gauge one needs an explicit calculation to check this statement.)

Notice that this result is invariant under rescalings of Z , which can be seen as rescalings of the collective coordinates. More generally, the matrix U_{ij} can be interpreted as a metric on the moduli space of collective coordinates. The measure is then invariant under general coordinate transformations on the moduli space.

One can repeat the analysis for gauge theories to show that (6.13) also holds for Yang-Mills instantons in singular gauge. For regular gauges, there are some complications due to the fact that neither of the two terms in (5.5) does fall off fast at infinity, but only their sum is convergent. In singular gauge, each term separately falls off fast at infinity. For this reason, it is more convenient to work in singular gauge. The measure for the bosonic collective coordinates for $k = 1$ $SU(N)$ YM theories, without the determinant from integrating out the quantum fluctuations which will be analyzed in the next section, becomes

$$\frac{2^{4N+2}\pi^{4N-2}}{(N-1)!(N-2)!g^{4N}} \int d^4x_0 \frac{d\rho}{\rho^5} \rho^{4N} . \quad (6.14)$$

This formula contains the square-root of the determinant of U in (5.26), $4N$ factors of $1/\sqrt{2\pi}$, and we have also integrated out the gauge orientation zero modes. This may be done only if we are evaluating gauge invariant correlation functions. The result of this integration follows from the volume of the coset space

$$\text{Vol} \left\{ \frac{SU(N)}{SU(N-2) \times U(1)} \right\} = \frac{2^{4N-5}\pi^{2N-2}}{(N-1)!(N-2)!} , \quad (6.15)$$

which is a factor 2^{4N-5} larger than in [52, 1], because we have used the normalization $\text{tr}(T_a T_b) = -\frac{1}{2}\delta_{ab}$, while in [52, 1] $\text{tr}(T_a T_b) = -2\delta_{ab}$ was used. We found in (5.26) another factor $2^{-(4N-5)}$, and indeed the result for the total measure in (6.14) is the same as in [52, 1]. The derivation of this formula can be found in Appendix C, which is a detailed version of [52].

6.2 The measure for the fermionic collective coordinates

We must also construct the measure on the moduli space of fermionic collective coordinates. Consider (5.27). The fermionic zero modes are linear in the Grassmann parameters ξ^α and $\bar{\eta}_{\alpha'}$. Thus these ξ^α and $\bar{\eta}_{\alpha'}$ correspond to the coefficients ξ^α in (6.5). One obtains the zero modes by differentiating λ^α in (5.27) w.r.t. ξ^α and $\bar{\eta}_{\alpha'}$, and for this reason one often calls these ξ^α and $\bar{\eta}_{\alpha'}$ the fermionic collective coordinates. This is not quite correct, because collective coordinates appear in the classical solution (the instanton) but we shall use this terminology nevertheless because it is common practice. We use again the measure in (6.8). There are in this case no factors $\frac{1}{\sqrt{2\pi}}$ because of the Grassmann integration, and instead of $(\det M')^{-1/2}$ we now obtain $(\det M')^{1/2}$ in (6.9). Because the parameters

$\xi^\alpha, \bar{\eta}_\alpha$, etc. appear linearly in the zero modes, we do not need the Faddeev-Popov trick to convert the integration over zero modes into an integration over collective coordinates. So for fermions the Grassmannian coefficients of the zero modes are at the same time the collective coordinates.

We shall discuss these issues in more detail when we come to supersymmetric gauge theories, but now we turn to computing the norms of the fermionic zero modes.

For the zero modes with ξ in (5.27), one finds

$$Z_{(\beta)}^\alpha = \frac{\partial \lambda^\alpha}{\partial \xi^\beta} = -\frac{1}{2} \sigma_{\mu\nu}{}^\alpha{}_\beta F_{\mu\nu} . \quad (6.16)$$

The norms of these two zero modes are given by

$$(U_\xi)_\beta{}^\gamma = -\frac{2}{g^2} \int d^4x \operatorname{tr} \{ Z_{\alpha(\beta)} Z^{\alpha(\gamma)} \} = 4S_{\text{cl}} \delta_\beta{}^\gamma , \quad (6.17)$$

where we have used the definition in (5.8) and contracted the spinor indices with the usual metric for spinors. This produces a term in the measure²⁸

$$\int d\xi^1 d\xi^2 (4S_{\text{cl}})^{-1} . \quad (6.18)$$

The result (6.18) actually holds for any k . We get the square root of the determinant in the denominator for fermions. One really gets the square root of the super determinant of the matrix of inner product, but because there is no mixing between bosonic and fermionic moduli, the superdeterminant factorizes into the bosonic determinant divided by the fermionic determinant.

For the $\bar{\eta}$ zero modes, we obtain, using some algebra for the σ -matrices,

$$Z^{\alpha\beta'} = \partial \lambda^\alpha / \partial \bar{\eta}_{\beta'} = \frac{1}{2} (\sigma_{\mu\nu} \sigma_\rho)^{\alpha\beta'} F_{\mu\nu} x_\rho , \quad (6.19)$$

and

$$(U_{\bar{\eta}})_{\alpha'}{}^{\beta'} = 8S_{\text{cl}} \delta_{\alpha'}{}^{\beta'} \rho^2 , \quad (6.20)$$

so that the corresponding measure is

$$\int d\bar{\eta}_1 d\bar{\eta}_2 (8\rho^2 S_{\text{cl}})^{-1} , \quad (6.21)$$

which only holds for $k = 1$.

Finally we compute the Jacobian for the fermionic “gauge orientation” zero modes. For convenience, we take the solutions in regular gauge (the Jacobian is gauge invariant anyway), and find from (5.29)

$$(Z_{(\mu^w)}^\alpha)_v{}^u = \frac{\rho}{\sqrt{(x^2 + \rho^2)^3}} \delta_v^\alpha \Delta^u{}_w , \quad (Z_{(\bar{\mu}^w)}^\alpha)_v{}^u = \frac{\rho}{\sqrt{(x^2 + \rho^2)^3}} \epsilon^{\alpha u} \Delta^w{}_v , \quad (6.22)$$

²⁸Sometimes one finds in the literature that $U_\xi = 2S_{\text{cl}}$. This is true when one uses the conventions for Grassmann integration $\int d^2\xi \xi^\alpha \xi^\beta = \frac{1}{2} \epsilon^{\alpha\beta}$. In our conventions $d^2\xi \equiv d\xi^1 d\xi^2$.

where the N by N matrix Δ is the unity matrix in the $(N-2)$ by $(N-2)$ upper diagonal block, and zero elsewhere. So Δ restricts the values of u, w and v to up to $N-2$ while in δ^α_v and $\epsilon^{\alpha v}$ the index runs over the next two values. Consequently, the norms of Z_μ and $Z_{\bar{\mu}}$ are easily seen to be zero, but the nonvanishing inner product is

$$(U_{\mu\bar{\mu}})^u_v = -\frac{2}{g^2} \int d^4x \operatorname{tr} Z_{(\bar{\mu}u)}^\alpha Z_{\alpha(\mu v)} = \frac{2\pi^2}{g^2} \Delta^u_v, \quad (6.23)$$

where we have used the integral (B.21). It also follows from the index structure that the ξ and $\bar{\eta}$ zero modes are orthogonal to the μ zero modes, so there is no mixing in the Jacobian.

Putting everything together, the fermionic part of the measure for $\mathcal{N}$ adjoint fermions coupled to $SU(N)$ YM theory, with $k=1$, is given by

$$\int \left(\prod_{A=1}^{\mathcal{N}} d^2 \xi^A \right) \left(\frac{g^2}{32\pi^2} \right)^{\mathcal{N}} \left(\prod_{A=1}^{\mathcal{N}} d^2 \bar{\eta}^A \right) \left(\frac{g^2}{64\pi^2 \rho^2} \right)^{\mathcal{N}} \prod_{A=1}^{\mathcal{N}} \left(\prod_{u=1}^{N-2} d\mu^{A,u} d\bar{\mu}_u^A \right) \left(\frac{g^2}{2\pi^2} \right)^{\mathcal{N}(N-2)}. \quad (6.24)$$

Similarly, one can include fermions in the fundamental representation, for which the Jacobian factor is

$$U_{\mathcal{K}} \equiv \int d^4x Z^\alpha_u Z_\alpha^u = \pi^2, \quad (6.25)$$

for each species. Here $\mathcal{K}$ is the Grassmann collective coordinate of (5.36). Hence in this case the fermionic part of the measure is

$$\int \left(\prod_{A=1}^{N_f} d\mathcal{K}^A \right) \left(\sqrt{\frac{1}{\pi^2}} \right)^{N_f} \quad (6.26)$$

for N_f fundamental Weyl spinors coupled to $SU(N)$ YM theory with $k=1$.

Note that we did not put a factor $\frac{1}{g^2}$ in front of the integral in (6.25), whereas we used such a factor for fermions in the adjoint representation. The reason we do not use such a factor for fermions in the fundamental representation has to do with the action. One finds a factor $\frac{1}{g^2}$ in front of the Yang-Mills action, and therefore also, by susy, in front of the Dirac action for gluinos. However, in the matter action the g -dependence has been absorbed by the gluons, so there is no factor $\frac{1}{g^2}$ in front of the matter fermions. The measure of the zero modes uses the metric of the collective coordinates. In soliton physics (and instantons can be considered as solitons in one higher dimension) one obtains this metric if one lets the collective coordinates become time dependent and integrates over d^4x in the action one ends up with a quantum mechanical action of the

$$\mathcal{L} = (U^{ij} \dot{\gamma}_i \dot{\gamma}_j + U_{\alpha\beta} \dot{\xi}^\alpha \dot{\xi}^\beta + U^{\alpha'\beta'} \dot{y}_{\alpha'} \dot{y}_{\beta'} + U_{AB} \dot{K}^A \dot{K}^B) \quad (6.27)$$

Since U^{ij} , $U_{\alpha\beta}$ and $U^{\alpha'\beta'}$ are produced by the Yang-Mills action and its susy partner, while U_{AB} is due to the matter action, there is no g -dependence in (6.26).

7 One loop determinants

Having determined the measure on the moduli space of collective coordinates, we now compute the determinants that arise by Gaussian integration over the quantum fluctuations. Before doing so, we extend the model by adding real scalar fields and Majorana fermions in the adjoint representation. The action is

$$S = -\frac{1}{g^2} \int d^4x \operatorname{tr} \left\{ \frac{1}{2} F_{\mu\nu} F_{\mu\nu} + (D_\mu \phi) (D_\mu \phi) - i \bar{\lambda} \not{D} \lambda - i \lambda \not{D} \bar{\lambda} \right\} . \quad (7.1)$$

Here, λ is a two-component Weyl spinor which we take in the adjoint representation²⁹. In Minkowski space there is a reality condition between the two complex 2-component spinors λ and $\bar{\lambda}$, and as a result $\bar{\lambda}_{\dot{\alpha}}$ transforms in the complex conjugate of the representation of λ_α , but in Euclidean space this reality condition is dropped. So λ^α and $\bar{\lambda}_{\alpha'}$ are independent complex variables. For the Grassmann integration this makes no difference. Written with indices the Euclidean Dirac action in (7.1) reads $\{-i\lambda_\alpha(\sigma^\mu)^{\alpha\beta'} D_\mu \bar{\lambda}_{\beta'} - i\bar{\lambda}^{\alpha'}(\bar{\sigma}^\mu)_{\alpha'\beta} D_\mu \lambda^\beta\}$ where $\lambda_\alpha = \lambda^\beta \epsilon_{\beta\alpha}$ and $\bar{\lambda}^{\alpha'} = \epsilon^{\alpha'\beta'} \bar{\lambda}_{\beta'}$. Generalization to fundamental fermions is straightforward. The anti-instanton solution around which we will expand is

$$A_\mu^{\text{cl}} , \quad \phi_{\text{cl}} = 0 , \quad \lambda_{\text{cl}} = 0 , \quad \bar{\lambda}_{\text{cl}} = 0 , \quad (7.2)$$

where A_μ^{cl} is the anti-instanton. This background represents an exact solution to the field equations. The bosonic and fermionic zero modes are taken care of by the measure for the collective coordinates, while in the orthogonal space of nonzero modes, one can define propagators and vertices, and perform perturbation theory around the (anti-) instanton.

After expanding $A_\mu = A_\mu^{\text{cl}} + A_\mu^{\text{qu}}$, and similarly for the other fields, we add gauge fixing and ghost terms

$$S_{\text{gf}} = -\frac{1}{g^2} \int d^4x \operatorname{tr} \left\{ (D_\mu^{\text{cl}} A_\mu^{\text{qu}})^2 - 2b D_{\text{cl}}^2 c \right\} , \quad (7.3)$$

such that the total gauge field action is given by (5.3). The integration over A_μ gives

$$[\det' \Delta_{\mu\nu}]^{-1/2} , \quad \Delta_{\mu\nu} = -D^2 \delta_{\mu\nu} - 2F_{\mu\nu} , \quad (7.4)$$

where the prime stands for the amputated determinant, with zero eigenvalues left out. We have suppressed the subscript ‘cl’ and Lie algebra indices. Integration over the scalar fields results in

$$[\det \Delta_\phi]^{-1/2} , \quad \Delta_\phi = -D^2 , \quad (7.5)$$

²⁹As before $\lambda = \begin{pmatrix} \lambda^\alpha \\ \bar{\lambda}_{\dot{\alpha}} \end{pmatrix}$, but the 4-component Majorana spinor $\bar{\lambda}$ is defined by $\lambda^T C$ both in Minkowski and in Euclidean space, where C is the charge conjugation matrix, $C = \begin{pmatrix} \epsilon_{\alpha\beta} & 0 \\ 0 & \epsilon^{\dot{\alpha}\dot{\beta}} \end{pmatrix}$. Then $\bar{\lambda} = (\lambda_\alpha, -\bar{\lambda}^{\dot{\alpha}})$ and Lorentz (or rather $SO(4)$) invariance is preserved in Euclidean space because the relation $C\gamma^\mu = -\gamma^{\mu,T} C$ holds in both spaces. In Euclidean space we denote the indices of $\bar{\lambda}$ by α' instead of $\dot{\alpha}$.

and the ghost system yields similarly

$$[\det \Delta_{\text{gh}}] , \quad \Delta_{\text{gh}} = -D^2 . \quad (7.6)$$

For the fermions λ and $\bar{\lambda}$, we need a bit more explanation. Since neither $\not{D}$ nor $\bar{\not{D}}$ is hermitean (even worse, $\not{D}$ maps antichiral spinors into chiral spinors), we cannot evaluate the determinants in terms of their eigenvalues. But both products

$$\Delta_- = -\not{D} \bar{\not{D}} = -D^2 - \frac{1}{2} \sigma_{\mu\nu} F_{\mu\nu} , \quad \Delta_+ = -\bar{\not{D}} \not{D} = -D^2 , \quad (7.7)$$

with spinor indices still suppressed, are hermitean. Let us label the nonzero modes by a subscript i . Then we can expand λ in terms of commuting eigenfunctions F_i of Δ_- with anticommuting coefficients ξ_i , and $\bar{\lambda}$ in terms of eigenfunctions $\bar{F}_i$ of Δ_+ with coefficients $\bar{\xi}_i$. We have seen before that both operators have the same spectrum of non-zero eigenvalues ϵ_i , and the relation between the eigenfunctions is $\bar{F}_i = \frac{1}{\sqrt{\epsilon_i}} \bar{\not{D}} F_i$ and $F_i = \frac{-1}{\sqrt{\epsilon_i}} \not{D} \bar{F}_i$. (The minus sign is needed in order that $\bar{F}_i = \frac{1}{\sqrt{\epsilon_i}} \bar{\not{D}} F_i = \frac{1}{\epsilon_i} (-\not{D} \not{D}) \bar{F}_i = \bar{F}_i$). Defining the path integral over λ and $\bar{\lambda}$ as the integration over ξ_i and $\bar{\xi}_i$, one gets the determinant over the nonzero eigenvalues³⁰. The result for the integration over the fermions can be written in symmetrized form as

$$[\det' \Delta_-]^{1/4} [\det \Delta_+]^{1/4} . \quad (7.9)$$

As stated before, since all the eigenvalues of both Δ_- and Δ_+ are the same, the determinants are formally equal. This result can also be obtained by writing the spinors in terms of Dirac fermions; the determinant we have to compute is then

$$[\det' \Delta_D^2]^{1/2} , \quad \Delta_D = \begin{pmatrix} 0 & \not{D} \\ \bar{\not{D}} & 0 \end{pmatrix} . \quad (7.10)$$

³⁰Namely, the action becomes

$$-\frac{1}{g^2} \text{tr} \int d^4x \left[-i \left(\bar{\xi}_i \bar{F}_i \bar{\not{D}} \xi_j \frac{(-\not{D} \bar{F}_j)}{\sqrt{\epsilon_j}} \right) - i \xi_j F_j \not{D} \bar{\xi}_i \frac{\bar{\not{D}} F_i}{\sqrt{\epsilon_i}} \right] = -\frac{i}{2} \bar{\xi}_i \xi_j \sqrt{\epsilon_j} \langle \bar{F}_i | \bar{F}_j^a \rangle + \frac{i}{2} \xi_j \bar{\xi}_i \langle F_j^a | F_i^a \rangle \sqrt{\epsilon_i} . \quad (7.8)$$

Next we use that the norms of $\bar{F}_i$ and F_i are equal:

$$\begin{aligned} \langle \bar{F}_i | \bar{F}_j \rangle &= \frac{1}{\sqrt{\epsilon_i}} \left\langle \bar{\not{D}} F_i \left| \frac{1}{\sqrt{\epsilon_j}} \bar{\not{D}} F_j \right. \right\rangle = \frac{1}{\sqrt{\epsilon_i \epsilon_j}} \langle F_i | -\not{D} \bar{\not{D}} F_j \rangle \\ &= \sqrt{\frac{\epsilon_j}{\epsilon_i}} \langle F_i | F_j \rangle = \frac{1}{\sqrt{\epsilon_i \epsilon_j}} \langle -\not{D} \bar{\not{D}} F_i | F_j \rangle = \sqrt{\frac{\epsilon_i}{\epsilon_j}} \langle F_i | F_j \rangle . \end{aligned}$$

Hence, as expected, the F_i and $\bar{F}_j$ for different eigenvalues are orthogonal to each other, and the norms of F_i and $\bar{F}_j$ are the same. Denoting $\frac{1}{g^2} \int d^4x (F_i^a)^* F_i^a = \langle F_i^a | F_i^a \rangle$ by u_i , one finds for the path integral

$$\int d\bar{\xi}_i d\xi_i e^{i\bar{\xi}_i \xi_i u_i \sqrt{\epsilon_i}} = i u_i \sqrt{\epsilon_i} .$$

Hence the measure is $\frac{d\xi_i}{\sqrt{u_i}} \frac{d\bar{\xi}_j}{\sqrt{u_j}}$, and the one-loop determinant is $\prod_i (\epsilon_i)^{1/2}$.

One would expect that in a supersymmetric model with vectors, spinors and scalars, the sum of all zero point energies cancel. These zero point energies correspond to the one-loop determinants in an external Yang-Mills field. So this suggests that all one-loop determinants are related, and since the one-loop determinants of fermions depend on Δ_+ and Δ_- , one would expect that the determinants for the bosons can be expressed in terms of the determinants of Δ_- and Δ_+ . For the ghosts and adjoint scalars this is obvious,

$$\det \Delta_\phi = \det \Delta_{\text{gh}} = [\det \Delta_+]^{1/2} . \quad (7.11)$$

We get $\det \Delta_\phi = \det(-D^2) = \det \Delta_+^{1/2}$ and $\det \Delta_{\text{gh}} = \det(-D^2) = \det \Delta_+^{1/2}$ because the spinor space is two-dimensional.

For the vector fields, we rewrite the operator $\Delta_{\mu\nu}$ in (7.4) in terms of the fermion operator Δ_- . Using $\text{tr}(\bar{\sigma}_\mu \sigma_\nu) = 2\delta_{\mu\nu}$ and $\text{tr}(\bar{\sigma}_\mu \sigma_{\rho\sigma} \sigma_\nu) = 2(\delta_{\mu\rho}\delta_{\sigma\nu} - \delta_{\mu\sigma}\delta_{\rho\nu} - \epsilon_{\mu\rho\sigma\nu})$ we obtain the following identity for $\Delta_{\mu\nu} = -\delta_{\mu\nu}D^2 - 2F_{\mu\nu}$,

$$\begin{aligned} \Delta_{\mu\nu} &= \frac{1}{2} \text{tr} \{ \bar{\sigma}_\mu \Delta_- \sigma_\nu \} = \frac{1}{2} \bar{\sigma}_{\mu\alpha'\beta} \left(\Delta_-^{\beta}{}_{\gamma} \right) \sigma_{\nu}^{\gamma\alpha'} \\ &= \frac{1}{2} (\bar{\sigma}_{\mu\alpha'\beta}) (\Delta_-^{\beta}{}_{\gamma} \delta^{\alpha'}{}_{\delta'}) (\bar{\sigma}_{\nu}^{\gamma\delta'}) \end{aligned} \quad (7.12)$$

where $(\Delta_-)^{\beta}{}_{\gamma} \delta^{\alpha'}{}_{\delta'}$ is block-diagonal on the basis $\beta\alpha' = \gamma\delta' = (11), (21), (12), (22)$.

$$\Delta_-^{\beta}{}_{\gamma} \delta^{\alpha'}{}_{\delta'} = \begin{pmatrix} \Delta_-^1{}_1 & \Delta_-^1{}_2 & 0 & 0 \\ \Delta_-^2{}_1 & \Delta_-^2{}_2 & 0 & 0 \\ 0 & 0 & \Delta_-^1{}_1 & \Delta_-^1{}_2 \\ 0 & 0 & \Delta_-^2{}_1 & \Delta_-^2{}_2 \end{pmatrix} \quad (7.13)$$

This proves that³¹

$$\det' \Delta_{\mu\nu} = [\det' \Delta_-]^2 . \quad (7.14)$$

Now we can put everything together. The one-loop determinant for a Yang-Mills system, including the ghosts, coupled to n real adjoint scalars and $\mathcal{N}$ Weyl spinors (or Majorana spinors) also in the adjoint representation is

$$[\det' \Delta_-]^{-1+\mathcal{N}/4} [\det \Delta_+]^{\frac{1}{4}(2+\mathcal{N}-n)} . \quad (7.15)$$

This expression simplifies to the ratio of the determinants when $\mathcal{N} - \frac{n}{2} = 1$. Particular cases are

$$\mathcal{N} = 1 \quad n = 0 \quad \rightarrow \quad \left[\frac{\det \Delta_+}{\det' \Delta_-} \right]^{3/4} ,$$

³¹Consider $\bar{\sigma}_{\mu,\alpha'\beta}$ and $\sigma_{\nu}^{\gamma\delta'}$ as 4×4 matrices. Then on the right-hand side of (7.12) one has the product of three 4×4 matrices. For fixed μ and ν one has $\bar{\sigma}_{\alpha'\beta}^{\mu} \sigma_{\nu}^{\beta\alpha'} = 2\delta_{\mu}{}^{\nu}$, hence $\det[\bar{\sigma}_{\mu,\alpha'\beta}] = 4$.

$$\begin{aligned}
\mathcal{N} = 2 \quad n = 2 &\rightarrow \left[\frac{\det \Delta_+}{\det' \Delta_-} \right]^{1/2}, \\
\mathcal{N} = 4 \quad n = 6 &\rightarrow \left[\frac{\det \Delta_+}{\det' \Delta_-} \right]^0.
\end{aligned} \tag{7.16}$$

These cases correspond to supersymmetric Yang-Mills theories with $\mathcal{N}$ -extended supersymmetry. Notice that for $\mathcal{N} = 4$, the determinants of Δ_+ and Δ_- separately cancel, so there is no one-loop contribution.

For $\mathcal{N} = 1, 2$ the determinants formally give unity since the non-zero eigenvalues are the same. However, one must first regularize the theory to define the determinants properly. After regularization, the renormalization procedure must be carried out and counterterms must be added. The counterterms are the same as in the theory without instantons and their finite as well as infinite parts must be specified by physical renormalization conditions. The ratios of products of non-zero eigenvalues can be written as the exponent of the difference of two infinite sums

$$\frac{\det \Delta_+}{\det' \Delta_-} = \exp \left(\sum_n \omega_n^{(+)} - \sum_n \omega_n^{(-)} \right), \tag{7.17}$$

with eigenvalues $\lambda_n = \exp \omega_n$. The frequencies $\omega_n^{(+)}$ and $\omega_n^{(-)}$ can be discretized by putting the system in a box of size R and imposing suitable boundary conditions on the quantum fields at R (for example, $\phi(R) = 0$, or $\frac{d}{dR}\phi(R) = 0$, or a combination thereof [4]). These boundary conditions may be different for different fields. The sums over $\omega_n^{(+)}$ and $\omega_n^{(-)}$ are divergent; their difference is still divergent (although less divergent than each sum separately) but after adding counterterms ΔS one obtains a finite answer. The problem is that one can combine the terms in both series in different ways, giving different answers. By combining $\omega_n^{(+)}$ with $\omega_n^{(-)}$ for each fixed n , one would find that the ratio $(\det \Delta_+ / \det' \Delta_-)$ equals unity. However, other values could result by using different ways to regulate these sums. We have discussed before that for susy instantons the densities of nonzero modes are equal, hence for susy instantons the contributions in (7.16) from the one-loop determinants cancel. This makes these models simpler to deal with than non-susy models. For ordinary (nonsusy) Yang-Mills theory, the results for the effective action due to different regularization schemes differ at most by a local finite counterterm. In the background field formalism we are using, this counterterm must be background gauge invariant, and since we consider only vacuum expectation values of the effective action, only one candidate is possible: it is proportional to the gauge action $\int d^4x \operatorname{tr} F^2$ and multiplied by the one-loop beta-function for the various fields which can run in the loop,

$$\Delta S \propto \beta(g) \int d^4x \operatorname{tr} F^2 \ln \frac{\mu^2}{\mu_0^2}. \tag{7.18}$$

The factor $\ln(\mu^2/\mu_0^2)$ parametrizes the freedom in choosing different renormalization schemes.

A particular regularization scheme used in [4] is Pauli-Villars regularization. In this case 't Hooft first used x -dependent regulator masses to compute the ratios of the one-loop determinants Δ in the instanton background and $\Delta^{(0)}$ in the trivial vacuum. Then he argued that the difference between using the x -dependent masses and using the more usual constant masses, was of the form ΔS given above. The final result for pure YM $SU(N)$ in the $|k=1|$ sector is [4, 52]

$$\left[\frac{\det' \Delta_-}{\det \Delta_-^{(0)}} \right]^{-1} \left[\frac{\det \Delta_+}{\det \Delta_+^{(0)}} \right]^{1/2} = \exp \left\{ \frac{2}{3} N \ln(\mu \rho) - \alpha(1) - 2(N-2)\alpha\left(\frac{1}{2}\right) \right\} . \quad (7.19)$$

Here we have normalized the determinants against the vacuum, indicated by the superscript (0). From the unregularized zero mode sector one obtains a factor ρ^{4N} , see (6.14), and Pauli-Villars regularization of the $4N$ zero modes yields a factor M_{PV}^{4N} . All together one obtains $\frac{8\pi^2}{g_0^2} + \frac{22}{3} \ln(M_{PV}\rho)$ in the exponent for $SU(2)$, where g_0 is the unrenormalized coupling constant. Subtracting $\frac{22}{3} \ln(M_{PV}\rho_0)$ to renormalize at mass scale $1/\rho_0$, one is left for the effective action with $\frac{8\pi^2}{g_0^2} - \frac{11}{3} \ln(\rho/\rho_0) \equiv \frac{8\pi^2}{g^2(\rho)}$. Replacing $\ln(\rho/\rho_0)$ by $\ln(\mu/\mu_0)$, this is the correct one-loop renormalization equation for the running of the coupling constant. For supersymmetric theories, the nonzero mode corrections to the effective action cancel, and performing the same renormalization procedure as for the non-supersymmetric case, one now obtains only from the zero modes the correct β function. For $\mathcal{N} = 4$ one finds a vanishing β function.

The fluctuations of the $SU(2)$ part of the gauge fields and the Faddeev-Popov ghosts yield the term $\alpha(1)$ in (7.19), while the fluctuations of the $2(N-2)$ doublets (corresponding to $\lambda_4, \dots, \lambda_7$ for $SU(3)$) yields the term with $\alpha(\frac{1}{2})$. The numerical values of the function $\alpha(t)$ are related to the Riemann zeta function, and take the values $\alpha(\frac{1}{2}) = 0.145873$ and $\alpha(1) = 0.443307$. Notice that this expression for the determinant depends on ρ , and therefore changes the ρ -dependence of the integrand of the collective coordinate measure. Combined with (6.14) one correctly reproduces the β -function of $SU(N)$ YM theory. The calculation of the contribution of the nonzero modes can be simplified by using a so-called $O(5)$ formalism [59] which uses the conformal symmetries of instantons, in addition to the nonconformal symmetries. One still has to regulate the sums over zero-point energies, and both Pauli-Villars regularization [59] and zeta-function regularization [60] have been applied to the $O(5)$ formulation.

7.1 The exact β function for SYM theories

In supersymmetric gauge theories, the contributions to the one-loop partition function by the nonzero modes in the bosonic and fermionic loops cancel each other [61]. Although this

has only been shown to occur in a gravitational background without winding, we assume here that still occurs in an instanton background. Actually, all contributions from the nonzero mode sector cancel: higher-loops as well as possible nonperturbative corrections. The zero mode sector can be regularized by Pauli-Villars fields, and since the partition function yields a physical observable, namely the cosmological constant (the sum over zero-point energies), the result for the partition function should not depend on the regularization parameter M_{PV} (the Pauli-Villars mass). From this observation one can derive a differential equation for the coupling constant $g(M_{PV})$, which yields the exact β function: it contains all perturbative contributions [62].

Before going on we should comment on the fact that from the 3-loop level on the result for the β function depends on the regularization scheme chosen. It is sometimes claimed that therefore higher-loop results for the β function have no meaning. This is incorrect: given a particular scheme, all orders in perturbation theory of β have meaning. In the derivation below of the β function we shall find an all-order result, but it is not (yet?) known which regularization scheme for Feynman graphs would reproduce these results. So the all-order expression for β has in principle meaning, but in practice one cannot do much with it. One can only say: there must exist a regularization scheme which, if used for the calculation of higher-loop Feynman graphs, will produce the all-order result for β obtained below.

We begin with pure supersymmetric gauge theory. We recall that the measure of the zero modes of a single instanton or anti-instanton ($k = \pm 1$) for $N = 1$ susy with gauge group $SU(N)$ and one Majorana or Weyl fermion in the adjoint representation is given by

$$\begin{aligned} d\mathcal{M}_{k=\pm 1} = e^{-\frac{8\pi^2}{g^2}} & \left[\frac{d^4x d\rho}{\rho^5} 2^{4N+2} \left(\frac{\rho}{g}\right)^{4N} \left(\frac{M_{PV}}{\sqrt{2\pi}}\right)^{4N} \frac{\pi^{4N-2}}{(N-1)!(N-2)!} \right] \times \\ & \left[\frac{d\xi_1 d\xi_2}{4S_{cl} M_{PV}} \frac{d\bar{\eta}_1 d\bar{\eta}_2}{8\rho^2 S_{cl} M_{PV}} \frac{\prod_{u=1}^{N-2} d\mu^u d\bar{\mu}_u}{(\frac{1}{4}S_{cl} M_{PV})^{N-2}} \right] \text{Vol} \left\{ \frac{SU(N)}{SU(N-2) \otimes U(1)} \right\} \end{aligned} \quad (7.20)$$

where $S_{cl} = 8\pi^2/g^2$. The volume of the gauge group was given in (6.15) but because it does not depend on g or M_{PV} it will play no role below. Note that this measure is dimensionless; $d^4x d\rho/\rho^5$ is dimensionless, and the remaining ρ and M_{PV} occur only in the combination ρM_{PV} . Also $d^2\xi/M_{PV}$ and $d^2\bar{\eta}/(\rho^2 M_{PV})$ are dimensionless. The prefactor $e^{-8\pi^2/g^2}$ is of course the classical action for the one-instanton background, and we have left out the term with the term with the theta-angle. In the first square brackets we find the product of the measure for the bosonic zero modes in (6.14) and factors $\sqrt{\frac{M_{PV}^2}{2\pi}}$ for each bosonic zero mode from the corresponding Pauli-Villars modes.³² The second expression in square brackets contains the contribution to the measure from the fermionic zero modes

³²If the one-loop determinant for the bosonic fields is $(\det M_b)^{-1/2}$ and for the fermionic fields $\det M_f$,

given by (6.24), with factors $\frac{1}{\sqrt{M_{PV}}}$ for each fermionic zero mode. Clearly, each bosonic zero mode contributes a factor M_{PV}/g and each fermionic zero mode contributes a factor $g/\sqrt{M_{PV}}$.

The dependence of $d\mathcal{M}$ on M_{PV} and g is thus as follows

$$d\mathcal{M} \propto e^{-\frac{8\pi^2}{g^2}} (M_{PV})^{3N} \left(\frac{1}{g}\right)^{2N}, \quad (7.21)$$

where g depends on M_{PV} , so $g = g(M_{PV})$. So g is the bare coupling constant in the regularized theory, and g and M_{PV} vary such that the renormalized coupling constant g_R is kept fixed. Usually one considers the renormalized coupling constant as a function of the renormalization mass μ , and then the bare coupling constant g satisfies $\mu \frac{\partial}{\partial \mu} g = 0$. Using dimensional regularization and $g = Z_g(g_{\text{ren}}) g_{\text{ren}} \mu^{\epsilon/2}$ with $\epsilon = 4 - n$ yields then the β function. If one uses Pauli-Villars regularization there are two masses which play a role: the cut-off (regulator) mass M_{PV} and the physical renormalization mass μ . The bare coupling depends on one of them, the renormalized coupling on the other.

$$\begin{aligned} g &= g(M_{PV}) & g_R &= g_R(\mu) \\ M_{PV} \frac{\partial}{\partial M_{PV}} g(M_{PV}) &= \beta(g) & \mu \partial / \partial \mu g_R(\mu) &= \beta(g_R) \\ \mu \frac{\partial}{\partial \mu} g(M_{PV}) &= 0 & M_{PV} \frac{\partial}{\partial M_{PV}} g_R(\mu) &= 0 \end{aligned} \quad (7.22)$$

Physical quantities depend on μ but not on M_{PV} . If one wants to apply the renormalization group to the measure, one must use the approach based on $(M_{PV} \partial / \partial M_{PV}) g_R(\mu) = 0$ because the regularized measure depends on M_{PV} , not on μ . The results for the β function obtained from both schemes differ by a sign, because the logarithms in the theory depend on $\ln(M_{PV}/\mu)$.

Equating the derivative of the logarithm of the measure w.r.t. in (7.21) M_{PV} to zero yields then

$$M_{PV} \frac{\partial}{\partial M_{PV}} \left(-\frac{8\pi^2}{g^2} + 3N \ln M_{PV} - 2N \ln g \right) = 0. \quad (7.23)$$

Hence

$$M_{PV} \frac{\partial}{\partial M_{PV}} g \equiv \beta = \left(\frac{3N}{\frac{2N}{g} - \frac{16\pi^2}{g^3}} \right), \quad (7.24)$$

or, written in terms of $\alpha = \frac{g^2}{4\pi}$

$$\frac{g}{2\pi} \beta = M_{PV} \frac{\partial}{\partial M_{PV}} \alpha = \frac{-3N\alpha^2}{2\pi} \frac{1}{1 - \frac{\alpha N}{2\pi}}. \quad (7.25)$$

then the Pauli-Villars method yields further determinants $\det(M_b + M_{PV}^2)^{+1/2}$ and $\det(M_f + M_{PV})^{-1}$. The zero modes are eigenfunctions of M_b and M_f with eigenvalue zero, so their Pauli-Villars counterparts become nonzero modes with eigenvalues M_{PV}^2 and M_{PV} .

This is the β -function for pure $\mathcal{N} = 1$ supersymmetric Yang-Mills theory. It is straightforward to extend this result to pure $\mathcal{N}$ -extended supersymmetry with $\mathcal{N}$ Majorana or Weyl fermions in the adjoint representation. One finds for $SU(N)$

$$M_{PV} \frac{\partial}{\partial M_{PV}} \alpha = -\frac{\alpha^2}{2\pi} \frac{4N - \mathcal{N}N}{1 - \frac{\alpha}{2\pi}(2N - \mathcal{N}N)} . \quad (7.26)$$

It is clear that for $\mathcal{N} = 2$ there is only a one-loop contribution to β , and for $\mathcal{N} = 4$ the β function vanishes altogether. These are well-known properties of pure extended susy gauge theories. For $\mathcal{N} = 1$ one finds agreement for one- and two- loops. Beyond two loops the result for the beta function becomes scheme dependent, so it becomes then pointless to investigate whether agreement holds.

Let us now add matter. In susy QCD with N_f flavours the matter part consists of N_f pairs of chiral superfields Q^i and $\tilde{Q}_i$ with $i = 1, N_f$ in the $\underline{N}$ and $\underline{N}^*$ representations of $SU(N)$. Each fermion in Q and $\tilde{Q}$ has one zero mode, see (4.37) and (5.35), while the scalars do not have any zero modes. So the zero mode measure for the matter part is according to (6.26)

$$d\mathcal{M}(\text{matter}) = \left(\frac{1}{\pi^2}\right)^{2N_f} \frac{1}{(M_{PV})^{2N_f}} \prod_{u=1}^{N_f} dK^u d\tilde{K}_u . \quad (7.27)$$

Renormalization leads to a further term in the measure, and thus in the β function. In susy only the kinetic term $\bar{\phi} e^V \phi$ of the matter fields gets a Z factor

$$\mathcal{L} = Z \bar{\phi}_{\text{ren}} e^{V^{\text{ren}}} \phi_{\text{ren}} , \quad \phi = \sqrt{Z} \phi_{\text{ren}} . \quad (7.28)$$

and rather than a factor $\frac{1}{\sqrt{M_{PV}}}$ for each fermion with one flavor, we now get in the measure a factor $(Z M_{PV})^{-1/2}$ for each zero mode. (The Pauli-Villars field operator becomes $Z M_f + M_{PV}$, so the zero modes continue to produce a factor $M_{PV}^{-1/2}$ in the Pauli-Villars sector, **not** $(Z M_{PV})^{-1/2}$. In the nonzero mode sector one can neglect the dependence on M_{PV} , and here the Z factors of bosons and fermions cancel due to susy).

For the gauge multiplet we factorized out a factor $1/g^2$ in front of the action of all fields of the gauge multiplet, so that the fields $gA_\mu = \tilde{A}_\mu$ do not renormalize. (We use here the background formalism in which $Z_g = Z_A^{-1/2}$, where Z_A is the wave function renormalization constant for the background fields.) Thus the renormalization of the gauge multiplet is taken care of by the renormalization of the factor $1/g^2$ in (7.23).

From here on we proceed as before. The measure for gauge group $SU(N)$ with N_f flavors is now given by

$$dM_{k=\pm 1} = e^{-8\pi^2/g^2} (M_{PV})^{3N} \left(\frac{1}{g}\right)^{2N} \left(\frac{1}{Z M_{PV}}\right)^{N_f} \quad (7.29)$$

We denote the anomalous dimension by γ_i where³³

$$\begin{aligned}\gamma_i &= \mu \frac{\partial}{\partial \mu} \ln Z_i \\ &= -M_{PV} \frac{\partial}{\partial M_{PV}} \ln Z_i = \gamma \quad (\text{the same for } i = 1, \dots, N_f) ,\end{aligned}\tag{7.30}$$

and obtain

$$\frac{g}{2\pi} \beta = M_{PV} \frac{\partial}{\partial M_{PV}} \alpha = -\frac{\alpha^2}{2\pi} \left(\frac{3N - N_f(1 - \gamma)}{1 - \frac{\alpha N}{2\pi}} \right) .\tag{7.31}$$

Expanding in terms of α , the result agrees with the results in the literature for the one-loop and two-loop β functions for $N = 1$ susy QCD with N_f pairs of chiral fields Q^i and $\tilde{Q}_i$. Namely, the one- and two-loop β function for an $N = 1$ vector multiplet coupled to a chiral multiplet in a representation R , including the effects of the Yukawa couplings whose coupling constant is also g (in fact, the renormalized coupling constant g_R), is given by [64]

$$\begin{aligned}\frac{g}{2\pi} \mu \frac{\partial}{\partial \mu} g &= \frac{\alpha^2}{2\pi} (-3C_2(G) + T(R)) \\ &+ \frac{\alpha^3}{8\pi^2} (-6C_2^2(G) + 2C_2(G)T(R) + 4C_2(R)T(R))\end{aligned}\tag{7.32}$$

For N_f pairs of chiral matter fields $\sum T(R) = N_f$, and $C_2(G) = N$ for $SU(N)$. Using also that the anomalous dimension $\gamma = \mu \frac{\partial}{\partial \mu} \ln Z$ for a complex fermion in the fundamental representation $\mathbf{N}$ of $SU(N)$ is equal to $-\alpha C_2(R)/\pi$, we indeed find agreement.³⁴

The β function in (7.31) can be rewritten such that only the numbers of zero modes appear.

$$\beta(\alpha) = -\frac{\alpha^2}{2\pi} \left(n_g - \frac{1}{2}n_f - \frac{1}{2} \sum_g \gamma_g + \frac{1}{2} \sum_f \gamma_f \right) .\tag{7.33}$$

Here n_g is the number of bosonic zero modes ($4N$), n_f the total (gluino and matter) number of fermionic zero modes ($2N + 2N_f$), and the sums $\sum_g$ and $\sum_f$ run over the gluon and fermion zero modes. For gluons and gluinos λ , the anomalous dimension is the same (due to susy) and proportional to the β function

$$\gamma_g = \gamma_\lambda = \beta/\alpha .\tag{7.34}$$

Substitution of this result yields back (7.31). This result does not yet agree with the results in the literature for the β -function of gauge fields minimally coupled to scalars and

³³Often one defines $\gamma = \mu \frac{\partial}{\partial \mu} \ln \sqrt{Z}$; here we follow [62]

³⁴With the usual normalization $\gamma = \mu \frac{\partial}{\partial \mu} \ln Z^{1/2}$ is equal to $\gamma = \frac{-\alpha}{2\pi} C_2(R)$ [64].

fermions, because in supersymmetry the Yukawa couplings between scalars and fermions have not an independent coupling constant λ but rather $\lambda = g^2$. At the two-loop level one therefore gets extra contributions which one must add to the results from the literature, and then one gets complete agreement.

8 $\mathcal{N} = 4$ supersymmetric Yang-Mills theory

An interesting field theory with instantons is the $\mathcal{N} = 4$ super Yang-Mills theory [65]. The action is of course well known in Minkowski space, but instantons require the formulation in $\mathcal{N} = 4$ Euclidean space. Due to absence of a real representation of Dirac matrices in four-dimensional Euclidean space, one cannot straightforwardly define Majorana spinors in Euclidean space. This complicates the construction of Euclidean Lagrangians for supersymmetric models [66, 67, 68]. For $\mathcal{N} = 2, 4$ theories, one can replace the Majorana condition by the so-called symplectic Majorana condition and then one can define (symplectic) Majorana spinors in Euclidean space. Equivalently, one can work with complex (Dirac) spinors [69, 70]. In the following subsection we write down the action in Minkowski space-time and discuss the reality conditions on the fields. Next we construct the hermitean $\mathcal{N} = 4$ Euclidean model via the dimensional reduction of ten-dimensional $\mathcal{N} = 1$ super Yang-Mills theory along the time direction. One can also define a continuous Wick rotation for the spinors directly in four dimensions [68].

8.1 Minkowskian $\mathcal{N} = 4$ SYM

The $\mathcal{N} = 4$ action in Minkowski space-time with the signature $\eta^{\mu\nu} = \text{diag}(-, +, +, +)$ is given by

$$S = \frac{1}{g^2} \int d^4x \text{tr} \left\{ \frac{1}{2} F_{\mu\nu} F^{\mu\nu} - i \bar{\lambda}_A^{\dot{\alpha}} \bar{\mathcal{D}}_{\dot{\alpha}\beta} \lambda^{\beta,A} - i \lambda_\alpha^A \mathcal{D}^{\alpha\dot{\beta}} \bar{\lambda}_{A\dot{\beta}} + \frac{1}{2} (D_\mu \bar{\phi}_{AB}) (D^\mu \phi^{AB}) \right. \\ \left. - \sqrt{2} \bar{\phi}_{AB} \{ \lambda^{\alpha,A}, \lambda_\alpha^B \} - \sqrt{2} \phi^{AB} \{ \bar{\lambda}_A^{\dot{\alpha}}, \bar{\lambda}_{\dot{\alpha},B} \} + \frac{1}{8} [\phi^{AB}, \phi^{CD}] [\bar{\phi}_{AB}, \bar{\phi}_{CD}] \right\}. \quad (8.1)$$

The on-shell $\mathcal{N} = 4$ supermultiplet consists of a real gauge field A_μ , four complex Weyl spinors $\lambda^{\alpha,A}$ (equivalently, four Majorana spinors) and an antisymmetric complex scalar ϕ^{AB} with labels $A, B = 1, \dots, 4$ of the internal R symmetry group $SU(4)$. The reality conditions on the components of this multiplet are³⁵ the Majorana conditions $(\lambda^{\alpha,A})^* = -\bar{\lambda}_A^{\dot{\alpha}}$ and $(\lambda_\alpha^A)^* = \bar{\lambda}_{\dot{\alpha},A}$ and

$$\bar{\phi}_{AB} \equiv (\phi^{AB})^* = \frac{1}{2} \epsilon_{ABCD} \phi^{CD}. \quad (8.2)$$

³⁵Unless specified otherwise, equations which involve complex conjugation of fields will be understood as not Lie algebra valued, i.e. they hold for the components $\lambda^{a,\alpha,A}$, etc.

These conditions are invariant under $SU(4)$ transformations. The sigma matrices are defined by $\sigma^{\mu\alpha\dot{\beta}} = (1, \tau^i)$, $\bar{\sigma}^{\mu}_{\dot{\alpha}\beta} = (-1, \tau^i)$ for $\mu = 0, 1, 2, 3$ and complex conjugation gives $(\sigma^{\alpha\dot{\beta}}_{\mu})^* = \sigma^{\beta\dot{\alpha}}_{\mu} = \bar{\sigma}^{\dot{\alpha}\beta}_{\mu} = \epsilon^{\dot{\alpha}\dot{\gamma}}\epsilon^{\beta\delta}\bar{\sigma}_{\mu\dot{\gamma}\delta}$, with $\epsilon^{\dot{\alpha}\dot{\beta}} = \epsilon_{\dot{\alpha}\dot{\beta}} = -\epsilon^{\alpha\beta} = -\epsilon_{\alpha\beta}$. Since ϕ^{AB} is antisymmetric, one can express it on a basis spanned by the real eta-matrices (see Appendix B)

$$\phi^{AB} = \frac{1}{\sqrt{2}} \{S^i \eta^{iAB} + iP^i \bar{\eta}^{iAB}\}, \quad \bar{\phi}_{AB} = \frac{1}{\sqrt{2}} \{S^i \eta^i_{AB} - iP^i \bar{\eta}^i_{AB}\}, \quad (8.3)$$

in terms of real scalars S^i and real pseudoscalars P^i , $i = 1, 2, 3$. Because η^{iAB} is selfdual and $\bar{\eta}^{iAB}$ anti-selfdual, $\eta^{iAB} = \eta^i_{AB}$ and $\bar{\eta}^{iAB} = -\bar{\eta}^i_{AB}$. Then the reality conditions are fulfilled and the kinetic terms for the (S, P) fields take the standard form. The action in (8.1) is invariant under the following supersymmetry transformation laws with parameters ζ^A_{α} and $\bar{\zeta}_{\dot{\alpha}A}$

$$\begin{aligned} \delta A_{\mu} &= -i\bar{\zeta}^{\dot{\alpha}}_A \bar{\sigma}_{\mu\dot{\alpha}\beta} \lambda^{\beta,A} + i\bar{\lambda}_{\dot{\beta},A} \sigma^{\alpha\dot{\beta}}_{\mu} \zeta^A_{\alpha}, \\ \delta \phi^{AB} &= \sqrt{2} \left(\zeta^{\alpha,A} \lambda^B_{\alpha} - \zeta^{\alpha,B} \lambda^A_{\alpha} + \epsilon^{ABCD} \bar{\zeta}^{\dot{\alpha}}_C \bar{\lambda}_{\dot{\alpha},D} \right), \\ \delta \lambda^{\alpha,A} &= -\frac{1}{2} \sigma^{\mu\nu\alpha}_{\beta} F_{\mu\nu} \zeta^{\beta,A} - i\sqrt{2} \bar{\zeta}_{\dot{\alpha},B} \not{D}^{\alpha\dot{\alpha}} \phi^{AB} + [\phi^{AB}, \bar{\phi}_{BC}] \zeta^{\alpha,C}, \end{aligned} \quad (8.4)$$

which are consistent with the reality conditions. Let us turn now to the discussion of the Euclidean version of this model and discuss the differences with the Minkowski theory.

8.2 Euclidean $\mathcal{N} = 4$ SYM

To find out the $\mathcal{N} = 4$ supersymmetric YM model in Euclidean $d = (4, 0)$ space, we follow the same procedure as in [65]. We start with the $\mathcal{N} = 1$ SYM model in $d = (9, 1)$ Minkowski space-time, but contrary to the original papers we reduce it on a six-torus with one time and five space coordinates [69, 70]. As opposed to the action in (8.1) with the $SU(4) = SO(6)$ R -symmetry group, this reduction leads to a model with an internal non-compact $SO(5, 1)$ R -symmetry group in Euclidean space. As we will see, the reality conditions on bosons and fermions will both use an internal metric for this non-compact internal symmetry group.

The $\mathcal{N} = 1$ Lagrangian in $d = (9, 1)$ dimensions reads

$$\mathcal{L}_{10} = \frac{1}{g_{10}^2} \text{tr} \left\{ \frac{1}{2} F_{MN} F^{MN} + \bar{\Psi} \Gamma^M D_M \Psi \right\}, \quad (8.5)$$

with the field strength $F_{MN} = \partial_M A_N - \partial_N A_M + [A_M, A_N]$ and the Majorana-Weyl spinor Ψ defined by the conditions

$$\Gamma^{11} \Psi = \Psi, \quad \Psi^T C_{10}^{-} = \Psi^{\dagger} i \Gamma^0 \equiv \bar{\Psi}. \quad (8.6)$$

Here the hermitean matrix $\Gamma^{11} \equiv \star \Gamma$ is a product of all Dirac matrices, $\Gamma^{11} = \Gamma^0 \dots \Gamma^9$, normalized to $(\star \Gamma)^2 = +1$. Furthermore, C_{10}^- is the charge conjugation matrix, satisfying $C_{10}^- \Gamma_M = -\Gamma_M^T C_{10}^-$. The Γ -matrices obey the Clifford algebra $\{\Gamma^M, \Gamma^N\} = 2\eta^{MN}$ with metric $\eta^{MN} = \text{diag}(-, +, \dots, +)$. The Lagrangian transforms into a total derivative under the standard transformation rules³⁶

$$\delta A_M = \bar{\zeta} \Gamma_M \Psi, \quad \delta \Psi = -\frac{1}{2} F_{MN} \Gamma^{MN} \zeta, \quad (8.7)$$

with $\Gamma^{MN} = \frac{1}{2}[\Gamma^M \Gamma^N - \Gamma^N \Gamma^M]$. The susy parameter is a Majorana-Weyl spinor, $\bar{\zeta} = \zeta^T C_{10}^- = \zeta^\dagger i \Gamma^0$ and $\star \Gamma \zeta = \zeta$. To proceed with the dimensional reduction we choose a particular representation of the gamma matrices in $d = (9, 1)$, namely

$$\Gamma^M = \{\hat{\gamma}^a \otimes \gamma^5, \mathbb{1}_{[8] \times [8]} \otimes \gamma^\mu\}, \quad \Gamma^{11} = \Gamma^0 \dots \Gamma^9 = \hat{\gamma}^7 \otimes \gamma^5, \quad (8.8)$$

where the 8×8 Dirac matrices $\hat{\gamma}^a$ and $\hat{\gamma}^7$ of $d = (5, 1)$ with $a = 1, \dots, 6$ can be conveniently defined by means of 't Hooft symbols as follows

$$\hat{\gamma}^a = \begin{pmatrix} 0 & \Sigma^{a,AB} \\ \bar{\Sigma}_{AB}^a & 0 \end{pmatrix}, \quad \hat{\gamma}^7 = \hat{\gamma}^1 \dots \hat{\gamma}^6 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (8.9)$$

In Euclidean $d = (6, 0)$ one defines

$$\begin{aligned} \Sigma^{a,AB} &= (\eta^{kAB}, i\bar{\eta}^{k,AB}) \\ \bar{\Sigma}_{AB}^a &= (-\eta^k_{AB}, i\bar{\eta}^k_{AB}) \end{aligned} \quad (8.10)$$

but in Minkowski space one puts a factor $-i$ in front of the first one. So explicitly $\Sigma^{a,AB} = \{-i\eta^{1,AB}, \eta^{2,AB}, \eta^{3,AB}, i\bar{\eta}^{k,AB}\}$, $\bar{\Sigma}_{AB}^a = \{i\eta_{AB}^1, -\eta_{AB}^2, -\eta_{AB}^3, i\bar{\eta}_{AB}^k\}$ so that $\frac{1}{2}\epsilon_{ABCD}\Sigma^{a,CD} = -\bar{\Sigma}_{AB}^a$. The first three matrices $\hat{\gamma}^1, \hat{\gamma}^2, \hat{\gamma}^3$ are symmetric while the latter three matrices $\hat{\gamma}^4, \hat{\gamma}^5, \hat{\gamma}^6$ are antisymmetric. Meanwhile γ^μ and γ^5 are the usual Dirac matrices of $d = (4, 0)$ introduced in (4.15). Note that in this construction we implicitly associated one of the Dirac matrices, namely $\hat{\gamma}^1$, in 6 dimensions with the time direction and thus it is anti-hermitean and has square -1 ; all others (as well as all Dirac matrices in $d = (4, 0)$) are again hermitean with square $+1$. Let us briefly discuss the charge conjugation matrices in $d = (9, 1)$, $d = (5, 1)$ and $d = (4, 0)$. One can prove by means of finite group theory [71] that all their properties are representation independent. In general there are two charge conjugation matrices C^+ and C^- in even dimensions, satisfying $C^\pm \Gamma^\mu = \pm (\Gamma^\mu)^T C^\pm$, and

³⁶After partial integrations, the Yang-Mills action transforms into $[\zeta \Gamma_N \psi^a] D_M F^{a,MN}$ and the Dirac action varies into $-\bar{\psi}^a \Gamma^M (-\frac{1}{2} \Gamma^{PQ} D_M F_{PQ} \zeta)$. The sum of these two variations cancels if one uses the Bianchi identity $D_{[M} F_{PQ]} = 0$. The variation of A_M in the covariant derivative in the Dirac action cancels separately due to the 3-spinor identity $(\bar{\psi}^a \Gamma^M \psi^b)(\bar{\zeta} \Gamma_M \psi^c) f_{abc} = 0$ which holds in 3,4,6 and 10 dimensions.

$C^+ = C^- * \Gamma$. These charge conjugation matrices do not depend on the signature of space-time and obey the relation $C^- * \Gamma = \pm (*\Gamma)^T C^-$ with $-$ sign in $d = 10, 6$ and $+$ sign in $d = 4$. The transposition depends on the dimension and leads to $(C^\pm)^T = \pm C^\pm$ in $d = 10$, $(C^\pm)^T = \mp C^\pm$ in $d = 6$, and finally $(C^\pm)^T = -C^\pm$ for $d = 4$. Explicitly, the charge conjugation matrix C_{10}^- is given by $C_6^- \otimes C_4^-$ where

$$C_4^- = \gamma^4 \gamma^2 = \begin{pmatrix} \epsilon_{\alpha\beta} & 0 \\ 0 & \epsilon^{\alpha'\beta'} \end{pmatrix}, \quad C_6^- = i\hat{\gamma}^4 \hat{\gamma}^5 \hat{\gamma}^6 = \begin{pmatrix} 0 & \delta_{AB} \\ \delta^{AB} & 0 \end{pmatrix}. \quad (8.11)$$

Upon compactification to Euclidean $d = (4, 0)$ space, the 10-dimensional Lorentz group $SO(9, 1)$ reduces to $SO(4) \times SO(5, 1)$ with compact space-time group $SO(4)$ and R -symmetry group $SO(5, 1)$. In these conventions a Weyl spinor Ψ in ten dimensions with 16 (complex) nonvanishing components decomposes as follows into 8 and 4 component chiral-chiral and antichiral-antichiral spinors

$$\Psi = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} \lambda^{\alpha, A} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ \bar{\lambda}_{\alpha', A} \end{pmatrix}, \quad (8.12)$$

or more explicitly

$$\begin{aligned} \psi^T &= [(\lambda^{\alpha 1}, 0); (\lambda^{\alpha 2}, 0); (\lambda^{\alpha 3}, 0); (\lambda^{\alpha 4}, 0); (0, 0); (0, 0); (0, 0), (0, 0)] \\ &+ [(0, 0); (0, 0); (0, 0); (0, 0); (0\bar{\lambda}_{\alpha' 1}); (0, \bar{\lambda}_{\alpha' 2}); (0, \bar{\lambda}_{\alpha' 3}); (0, \bar{\lambda}_{\alpha' 4})]. \end{aligned} \quad (8.13)$$

Here $\lambda^{\alpha, A}$ ($\alpha = 1, 2$) transforms only under the first $SU(2)$ in $SO(4) = SU(2) \times SU(2)$, while $\bar{\lambda}_{\alpha', A}$ changes only under the second $SU(2)$. Furthermore, $\bar{\lambda}_{\alpha', A}$ transforms in the complex conjugate of the $SO(5, 1)$ representation of $\lambda^{\alpha, A}$. To understand this latter statement, note that the $SO(5, 1)$ generators are $\hat{M}^{ab} = \frac{1}{2}(\hat{\gamma}^a \hat{\gamma}^b - \hat{\gamma}^b \hat{\gamma}^a)$, and $\hat{\gamma}^1$ is antihermitian. Furthermore, $\hat{\gamma}^1, \hat{\gamma}^4, \hat{\gamma}^5$ and $\hat{\gamma}^6$ are purely imaginary. Thus

$$(\hat{\gamma}^a)^* = -\hat{S} \hat{\gamma}^a \hat{S}^{-1}; \quad \hat{S} = \hat{\gamma}^2 \hat{\gamma}^3. \quad (8.14)$$

This matrix $\hat{S}$ is not the charge conjugation matrix. The Lorentz generators $\hat{M}^{ab}$ and $\hat{S}$ are block diagonal

$$\hat{M}^{ab} = \begin{pmatrix} \Sigma^{ab} & 0 \\ 0 & \bar{\Sigma}^{ab} \end{pmatrix}; \quad \hat{S} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}, \quad (8.15)$$

where $\Sigma^{ab} = \frac{1}{2}(\Sigma^a \bar{\Sigma}^b - \Sigma^b \bar{\Sigma}^a)$ and $\bar{\Sigma}^{ab} = \frac{1}{2}(\bar{\Sigma}^a \Sigma^b - \bar{\Sigma}^b \Sigma^a)$ while $S = -\eta^2 \eta^3 = \eta^1$. It follows that

$$S \Sigma^a S^{-1} = -(\Sigma^a)^* \Rightarrow S \Sigma^{ab} S^{-1} = (\Sigma^{ab})^*,$$

$$S\bar{\Sigma}^b S^{-1} = -(\Sigma^b)^* \Rightarrow S\bar{\Sigma}^{ab} S^{-1} = (\bar{\Sigma}^{ab})^* . \quad (8.16)$$

Thus the two spinor representations of $SO(5,1)$ are each pseudoreal (they are not real since S is antisymmetric), but they are not equivalent to each other. For $SO(6) \simeq SU(4)$, the two spinor representations are of course complex and inequivalent to each other. For $SO(3,1)$ the opposite is the case: there the two spinor representations are complex, and equivalent to each other under complex conjugation, $(\sigma_{\mu\nu})^* = \sigma_2 \bar{\sigma}^{\mu\nu} \sigma_2$ because $\hat{S} = \gamma_2$ is off-diagonal.

Substituting these results, the Lagrangian reduces to

$$\begin{aligned} \mathcal{L}_E^{\mathcal{N}=4} = \frac{1}{g^2} \text{tr} \left\{ \frac{1}{2} F_{\mu\nu} F_{\mu\nu} - i \bar{\lambda}_A^{\alpha'} \bar{\mathcal{D}}_{\alpha'\beta} \lambda^{\beta,A} - i \lambda_A^{\alpha'} \mathcal{D}^{\alpha\beta'} \bar{\lambda}_{\beta',A} + \frac{1}{2} (D_\mu \bar{\phi}_{AB}) (D_\mu \phi^{AB}) \right. \\ \left. - \sqrt{2} \bar{\phi}_{AB} \{ \lambda^{\alpha,A}, \lambda_\alpha^B \} - \sqrt{2} \phi^{AB} \{ \bar{\lambda}_A^{\alpha'}, \bar{\lambda}_{\alpha',B} \} + \frac{1}{8} [\phi^{AB}, \phi^{CD}] [\bar{\phi}_{AB}, \bar{\phi}_{CD}] \right\} , \end{aligned} \quad (8.17)$$

where we still use the definition $\bar{\phi}_{AB} \equiv \frac{1}{2} \epsilon_{ABCD} \phi^{CD}$. These scalars come from the ten-dimensional gauge field, and can be grouped into $\phi^{AB} = \frac{1}{\sqrt{2}} \Sigma^{aAB} A_a$, where A_a are the first six real components of the ten dimensional gauge field A_M . Using $\eta^{ab} = \{ \frac{1}{\sqrt{2}} \Sigma^{aAB}, \frac{1}{\sqrt{2}} \bar{\Sigma}_{AB}^b \}$ with $\eta^{ab} = (-1, +1, +1, +1, +1, +1)$ the vector indices are turned into $SU(4)$ indices. Writing the action in terms of the 6 scalars A_a , one of these fields, say A_0 , has a different sign in the kinetic term, which reflects the $SO(5,1)$ symmetry of the theory. In the basis with the ϕ^{AB} fields, we obtain formally the same action for the Minkowski case by reducing on a torus with 6 space coordinates, but the difference hides in the reality conditions which we will discuss in the next subsection. The action is invariant under the dimensionally reduced supersymmetry transformation rules

$$\begin{aligned} \delta A_\mu &= -i \bar{\zeta}_A^{\alpha'} \bar{\sigma}_{\mu\alpha'\beta} \lambda^{\beta,A} + i \bar{\lambda}_{\beta',A} \sigma_\mu^{\alpha\beta'} \zeta_A^\alpha , \\ \delta \phi^{AB} &= \sqrt{2} \left(\zeta^{\alpha,A} \lambda_\alpha^B - \zeta^{\alpha,B} \lambda_\alpha^A + \epsilon^{ABCD} \bar{\zeta}_C^{\alpha'} \bar{\lambda}_{\alpha',D} \right) , \\ \delta \lambda^{\alpha,A} &= -\frac{1}{2} \sigma^{\mu\nu\alpha}{}_\beta F_{\mu\nu} \zeta^{\beta,A} - i \sqrt{2} \bar{\zeta}_{\alpha',B} \mathcal{D}^{\alpha\beta'} \phi^{AB} + [\phi^{AB}, \bar{\phi}_{BC}] \zeta^{\alpha,C} , \\ \delta \bar{\lambda}_{\alpha',A} &= -\frac{1}{2} \bar{\sigma}^{\mu\nu}{}_{\alpha'}{}^{\beta'} F_{\mu\nu} \bar{\zeta}_{\beta',A} + i \sqrt{2} \zeta^{\alpha,B} \bar{\mathcal{D}}_{\alpha'\alpha} \bar{\phi}_{AB} + [\bar{\phi}_{AB}, \phi^{BC}] \bar{\zeta}_{\alpha',C} . \end{aligned} \quad (8.18)$$

Again, these rules are formally the same as in (8.4). Note that the indices A, B are lowered by complex conjugation, but the spinor indices α and α' are lowered by ϵ -symbols.

8.3 Involution in Euclidean space

The Majorana-Weyl condition (8.6) on Ψ leads in four-dimensional Euclidean space to reality conditions on λ^α which are independent of those on $\bar{\lambda}_{\alpha'}$, namely,

$$(\lambda^{\alpha,A})^* = -\lambda^{\beta,B} \epsilon_{\beta\alpha} \eta_{BA}^1 , \quad (\bar{\lambda}_{\alpha',A})^* = -\bar{\lambda}_{\beta',B} \epsilon^{\beta'\alpha'} \eta^{1,BA} . \quad (8.19)$$

These reality conditions are consistent and define a symplectic Majorana spinor in Euclidean space. The $SU(2) \times SU(2)$ covariance of (8.19) is obvious from the pseudoreality of the $\mathbf{2}$ of $SU(2)$, but covariance under $SO(5, 1)$ can also be checked (use $[\eta^a, \bar{\eta}^b] = 0$). Since the first Σ matrix has an extra factor i in order that $(\Gamma^0)^2 = -1$, see (8.8), the reality condition on ϕ^{AB} involves η_{AB}^1

$$(\phi^{AB})^* = \eta_{AC}^1 \phi^{CD} \eta_{DB}^1 . \quad (8.20)$$

The Euclidean action in (8.17) is hermitean under the reality conditions in (8.19) and (8.20). For the σ -matrices, we have under complex conjugation

$$\left(\sigma_\mu^{\alpha\beta'}\right)^* = \sigma_{\mu\alpha\beta'} , \quad (\bar{\sigma}_{\mu\alpha'\beta})^* = \bar{\sigma}_\mu^{\alpha'\beta} . \quad (8.21)$$

Due to the nature of the Lorentz group the involution cannot change one type of indices into another, as opposed to the Minkowskian case.

9 Large instantons and the Higgs effect

We have seen in previous sections that the instanton measure on the moduli space for pure $SU(2)$ gauge theory with one anti-instanton ($k = -1$) is given by (dropping overall multiplicative factors of two and π)

$$d\mathcal{M} \propto d^4x_0 \frac{d\rho\rho^3}{g^8} M_{PV}^8 e^{-\frac{8\pi^2}{g^2}} = \left(\frac{d^4x_0 d\rho}{\rho^5 g^8} \right) e^{-\frac{8\pi^2}{g^2} + 4 \ln(\rho M_{PV})^2} . \quad (9.1)$$

The one-loop corrections coming from the determinants further modify the factor 4 into $4 - \frac{1}{3} = \frac{11}{3}$, see (7.19), and in addition yield some constants in the exponent. The integral over ρ , the instanton size, is clearly nonsingular for small ρ as long as asymptotic freedom holds³⁷, but for large ρ it diverges severely. However, in a Higgs model, the mass term for the gauge bosons ($\mathcal{L} = -\frac{1}{2}A_\mu^2 g^2 v^2$ if there are no instantons) yields further terms of the form

$$-\frac{1}{\hbar} \mathcal{L}_{\text{cl}}(\text{Higgs}) = -\frac{1}{\hbar} 2\pi^2 v^2 \rho^2 + \dots . \quad (9.2)$$

Thus for spontaneously broken gauge theories the ρ integral acquires a Gaussian cut-off, and yields a finite result. This solves the large- ρ problem for the electroweak interactions. For QCD the situation is more complicated; in fact, the large- ρ problem is presumably intimately related to confinement. We now give some details.

³⁷One integrates ρ up to the renormalization scale μ , and instantons with scale ρ yield the prefactor $\exp(-8\pi^2/g^2)$. The g^2 in this prefactor depends on ρ , not on μ . One finds then that $g^2(\rho) = 8\pi^2/(-\beta_1 \ln(\rho\Lambda))$ where $\beta_1 = -\frac{11}{3}C_2(G) + \dots$ is negative if asymptotic freedom holds. So, if $-\beta_1 + 3 \geq 0$, there is no singularity at $\rho = 0$.

The Higgs action for an $SU(2)$ Higgs doublet is given by

$$\begin{aligned}\mathcal{L}_H &= D_\mu \varphi^* D_\mu \varphi + \lambda(\varphi^* \varphi - v^2)^2 \\ D_\mu \varphi &= \partial_\mu \varphi + A_\mu \varphi; \quad \varphi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}; \quad A_\mu = A_\mu^a \frac{\tau_a}{2i}.\end{aligned}\quad (9.3)$$

With $\langle \varphi^0 \rangle = v$, the ordinary Higgs effect in Minkowski space gives a mass term $\mathcal{L} = -\frac{1}{4}A_\mu^2 v^2$ for the vector bosons³⁸. We could also discuss other representations for the Higgs field but the analysis is very similar, and a doublet is of course the most interesting case. In Euclidean space we take for A_μ the regular selfdual instanton solution with $k = 1$

$$A_\mu = -\frac{\bar{\sigma}_{\mu\nu} x^\nu}{x^2 + \rho^2}.\quad (9.4)$$

We next solve the φ field equation in this instanton background. For general λ , an exact solution to the coupled equations seems out of reach. We therefore drop the potential term and only require that $|\varphi| \rightarrow |v|$ at large $|x|$ ³⁹. So this is not an exact solution, but the first term in an approximate solution. We shall discuss the higher-order terms later. As we shall show, the solution to the equation $D^\mu D_\mu \varphi = 0, |\varphi(|x| \rightarrow \infty)| = v$ is of the form $\varphi = f(r^2) \begin{pmatrix} -ix_4 + x_3 \\ x^1 + ix^2 \end{pmatrix}$. This clearly looks awkward, and a more covariant way to construct the solution is to write φ as

$$\varphi = \begin{pmatrix} \varphi^+ & -(\varphi^0)^* \\ \varphi^0 & (\varphi^+)^* \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix},\quad (9.5)$$

and to make the ansatz

$$\varphi = v f(x^2) \left(\bar{\sigma}_\mu x_\mu / \sqrt{x^2} \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix},\quad (9.6)$$

with $f(x^2) \rightarrow 1$ as $x^2 \rightarrow \infty$. (Recall that one can always write $\begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}$ as $\frac{1}{\sqrt{2}}(\sigma + i\vec{\chi} \cdot \vec{\sigma}) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and this yields the form of φ given in (9.5) up to an inessential factor i).

The function $f(x^2)$ satisfies a second-order differential equation, but we do not analyze this equation, but present the result and check that it solves $D^\mu D_\mu \varphi = 0$:

$$\varphi = v \sqrt{\frac{x^2}{x^2 + \rho^2}} \frac{\bar{\sigma}_\mu x_\mu}{\sqrt{x^2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{v}{\sqrt{x^2 + \rho^2}} \bar{\sigma}_\mu x_\mu \begin{pmatrix} 1 \\ 0 \end{pmatrix}.\quad (9.7)$$

³⁸One usually decomposes φ^0 into $\varphi^0 = \frac{1}{\sqrt{2}}(\sigma - i\chi_3)$, see below, with $\langle \sigma \rangle = v_\sigma$. Then $v^2 = \frac{1}{2}v_\sigma^2$, and the mass of the vector boson is $m_A = \frac{1}{2}gv$.

³⁹This is 't Hooft's approach [4]. Note that the field equation for A_μ is not restricted due to the backreaction of the Higgs field. Affleck [72] considered instead the case $v = 0$, λ arbitrary, in which case the usual instanton solution together with $\varphi = 0$ solves the coupled equations. Both approaches yield equivalent results.

The boundary condition is clearly satisfied because $\frac{\bar{\sigma}_\mu x_\mu}{\sqrt{x^2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ has unit norm. It is straightforward to check that this expression for φ satisfies the field equation. Namely, omitting the overall factor v and the spinor $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ one finds

$$\begin{aligned} D_\mu \varphi &= \partial_\mu \varphi + A_\mu \varphi = \frac{\bar{\sigma}_\mu (x^2 + \rho^2) - x_\mu x_\nu \bar{\sigma}_\nu}{(x^2 + \rho^2)^{3/2}} - \frac{(\bar{\sigma}_{\mu\nu} x_\nu)(\bar{\sigma}_\rho x^\rho)}{(x^2 + \rho^2)^{3/2}} \\ &= \bar{\sigma}_\mu \rho^2 / (x^2 + \rho^2)^{3/2} \\ D_\mu D_\mu \varphi &= \partial_\mu D_\mu \varphi + A_\mu D_\mu \varphi = \frac{-3\bar{\sigma}_\mu x_\mu \rho^2 - (\bar{\sigma}_{\mu\nu} x_\nu) \bar{\sigma}_\mu \rho^2}{(x^2 + \rho^2)^{5/2}} = 0 . \end{aligned} \quad (9.8)$$

Having found the solution of the field equation of the Higgs scalar in the background of an instanton, we now substitute it into the action to find the corrections to the classical action. The kinetic term only yields a surface integral due to partial integration

$$\begin{aligned} \int D_\mu \varphi^\dagger D_\mu \varphi \, d^4x &= \int d\Omega_\mu (\varphi^\dagger D_\mu \varphi) \\ &= \lim_{x^2 \rightarrow \infty} 2\pi^2 (x^2)^{3/2} v^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \frac{\sigma_\nu x_\nu}{\sqrt{x^2 + \rho^2}} \frac{1}{\sqrt{x^2}} \frac{x_\tau \bar{\sigma}_\tau \rho^2}{(x^2 + \rho^2)^{3/2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2\pi^2 v^2 \rho^2 . \end{aligned} \quad (9.9)$$

This is the extra term mentioned in (9.2).

However, the contribution of the term with the potential is divergent

$$\lambda \int (\varphi^* \varphi - v^2)^2 d^4x = \lambda \int \left(\frac{v^2 \rho^2}{x^2 + \rho^2} \right)^2 d^4x = \infty . \quad (9.10)$$

The reason for this divergence is clear: we did not solve the full field equation, but rather took the instanton solution of pure Yang-Mills theory, and solved the field equation for the scalar in this background, omitting the potential term.

We enter here the difficult area of “constrained instantons” [72, 73]. There does not exist an exact and stable solution of the coupled field equations, as can be shown as follows. Suppose there was a solution with $\varphi \neq 0$, and a finite but nonvanishing action for the scalars. If one replaces $A_\mu(x)$ by $aA_\mu(ax)$ and $\varphi(x)$ by $\varphi(ax)$ (which preserves the boundary condition $|\varphi| \rightarrow v$) then the action becomes upon also setting $ax = y$

$$\begin{aligned} S_{\text{cl}}(a) &= \int d^4y \left[-\frac{1}{2g^2} \text{tr} F_{\mu\nu}^2(y) + \frac{1}{a^2} |D_\mu \varphi(y)|^2 \right. \\ &\quad \left. + \frac{1}{a^4} \lambda (\varphi^*(y) \varphi(y) - v^2)^2 \right] . \end{aligned} \quad (9.11)$$

Note that all three terms in the action are positive. Replacing $A_\mu(x)$ by $aA_\mu(ax)$ for a near unity amounts to a particular small variation of A_μ , and similarly for φ . So one can make

the value of the action slightly smaller by making a slightly larger than unity. This proves that no solution exists. In fact, if a tends to infinity, we approach the bound $S = 8\pi^2/g^2$, but this bound can never be reached. The expression for $aA_\mu(ax)$ is equal to the instanton solution with ρ replaced by ρ/a , and for $a \rightarrow \infty$ we get a zero-size instanton. That leaves open the possibility that a local minimum might still exist, but detailed analysis shows that this is not the case. This scaling argument is called Derrick's theorem [74], and often yields valuable information without having to perform integrals.

One can still use an approximate solution to find a large part of the contributions to the path integral, and this approximate solution is obtained by first inserting a constraint into the path integral which yields an exact solution, and then to integrate over this constraint. The idea is as follows. There are one or at most a finite number of directions in field space along which the action decreases ("destabilizing directions", in our $SU(2)$ model the directions parametrized by a). Deformations in all other directions increase the action. The constraint prevents deformations in the destabilizing directions, and one first minimizes the action with the constraint present. The solution is called the constrained instanton. It looks like the instanton for pure Yang-Mills theory at short distances but decays exponentially at large distances. It has a particular value of ρ . Finally, one integrates with the measure for the zero modes over all values of ρ . The expectation is that this should capture most of the path integral, even though one is not expanding around a solution of the theory without constraint. For the $SU(2)$ instanton one may add a term $\sigma_1 \int d^4x [\text{tr} F^3 - c_1 \rho^{-2}]$ to the action to constrain deformations in the direction of the gauge zero mode $(\partial/\partial\rho)A_\mu^{\text{cl}}$, and a term $\sigma_2 [\int d^4x (\varphi^* \varphi - v^2)^3 - c_2 \rho^{-2}]$ to freeze deformations in the directions of the matter zero mode $(\partial/\partial\rho)\varphi^{\text{cl}}$, with φ^{cl} given by (9.7). One might fix the values of c_1 and c_2 such that the constraint is satisfied for the instanton solution and φ in (9.7). The Lagrange multipliers σ_1 and σ_2 are then fixed order by order in perturbation theory, by requiring suitable boundary conditions for the deformations.

The result is that one can make an expansion of the full approximate solution in terms of ρv and finds then the following results in the singular gauge [72, 73, 32]:

- (i) inside a core of radius $\rho = \frac{1}{m_W}$ where $m_W = gv$, the approximate solution given in (9.7) is still valid
- (ii) far away the solution decays exponentially, $A_\mu \sim \exp(-m_W|x|)$ and $|\varphi - v| \sim \exp(-m_H|x|)$ with $m_H = 2\sqrt{\lambda}v$.
- (iii) the integral over $|D_\mu\varphi|^2$ has the same leading term $2\pi^2\rho^2v^2 + \mathcal{O}\left(\lambda(v\rho)^4 \ln(v\rho\sqrt{\lambda})\right)$, but the potential term is now convergent and yields a result $\mathcal{O}\left(\lambda(v\rho)^4 \ln(v\rho\sqrt{\lambda})\right)$.

Hence, the Higgs effect indeed solves the large ρ problem, and asymptotic freedom solves the small ρ problem. Constrained instantons are also relevant for $\mathcal{N} = 1, 2$ SYM theories. They can also be studied in the context of topological YM theories, as was discussed e.g.

in [75].

10 Instantons as most probable tunnelling paths

Instantons of nonabelian gauge theories can be interpreted as amplitudes for tunnelling between vacua in Minkowski space with different winding numbers Q . We shall determine a path in Minkowski spacetime which yields the “most probable barrier tunnelling amplitude”. We follow closely [76], but related work is found in [77, 78].

We begin with one particular path $A_{I,\mu} = \{\vec{A}_I(\vec{x}, t), A_{I,0}(\vec{x}, t)\}$ from which we construct a class of paths which all differ by how fast one goes from one configuration at t_1 to the next at t_2 . Namely, we make a coordinate transformation from t to $\lambda(t)$ **in Minkowski spacetime** and consider the following collection of paths

$$\vec{A}_I^{(\lambda)}(\vec{x}, t) = \vec{A}_I(\vec{x}, \lambda(t)) ; \quad A_{I,0}^{(\lambda)}(\vec{x}, t) = A_{I,0}(\vec{x}, \lambda(t)) \dot{\lambda}(t) \quad (10.1)$$

(Often one works in the temporal gauge $A_0^{(\lambda)} = 0$ because this makes the physical interpretation clearer. All our results are, however, gauge invariant). The case $\lambda(t) = t$ yields the original path, but different $\lambda(t)$ yield paths which all run through the same sequence of 3-geometries $\vec{A}_I(\vec{x}, t_1), \vec{A}_I(\vec{x}, t_2), \vec{A}_I(\vec{x}, t_3) \dots$ but at different speeds. The variable $\lambda(t)$ can be considered as a kind of collective coordinate which measures a kind of continuous winding number because we will start with one winding number and end up with another winding number. For t between t_1 and t_2 this continuous winding number is due to an integral $\int d^3x \int_{t_1}^t dt' \partial_\mu j^\mu$ over a surface where A_μ is not everywhere pure gauge. Only for $t = t_1$ and $t = t_2$ does A_μ everywhere on the surface become pure gauge and only at these times the winding number is an integer. These initial and final configurations describe vacua of the theory in Minkowski spacetime. We can also consider another particular path $A_{II,\mu} = \{\vec{A}_{II}(\vec{x}, t), A_{II,0}(\vec{x}, t)\}$, and then we can in the same way create a second class of paths, parametrized again by the function $\lambda(t)$. In this way we generate an infinite collection of classes of paths.

For a given class $A_\mu^{(\lambda)}(\vec{x}, t)$, we can substitute $\vec{A}^{(\lambda)}$ and $A_0^{(\lambda)}$ into the action, and then we obtain, as we shall show, the Lagrangian for a point particle (one dynamical degree of freedom)

$$L = \frac{1}{2} m(\lambda) \dot{\lambda}^2 - V(\lambda) \quad (10.2)$$

where $m(\lambda)$ and $V(\lambda)$ depend on the choice for A_μ . We shall then determine for which $m(\lambda)$ and $V(\lambda)$ the tunnelling rate is maximal. The solution of this problem in Minkowski space involves instantons in Euclidean space. A crucial role is played by the notion of a winding number in Minkowski space, so we first discuss this subject.

One can define a winding number Q in Minkowski space in the same way as in Euclidean space because Q does not depend on the metric (in technical terms it is an affine quantity)

$$\begin{aligned}
Q &= \frac{-1}{64\pi^2} \int_{\sigma_1}^{\sigma_2} F_{\mu\nu}^a F_{\rho\sigma}^a \epsilon^{\mu\nu\rho\sigma} d^4x \\
&= \frac{1}{32\pi^2} \int (\text{tr } F_{\mu\nu} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}) d^4x \\
&= \frac{-1}{4\pi^2} \int \text{tr } \vec{E} \cdot \vec{B} d^4x
\end{aligned} \tag{10.3}$$

where $d^4x = d^3x dt$ and $\epsilon^{0123} = +1$, and we used that $A_\mu = A_\mu^a T_a$ with $T_a = -\frac{i}{2}\sigma_a$ so that $\text{tr}(T_a T_b) = -\frac{1}{2}\delta_{ab}$ and the structure constants are given by $[T_a, T_b] = \epsilon_{ab}^c T_c$, so $f_{ab}^c = \epsilon_{ab}^c$. Furthermore, by definition $E_j = -F_{0j}$ and $B_j = \frac{1}{2}\epsilon_{jkl} F_{kl}$. Because we are (and stay all the time) in Minkowski space, $\epsilon^{0123} = -\epsilon_{0123} = +1$ and $-F_{\mu\nu} F^{\mu\nu} = 2F_{0i}^2 - F_{ij}^2$. The integral is taken between two 3-dimensional hypersurfaces σ_1 and σ_2 at t_1 and t_2 .

If at t_1 the configuration $A_\mu(\vec{x}, t)$ describes a vacuum, it has by definition vanishing energy. Since the energy density⁴⁰ is given by $\mathcal{H} = \frac{1}{2}(\vec{E}^a)^2 + \frac{1}{2}(\vec{B}^a)^2$, vanishing energy means $F_{\mu\nu}^a = 0$, hence A_μ is pure gauge at $t = t_1$

$$A_\mu(\vec{x}, t_1) = e^{-\alpha(\vec{x}, t_1)} \partial_\mu e^{\alpha(\vec{x}, t_1)} . \tag{10.4}$$

Similarly, at t_2 we have $A_\mu(\vec{x}, t_2) = e^{-\beta(\vec{x}, t_2)} \partial_\mu e^{\beta(\vec{x}, t_2)}$. We now choose the temporal gauge

$$A_0(\vec{x}, t) = 0 . \tag{10.5}$$

Having fixed $A_0 = 0$, there are still residual space-dependent gauge transformations possible because they preserve the gauge $A_0 = 0$. To check this statement is easy:

$$A'_0(\vec{x}, t) = e^{-g(\vec{x})} \partial_0 e^{g(\vec{x})} = 0 . \tag{10.6}$$

We use these residual gauge transformations to set $\alpha(\vec{x}, t_1) = 0$.⁴¹ Then $A_\mu(\vec{x}, t_1) = 0$ for all μ and all $\vec{r}$.

⁴⁰The gravitational stress tensor is $T_{\mu\nu} = F_{\mu\rho}^a F_{\nu}^{a\rho} - \frac{1}{4}\eta_{\mu\nu} F_{\rho\sigma}^a F^{a,\rho\sigma}$ and $T_{00} = \frac{1}{2}(E^a)^2 + \frac{1}{2}(B^a)^2$. One can also obtain $T_{\mu\nu}$ from canonical methods as follows. Evaluating $H = p\dot{q} - L$ with $q = A_j$ and $p = -E_j$ one finds upon using that $\dot{A}_j = F_{0j} + D_j A_0$ and partially integrating that $H = \int [\frac{1}{2}\{(E_j^a)^2 + (B_j^a)^2\} + A_0^a (D^j E_j^a)] d^3x$ plus a boundary term. For solutions of the field equations such as the vacuum, $D^j E_j = 0$. For configurations with finite energy ($E = \mathcal{O}(\frac{1}{r^2})$) the boundary term vanishes when A_0 falls off like $\mathcal{O}(\frac{1}{r})$. Moreover in the temporal gauge the last term vanishes. Actually, according to the Dirac formalism, the Gauss operator $D^j E_j$ is a first-class constraint, and should be omitted from the Hamiltonian. Thus, $H = \int [\frac{1}{2}(E_j^a)^2 + \frac{1}{2}(B_j^a)^2] d^3x$ also according to canonical methods.

⁴¹With $A_j(\vec{x}, t_1) = e^{-\alpha(\vec{x}, t_1)} \partial_j e^{\alpha(\vec{x}, t_1)}$ we get $A'_j = e^{-g(\vec{x})} e^{-\alpha(\vec{x}, t_1)} \partial_j (e^{\alpha(\vec{x}, t_1)} e^{g(\vec{x})})$ and clearly $A'_j = 0$ if we take $e^{g(\vec{x})}$ to be the inverse of $e^{\alpha(\vec{x}, t_1)}$.

Note that even if there is winding in the vacuum at $t = t_1$ (such winding at one fixed time is discussed below (10.13)), one can still gauge it away by a time-independent gauge transformation, but then the winding at $t = t_2$ increases by just the same amount. This is as it should be, because the total winding is gauge-invariant.

We shall consider paths from σ_1 to σ_2 which at every time t have finite energy (finite integral $\int (E^2 + B^2) d^3x$). This means that the energy density for fixed t must tend to zero for $|\vec{x}| \rightarrow \infty$ (to make the integral $\int (E^2 + B^2) d^3x$ convergent), hence at large $|\vec{x}|$ the gauge fields become pure gauge

$$A_\mu(\vec{x}, t) \xrightarrow{|\vec{x}| \rightarrow \infty} e^{-\alpha(\vec{x}, t)} \partial_\mu e^{\alpha(\vec{x}, t)} . \quad (10.7)$$

But since $A_0(\vec{x}, t) = 0$, we see that $\alpha(\vec{x}, t)$ is independent of t . Because $\alpha(\vec{x}, t_1) = 0$ we obtain $\alpha(\vec{x}, t) = 0$ for all t and $|\vec{x}| \rightarrow \infty$. This means in particular that at t_2 for large $|\vec{x}|$ the gauge fields tend to zero

$$A_j(\vec{x}, t_2) \xrightarrow{|\vec{x}| \rightarrow \infty} 0 \quad (10.8)$$

The fact that for large $|\vec{x}|$ all A_j vanish allows us to compactify the 3-dimensional spacelike hypersurfaces at fixed t into spheres S_3 . The north pole of each sphere corresponds to all points with $|\vec{x}| = \infty$, and at this point on S_3 all A_j vanish. Thus, **all 3-spaces at fixed t compactify to a sphere S_3** . We summarize the results in a figure

$$A_j = e^{-\alpha} \partial_j e^{\alpha} \quad (10.9)$$

Everywhere on the boundary of this cylinder the gauge fields vanish, except at the disk at $t = t_2$, but there $A_0 = 0$ and A_j are only pure gauge.

We now return to Q . First of all, Q can be written as a total derivative, using the same algebra as in Euclidean space

$$Q = \frac{1}{8\pi^2} \epsilon^{\mu\nu\rho\sigma} \int \partial_\mu \text{tr} [A_\nu \partial_\rho A_\sigma + \frac{2}{3} A_\nu A_\rho A_\sigma] d^4x . \quad (10.10)$$

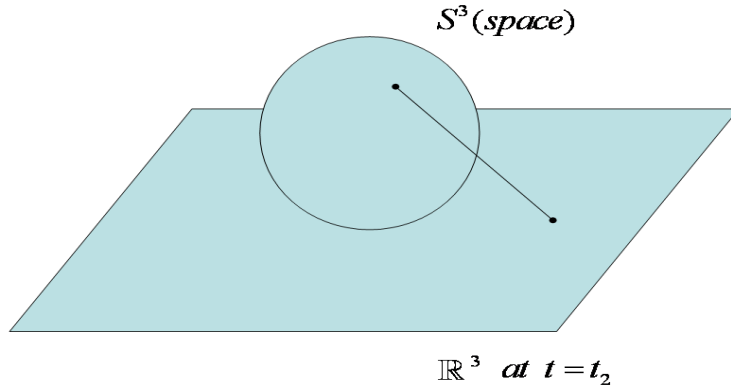
(we recall that $\text{tr}[A_\mu A_\nu A_\rho A_\sigma \epsilon^{\mu\nu\rho\sigma}] = 0$). Furthermore, since on the boundary $F_{\mu\nu} = 0$, we can replace $\partial_\rho A_\sigma$ by $-A_\rho A_\sigma$ in (10.10). We then find

$$Q = \frac{-1}{24\pi^2} \epsilon^{\mu\nu\rho\sigma} \int d\sigma_\mu \text{tr}[A_\nu A_\rho A_\sigma] \quad A_\nu = e^{-\alpha} \partial_\nu e^\alpha . \quad (10.11)$$

Since $A_0 = 0$, there is no contribution from the sides of the cylinder, and since $A_j = 0$ at the bottom, there is also no contribution from the bottom. Hence in the gauge we have chosen, all contributions to the winding come from the top of the cylinder:

$$Q = \frac{-1}{24\pi^2} \epsilon^{0ijk} \int \text{Tr}(e^{-\alpha} \partial_i e^\alpha)(e^{-\alpha} \partial_j e^\alpha)(e^{-\alpha} \partial_k e^\alpha) d^3x . \quad (10.12)$$

At the top of the cylinder the 3-space $t = t_2$ compactifies to a sphere S_3 (space). The map from this 3-sphere into the group $SU(2)$ is a map from one S_3 to another S_3 ⁴² because (i) we can always compactify the $\mathbf{R}^3$ with coordinates $\vec{x}$ to an S_3 and (ii) the gauge fields at $|\vec{x}| = \infty$ are equal (and vanish)



$\mathbb{R}^3$ at $t = t_2$

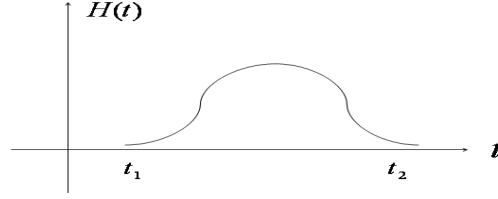
(10.13)

The maps S_3 (space) $\rightarrow S_3$ (group) in Minkowski space fall into equivalence classes with a winding number $k \in \mathbf{Z}$, just as the maps of instantons in Euclidean space give maps from S_3 (space) $\rightarrow S_3$ (group). In the latter case S_3 (space) is the boundary of all of $\mathbf{R}^4$ while here it is the compactification of the whole $\mathbf{R}^3$ at $t = t_2$. It follows that

$$Q = \pm k \quad k \in \mathbf{Z} . \quad (10.14)$$

⁴²The matrix elements of any 2×2 complex matrix can be written as $g = a_\mu \sigma^\mu$ with $\sigma^\mu = \{\vec{\sigma}, I\}$ and $\mu = 1, 2, 3, 0$. Unitarity requires that $g^\dagger = a_\mu^* \sigma^\mu$ equals g^{-1} , hence $g^\dagger g = \sum |a_\mu|^2 + (a_j^* a_k i \epsilon_{jkl} + a_0^* a_l + a_l^* a_0) \sigma^l = 1$. Hence $|a_0|^2 + |a_k|^2 = 1$ and the coefficients of σ_l must vanish. The determinant yields $\det g = a_0^2 - a_k^2$, and since also $|a_0|^2 + |a_k|^2 = 1$, requiring $\det g = 1$ leads to $a_k = \pm i |a_k|$ and $a_0 = \pm |a_0|$. Then we are left with $g = a_0 I + i a_k \sigma_k$ with real a_0 and a_k satisfying $a_0^2 + a_k^2 = 1$ which defines S_3 .

We now can draw a picture of the energy $H = \int \mathcal{H} d^3x$ at times t as we move from $t = t_1$ to $t = t_2$. Initially and at the end one has $H = 0$, but in between we must have $H > 0$ (note that $\mathcal{H} \geq 0$) for the following reason.



(10.15)

There are no paths possible which connect the vacuum at t_1 to the vacuum at t_2 which are solutions of the field equations because if $F_{\mu\nu} = 0$ on σ_1 (or σ_2) and the field equations are satisfied, one has $F_{\mu\nu} = 0$ everywhere⁴³. But, if $F_{\mu\nu}$ would vanish everywhere, $Q \sim \int E \cdot B d^4x$ would vanish, hence one could not change the winding number. The conclusion is that paths which go from one vacuum with winding number zero to another vacuum with nonvanishing winding number necessarily have positive energy at some intermediate times.

We are now ready to define a subset of paths which depend on one collective coordinate, and to which (we claim) we can restrict our attention. Consider first one given path corresponding to a fixed field configuration $A_j(\vec{x}, t)$. Instead of this single path, we consider the set of paths $A_j^{(\lambda)}(\vec{x}, t)$, as defined in (10.1). Each path is labeled by a different function $\lambda(t)$, and is defined by

$$A_j^{(\lambda)}(\vec{x}, t) = A_j(\vec{x}, \lambda(t)) . \quad (10.16)$$

As we already discussed, for $\lambda(t) = t$ we recover the original path, but for different $\lambda(t)$ we obtain paths which run through the same 3-dimensional configurations $\vec{A}(\vec{x}, t_1), \vec{A}(\vec{x}, t_2), \vec{A}(\vec{x}, t_3) \dots$

⁴³For the proof, note that if at t_1 one has $F_{\mu\nu} = 0$ and at all t one has $D^\mu F_{\mu\nu} = 0$, then $\partial^\mu F_{\mu j} = \partial_0 F_{0j} = 0$ at t_1 . Furthermore, the Bianchi identity $D_0 F_{ij} + D_i F_{j0} + D_j F_{0i} = 0$ yields $\partial_0 F_{ij} = 0$. Hence $\partial_0 F_{\mu\nu} = 0$ at t_1 . Also $\partial_j F_{\mu\nu} = 0$ because $F_{\mu\nu} = 0$ at $t = t_1$ for all x . Hence $\partial_\rho F_{\mu\nu} = 0$ at $t = t_1$ for all ρ, μ, ν . We can rewrite this as $(D_\rho F_{\mu\nu}) = 0$ at $t = t_1$. Next we repeat this analysis by noting that also $D_\mu(D_\rho F^{\mu\nu}) = 0$ at $t = t_1$, because $D_\mu(D_\rho F^{\mu\nu}) = [D_\mu, D_\rho]F^{\mu\nu} + D_\rho(D_\mu F^{\mu\nu})$ and $D_\mu F^{\mu\nu} = 0$ everywhere. This shows that $\partial_0(D_\rho F^{0j}) = 0$ at $t = t_1$. To also show that $\partial_0(D_\rho F_{ij}) = 0$ at $t = t_1$ we rewrite $\partial_0(D_\rho F_{ij}) = -\partial_0 D_i F_{j\rho} - \partial_0 D_j F_{\rho i}$ and then use $D_0(D_i F_{\mu\nu}) = [D_0, D_i]F_{\mu\nu} + D_i(D_0 F_{\mu\nu}) = 0$. In this way we get $\partial_0^n F_{\mu\nu} = 0$ for any n . Hence $F_{\mu\nu} = 0$ at all t .

at different speeds. For example if $\lambda(t)$ is constant for some time interval, the corresponding $\vec{A}(\vec{x}, t)$ do not change, but if $\lambda(t)$ changes rapidly, the sequence of $A(\vec{x}, t)$ is traversed rapidly.

Each path $A_j(\vec{x}, \lambda(t))$ should begin at $A_j(\vec{x}, t_1)$ and end at $A_j(\vec{x}, t_2)$, so we require $\lambda(t_1) = t_1$, and $\lambda(t_2) = t_2$, but between t_1 and t_2 the function $\lambda(t)$ is arbitrary. We shall later take $t_1 = -\infty$ and $t_2 = +\infty$, and then also require that $\lambda(t_1) = -\infty$ and $\lambda(t_2) = +\infty$. Given a path $A_j^{(\lambda)}(\vec{x}, t)$ we can compute the electric and magnetic fields

$$\begin{aligned} -E_j &= F_{0j} = \partial_0 A_j^{(\lambda)}(\vec{x}, t) = \frac{\partial A_j}{\partial \lambda}(\vec{x}, \lambda(t)) \dot{\lambda} \text{ because } A_0(\vec{x}, t) = 0 \\ B_i &= \frac{1}{2} \epsilon_{ijk} F_{jk} = \frac{1}{2} \epsilon_{ijk} (\partial_j A_k(\vec{x}, \lambda(t)) - \partial_k A_j(\vec{x}, \lambda(t))) - j \leftrightarrow k) . \end{aligned} \quad (10.17)$$

The Lagrangian $L = \int \mathcal{L} d^3x$ with $\mathcal{L} = \frac{1}{2g^2} \text{tr } F_{\mu\nu}^2 = \frac{-1}{g^2} \text{tr } (\vec{E}^2 - \vec{B}^2)$ can then be written as

$$\begin{aligned} L &= \frac{1}{2} m(\lambda) \dot{\lambda}^2 - V(\lambda) , \\ m(\lambda) &= \frac{-2}{g^2} \int \text{tr} \left(\frac{\partial \vec{A}}{\partial \lambda} \right)^2 d^3x \geq 0 , \\ V(\lambda) &= -\frac{1}{g^2} \int \text{tr } \vec{B}^2 d^3x \geq 0 \end{aligned} \quad (10.18)$$

The momentum conjugate to $\lambda(t)$ is $p(\lambda) = \frac{\partial}{\partial \dot{\lambda}} L = m(\lambda) \dot{\lambda}$. Hence

$$H = \frac{(p(\lambda))^2}{2m(\lambda)} + V(\lambda) . \quad (10.19)$$

For a given path $A_j^{(\lambda)}(\vec{x}, t)$ one can plot H as a function of t , and one finds then the profile in figure (10.15).

We have thus isolated a class of paths $A^{(\lambda)}(\vec{x}, t)$ which depends on one collective coordinate $\lambda(t)$. For one given $A(\vec{x}, t)$, this still yields an infinite set of paths, but all these paths run through the same set of 3-configurations $A_j(\vec{x}, t_1), A_j(\vec{x}, t_2), \dots$. These are, of course, infinitely many other collective coordinates which describe a general path $A_j(x, t)$, but the idea is that $\lambda(t)$ is the relevant coordinate to describe tunnelling, while the other collective coordinates describe variations away from the paths $A_j^{(\lambda)}(\vec{x}, t)$ which give only small corrections to the results obtained from $\lambda(t)$. It is, of course, difficult to prove this assertion; one could begin with two collective coordinates as a start, but even this would lead to a complicated analysis.

The action for $\lambda(t)$ in (10.18) can be viewed as the action for one point particle. This particle feels the potential barrier $V(\lambda)$, and to go from the vacuum at $t = t_1$ with $V(\lambda) =$

$m(\lambda) = 0$ to the vacuum at t_2 with also $V(\lambda) = m(\lambda) = 0$, we need tunnelling. The tunnelling rate R in quantum mechanics is proportional to e^{-2R} where

$$R = \int_{\lambda_1}^{\lambda_2} d\lambda \sqrt{2m(\lambda)(V(\lambda) - E)} , \quad (10.20)$$

with $\lambda(t_1) \equiv \lambda_1 = t_1$, $\lambda(t_2) \equiv \lambda_2 = t_2$ and $V(\lambda(t_1)) = V(\lambda(t_2)) = 0$ and $m(\lambda(t_1)) = m(\lambda(t_2)) = 0$. We also set $E = 0$ because we consider tunnelling from one vacuum (with $E = 0$) to another.

There is, of course, an important difference with ordinary quantum mechanics. The point particle $\lambda(t)$ feels a potential $V(\lambda)$, but both are derived from the same object, the fields $A_j(x, \lambda(t))$. In addition the mass is here “position”-dependent, $m = m(\lambda)$. One can show that in quantum mechanics the formula for R also holds if the mass $m(\lambda)$ depends on the point particle $\lambda(t)$. The crucial step is now to pose the question: **for which set of paths $\vec{A}(\vec{x}, \lambda(t))$ is the tunnelling rate maximal?** The tunnelling rate for the quantum mechanical particle $\lambda(t)$ can be described by Minkowski path integrals, so we ask: for which $\vec{A}(\vec{x}, t)$ is there least destructive interference of the associated paths $A(\vec{x}, \lambda(t))$ in the path integral? Clearly, $V(\lambda)$ should be as small as possible, but it cannot be too small because it must produce winding.

The tunnelling rate is e^{-2R} where according to (10.20)

$$\begin{aligned} R &= \int_{\lambda_1}^{\lambda_2} d\lambda \, 2 \left[\left(\frac{1}{g^2} \int \text{tr} \left(\frac{\partial \vec{A}}{\partial \lambda} \right)^2 d^3x \right) \left(\frac{1}{g^2} \int \text{tr} \vec{B}^2 d^3x \right) \right]^{1/2} \\ &= \frac{2}{g^2} \int_{t_1}^{t_2} dt [(\text{tr} \int \vec{E}^2 d^3x)(\text{tr} \int \vec{B}^2 d^3x)]^{1/2} . \end{aligned} \quad (10.21)$$

We replaced $d\lambda$ by $dt\dot{\lambda}$ and brought $\dot{\lambda}$ inside the square root. The fields $\vec{E}$ and $\vec{B}$ still depend on $\lambda(t)$. Since $\text{tr} \int \vec{a}(\vec{x})\vec{b}(\vec{x})d^3x$ is an inner product, while $\int \text{tr} \vec{E} \cdot \vec{B}$ is proportional to the winding number according to (10.3), we have the triangle inequality

$$R \geq \frac{2}{g^2} \left| \int_{t_1}^{t_2} (\text{tr} \vec{E} \cdot \vec{B}) d^4x \right| = \frac{8\pi^2}{g^2} |Q| . \quad (10.22)$$

Hence the tunnelling amplitude is bounded from above by

$$e^{-R} \leq -e^{-\frac{8\pi^2}{g^2}|Q|} . \quad (10.23)$$

The inequality is saturated when $\vec{E}$ is parallel to $\vec{B}$ at each vector $\vec{x}$ and at each time t : $\vec{E}(\vec{x}, t) = \alpha(t)\vec{B}(\vec{x}, t)$. The claim is that among all paths with the same Q , the paths with the smallest R are the paths with $\vec{E}$ parallel to $\vec{B}$.

Let us discuss the meaning of this result. Paths which interpolate between vacua with different winding number must produce electric and magnetic fields $\vec{E}$ and $\vec{B}$ in between at finite $\vec{x}$ and t which cannot be too small, namely $|\int (E_j^a B_j^a) d^4x|$ should be equal to $8\pi^2|Q|$. On the other hand, the tunnelling rate is proportional to the length of E^a times the length of B^a , so to make the tunnelling rate as large as possible, the product of these lengths should be as small as possible. One could set up a variational problem for R under the constraint that $\int \text{tr } \vec{E} \cdot \vec{B} d^4x$ be equal to $4\pi^2 Q$, but we shall not work this out.

The bound is reached, namely the tunnelling rate is maximal, when the set of paths $A_j(\vec{x}, \lambda(t))$ produces parallel electric and magnetic fields

$$\vec{E}(\vec{x}, \lambda(t)) = \alpha(t) \vec{B}(\vec{x}, \lambda(t)) . \quad (10.24)$$

Of course, $\alpha(t)$ can also be viewed as a function of $\lambda(t)$ because $\lambda(t)$ is just another parametrization of the time interval. Note that this condition does not change if one changes the parametrization from $\lambda(t)$ to another function $\lambda'(t)$, because under such reparametrizations $\vec{E}$ scales by a constant factor $\partial\lambda'/\partial\lambda$, which cancels the Jacobian in (10.20) for this change of integration variables. We use this scaling property to select a particular $\lambda_0(t)$ such that $\vec{E}(\vec{x}, \lambda_0(t)) = \pm \vec{B}(\vec{x}, \lambda_0(t))$. The property of $\vec{E}$ and $\vec{B}$ being parallel is also a gauge-invariant property, and $\mathcal{L}$ and R are of course gauge-invariant. So, our characterization of paths with maximal tunnelling rate is gauge-invariant, as it should be. Thus the use of temporal gauge did not restrict the generality of the results.

We now can establish the connection between tunnelling and instantons. The fields for which $\vec{E}$ and $\vec{B}$ in Minkowski space are parallel are closely connected to instantons in Euclidean space. Namely, among the class of paths $\vec{A}(\vec{x}, \lambda(t))$ parametrized by $\lambda(t)$, there is the path $\vec{E}(\vec{x}, \lambda_0(t)) = \vec{B}(\vec{x}, \lambda_0(t))$ (and another path with another $\lambda'_0(t)$ such that $\vec{E}(\vec{x}, \lambda'_0(t)) = -\vec{B}(\vec{x}, \lambda'_0(t))$). If we then define Euclidean gauge fields $A_\mu^E(x, t)$ by $A_j^E(\vec{x}, t) = A_j(\vec{x}, \lambda_0(t))$ and $A_4^E(\vec{x}, t) = A_0(\vec{x}, \lambda(t)) \frac{d\lambda}{dt}$ then this $A_\mu^E(\vec{x}, t)$ is self dual. The parameter t is Minkowski time, but in the expressions for $A_j^E(\vec{x}, t)$ we should interpret t as the Euclidean time.

Summarizing: the most probable tunnelling paths are given by the set of paths $A_j(\vec{x}, \lambda(t))$ with parallel $\vec{E}$ and $\vec{B}$ fields. A given class of paths with $\vec{E}$ parallel to $\vec{B}$ contains one path which, when viewed as a configuration in Euclidean space, is an instanton. Conversely, given an instanton $A_\mu^E(\vec{x}, t)$ in Euclidean space, one can construct a corresponding set of paths $A_\mu^M(x, \lambda(t))$ in Minkowski space by setting

$$\begin{aligned} A_j^{M,(\lambda)}(x, t) &= A_j^E(\vec{x}, \lambda(t)) \\ A_0^{(M,(\lambda))}(x, t) &= A_4^E(\vec{x}, \lambda(t)) \dot{\lambda} . \end{aligned} \quad (10.25)$$

As an example we take the $Q = -1$ anti-instanton solution in regular gauge, $A_\mu =$

$-\sigma_{\mu\nu}x^\nu/(x^2 + \rho^2)$, see (3.41), which yields the following set of paths in Minkowski space

$$\left. \begin{aligned} A_0^{(\lambda)}(\vec{x}, t) &= \frac{-i\vec{x}\cdot\vec{\sigma}}{\vec{x}^2 + \lambda(t)^2 + \rho^2} \dot{\lambda}(t) \\ \vec{A}^{(\lambda)}(\vec{x}, t) &= \frac{i\lambda(t)\vec{\sigma} - i\vec{x}\times\vec{\sigma}}{\vec{x}^2 + \lambda(t)^2 + \rho^2} \end{aligned} \right\} \begin{aligned} \lambda(t \rightarrow -\infty) &= -\infty \\ \lambda(t \rightarrow +\infty) &= +\infty . \end{aligned} \quad (10.26)$$

We are clearly not in the temporal gauge, but since our results are gauge-invariant, it does not matter which gauge we use. We still have $A_\mu \rightarrow 0$ at large $|\vec{x}|$, so that we still have the notion of winding as a map from S_3 (space) into S_3 (group) at each time.

Straightforward calculation yields for the curvatures in Minkowski space

$$\begin{aligned} F_{01} &= \partial_0 A_1 - \partial_1 A_0 + [A_0, A_1] = \frac{2i\rho^2\sigma_1}{(\vec{x}^2 + \lambda^2 + \rho^2)^2} \dot{\lambda} , \\ F_{23} &= \partial_2 A_3 - \partial_3 A_2 + [A_2, A_3] = \frac{2i\rho^2\sigma_1}{(\vec{x}^2 + \lambda^2 + \rho^2)^2} . \end{aligned} \quad (10.27)$$

Hence

$$\vec{E} = \frac{-2i\rho^2\vec{\sigma}}{(\vec{x}^2 + \lambda^2 + \rho^2)^2} \dot{\lambda} ; \quad \vec{B} = \frac{2i\rho^2\vec{\sigma}}{(\vec{x}^2 + \lambda^2 + \rho^2)^2} , \quad (10.28)$$

which depend on $x^2 = \vec{x}^2 + \lambda(t)^2$ (**not** on $\vec{x}^2 - t^2$). Hence, $\vec{E}$ is indeed parallel to $\vec{B}$ (in fact, anti-parallel).

The winding number Q can be written in two ways

$$\begin{aligned} Q &= \frac{-1}{4\pi^2} \int_{-\infty}^{\infty} [\text{tr } \vec{E} \cdot \vec{B} d^3x] dt \\ &= \frac{-1}{24\pi^2} \epsilon^{\mu\nu\rho\sigma} \int \partial_\mu \text{tr} [A_\nu A_\rho A_\sigma] d^4x . \end{aligned} \quad (10.29)$$

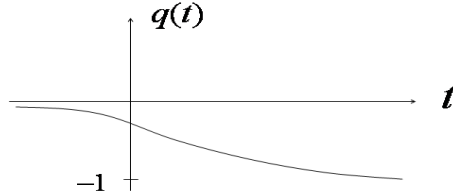
In the latter expression Q receives only a contribution from the boundary,⁴⁴ but in the former expression we compute Q by integrating over all space and time. It is then natural to define a t -dependent function by integrating only up to a time t

$$\begin{aligned} q(t) &= \frac{-1}{4\pi^2} \int_{-\infty}^t \left[\int \text{tr } \vec{E} \cdot \vec{B} d^3x \right] \\ &= -\frac{1}{4\pi^2} \int_{-\infty}^{\lambda} d\lambda \int d^3x \frac{24\rho^4}{[\vec{x}^2 + \lambda^2 + \rho^2]^4} \\ &= -\frac{3}{4} \int_{-\infty}^{\lambda} \frac{\rho^4 d\lambda}{(\lambda^2 + \rho^2)^{5/2}} \end{aligned}$$

⁴⁴For example, the contribution to Q from the surface at $t = t_1$ is proportional to $\int \frac{t(\vec{x}^2 + t^2) d^3x}{(t^2 + \vec{x}^2 + \rho^2)^3}$ which is nonvanishing. On the other hand, the contribution to Q from the sides of the cylinder converges for large $|t|$.

$$\begin{aligned}
&= -\frac{3}{4} \int_{-\infty}^{\lambda/\rho} \frac{dy}{(y^2 + 1)^{5/2}} \\
&= -\frac{3}{4} \left(t - \frac{1}{3} t^3 \right) \Big|_{-1}^x \quad \text{with } x = \frac{\lambda}{\sqrt{\lambda^2 + \rho^2}} .
\end{aligned} \tag{10.30}$$

Clearly, $q(t)$ is gauge-invariant and has the following form



(10.31)

It only receives contributions from regions where $\vec{E}$ and $\vec{B}$ are nonvanishing, hence where A_μ^a is not pure gauge.

To obtain the action for $\lambda(t)$ in this example we evaluate

$$\mathcal{L} = -\frac{1}{g^2} \text{tr} (\vec{E}^2 - \vec{B}^2) = \frac{24}{g^2} \rho^4 \left[\frac{\dot{\lambda}^2}{(x^2 + \rho^2)^4} - \frac{1}{(x^2 + \rho^2)^4} \right] . \tag{10.32}$$

Doing the space integral we obtain

$$L = \frac{1}{2} m(\lambda) \dot{\lambda}^2 - V(\lambda) = \frac{3\pi^2 \rho^4}{g^2 (\lambda^2 + \rho^2)^{5/2}} (\dot{\lambda}^2 - 1) , \tag{10.33}$$

where we used

$$\int \frac{d^3x}{(\vec{x}^2 + \lambda^2 + \rho^2)^4} = \frac{4\pi}{(\lambda^2 + \rho^2)^{5/2}} \frac{1}{2} \int_{-\infty}^{\infty} \frac{y^2 dy}{(y^2 + 1)^4} = \frac{4\pi^2}{32} \frac{1}{(\lambda^2 + \rho^2)^{5/2}} . \tag{10.34}$$

In this example, we were dealing with a gauge with $A_0 \neq 0$. We can map to a gauge in which $A_0 = 0$ by a suitable large gauge transformation

$$\begin{aligned}
A'_\mu &= U^{-1} (\partial_\mu + A_\mu) U \\
U &= \exp \left[\frac{i \vec{x} \cdot \vec{\sigma}}{\sqrt{\vec{x}^2 + \rho^2}} \text{arctg} \frac{\lambda(t)}{\sqrt{\vec{x}^2 + \rho^2}} \right] .
\end{aligned} \tag{10.35}$$

Indeed, using the expression for A_0 in (10.26)

$$A_0 = \frac{-i\vec{x} \cdot \vec{\sigma}}{\vec{x}^2 + \lambda(t)^2 + \rho^2} \dot{\lambda}(t) , \quad (10.36)$$

one finds that $A'_0 = U^{-1}(\partial_t + A_0)U = U^{-1}\partial_t U + A_0$ vanishes

$$A'_0 = \frac{i\vec{x} \cdot \vec{\sigma}}{\sqrt{\vec{x}^2 + \rho^2}} \frac{1}{1 + \frac{\lambda(t)^2}{\vec{x}^2 + \rho^2}} \frac{\dot{\lambda}(t)}{\sqrt{\vec{x}^2 + \rho^2}} + A_0 = 0 , \quad (10.37)$$

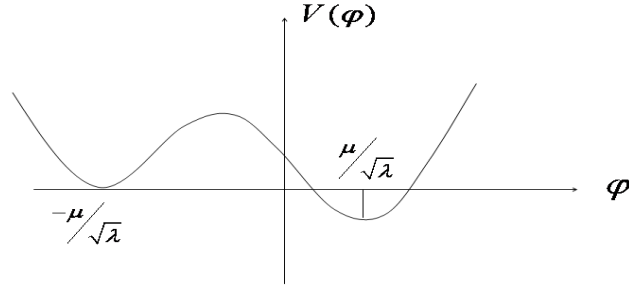
where we used that A_0 commutes with U . Of course, Q is gauge invariant because it can be written as a trace over $\vec{E} \cdot \vec{B}$ but it is instructive to see what happens if one writes Q as a surface integral and makes a gauge transformation with U . On the boundary of Minkowski space the $A_\mu = V^{-1}\partial_\mu V$ transform into $(VU)^{-1}\partial_\mu VU$ and the winding number of VU is the sum of the winding numbers of V and U . However, U is connected to the identity element: $U \equiv \exp \alpha [\frac{i\vec{x} \cdot \vec{\sigma}}{\sqrt{\vec{x}^2 + \rho^2}} \arctg \frac{\lambda(t)}{\sqrt{\vec{x}^2 + \rho^2}}]$ traces an orbit as α runs from 0 to 1 which begins at the identity element and ends at U . Thus U does not produce any winding, and thus the answer for Q from the total derivative is the same, whether one uses a gauge in which A_0 vanishes or a gauge in which A_0 is nonvanishing. Note, however, that when $A_0 \neq 0$ one gets contributions from the timelike part of the boundary of the spacetime cylinder.

11 False vacua and phase transitions

In spontaneously broken gauge theories, the potential has a local maximum and an absolute minimum. These extrema form a metastable and a stable vacuum, respectively. If a system is in the metastable vacuum at all points in spacetime, it could at some point and at some time, say $\vec{x} = 0$ and $t = 0$, make a quantum fluctuation to the stable vacuum. This transition costs energy, but if the region around x (“the bubble”) is large enough, the energy needed for creation of a bubble (this energy is located in the boundaries of the bubble) is less than the energy gained by tunnelling to the lower vacuum (this energy is liberated in the volume of the bubble), and then the bubble will rapidly expand. In fact, since the rate of energy production increases the larger the bubble, the bubble will spread through space, with accelerating speed, converting the false vacuum to a true vacuum. As an application of this process one may consider the universe just after the Big Bang; at high temperature the universe is in the symmetric vacuum, but as cooling due to expansion sets in the potential develops a lower (true) vacuum, and if for some reason the universe remains stuck in the false vacuum, one can study the decay of the universe towards the true (asymmetric) vacuum. We shall consider another example: the perturbed double-well potential, with two classically stable minima, but one minimum (the true vacuum) below

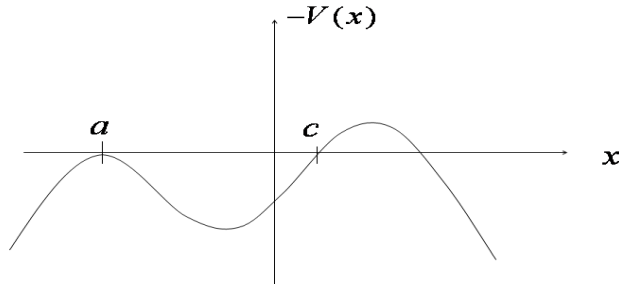
the other minimum (the false vacuum). We shall study the decay of the false vacuum in this model into the true vacuum [79, 80]. We follow [81].

As a preliminary to the calculation of the phase transition in field theory, we first revert to quantum mechanics and study the double-well. Let us pretend that we do not know that there are big differences between the double-well potential and the following potential.



(11.1)

We can then repeat the calculation of the nonperturbative corrections to the energy of the ground state. Already at this point it is clear that we should not blindly repeat all steps because previously we were dealing with two perturbatively degenerate vacua, and the kink-instantons provided the energy shift between both vacua. In the present case, the degeneracy is already broken at the classical level. Proceeding nevertheless we find a classical solution of the Euclidean equation $-\frac{\partial^2 x}{\partial t^2} + \frac{\partial V}{\partial x} = 0$ describing a point particle $x(t)$ in the inverted potential and use path integral methods.



(11.2)

The particle starts at $t = -\infty$ in the point $x = a$, rolls to the point $x = c$, “bounces” at time $t = X$, and ends up at $t = +\infty$ at the same point $x = a$. Clearly, X is the collective coordinate for this classical solution $x_{cl}(t)$. We then get for the “one-bounce solution”

$$\begin{aligned} T_{00} &\equiv \langle x = a | e^{-\frac{1}{\hbar} H \tau_0} | x = a \rangle = e^{\frac{1}{\hbar} S_{cl}} \tau_0 \sqrt{-S_{cl}} I_0 , \\ I_0 &= \mathcal{N} \int_{n.z.} dq(\tau) e^{\frac{1}{\hbar} S_E^{(2)}} \quad \text{with } q(\pm \tau_{0/2}) = 0 , \end{aligned} \quad (11.3)$$

where we used the Faddeev-Popov trick, and “n.z.” indicates that the path integral is over the solutions of the field equation for the fluctuations about $x_{cl}(t)$ in the space orthogonal to the almost-zero mode. Assuming again that I_0 can be written as a factor K times the path integral of the harmonic oscillator we get

$$I_0 = K \sqrt{\frac{\omega}{\pi \hbar}} e^{-\frac{1}{2} \omega \tau_0} ; \quad K = \sqrt{\frac{\det(-\partial_t^2 + \omega^2)}{\det'(-\partial_t^2 + V''(x_{cl}))}} \quad (11.4)$$

Continuing without further thought we would sum over multi-bounces and obtain

$$\begin{aligned} T_{00} &= \sqrt{\frac{\omega}{\pi \hbar}} e^{-\frac{1}{2} \omega \tau_0} \sum_{n=0}^{\infty} \frac{(\sqrt{-S_{cl}} \tau_0 K e^{\frac{1}{\hbar} S_{cl}})^n}{n!} \\ &= \frac{\omega}{\pi \hbar} e^{-\frac{1}{2} \omega \tau_0} \exp(K \tau_0 e^{\frac{1}{\hbar} S_{cl}}) . \end{aligned} \quad (11.5)$$

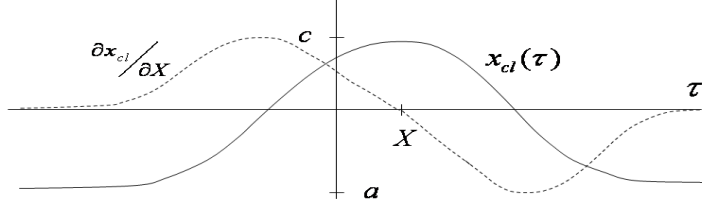
Using the same arguments as used before for the unperturbed double-well potential, we would conclude that the ground state energy is given by

$$E_0 = \frac{1}{2} \hbar \omega - \hbar K e^{\frac{1}{\hbar} S_{cl}} . \quad (11.6)$$

However, at this point we note that there are problems with this result

- (i) first a small problem: the nonperturbative correction is exponentially suppressed, hence it should be neglected compared to the perturbative correction.
- (ii) a more serious problem (actually a virtue, as we shall see) is that K has a negative eigenvalue. This is easy to prove: $\frac{d}{dX} x_{cl}(t - X)$ is the zero mode fluctuation. It has a mode because x_{cl} bounces: unlike the kink, $x_{cl}(\tau)$ moves first forward and then backwards,

yielding a kind of kink-antikink solution.



(11.7)

Hence there exists one mode for the fluctuations with lower eigenvalue and without a node, and since $\frac{\partial}{\partial X}x_{cl}(x-X)$ has zero eigenvalue, there exists an eigenfunction for the fluctuation with negative eigenvalue. Thus the nonperturbative correction is imaginary, reflecting the fact that the perturbative ground state near $x = 0$ is nonperturbatively unstable

$$Im E_0 = \hbar |K| e^{\frac{1}{\hbar} S_{cl}} \equiv \Gamma/2 . \quad (11.8)$$

So, instantons (or rather bounces, still solutions of the classical field equations with finite action) yield in this case the width Γ of the unstable state.

Having seen that in quantum mechanics the path integral approach to nonperturbative corrections to the vacuum energy leads to the correct result that the ground state is unstable, we now return to the problem of phase transitions.

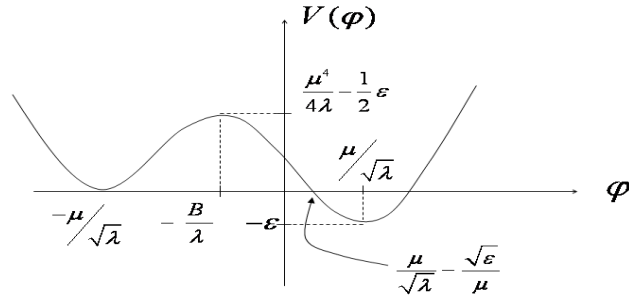
As a toy model for studying such decays we need a system with at least one space coordinate because bubbles have a finite extension in space. The simplest choice is a $1+1$ dimensional field theory. We choose the double-well potential with an extra term to destroy the degeneracy between both minima. Since the double-well potential is symmetric under $\varphi \rightarrow -\varphi$, the extra term should be antisymmetric, and if it is to be a small perturbation compared to the leading $\lambda\varphi^4$ term, we need either a term linear in φ or cubic in φ , or both. It simplifies the mathematics if we keep the local minima of the perturbed potential at the same place as the minima of the unperturbed potential, namely at $\varphi = \pm\mu/\sqrt{\lambda}$. We are then led to the following model

$$\mathcal{L} = \frac{1}{2}\dot{\varphi}^2 - \frac{1}{2}(\varphi')^2 - \frac{\lambda}{4}\left(\varphi^2 - \frac{\mu^2}{\lambda}\right)^2 - B\left(\frac{1}{3}\varphi^3 - \frac{\mu^2}{\lambda}\varphi\right) + \frac{2}{3}B\left(\frac{\mu}{\sqrt{\lambda}}\right)^3 , \quad (11.9)$$

where we take B small and positive. For constant φ , the solutions of the classical field equations occur at

$$\frac{\partial V}{\partial \varphi} = \lambda \varphi \left(\varphi^2 - \frac{\mu^2}{\lambda} \right) + B \left(\varphi^2 - \frac{\mu^2}{\lambda} \right) = 0 , \quad (11.10)$$

and from this result it is clear that the values $\varphi = \pm \mu/\sqrt{\lambda}$ are indeed extrema. The potential has the following form



(11.11)

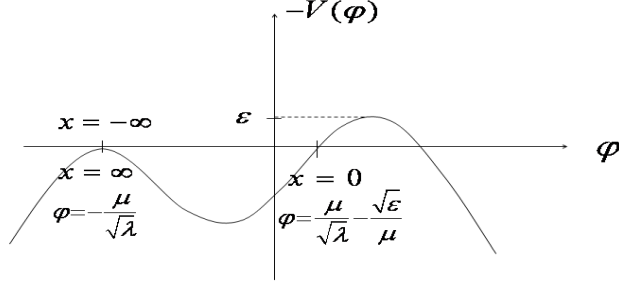
It vanishes at $\varphi = -\mu/\sqrt{\lambda}$ because we added the constant $\frac{2}{3}B(\mu/\sqrt{\lambda})^3$, but at $\varphi = \mu/\sqrt{\lambda}$ it is negative. Thus $\varphi = -\mu/\sqrt{\lambda}$ is the unstable vacuum and $\varphi = \mu/\sqrt{\lambda}$ is the stable vacuum. The value of the potential at the stable minimum is

$$V\left(\varphi = \mu/\sqrt{\lambda}\right) = -\epsilon = -\frac{4}{3}B\left(\mu/\sqrt{\lambda}\right)^3 . \quad (11.12)$$

There is a relative maximum a bit below the maximum of the symmetric potential $V(B = 0, \varphi)$ at $\varphi = 0$; for small B it occurs at $\varphi \simeq -B/\lambda$ and its value is $\frac{1}{4}\mu^4/\lambda - \frac{1}{2}\epsilon + \mathcal{O}(B^2)$. These results are intuitively clear: if one pulls φ down at $\mu/\sqrt{\lambda}$ by an amount ϵ , then the maximum at $\varphi = 0$ is pulled down half as much, and moves of course a bit to the left.

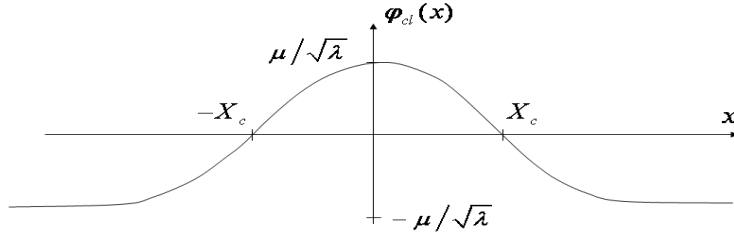
In addition to the three solutions of the classical field equations with constant φ ($\varphi = -\mu/\sqrt{\lambda}$, $\varphi = \mu/\sqrt{\lambda}$, and $\varphi \sim -B/\lambda$), there is an exact kink-antikink solution. This is clear

by inspection of the inverted potential



(11.13)

A ball at rest at $\varphi = -\mu/\sqrt{\lambda}$ at $x = -\infty$ starts rolling down to the hill and up the other hill; it reaches the point where $V(\varphi) = 0$ at $x = 0$ and then returns and comes to rest at $\varphi = -\mu/\sqrt{\lambda}$ at $x = +\infty$. The classical solution $\varphi_{cl}(x)$ is thus a soliton of the following form



(11.14)

We approximate $\varphi_{cl}(x)$ by the following expression

$$\varphi_{cl}(x) = \frac{\mu}{\sqrt{\lambda}} \left[\tanh \left(\frac{m}{2}(x + X_c) \right) - \tanh \left(\frac{m}{2}(x - X_c) \right) - 1 \right] . \quad (11.15)$$

This is a static soliton in $1 + 1$ dimensions, which can also be viewed as an instanton in x -space. (In the quantum mechanical models we considered previously, we dealt with instantons in Euclidean time). Near $x = -X_c$ the antikink is exponentially suppressed

and the mass of the kink is M . Between the kink and antikink φ is equal to $\mu/\sqrt{\lambda}$ (up to exponentially suppressed corrections), and near $x = X_c$ we have an antikink with mass M . For large x we find the correct asymptotic value $\varphi_d(x \rightarrow \pm\infty) = -\mu/\sqrt{\lambda}$. We fix the value of X_c such that the total energy of $\varphi_d(x)$ (which is the energy of the ball rolling up and down the hills in (11.13)) vanishes

$$E = 2M - 2\epsilon X_c = 0 , \quad (11.16)$$

where $M = \frac{m^3}{3\lambda}$ is the classical mass of a single kink. Hence, the separation between the kink and antikink is given by $2X$ with $X = M/\epsilon$.

The exact solution begins at $V = 0$, climbs the hill, and comes down on the other side where it reaches the value $V = 0$, and then it returns, climbing the hill once more, and ending at $V = 0$. The approximate solution comes down to $V = -\epsilon$ after climbing the hill, but it has more energy in the kink (and antikink) region, such that in both cases the total energy is zero.

We now compute the transition amplitude from the unstable vacuum $\varphi = -\mu/\sqrt{\lambda}$ to the kink-antikink solution (the bubble). Once a bubble has formed, it will rapidly grow (the kink and antikink move increasingly fast away from each other, i.e., X exponentially increase).

This is a tunnelling process because classically it is forbidden but quantum mechanically allowed. If the field φ at $x = 0$ starts making a transition from the metastable vacuum to the stable vacuum, it must first climb the potential barrier, but when it comes down in the true vacuum energy density $-\epsilon$ is gained. However, as we already mentioned, it takes energy to distort the field in order to go from one vacuum to another; this is just the energy (mass) of a kink and of an antikink. These energies are located at the boundary of the bubble (around the centers of the kink and the antikink). Once in a while there occurs a quantum mechanical transition to a bubble which is large enough that $\epsilon 2X$ is larger than $2M$; in that case the bubble does not collapse but grows increasingly rapidly.

Note that we do not tunnel from the state $\varphi(x) = -\mu/\sqrt{\lambda}$ to the state $\varphi(x) = \mu/\sqrt{\lambda}$ because the energy difference of these states is infinite (namely ϵ times the volume of x -space, so $2L\epsilon$ with $L \rightarrow \infty$). When we discussed the unperturbed kink, the vacua $\varphi = \pm\mu/\sqrt{\lambda}$ were exactly degenerate, and in such cases the true vacuum is a linear combination of these vacua which can be determined by tunnelling from one vacuum to another.

The intermediate configuration with the kink and antikink moving away from each other can be described by Lorentz boosting the kink to a velocity $-v$ and the antikink to a velocity $+v$

$$\varphi_d(x, t) = \frac{\mu}{\sqrt{\lambda}} \left[\tanh \frac{m}{2} \left(\frac{x + X_c + vt}{\sqrt{1 - v^2}} \right) - \tanh \frac{m}{2} \left(\frac{x - X_c - vt}{\sqrt{1 - v^2}} \right) - 1 \right] . \quad (11.17)$$

For constant $\dot{X}$ the boost of the kink is again a solution because the field equation use relativistically invariant. However, since $\dot{X}$ itself is expected to change with time, we denote $X + \dot{X}t$ by $\lambda(t)$ and obtain then

$$\varphi_{cl}(x, t) = \frac{\mu}{\sqrt{\lambda}} \left[\tanh \frac{m}{2} \left(\frac{x + \lambda(t)}{\sqrt{1 - \dot{\lambda}^2}} \right) - \tanh \frac{m}{2} \left(\frac{x - \lambda(t)}{\sqrt{1 - \dot{\lambda}^2}} \right) - 1 \right]. \quad (11.18)$$

The distance between the kink and antikink is now $2\lambda(t)$. The Lagrangian for this approximate solution is obtained by substituting φ_{cl} into the action. The calculation of the first two terms is straightforward. Taking twice the result for a single kink yields

$$\begin{aligned} \int_{-\infty}^{\infty} \left[\frac{1}{2} \dot{\varphi}^2 - \frac{1}{2} \left(\frac{\partial \varphi}{\partial x} \right)^2 \right] dx &= 2 \int_{-\infty}^{\infty} dx \frac{1}{2} \frac{1}{\cosh^4 \left(\frac{m}{2} \frac{x + \lambda}{\sqrt{1 - \dot{\lambda}^2}} \right)} \\ &\left[\frac{\mu^2 m^2}{\lambda} \frac{1}{4} \left(\frac{\dot{\lambda}}{\sqrt{1 - \dot{\lambda}^2}} + \frac{\dot{\lambda} \ddot{\lambda} (x + \lambda)}{(1 - \dot{\lambda}^2)^{3/2}} \right)^2 - \frac{\mu^2 m^2}{\lambda} \frac{1}{4} \frac{1}{1 - \dot{\lambda}^2} \right]. \end{aligned} \quad (11.19)$$

The calculation of the contribution from the nonderivative terms splits into two parts: from the region between the kink and antikink we obtain a term $\epsilon 2\lambda$, while from each of the two walls we find a term $\frac{1}{2} M \sqrt{1 - \dot{\lambda}^2}$ as we now explain. Around $x = -\lambda$ and $x = +\lambda$, the integral $\int V(\varphi) dx$ with $\varphi = \frac{\mu}{\sqrt{\lambda}} \tanh \frac{m}{2} \frac{x + \lambda}{\sqrt{1 - \dot{\lambda}^2}}$ can be evaluated as follows. The integral $\int_{-\infty}^{\infty} \frac{1}{2} U^2(\varphi) dx$ with $\varphi = \frac{\mu}{\sqrt{\lambda}} \tanh \frac{m}{2} (x^* + \lambda^*)$ with $x^* = \frac{x}{\sqrt{1 - \dot{\lambda}^2}}$ and $\lambda^* = \frac{\lambda}{\sqrt{1 - \dot{\lambda}^2}}$ is equal to $(\int \frac{1}{2} U^2(\varphi(x^*)) dx^*) \sqrt{1 - \dot{\lambda}^2}$. From equipartition of energy for a static kink we know that the integral $\int \frac{1}{2} U^2(\varphi(y)) dy$ equals $\frac{1}{2} M$. Thus $\int_{\text{around kink}} (\frac{1}{2} U^2)(\varphi_{cl}) dx = \frac{1}{2} M \sqrt{1 - \dot{\lambda}^2}$.

Hence, neglecting term with $\ddot{\lambda}$, we find

$$\begin{aligned} L &= -\frac{m^4}{8\lambda} \left(\frac{2}{m} \sqrt{1 - \dot{\lambda}^2} \int_{-\infty}^{\infty} \frac{dy}{\cosh^4 y} \right) + \epsilon 2\lambda - M \sqrt{1 - \dot{\lambda}^2} \\ &= -2M \sqrt{1 - \dot{\lambda}^2} + \epsilon 2\lambda. \end{aligned} \quad (11.20)$$

The Hamiltonian follows from $p = \frac{\partial L}{\partial \dot{\lambda}} = \frac{2M\dot{\lambda}}{\sqrt{1 - \dot{\lambda}^2}}$ and reads

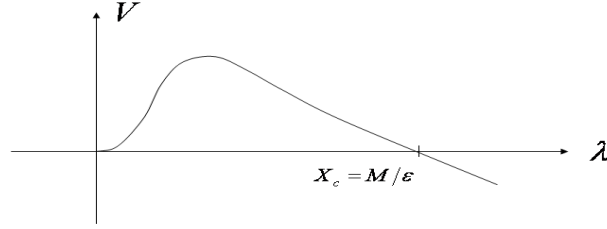
$$H = \frac{2M}{\sqrt{1 - \dot{\lambda}^2}} - \epsilon 2\lambda = \sqrt{p^2 + 4M^2} - \epsilon 2\lambda. \quad (11.21)$$

We can split H into a kinetic term K and a potential term V

$$K = \sqrt{p^2 + 4M^2} - 2M = \frac{1}{2} p^2 / M + \mathcal{O}(p^4),$$

$$V(\lambda) = 2M - \epsilon 2\lambda . \quad (11.22)$$

This formula for $V(\lambda)$ is valid when the bubble is reasonably large: when λ is larger than the kink size (when the bubble is larger than the thickness of its walls). For smaller x we expect that $V(\lambda)$ rises from 0 till a maximum value when the bubble is formed, and then decreases as the bubble gets larger



$$(11.23)$$

The value X_c corresponds to the classical solution, with energy $E = 0$ and constant X , corresponding to the ball rolling in the inverted potential. For this case, $p = 0$. Quantum fluctuations with $X < X_c$ produce only bubbles which collapse since their potential energy is positive, but bubbles with $X = X_c$ are metastable (they have constant $X = X_c$ so $p = 0$), while for $X > X_c$ the bubble expands.

We now treat H as the Hamiltonian of a point particle which sees the potential $V(\lambda)$ and has energy zero. We find with the WKB approximation for the tunneling amplitude

$$\begin{aligned} A &= \exp \left[- \int_0^{X_c} |p| d\lambda \right] \\ &= \exp \left[- \int_0^{X_c} \sqrt{4M^2 - (\epsilon 2\lambda)^2} d\lambda \right] , \end{aligned} \quad (11.24)$$

where we used that $H = 0 = \sqrt{p^2 + 4M^2} - \epsilon 2\lambda$. Since $X_c = \frac{M}{\epsilon}$, we have

$$\begin{aligned} A &= \exp \left[- \frac{2M^2}{\epsilon} \int_0^1 \sqrt{1 - \frac{\epsilon^2}{M^2} \lambda^2} d \left(\frac{\epsilon}{M} \lambda \right) \right] \\ &= \exp \left[- \frac{2M^2}{\epsilon} \int_0^1 \sqrt{1 - y^2} dy \right] = \exp \left(- \frac{\pi M^2}{2\epsilon} \right) . \end{aligned} \quad (11.25)$$

Hence, the rate of the transition to the true vacuum is $\exp -\frac{\pi M^2}{\epsilon}$ per second and per unit volume. (To evaluate the integral we set $y = \cos \varphi$).

We end this section with a few comments. 1. The decay of the false vacuum per unit time and per unit volume is of the form $\Gamma/V = Ae^{-B/\hbar}(1 + \mathcal{O}(\hbar))$. We computed B . For A see [80, 76].

2. We used energy conservation to determine how fast a bubble expands. However, we neglected radiation of mesons. In general, when the false vacuum collapses to the true vacuum, mesons will be created, and thus the bubble will expand less rapidly.

3. Above we considered the critical bubble: a static solution of the classical field equations which describes a bubble which has just the correct form and size that it is metastable. For larger sizes there is no static solution, but one can consider the creation at $t = 0$ of a large bubble which then expands. This is an initial value problem: $\varphi(x)$ is given and also $\frac{\partial \varphi}{\partial t} = 0$ at $t = 0$. One can define the size of a bubble for example as $Q = \int_{-\infty}^{\infty} (\varphi + \frac{\mu}{\sqrt{\lambda}})^2 dx$. Far away, $\varphi = -\frac{\mu}{\sqrt{\lambda}}$, so Q is finite for bubbles. A problem we now want to solve is: given the size Q of a bubble, for which shape is its action minimal. (Minimal action in Euclidean space means maximal tunnelling rate). This will yield a one-parameter parametrization of bubbles; the parameter is a collective coordinate $\lambda(t)$, and having found the solution, we can then compare our ansatz in (11.15) and see how good the ansatz was. Mathematically, we can formulate this problem as a variational problem with a constraint. Introducing a constant Lagrange multiplier α we consider the action for the variational problem

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}(\partial_x \varphi)^2 - \frac{\lambda}{4} \left(\varphi^2 - \frac{\mu^2}{\lambda} \right)^2 - B \left(\frac{1}{3} \varphi^3 - \frac{\mu^2}{\lambda} \varphi \right) \\ & + \frac{2}{3} B \left(\frac{\mu}{\sqrt{\lambda}} \right)^3 + \frac{1}{2} \alpha \left(\varphi + \frac{\mu}{\sqrt{\lambda}} \right)^2 . \end{aligned} \quad (11.26)$$

The equation of motion

$$\frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial \varphi_x} - \frac{\partial \mathcal{L}}{\partial \varphi} = 0 , \quad (11.27)$$

has a first integral due to equipartition of energy

$$\frac{1}{2} \left(\frac{d\varphi}{dx} \right)^2 = \frac{\lambda}{4} \left(\varphi^2 - \frac{\mu^2}{\lambda} \right)^2 + B \left(\frac{1}{3} \varphi^3 - \frac{\mu^2}{\lambda} \varphi \right) - \frac{2}{3} B \left(\frac{\mu}{\sqrt{\lambda}} \right)^3 - \frac{1}{2} \alpha \left(\varphi + \frac{\mu}{\sqrt{\lambda}} \right)^2 . \quad (11.28)$$

The integration constant vanishes for bubbles. Introducing a field $\tilde{\varphi}$ which vanishes for large x

$$\tilde{\varphi} = \varphi + \frac{\mu}{\sqrt{\lambda}} , \quad \varphi = \tilde{\varphi} - \frac{\mu}{\sqrt{\lambda}} , \quad (11.29)$$

we obtain

$$\frac{d\tilde{\varphi}}{dx} = \sqrt{\frac{\lambda}{2}} \tilde{\varphi} \sqrt{\left(\tilde{\varphi} - \frac{2\mu}{\sqrt{\lambda}}\right)^2 + \frac{4B}{\lambda} \left(\frac{1}{3}\tilde{\varphi} - \frac{\mu}{\sqrt{\lambda}}\right) - \frac{2\alpha}{\lambda}}. \quad (11.30)$$

For $\alpha = B = 0$ the solution is the kink, but for $\alpha \neq 0$ we get bubbles. One can actually solve this equation exactly by using (see Gradhstein and Resznik, page 84, 2.266)

$$\int \frac{dy}{y\sqrt{a+by+cy^2}} = \frac{1}{\sqrt{a}} \operatorname{arccosh} \frac{2a+by}{y\sqrt{-4ac}}, \quad (11.31)$$

which holds if $a > 0$ and $b^2 - 4ac > 0$. This corresponds to $0 < \alpha < 2\mu^2$. By writing the differential equation as

$$\int d\left(\sqrt{\frac{\lambda}{2}}x\right) = \int \frac{d\tilde{\varphi}}{\tilde{\varphi} \sqrt{\left(\frac{4\mu^2}{\lambda} - \frac{2\alpha}{\lambda} - \frac{4B\mu}{\lambda\sqrt{\lambda}}\right) + \left(-\frac{4\mu}{\sqrt{\lambda}} + \frac{4B}{3\lambda}\right)\tilde{\varphi} + \tilde{\varphi}^2}}, \quad (11.32)$$

we obtain for the bubble with fixed size and minimum action

$$\begin{aligned} \cosh\left(\sqrt{a}\sqrt{\frac{\lambda}{2}}x\right) &= \frac{2a+b\tilde{\varphi}}{\tilde{\varphi}\sqrt{4ac-b^2}} \\ \tilde{\varphi} &= \frac{2a}{\sqrt{4ac-b^2} \cosh\left(\sqrt{a}\sqrt{\frac{\lambda}{2}}x-x_0\right)-b} \simeq \frac{\frac{2}{\sqrt{\lambda}}(2\mu^2-\alpha)}{\sqrt{\alpha} \cosh(\sqrt{2}\mu(x-x_0))+2\mu}. \end{aligned} \quad (11.33)$$

The constant α lies in the domain $0 < \alpha < 2\mu^2$. For $\alpha = 0$ we find $\tilde{\varphi} = \frac{2\mu}{\sqrt{\lambda}}$ (or $\varphi = \frac{\mu}{\sqrt{\lambda}}$) while for $\alpha = 2\mu^2$ we find $\tilde{\varphi} = 0$ (or $\varphi = -\frac{\mu}{\sqrt{\lambda}}$). In between, we have bubbles of finite extent; for small α the function $\tilde{\varphi}$ remains constant for a long time (the bubble) and then it falls rapidly off to zero (due to the cosh). This is the same behaviour as displayed by our ansatz in (11.15).

12 The strong CP problem

The vacua $|n\rangle$ of Yang-Mills theory in Minkowski space with winding number n all have the same energy, namely zero (because they are vacua). We recall that at fixed time space was compactified to an S^3 which was mapped to the S^3 of the group manifold of $SU(2)$. Since there is tunnelling as we have discussed, the physical vacuum is a linear combination of all of them. Since they all appear on equal footing, we expect that the generator T for large gauge transformations which change the winding number, defined by $T|n\rangle = |n+1\rangle$,

commutes with the Hamiltonian. Hence T maps the physical vacuum into itself. It follows that $T|\text{vac}\rangle = e^{i\varphi}|\text{vac}\rangle$ with φ some phase. The solution of this equation is

$$|\text{vac}\rangle \equiv |\theta\rangle = \sum_n e^{in\theta} |n\rangle. \quad (12.1)$$

Indeed $T|\theta\rangle = \sum e^{in\theta} |n+1\rangle = e^{-i\theta} |\theta\rangle$.

Instead of using the infinite set of states in (12.1), we can work with the ordinary vacuum $|0\rangle$ if at the same time we add a term

$$\mathcal{L}_\theta = -\theta_{\text{QCD}} \frac{g^2}{16\pi^2} \text{tr} F_{\mu\nu} {}^*F_{\mu\nu} \quad (12.2)$$

to the action. This term yields a factor $e^{in\theta}$ in the action $e^{\frac{i}{\hbar}S}$ if one is in the vacuum with winding number n . We shall set $\hbar = 1$. Note that we are in Minkowski space and that $\mathcal{L}_\theta$ is hermitian.

Strictly speaking, we should first make a Wick rotation to Euclidean space because we can only discuss instantons in Euclidean space, but $\mathcal{L}_\theta$ has the same form in Euclidean space: one gets a factor i from d^4x and another factor i from F_{0i} . Together with the factor $\frac{i}{\hbar}$ in $e^{\frac{i}{\hbar}S}$ one gets the same factor $e^{in\theta}$ in Euclidean space. The θ -term is a total derivative, and usually one discards total derivatives in Lagrangians because fields vanish at infinity, but for instanton backgrounds one finds of course a nonvanishing contribution due to winding.

The θ -term clearly violates parity P . It conserves charge conjugation symmetry C , hence it violates CP . The strong interactions described by QCD are not supposed to violate P or PC , hence θ_{QCD} should be very small. However, the observed θ parameter contains more than only θ_{QCD} . There is a second origin for a θ -angle coming from the electroweak sector: the manipulations leading to the CKM matrix. Recall that the mass terms of the quarks in the Standard Model come from Yukawa couplings

$$\begin{aligned} \mathcal{L} = & - \sum_{m,n} \left[g(qu)_{mn} \begin{pmatrix} \bar{q}_{L,m} \\ \bar{q}'_{L,m} \end{pmatrix}^T \begin{pmatrix} (h^0)^* \\ -(h^+)^* \end{pmatrix} q_{R,n} \right. \\ & \left. + g'(qu)_{mn} \begin{pmatrix} \bar{q}_{L,m} \\ \bar{q}'_{L,m} \end{pmatrix}^T \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} q'_{R,n} \right] + h.c. , \end{aligned} \quad (12.3)$$

where $g(qu)$ are the Yukawa couplings to quarks, and h^+, h^0 are the two components of the complex $SU(2)$ Higgs doublet. Furthermore $m = 1, 2, 3$ labels the families, so q_1 denotes the up quark while q'_1 denotes the down quark. When h^0 gets a vacuum expectation value $\langle h^0 \rangle = \frac{1}{\sqrt{2}}v$, one obtains mass matrices M for the (u, c, t) quarks and M' for the (d, s, b)

quarks, where

$$M_{mn} = \frac{v}{\sqrt{2}} g_{mn} , \quad M'_{mn} = \frac{v}{\sqrt{2}} g'_{mn} . \quad (12.4)$$

These matrices are in general arbitrary complex 3×3 matrices. One diagonalizes them with 3×3 unitary matrices which are different for left- and right- handed quarks⁴⁵

$$\begin{aligned} U_L M U_R^{-1} &= \text{diag}(m_u, m_c, m_t) \equiv D , \\ U'_L M' U'^{-1}_R &= \text{diag}(m_d, m_s, m_b) \equiv D' . \end{aligned} \quad (12.5)$$

The mass matrix for the quarks becomes then diagonal with real masses

$$\bar{q}_{L,m} M_{mn} q_{Rn} = \overline{(U_L q_L)} D (U_R q_R) \quad (12.6)$$

and similarly for $\bar{q}'_{L,m} M'_{mn} q'_{Rn}$. So, the physical quarks are $Q_L = U_L q_L$ and $Q_R = U_R q_R$, and similarly for Q'_L and Q'_R .

If one rescales q_L to Q_L , and q_R to Q_R , three things happen

- (i) the quark mass terms are diagonalized as we have discussed, yielding real physical quark masses
- (ii) a phase δ appears in the CKM matrix which describes electroweak CP violation
- (iii) a new term is produced in the action by the Jacobian for these chiral rescalings. This new term is again proportional to $\int F_{\mu\nu}^* F_{\mu\nu} d^4x$ with a coefficient which we call $-\theta_{\text{EW}}$. Hence, now the action contains the sum $\theta = \theta_{\text{QCD}} + \theta_{\text{EW}}$

$$\mathcal{L}_\theta = -(\theta_{\text{QCD}} + \theta_{\text{EW}}) \frac{g^2}{32\pi^2} \int (F_{\mu\nu}^a F_{\mu\nu}^a) d^4x . \quad (12.7)$$

There is no reason that $\theta_{\text{strong}} = \theta_{\text{QCD}} + \theta_{\text{EW}}$ vanishes, yet, as we now discuss, this seems to be the case.

We can make a final chiral rescaling of the 3 light quarks (the u, d and s quarks) such that the θ -term is entirely removed. Rescaling the left-handed quarks by $U(1)$ factors $e^{i\varphi_u}, e^{i\varphi_d}$ and $e^{i\varphi_s}$, the Jacobians for these rescalings yield a term

$$-(\varphi_u + \varphi_d + \varphi_s) \frac{g_s^2}{16\pi^2} \text{tr} F_{\mu\nu}^* F_{\mu\nu} , \quad (12.8)$$

which cancels the θ -term if $\varphi_u + \varphi_d + \varphi_s = \theta_{\text{strong}}$. Because the action is invariant except for the mass terms, only the transformation of the mass terms yields a new term in the action. In the diagonal mass term

$$m_u \bar{u}u + m_d \bar{d}d + m_s \bar{s}s , \quad (12.9)$$

⁴⁵ A complex matrix M can always be written as VH where V is unitary and H hermitian. This is the generalization to matrices of the decomposition $z = e^{i\varphi} \rho$ of complex numbers. Then H can be diagonalized by a unitary matrix, $H = U_R D U_R^{-1}$, and U_L is given by $V U_R$.

the rescalings yield, to first order in $\varphi_u, \varphi_d, \varphi_s$, a new term in the action

$$\mathcal{L}_{\text{CP violation}} = i\varphi_u m_u \bar{u}\gamma_5 u + i\varphi_d m_d \bar{d}\gamma_5 d + i\varphi_s m_s \bar{s}\gamma_5 s . \quad (12.10)$$

The φ 's are only constrained by $\varphi_u + \varphi_d + \varphi_s = \theta_{\text{strong}}$, so we can still choose them such that this new term is $SU(3)_V$ invariant. Namely if $\varphi_u = \frac{\theta m_d m_s}{m_u m_d + m_u m_s + m_d m_s}$, and cyclically for φ_d and φ_s , then

$$\mathcal{L}_{\text{CP violation}} = \frac{i\theta_{\text{strong}} m_u m_d m_s}{m_u m_d + m_u m_s + m_d m_s} (\bar{u}\gamma_5 u + \bar{d}\gamma_5 d + \bar{s}\gamma_5 s) . \quad (12.11)$$

This term is hermitian and $SU(3)_V$ invariant, but it violates P, and since it conserves C, it also violates CP. The original θ -term in the action in (12.2) has been transformed into the masslike terms in (12.11). No longer does one have to deal with total derivatives, but an ordinary extra masslike term has appeared in the QCD action. There is no reason that θ_{strong} should be small, but one can compute the electric dipole moment of the neutron which is nonzero if θ_{strong} is nonzero, and since experimentally the electric dipole moment has a very small upper bound, one finds that θ_{strong} is incredibly small

$$\theta_{\text{strong}} < 10^{-9} . \quad (12.12)$$

The problem why θ_{strong} is so small is called the strong CP problem. Note that it has nothing to do with the CP violation due to the phase δ in the CKM matrix, which is an electroweak effect. Also the electroweak CP violation is very small; it can be parametrized by the area of the unitarity triangles (each of the 6 unitarity triangles has the same area $2J$ in the Standard Model)

$$J = (3.0 \pm 0.3) 10^{-5} . \quad (12.13)$$

13 The $U(1)$ problem

In this section we discuss an application of instantons in QCD.

In the 1960's, in the absence of a renormalizable theory of the strong interactions, current algebra was developed as a method to derive information about matrix elements of currents, mostly the vector and axial-vector Noether currents which correspond to the (approximate) rigid flavor symmetry of the up, down and strange quarks. In terms of modern QCD, the action for the strong interactions reads

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu}^a)^2 - \sum \bar{\psi}_i \not{D} \psi^i , \quad (13.1)$$

where $i = 1, \dots, N_f$ labels the flavors. One can consider either two very light quarks (u and d), or three rather light quarks (u , d and s). Decomposing the massless quarks into

left-handed and right-handed parts, their action becomes

$$\mathcal{L}(\text{quarks}) = \bar{\psi}_{i,L} \bar{D} \psi_L^i - \bar{\psi}_{i,R} \bar{D} \psi_R^i . \quad (13.2)$$

It has clearly a rigid $U_L(N_f) \times U_R(N_f)$ symmetry group, where U_L acts only on ψ_L^i and U_R only on ψ_R^i . Instead of U_L and U_R we consider the vector part $U_V(N_f)$ and the axial vector part $U_A(N_f)$. The vector part transforms ψ_L^i and ψ_R^i the same way, while they transform oppositely under U_A . The total number of symmetries and group parameters has not changed, but physically U_V and U_A are very different. The $SU(2)_V$ part of the symmetry is realized in Nature, and yields the $SU(2)$ classification scheme for quark hadroscopy. The $U(1)_V$ corresponds to baryon-number conservation which is also (very well) satisfied. The $SU(2)_A$ symmetry is spontaneously broken, and the corresponding Goldstone bosons form an $SU(2)$ multiplet of pseudoscalars (the pions and the η meson). One might be inclined to apply the same reasoning to the $U(1)_A$ symmetry, and argue that it, too, must be spontaneously broken because there is no doubling of multiplets with opposite parity observed in nature. However, the $U(1)_A$ symmetry is explicitly violated by the presence of instantons in QCD, leading to the instanton-induced six-fermion interaction in the effective action. This solves “the $U(1)$ problem” that no isoscalar Goldstone boson exists in Nature [82]. There is a pseudoscalar meson, the η with a mass of $478 MeV$. It cannot be the Goldstone boson because from current algebra S. Weinberg has shown that the mass of such a Goldstone boson has to be smaller than $\sqrt{3}m_\pi$, far below the mass of the η meson [83]. (The η meson can still be made of a quark and an antiquark, so the usual $SU(2)$ scheme is still applicable - only this η meson is not a Goldstone boson)⁴⁶.

The axial-vector isoscalar current associated with the $U_A(1)$ symmetry is $j_\mu^5 = \sum_i^2 \bar{\psi}_i \gamma_5 \gamma_\mu \psi^i$. It has an Adler-Bell-Jackiw chiral anomaly

$$\partial^\mu j_\mu^{(5)} = N_f \frac{g_3^2}{32\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^a , \quad (13.3)$$

where $F_{\mu\nu}^a$ denotes the field strengths of the gluons, and g_3 is the QCD coupling constant. In a QCD instanton background, integration over spacetime yields

$$\int \frac{\partial}{\partial t} Q^5 dt = N_f k , \quad (13.4)$$

where k is the winding number. To make sense of this equation one should first integrate in Euclidean space to obtain a non-vanishing expression for the right-hand side in terms of the winding number k of the QCD instanton, and then Wick-rotate so that the left-hand side can be written as $\int d^4x \partial_\mu j_5^\mu \propto \int \frac{\partial}{\partial t} Q^5 dt$. The conclusion is that Q^5 is not conserved because k can be different from zero. For further discussion of the $U(1)$ problem we refer to [82], [84, 85], or to the lecture notes by Coleman in [2].

⁴⁶One can extend this discussion to $U_L(3) \times U_R(3)$ with pions, kaons and η now 8 Goldstone bosons, and the η' with mass $958 MeV$ taking the place of η . This η' is an $SU(3)$ singlet.

14 Baryon decay

In this section we present an application of instantons to the gauge fields of the electroweak sector of the Standard Model.

In an instanton background with winding number k , massless (or approximately massless) fermions in the fundamental representation of $SU(N)$ have $|k|$ zero modes, see (4.37). In the electroweak $SU(2)_w \times U(1)$ theory (the subscript w stands for weak), quarks and leptons are in the fundamental representation (doublets) of $SU(2)_w$. In Euclidean space the integration over zero modes of these quarks and leptons has dynamical consequences which we shall derive, but of course real quarks and leptons live in Minkowski space and not in Euclidean space. We assume that the Green functions in Minkowski spacetime can be obtained from those in Euclidean space by analytic continuation. Ideally we should prove that the Euclidean results give the main contribution to processes in Minkowski space in the same way as this was shown for tunnelling, but as far as we know this has not been done. Since processes involving electroweak instantons are suppressed by a factor $\exp\left(-\frac{1}{\hbar} \frac{8\pi^2}{g_2^2} |k|\right)$ with g_2 the electroweak $SU(2)$ coupling constant, we only consider instantons with $|k| = 1$ made from W^+ , W^- and W^0 bosons. Then the left-handed quark doublets $\begin{pmatrix} u \\ d' \end{pmatrix}$ and $\begin{pmatrix} c \\ s' \end{pmatrix}$ each have 3 zero modes because there are 3 colors, while the lepton doublets $\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$ and $\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$ each have one zero mode. The primes on d' and s' denote Cabibbo-rotated quarks

$$\begin{aligned} d' &= d \cos \theta_c + s \sin \theta_c \\ s' &= s \cos \theta_c - d \sin \theta_c \end{aligned} \tag{14.1}$$

with $\theta_c = 13^\circ$ the Cabibbo angle. As we shall explain, this Cabibbo rotation makes it possible for a neutron and a proton (six quarks together) to decay into two antileptons [5]

$$p + n \rightarrow e^+ + \bar{\nu}_\mu \quad (\text{or} \quad \mu^+ + \bar{\nu}_e) . \tag{14.2}$$

In these instanton-induced processes, the electron number E , muon number M , up plus down number, and charm plus strangeness number change as follows

$$\Delta E = \Delta M = 1 , \quad \Delta u + \Delta d' = 3 , \quad \Delta c + \Delta s' = 3 . \tag{14.3}$$

The decay of a proton with (u, u, d) and neutron with (u, d, d) quarks into e^+ and $\bar{\nu}_\mu$, or into μ^+ and $\bar{\nu}_e$, can be described by a local vertex operator with 3 up-quark fields with different colors from $\begin{pmatrix} u \\ d' \end{pmatrix}$ doublets, and 3 down-quark fields also with different colors from $\begin{pmatrix} c \\ s' \end{pmatrix}$ doublets, and further one field from each of the two lepton doublets. This operator is of course nonrenormalizable, but it can be used in effective field theories for phenomenological purposes. Although this effective operator is derived from field theory in the sector with an

instanton, once it is obtained one can add it to the effective action and then forget about the existence of instantons. We now derive these results.

The $U(1)_A$ symmetry has at the perturbative level an anomaly. There are triangle graphs with an anomaly: one vertex of the triangle graph is given by $j_\mu^{(5)} = \sum_s \bar{\psi}^s \gamma_5 \gamma_\mu \psi^s$ (where $s = 1, \dots, N_f$ and N_f is the number of flavors, 3 in our case if we restrict our attention to the lightest quarks u, d and s). The one-loop perturbative chiral anomaly is then given by

$$\partial^\mu j_\mu^{(5)} = iN_f \frac{g_2^2}{16\pi^2} G_{\mu\nu}^a {}^*G_{\mu\nu}^a, \quad (14.4)$$

where $G_{\mu\nu}^a$ is the W -boson field strength and g_2 the coupling constant of the $SU(2)$ weak interactions. (This is thus the abelian flavor $U(1)_A$ anomaly. The nonabelian anomaly for the rigid flavor group vanishes because it is proportional to the trace of T_a of the flavor group, which vanishes).

If one integrates over space and time, the anomaly equation becomes

$$\int \frac{d}{dt} Q^{(5)} \equiv \int dt \frac{d}{dt} \int d^3x (i j_0^{(5)}) = 2N_f k. \quad (14.5)$$

The instanton number k counts the number of left-handed fermions minus the number of right-handed fermions, and in ordinary perturbation theory (with $k = 0$) for massless quarks, this difference is thus conserved. However, in an instanton background ($k \neq 0$), the chiral charge of the vacuum at $t = -\infty$ changes to a different chiral charge of the vacuum at $t = +\infty$: $\Delta Q^{(5)} = 2N_f k$. The conclusion is that the perturbative anomaly, and the violation of the axial charge which occurs when one tunnels from one vacuum to another, are related! Both are different aspects of the same chiral anomaly. The perturbative anomaly occurs when fields are small, so the winding vanishes and one is in the $k = 0$ sector. The nonperturbative anomaly is due to the same axial-vector current but now in the background of instantons which cannot be viewed as small and tending to zero at infinity, since they must produce winding.

One may at this point wonder whether the Higgs effect which gives the W -bosons a mass, in such a way that they vanish exponentially at large distances, does at the same time destroy the concept of winding. There is no contradiction. When we discussed the large instanton problem, we chose the regular gauge for the instanton to simplify the calculations. However, exponential fall-off only occurs in the singular gauge. In that case, the winding takes place at the origin, as we discussed in the introduction. In this section we use the regular gauge and then there is winding at infinity even in Higgs models.

To saturate the integrations over the Grassmann collective coordinates, one needs 6 chiral quark fields in a correlator (one for each zero mode). Each field has a mode expansion

into a zero mode and all nonzero modes, but the integration measure $d\mathcal{K}$ over the Grassmann variable $\mathcal{K}$ in the mode expansion picks out only the zero mode. Then the integration over collective coordinates gives as result the product of the 6 zero mode functions. If we put one $SU(2)_w$ doublet with one up quark and one down quark at a point x_1 , a second pair a x_2 , and a third pair at x_3 ,⁴⁷ and the instanton is at x_0 , we find from (5.35) for large separations ($x^2 \gg \rho^2$) the factor⁴⁸

$$\prod_{i=1}^3 \frac{1}{(x_i - x_0)^6} . \quad (14.6)$$

So if one computes some correlator in a theory with instantons, six quark fields from the correlator are needed to saturate the Grassmann integrals, and the remaining fields are then treated as in ordinary field theory (with propagators and vertices). Thus instantons induce a term proportional to $\prod_{i=1}^3 \frac{1}{(x_i - x_0)^6}$ in the effective action which describes the annihilation of 6 quarks. Further there are σ matrices and other constants which are also due to the zero mode function.

One can now construct an effective local 6-quark vertex V at a point x_0 which yields the same results in a theory without instantons as one obtains in a theory with instantons if one integrates over the fermionic collective coordinates of the quarks. This vertex must contain 6 quark fields which contain the 6 different collective coordinates, hence it has the form $V = u_L^{\alpha,1} u_L^{\beta,2} u_L^{\gamma,3} d_L^{\delta,1} d_L^{\epsilon,2} d_L^{\zeta,3} T_{\alpha\beta\gamma\delta\epsilon\zeta}$ where T is a numerical tensor. Contraction of 6 “probe-quarks” at positions x_1, x_2, x_3 with V at x_0 using ordinary flat space propagators $\frac{1}{(x_i - x_0)^3}$ for massless quarks in a trivial vacuum precisely reproduces the result for the correlation function of the 6 probe-quarks in an instanton background centered around x_0 , provided the form of T is correctly chosen.

These new vertices lead to anomalies in the baryon currents and lepton currents. In particular, the rigid $U(1)_A$ symmetry is **explicitly** broken by the presence of the interaction V in the action, and as we discussed in the previous section, this solves the $U(1)_A$ broken. As we already mentioned, a proton and a neutron (two baryons equal six quarks) may annihilate to form two antileptons (an e^+ or a μ^+ , and an anti-neutrino). However, due to

⁴⁷A massless complex Dirac spinor contains two Weyl spinors which are decoupled from each other $\bar{\psi}_D \not{D} \psi_D = \bar{\psi}_L \not{D} \psi_L + \bar{\psi}_R \not{D} \psi_R$. Each has a zero mode. However, since only left-handed quarks couple to the W gauge fields, only left-handed quarks feel the presence of instantons, and so we neglect the right-handed quarks in this discussion.

⁴⁸The down quark is contained in the s' of the doublet (c, s') . This is an $SU(2)_w$ doublet, and the instanton is an $SU(2)_w$ instanton. Although the c quark is heavier than the s quark, one can still view them as massless compared to the scale 250GeV of electroweak interactions. Massive spinors in an instanton background have no zero modes, as one may show by adding a mass term to (4.17) and (4.18). We assume that for such a broken $SU(2)_w$ doublet there still exists approximately a zero mode.

the incredibly small prefactor $\exp\left(-\frac{8\pi^2}{g_2^2}|k|\right)$, where g_2 is the $SU(2)$ weak coupling constant, these processes are not observable.

15 Discussion

In this chapter we have reviewed the general properties of single Yang-Mills instantons, and have given tools to compute non-perturbative effects in (non-) supersymmetric gauge theories. However, we have not discussed several other important or interesting topics:

- Perturbation theory around the instanton: the methods described here enable us to compute non-perturbative effects in the semi-classical approximation where the coupling constant is small. It is in many cases important to go beyond this limit, and to study sub-leading corrections that arise from higher order perturbation theory around the instanton [58, 57]. Apart from a brief discussion about the one-loop determinants in section 7, we have not really addressed these issues.

- Multi-instantons: we have completely omitted a discussion of multi-instantons. These can be constructed using the ADHM formalism [31]. The main difficulty lies in the explicit construction of the collective coordinates in an instanton solution and of the measure of collective coordinates beyond instanton number $k = 2$. However, it was demonstrated that certain simplifications occur in the large N limit of $\mathcal{N} = 4$ SYM theories [25], where one can actually sum over all multi-instantons to get exact results for certain correlation functions. For reviews on the ADHM construction in super Yang-Mills theories, see e.g. [32, 25, 86]. The same techniques were later applied for $\mathcal{N} = 2, 1$ SYM [87, 88], and it would be interesting to study the consequences of multi-instantons for large N non-supersymmetric theories. For a review on instantons in QCD, see for instance [89].

A Winding number

For a gauge field configuration with finite classical gauge action the field strength must tend to zero faster than x^{-2} at large x . For vanishing $F_{\mu\nu}$, the potential A_μ becomes then pure gauge, $A_\mu \xrightarrow{x \rightarrow \infty} U^{-1} \partial_\mu U$. All configurations of A_μ which become pure gauge at infinity fall into equivalence classes, where each class has a definite winding number. As we now show, this winding number is given by

$$k = -\frac{1}{16\pi^2} \int d^4x \operatorname{tr} {}^*F_{\mu\nu} F_{\mu\nu} , \quad (\text{A.1})$$

where ${}^*F_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F_{\rho\sigma}$ and T_a are the generators in the fundamental representation of $SU(N)$, antihermitean $N \times N$ matrices satisfying $\operatorname{tr} T_a T_b = -\frac{1}{2} \delta_{ab}$. This is the normalization

we adopt for the fundamental representation. The key observation is that $*F_{\mu\nu}F_{\mu\nu}$ is a total derivative of a gauge variant current⁴⁹

$$\text{tr } *F_{\mu\nu}F_{\mu\nu} = 2\partial_\mu \text{tr } \epsilon_{\mu\nu\rho\sigma} \left\{ A_\nu \partial_\rho A_\sigma + \frac{2}{3} A_\nu A_\rho A_\sigma \right\} . \quad (\text{A.2})$$

According to Stokes' theorem, the four-dimensional space integral becomes an integral over the three-dimensional boundary at infinity if one uses the regular gauge in which there are no singularities at the origin. Since $F_{\mu\nu}$ vanishes at large x , one may replace $\partial_\rho A_\sigma$ by $-A_\rho A_\sigma$, and since A_μ becomes a pure gauge at large x , one obtains

$$k = \frac{1}{24\pi^2} \oint_{S^3(\text{space})} d\Omega_\mu \epsilon_{\mu\nu\rho\sigma} \text{tr } \left\{ (U^{-1} \partial_\nu U) (U^{-1} \partial_\rho U) (U^{-1} \partial_\sigma U) \right\} , \quad (\text{A.3})$$

where the integration is over a large three-sphere, $S^3(\text{space})$, in four-dimensional Euclidean space. To each point x^μ on this large three-sphere in space corresponds a group element U in the gauge group G . If $G = SU(2)$, the group manifold is also a three-sphere⁵⁰ $S^3(\text{group})$. Then $U(x)$ maps $S^3(\text{space})$ into $S^3(\text{group})$,⁵¹ and as we now show, k is an integer which counts how many times $S^3(\text{space})$ is wrapped around $S^3(\text{group})$. Choose a parametrization of the group elements of $SU(2)$ in terms of group parameters⁵² $\xi^i(x)$ ($i = 1, 2, 3$). Then the functions $\xi^i(x)$ map x into $SU(2)$. Consider a small surface element of $S^3(\text{space})$. According to the chain rule

$$\begin{aligned} & \text{tr } \left\{ (U^{-1} \partial_\nu U) (U^{-1} \partial_\rho U) (U^{-1} \partial_\sigma U) \right\} \\ &= \frac{\partial \xi^i}{\partial x_\nu} \frac{\partial \xi^j}{\partial x_\rho} \frac{\partial \xi^k}{\partial x_\sigma} \text{tr } \left\{ (U^{-1} \partial_i U) (U^{-1} \partial_j U) (U^{-1} \partial_k U) \right\} , \end{aligned} \quad (\text{A.4})$$

and using⁵³

$$\Delta\Omega_\mu = \frac{1}{6} \epsilon_{\mu\alpha\beta\gamma} \Delta x_\alpha \Delta x_\beta \Delta x_\gamma , \quad (\text{A.5})$$

⁴⁹Note that $*F_{\mu\nu}F_{\mu\nu}$ is equal to $2\epsilon_{\mu\nu\rho\sigma} \{ \partial_\mu A_\nu \partial_\rho A_\sigma + 2\partial_\mu A_\nu A_\rho A_\sigma + A_\mu A_\nu A_\rho A_\sigma \}$ but the last term vanishes in the trace due to the cyclicity of the trace.

⁵⁰The elements of $SU(2)$ can be written in the fundamental representation as $U = a_0 \mathbb{1} + i \sum_k a_k \tau_k$ where τ_k are the Pauli matrices and a_0 and a_k are real coefficients satisfying the condition $a_0^2 + \sum_k a_k^2 = 1$. This defines a sphere $S^3(\text{group})$. (If the a 's are not real but carry a common phase, one obtains the elements of $U(2)$).

⁵¹There is actually a complication. Far away $A_\mu = U^{-1} \partial_\mu U$ but in order that U be only a function on $S^3(\text{space})$ it should only depend on the 3 polar angles but not on the radius. Hence $A_r = U^{-1} \partial_r U$ should vanish. We can make a gauge transformation with a group element V such that $A'_r = V^{-1} (\partial_r + A_r) V$ vanishes. The V which achieves this is the path ordered integral along the radius from the origin, $V = P \exp - \int^r A_r dr$. Note that U is only defined for large r , but V must be defined everywhere, and $V \neq U$. In fact, V does not have winding since it can be continuously deformed to the unit group element. The winding number is computed in the text for UV , but since k in (A.1) is gauge invariant, k is also the winding number of the original gauge field A_μ .

⁵²For example, Euler angles, or Lie parameters $U = a_0 \mathbb{1} + i \sum_k a_k \tau_k$ with $a_0 = \sqrt{1 - \sum_k a_k^2}$.

⁵³For example, if the surface element points in the x -direction we have $\Delta\Omega = \Delta y \Delta z \Delta \tau$ if $\epsilon_{1234} = 1$.

with $\frac{1}{6}\epsilon_{\mu\nu\rho\sigma}\epsilon_{\mu\alpha\beta\gamma} = \delta_{[\alpha\beta\gamma]}^{\nu\rho\sigma}$ and $\Delta\xi^{[i}\Delta\xi^j\Delta\xi^k] = \epsilon^{ijk}\Delta^3\xi$, we obtain for the contribution Δk of the small surface element to k

$$\Delta k = \frac{1}{24\pi^2}\epsilon^{ijk}\text{tr} \{ (U^{-1}\partial_i U) (U^{-1}\partial_j U) (U^{-1}\partial_k U) \} \Delta^3\xi, \quad (\text{A.6})$$

where $k = \oint_{S^3(\text{space})} \Delta k$. The elements $(U^{-1}(\xi)\partial_i U(\xi))$ lie in the Lie algebra, and define the group vielbein $e_i^a(\xi)$ by

$$(U^{-1}\partial_i U) = e_i^a(\xi)T_a. \quad (\text{A.7})$$

With $\epsilon^{ijk}e_i^a e_j^b e_k^c = (\det e) \epsilon^{abc}$, we obtain for the contribution to k from a surface element $\Delta\Omega_\mu$

$$\Delta k = \frac{1}{24\pi^2} (\det e) \text{tr} (\epsilon^{abc}T_a T_b T_c) \Delta^3\xi = -\frac{1}{16\pi^2} (\det e) \Delta^3\xi. \quad (\text{A.8})$$

We used that for $SU(2)$ we have $[T_a, T_b] = \epsilon_{abc}T_c$. As we have demonstrated, the original integral over the physical space is reduced to one over the group with measure $(\det e) d^3\xi$. The volume of a surface element of $S^3(\text{group})$ with coordinates $d\xi^i$ is proportional to $(\det e) d^3\xi$ (called the Haar measure). Since this expression is a scalar in general relativity,⁵⁴ we know that the value of the volume does not depend on which coordinates one uses except for an overall normalization. We fix this overall normalization of the group volume such that near $\xi = 0$ the volume is $\Delta^3\xi$. Since $e_i^a = \delta_i^a$ near $\xi = 0$, we have there the usual Euclidean measure $d^3\xi$. Each small patch on $S^3(\text{space})$ corresponds to a small patch on $S^3(\text{group})$, $\Delta k \sim \text{Vol}(\Delta^3\xi)$. Since the U 's fall into homotopy classes, integrating once over $S^3(\text{space})$ we cover $S^3(\text{group})$ an integer number of times. To check the proportionality factor in $\Delta k \sim \text{Vol}(\Delta^3\xi)$, we consider the fundamental map

$$U(x) = ix_\mu\sigma_\mu/\sqrt{x^2}, \quad U^{-1}(x) = -ix_\mu\bar{\sigma}_\mu/\sqrt{x^2}. \quad (\text{A.9})$$

where σ_μ denotes the 2×2 matrices $(\vec{\sigma}, i)$ with $\vec{\sigma}$ the Pauli matrices, and $\bar{\sigma}_\mu = (\vec{\sigma}, -i)$. This is clearly a one-to-one map from $S^3(\text{space})$ to $S^3(\text{group})$ and should therefore yield $|k| = 1$. Direct calculation gives

$$U^{-1}\partial_\mu U = \frac{-x_\mu}{x^2} + \frac{x_\nu\bar{\sigma}_\nu}{x^2}\sigma_\mu = -\sigma_{\mu\nu}x_\nu/x^2, \quad (\text{A.10})$$

where $\sigma_{\mu\nu}$ is defined in (B.8). Substitution into (A.3) leads to $k = -\frac{1}{2\pi^2} \oint d\Omega_\mu x_\mu/x^4 = -1$ making use of (B.18).⁵⁵ To obtain $k = 1$ one has to make the change $\sigma \leftrightarrow \bar{\sigma}$ or $x \leftrightarrow -x$ in Eq. (A.9).

⁵⁴Under a change of coordinates $\xi = \xi(\xi')$ at the point ξ , the vielbein transforms as $e_i^a(\xi) = \frac{\partial \xi'^j}{\partial \xi^i} e_j'^a(\xi(\xi'))$, hence $\det e(\xi) = \left(\det \frac{\partial \xi'}{\partial \xi}\right) \det(e'(\xi'))$, while $d^3\xi$ is equal to $|\det \frac{\partial \xi}{\partial \xi'}| d^3\xi'$. For small coordinate transformations $\det \partial \xi / \partial \xi'$ is positive, hence $\det e d^3\xi$ is invariant.

⁵⁵Only the commutator of the first two matrices in $\text{tr} (\sigma_{\nu\alpha}\sigma_{\rho\beta}\sigma_{\sigma\gamma})x^\alpha x^\beta x^\gamma$ contributes because the anti-commutator is proportional to the unit matrix. In the result only the anticommutator gives a nonvanishing result, because the commutator yields term proportional to $\sigma_{\alpha\beta}$ whose trace vanishes.

Let us comment on the origin of the winding number of the instanton in the singular gauge. In this case A_μ^{sing} vanishes fast at infinity, but becomes pure gauge near $x = 0$. In the region between a small sphere in the vicinity of $x = 0$ and a large sphere at $x = \infty$ we have an expression for k in terms of a total derivative, but now for A_μ^{sing} the only contribution to the topological charge comes from the boundary near $x = 0$:

$$k = -\frac{1}{24\pi^2} \oint_{S_{x \rightarrow 0}^3(\text{space})} d\Omega_\mu \epsilon_{\mu\nu\rho\sigma} \text{tr} \{ (U^{-1} \partial_\nu U) (U^{-1} \partial_\rho U) (U^{-1} \partial_\sigma U) \} . \quad (\text{A.11})$$

The extra minus sign is due to the fact that the normal to the $S^3(\text{space})$ at $x = 0$ points inward. Furthermore, $A_\mu^{\text{sing}} \sim U^{-1} \partial_\mu U = -\bar{\sigma}_{\mu\nu} x_\nu / x^2$ near $x = 0$, while $A_\mu^{\text{reg}} \sim U \partial_\mu U^{-1} = -\sigma_{\mu\nu} x_\nu / x^2$ for $x \sim \infty$. There is a second extra minus sign in the evaluation of k from the trace of Lorentz generators. As a result $k_{\text{sing}} = k_{\text{reg}}$, as it should be since k is a gauge invariant object. The gauge transformation which maps A_μ^{reg} to A_μ^{sing} transfers the winding from a large to a small $S^3(\text{space})$.

B 't Hooft symbols and Euclidean spinors

In this appendix we give a list of conventions and formulae useful for instanton calculus. Let us first discuss the structure of Lorentz algebra $so(3, 1)$ in Minkowski space-time. The generators can be represented by $L_{\mu\nu} = \frac{1}{2}(x_\mu \partial_\nu - x_\nu \partial_\mu)$ and form the algebra $[L_{\mu\nu}, L_{\rho\sigma}] = -\eta_{\mu\rho} L_{\nu\sigma} - \eta_{\nu\sigma} L_{\mu\rho} + \eta_{\mu\sigma} L_{\nu\rho} + \eta_{\nu\rho} L_{\mu\sigma}$, with the signature $\eta_{\mu\nu} = \text{diag}(-, +, +, +)$. The spatial rotations $J_i \equiv \frac{1}{2}\epsilon_{ijk} L_{jk}$ and boosts $K_i \equiv L_{0i}$ satisfy the algebra $[J_i, J_j] = -\epsilon_{ijk} J_k$, $[J_i, K_j] = [K_i, J_j] = -\epsilon_{ijk} K_k$ and $[K_i, K_j] = \epsilon_{ijk} J_k$.

There exist two 2-component spinor representations, which we denote by λ^α and $\bar{\chi}_{\dot{\alpha}}$ ($\alpha = 1, 2$ and $\dot{\alpha} = 1, 2$). The generators for these spinor representations are $\sigma^{\mu\nu}$ and $\bar{\sigma}^{\mu\nu}$, where $\sigma_{\mu\nu} \equiv \frac{1}{2}(\sigma_\mu \bar{\sigma}_\nu - \sigma_\nu \bar{\sigma}_\mu)$, $\bar{\sigma}_{\mu\nu} = \frac{1}{2}(\bar{\sigma}_\mu \sigma_\nu - \bar{\sigma}_\nu \sigma_\mu)$, with $\sigma_\mu^{\alpha\dot{\beta}} = (\vec{\tau}, I)$, $\bar{\sigma}_{\mu\dot{\alpha}\beta} = (\vec{\tau}, -I)$, $\mu = 1, 2, 3, 0$, and I denotes the identity matrix. The matrices τ^i with $i = 1, 2, 3$ are the usual Pauli matrices. They consist of $\sigma^{ij} = i\epsilon^{ijk} \tau^k$ and $\sigma^{0i} = \tau^i$ for λ^α , and $\bar{\sigma}^{ij} = i\epsilon^{ijk} \tau^k$ and $\bar{\sigma}^{0i} = -\tau^i$ for $\bar{\chi}_{\dot{\alpha}}$. The rotation generators σ^{ij} are clearly antihermitian, but the boost generators are hermitian.

Under a rotation or boost, both spinors simultaneously transform. Most importantly, the two spinor representations are complex. In fact, they are each other's complex conjugate up to a similarity transformation: $(\sigma^{\mu\nu})^* = \sigma_2 \bar{\sigma}^{\mu\nu} \sigma_2$. The matrices $i\tau^k$ and τ^k form the 2×2 defining representation of the group $Sl(2, C)$, which is the covering group of $SO(3, 1)$.

The situation differs for Euclidean space ($\delta_{\mu\nu} = \text{diag}(+, +, +, +)$) with $SO(4)$ instead of the Lorentz group $SO(3, 1)$. Now $[L_{\mu\nu}, L_{\rho\sigma}] = \delta_{\nu\rho} L_{\mu\sigma} + 3$ terms, and $[J_i, J_j] = -\epsilon_{ijk} J_k$, $[J_i, K_j] = -\epsilon_{ijk} K_k$ but $[K_i, K_j] = -\epsilon_{ijk} J_k$ where obviously $J_i \equiv \frac{1}{2}\epsilon_{ijk} L_{jk}$ and

boosts $K_i \equiv L_{i4}$. The linear combinations of (ij) and $(4, i)$ -plane rotations

$$M_i \equiv \frac{1}{2}(J_i + K_i) , \quad N_i \equiv \frac{1}{2}(J_i - K_i) , \quad (\text{B.1})$$

give the algebras of commuting $SU(2)$ subgroups of $SO(4) = SU(2) \times SU(2)$ in view of the anti-hermiticity $M_i^\dagger = -M_i$, $N_i^\dagger = -N_i$. We now denote the two spinor representations by λ^α and $\bar{\chi}_{\alpha'}$. Because M and N are represented by generators $i\vec{\sigma}_M$ and $i\vec{\sigma}_N$ which act in different spaces, one can transform λ^α while $\bar{\chi}_{\alpha'}$ stays fixed, or vice versa. The two spinor representations in Euclidean space are each pseudoreal: as we shall discuss $(\sigma^{\mu\nu})^* = \sigma_2 \sigma^{\mu\nu} \sigma_2$ and $(\bar{\sigma}_{\mu\nu})^* = \sigma_2 \bar{\sigma}_{\mu\nu} \sigma_2$.

It is an easy exercise to check that we can represent the operators M and N by

$$M_i = \bar{\eta}_{i\mu\nu} , \quad \text{and} \quad N_i = \eta_{i\mu\nu} , \quad (\text{B.2})$$

where we introduced 't Hooft symbols [4]

$$\begin{aligned} \eta_{a\mu\nu} &\equiv \epsilon_{a\mu\nu} + \delta_{a\mu}\delta_{\nu 4} - \delta_{a\nu}\delta_{4\mu}, \text{ or } \eta_{aij} = \epsilon_{aij}, \eta_{aj4} = \delta_{aj} \\ \bar{\eta}_{a\mu\nu} &\equiv \epsilon_{a\mu\nu} - \delta_{a\mu}\delta_{\nu 4} + \delta_{a\nu}\delta_{4\mu}, \text{ or } \bar{\eta}_{aij} = \epsilon_{aij}, \bar{\eta}_{aj4} = -\delta_{aj} \end{aligned} \quad (\text{B.3})$$

and $\bar{\eta}_{a\mu\nu} = (-1)^{\delta_{4\mu} + \delta_{4\nu}} \eta_{a\mu\nu}$. They form a basis of anti-symmetric 4 by 4 matrices and are (anti-)selfdual in vector indices ($\epsilon_{1234} = 1$)

$$\eta_{a\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\eta_{a\rho\sigma} , \quad \bar{\eta}_{a\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\bar{\eta}_{a\rho\sigma} . \quad (\text{B.4})$$

The η -symbols obey the following relations

$$\begin{aligned} \epsilon_{abc}\eta_{b\mu\nu}\eta_{c\rho\sigma} &= \delta_{\mu\rho}\eta_{a\nu\sigma} + \delta_{\nu\sigma}\eta_{a\mu\rho} - \delta_{\mu\sigma}\eta_{a\nu\rho} - \delta_{\nu\rho}\eta_{a\mu\sigma} , \\ \eta_{a\mu\nu}\eta_{a\rho\sigma} &= \delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\mu\sigma}\delta_{\nu\rho} + \epsilon_{\mu\nu\rho\sigma} , \\ \eta_{a\mu\rho}\eta_{b\mu\sigma} &= \delta_{ab}\delta_{\rho\sigma} + \epsilon_{abc}\eta_{c\rho\sigma} , \\ \epsilon_{\mu\nu\rho\tau}\eta_{a\sigma\tau} &= \delta_{\sigma\mu}\eta_{a\nu\rho} + \delta_{\sigma\rho}\eta_{a\mu\nu} - \delta_{\sigma\nu}\eta_{a\mu\rho} , \\ \eta_{a\mu\nu}\eta_{a\mu\nu} &= 12 , \quad \eta_{a\mu\nu}\eta_{b\mu\nu} = 4\delta_{ab} , \quad \eta_{a\mu\rho}\eta_{a\mu\sigma} = 3\delta_{\rho\sigma} . \end{aligned} \quad (\text{B.5})$$

The same holds for $\bar{\eta}$ except for the terms with $\epsilon_{\mu\nu\rho\sigma}$,

$$\begin{aligned} \bar{\eta}_{a\mu\nu}\bar{\eta}_{a\rho\sigma} &= \delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\mu\sigma}\delta_{\nu\rho} - \epsilon_{\mu\nu\rho\sigma} , \\ \epsilon_{\mu\nu\rho\sigma}\bar{\eta}_{a\sigma\tau} &= -\delta_{\sigma\mu}\bar{\eta}_{a\nu\rho} - \delta_{\sigma\rho}\bar{\eta}_{a\mu\nu} + \delta_{\sigma\nu}\bar{\eta}_{a\mu\rho} . \end{aligned} \quad (\text{B.6})$$

Obviously $\eta_{a\mu\nu}\bar{\eta}_{b\mu\nu} = 0$ due to different duality properties. In matrix notation, we have

$$\begin{aligned} [\eta_a, \eta_b] &= -2\epsilon_{abc}\eta_c , & [\bar{\eta}_a, \bar{\eta}_b] &= -2\epsilon_{abc}\bar{\eta}_c , \\ \{\eta_a, \eta_b\} &= -2\delta_{ab} , & \{\bar{\eta}_a, \bar{\eta}_b\} &= -2\delta_{ab} , \end{aligned} \quad (\text{B.7})$$

and the two sets of matrices commute, i.e. $[\eta_a, \bar{\eta}_b] = 0$ (this is equivalent to the statement that the generators M and N commute).

The two inequivalent spinor representations of the Euclidean Lorentz algebra are given by

$$\sigma_{\mu\nu} \equiv \frac{1}{2}[\sigma_\mu \bar{\sigma}_\nu - \sigma_\nu \bar{\sigma}_\mu] , \quad \bar{\sigma}_{\mu\nu} = \frac{1}{2}[\bar{\sigma}_\mu \sigma_\nu - \bar{\sigma}_\nu \sigma_\mu] , \quad (\text{B.8})$$

in terms of Euclidean matrices

$$\sigma_\mu^{\alpha\beta'} = (\tau^a, i) , \quad \bar{\sigma}_{\mu\alpha'\beta} = (\tau^a, -i) , \quad \mu = 1, 2, 3, 4 , \quad (\text{B.9})$$

obeying the Clifford algebra $\sigma_\mu \bar{\sigma}_\nu + \sigma_\nu \bar{\sigma}_\mu = 2\delta_{\mu\nu}$. Since $\sigma_{\mu\nu}$ contains $\sigma_{ij} = \epsilon_{ijk} i\tau^k$ and $\sigma_{i4} = -i\tau_i$, while $\bar{\sigma}_{\mu\nu}$ contains $\bar{\sigma}_{ij} = \epsilon_{ijk} i\tau^k$ and $\bar{\sigma}_{i4} = i\tau_i$, they are not each others complex conjugate, contrary to the Minkowski case. Rather, they are pseudo-real, meaning that their complex-conjugates are related to themselves by a similarity transformation

$$\sigma_{\mu\nu}^* = \sigma_2 \sigma_{\mu\nu} \sigma_2 ; (\bar{\sigma}_{\mu\nu})^* = \sigma_2 \bar{\sigma}_{\mu\nu} \sigma_2 . \quad (\text{B.10})$$

To prove these, and other, spinor relations, one needs some formulas which we now present. As in Minkowski space, also in Euclidean space σ_μ and $\bar{\sigma}_\mu$ are related by transposition

$$\sigma_\mu^{\alpha\alpha'} = \bar{\sigma}_\mu^{\alpha'\alpha} \quad (\text{B.11})$$

where $\bar{\sigma}_\mu^{\alpha'\alpha}$ is obtained from $\bar{\sigma}^\mu_{\beta'\beta}$ by raising indices

$$\bar{\sigma}_\mu^{\alpha'\alpha} \equiv \epsilon^{\alpha'\beta'} \epsilon^{\alpha\beta} \bar{\sigma}_{\beta'\beta}^\mu \quad (\text{B.12})$$

We use everywhere the north-west convention for raising and lowering the spinor indices

$$\epsilon^{\alpha\beta} \xi_\beta = \xi^\alpha , \quad \bar{\xi}^{\beta'} \epsilon_{\beta'\alpha'} = \bar{\xi}_{\alpha'} , \quad (\text{B.13})$$

with $\epsilon_{\alpha\beta} = -\epsilon_{\alpha'\beta'}$, $\epsilon_{\alpha\beta} = \epsilon^{\alpha\beta}$, and $\epsilon_{\alpha'\beta'} = \epsilon^{\alpha'\beta'}$. However, the relation between σ_μ and $\bar{\sigma}_\mu$ under complex conjugation is different (as expected because $\sigma^0 = I$ but $\sigma^4 = iI$). In Minkowski space we have $(\sigma_\mu^{\alpha\beta})^* = \bar{\sigma}_\mu^{\dot{\beta}\alpha}$, while in Euclidean space $(\sigma_\mu^{\alpha\beta'})^* = \bar{\sigma}_{\mu,\beta'\alpha} = \sigma_{\mu,\alpha\beta'}$ and $(\bar{\sigma}_{\mu,\alpha'\beta})^* = \sigma_\mu^{\beta\alpha'} = \bar{\sigma}_\mu^{\alpha'\beta}$.

Let us now apply these formulas to give another proof that $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ are pseudoreal in Euclidean space

$$\begin{aligned} ((\sigma_{\mu\nu})^\alpha_\beta)^* &= \frac{1}{2}(\sigma_\mu^{\alpha\beta'})^* (\bar{\sigma}_{\nu,\beta'\beta})^* - \mu \leftrightarrow \nu \\ &= \frac{1}{2}\sigma_{\mu,\alpha\beta'} \bar{\sigma}_\nu^{\beta'\beta} - \mu \leftrightarrow \nu = -\frac{1}{2}\sigma_{\mu,\alpha}^{\beta'} \bar{\sigma}_{\nu,\beta'}^\beta - \mu \leftrightarrow \nu \\ &= -\epsilon_{\gamma\alpha} (\sigma_{\mu\nu})^\gamma_\delta \epsilon^{\beta\delta} = (-i\sigma_2)(-\sigma_{\mu\nu})(-i\sigma_2) = \sigma_2 \sigma_{\mu\nu} \sigma_2 \end{aligned} \quad (\text{B.14})$$

and idem for $\bar{\sigma}_{\mu\nu}$.

The two spinor and vector representations of the $su(2)$ algebra are all given in terms of anti-hermitian 2x2 matrices $\sigma_{\mu\nu}, \bar{\sigma}_{\mu\nu}$ and $i\tau^a$ and they are related by the 't Hooft symbols,

$$\bar{\sigma}_{\mu\nu} = i\eta_{a\mu\nu}\tau^a, \quad \sigma_{\mu\nu} = i\bar{\eta}_{a\mu\nu}\tau^a. \quad (\text{B.15})$$

Furthermore, $\bar{\sigma}_{\mu\nu}$ is selfdual whereas $\sigma_{\mu\nu}$ is anti-selfdual. Some frequently used identities are

$$\begin{aligned} \bar{\sigma}_\mu \sigma_{\nu\rho} &= \delta_{\mu\nu} \bar{\sigma}_\rho - \delta_{\mu\rho} \bar{\sigma}_\nu - \epsilon_{\mu\nu\rho\sigma} \bar{\sigma}_\sigma, & \sigma_\mu \bar{\sigma}_{\nu\rho} &= \delta_{\mu\nu} \sigma_\rho - \delta_{\mu\rho} \sigma_\nu + \epsilon_{\mu\nu\rho\sigma} \sigma_\sigma, \\ \sigma_{\mu\nu} \sigma_\rho &= \delta_{\nu\rho} \sigma_\mu - \delta_{\mu\rho} \sigma_\nu + \epsilon_{\mu\nu\rho\sigma} \sigma_\sigma, & \bar{\sigma}_{\mu\nu} \bar{\sigma}_\rho &= \delta_{\nu\rho} \bar{\sigma}_\mu - \delta_{\mu\rho} \bar{\sigma}_\nu - \epsilon_{\mu\nu\rho\sigma} \bar{\sigma}_\sigma. \end{aligned} \quad (\text{B.16})$$

The Lorentz generators are antisymmetric in vector and symmetric in spinor indices

$$\sigma_{\mu\nu\alpha\beta} = -\sigma_{\nu\mu\alpha\beta}, \quad \sigma_{\mu\nu\alpha\beta} = \sigma_{\mu\nu\beta\alpha}, \quad (\text{B.17})$$

and obey the algebra

$$\begin{aligned} [\sigma_{\mu\nu}, \sigma_{\rho\sigma}] &= -2 \{ \delta_{\mu\rho} \sigma_{\nu\sigma} + \delta_{\nu\sigma} \sigma_{\mu\rho} - \delta_{\mu\sigma} \sigma_{\nu\rho} - \delta_{\nu\rho} \sigma_{\mu\sigma} \}, \\ \{ \sigma_{\mu\nu}, \sigma_{\rho\sigma} \} &= -2 \{ \delta_{\mu\rho} \delta_{\nu\sigma} - \delta_{\mu\sigma} \delta_{\nu\rho} - \epsilon_{\mu\nu\rho\sigma} \}. \end{aligned} \quad (\text{B.18})$$

The same relations hold for $\bar{\sigma}_{\mu\nu}$ but with $+\epsilon_{\mu\nu\rho\sigma}$. In spinor algebra the following contractions are frequently used

$$\sigma_\mu^{\alpha\alpha'} \bar{\sigma}_{\mu\beta'\beta} = 2\delta_\alpha^\beta \delta_{\alpha'}^{\beta'}, \quad \sigma_{\rho\sigma}^\alpha \sigma_{\rho\sigma}^\gamma{}_\delta = 4 \left\{ \delta_\beta^\alpha \delta_\delta^\gamma - 2\delta_\delta^\alpha \delta_\beta^\gamma \right\}. \quad (\text{B.19})$$

so that $\xi_{(1)}^\alpha \xi_{(2)\alpha} = \xi_{(2)}^\alpha \xi_{(1)\alpha}$. For hermitean conjugation we define $\left(\xi_{(1)}^\alpha \xi_{(2)\alpha} \right)^\dagger = (\xi_{(2)\alpha})^\dagger (\xi_{(1)}^\alpha)^\dagger$

$$\left(\sigma_\mu^{\alpha\beta'} \right)^* = \sigma_{\mu\alpha\beta'}, \quad \left(\bar{\sigma}_{\mu\alpha'\beta} \right)^* = \bar{\sigma}_\mu^{\alpha'\beta}. \quad (\text{B.20})$$

Throughout the paper we frequently use the following integral formula

$$\int d^4x \frac{(x^2)^n}{(x^2 + \rho^2)^m} = \pi^2 (\rho^2)^{n-m+2} \frac{\Gamma(n+2)\Gamma(m-n-2)}{\Gamma(m)}, \quad (\text{B.21})$$

which converges for $m - n > 2$.

C The volume of the gauge orientation moduli space

The purpose of this appendix⁵⁶ is to prove equation (6.15). Let us consider an instanton in $SU(N)$ gauge theory. Deformations of this configuration which are still self-dual and

⁵⁶We thank R. Roiban for help in writing this appendix.

not a gauge transformation are parametrized by collective coordinates. Constant gauge transformations $A_\mu \rightarrow U^{-1}A_\mu U$ preserve self-duality and transversality but not all constant $SU(N)$ matrices U change A_μ . Those U which keep A_μ fixed form the stability subgroup H of the instanton, hence we want to determine the volume of the coset space $SU(N)/H$. If the instanton is embedded in the lower-right 2×2 submatrix of the $N \times N$ $SU(N)$ matrix, then H contains the $SU(N-2)$ subgroup in the left-upper part, and a $U(1)$ subgroup with elements $\exp(\theta A)$ where A is the diagonal matrix

$$A = \frac{i}{2} \sqrt{\frac{N-2}{N}} \text{diag} \left(\frac{2}{2-N}, \dots, \frac{2}{2-N}, 1, 1 \right). \quad (\text{C.1})$$

All generators of $SU(N)$ (and also all generators of $SO(N)$ discussed below) are normalized according to $\text{tr } T_a T_b = -\frac{1}{2} \delta_{ab}$, as in the main text.

At first sight one might expect the range of θ to be such that the exponents of all entries cover the range 2π an integer number of times. However, this is incorrect: only for the last two entries of $\exp(\theta A)$ we must require periodicity, because whatever happens in the other $N-2$ diagonal entries is already contained in the $SU(N-2)$ part of the stability subgroup. Thus all elements h in H are of the form [52]

$$h = e^{\theta A} g, \quad \text{with} \quad g \in SU(N-2) \quad \text{and} \quad 0 \leq \theta \leq \theta_{\max} = 4\pi \sqrt{\frac{N}{N-2}}. \quad (\text{C.2})$$

For $N=3$ the range of θ is larger than required by periodicity of the first $N-2$ entries, for $N=4$ it corresponds to periodicity of all entries, but for $N \geq 5$ the range of θ is less than required for periodicity of the first $N-2$ entries.⁵⁷ Thus $H \neq SU(N) \times U(1)$ for $N \geq 5$. The first $N-2$ entries of $\exp(k\theta_{\max} A)$ with integer k are given by $\exp(-ik \frac{4\pi}{N-2})$ and lie therefore in the center Z_N of $SU(N-2)$. So, the $SU(N)$ group elements $h = \exp(\theta A) g$ with $0 \leq \theta \leq \theta_{\max}$ and g in $SU(N-2)$ form a subgroup H . We shall denote H by $SU(N-2) \times "U(1)"$ where " $U(1)$ " denotes the part of the $U(1)$ generated by A which lies in H . We now use three theorems to evaluate the volume of $SU(N)/H$:

$$\begin{aligned} \text{(I)} \quad & \text{Vol} \frac{SU(N)}{SU(N-2) \times "U(1)"} = \frac{\text{Vol}(SU(N)/SU(N-2))}{\text{Vol} "U(1)"}, \\ \text{(II)} \quad & \text{Vol} \frac{SU(N)}{SU(N-2)} = \text{Vol} \frac{SU(N)}{SU(N-1)} \text{Vol} \frac{SU(N-1)}{SU(N-2)}, \end{aligned} \quad (\text{C.3})$$

⁵⁷For example, consider $SU(5)$ with $\exp[\frac{i\theta}{2} \sqrt{\frac{3}{5}} \text{diag}(-\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3}, 1, 1)]$. When θ runs from 0 to $\sqrt{\frac{5}{3}}4\pi$, last two entries repeat, but the first three entries only reach $\exp(-4\pi i/3)$. The first three entries form then an element of $SU(N-2) = SU(3)$, namely they yield an element z of the center Z_3 . So when θ ranges beyond $\sqrt{\frac{5}{3}}4\pi$, these $SU(5)$ elements can be written as a product of z and $\exp i\theta A$ with θ smaller than $\sqrt{\frac{5}{3}}4\pi$. So, the range of θ is bounded by $\sqrt{\frac{5}{3}}4\pi$.

$$(III) \quad \text{Vol} \frac{SU(N)}{SU(N-1)} = \frac{\text{Vol } SU(N)}{\text{Vol } SU(N-1)} .$$

It is, in fact, easiest to first compute $\text{Vol}(SU(N)/SU(N-1))$ and then to use this result for the evaluation of $\text{Vol } SU(N)/H$ (with $\text{Vol } SU(N)$ as a bonus).

In general the volume of a coset manifold G/H is given by $V = \int \prod_{\mu} dx^{\mu} \det e_{\mu}^m(x)$ where x^{μ} are the coordinates on the coset manifold and $e_{\mu}^m(x)$ are the coset vielbeins. One begins with “coset representatives” $L(x)$ which are group elements $g \in G$ such that every group element can be decomposed as $g = L(x)h$ with $h \in H$. We denote the coset generators by K_m and the subgroup generators by H_i . Then $L^{-1}(x)\partial_{\mu}L(x) = e_{\mu}^m(x)K_m + \omega_{\mu}^i(x)H_i$. We shall take the generators K_m and H_i in the fundamental representation of $SU(N)$: antihermitian $N \times N$ matrices. Under a general coordinate transformation from x^{μ} to $y^{\mu}(x)$, the vielbein transforms as a covariant vector with index μ but also as a contravariant vector with index m at $x = 0$. Hence V *does (only) depend on the choice of the coordinates at the origin*. At the origin, $L^{-1}\partial_{\mu}L = e_{\mu}^m(0)K_m$, and we fix the normalization of K_m by $\text{tr } K_m^2 = -\frac{1}{2}$ for K_m in the $N \times N$ matrix representation of $SU(N)$.

To find the volume of $SU(N)/SU(N-1)$ we note that the group elements of $SU(N)$ have a natural action on the space $\mathbf{C}^N$ and map a point $(z^1, \dots, z^N) \in \mathbf{C}^N$ on the complex hypersphere $\sum_{i=1}^N |z^i|^2 = 1$ into another point on the complex hypersphere. The “south-pole” $(0, \dots, 0, 1)$ is kept invariant by the subgroup $SU(N-1)$, and points on the complex hypersphere are in one-to-one correspondence with the coset representatives $L(z)$ of $SU(N)/SU(N-1)$. We use as generators for $SU(N)$ the generators for $SU(N-1)$ in the upper-left block, and further the following coset generators: $N-1$ pairs T_{2k} and T_{2k+1} each of them containing only two non-zero elements

$$\begin{pmatrix} 0 & \dots & 0 \\ & & \vdots \\ \vdots & & i/2 \\ & \ddots & \vdots \\ 0 & i/2 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & \dots & 0 \\ & & \vdots \\ \vdots & & 1/2 \\ & \ddots & \vdots \\ 0 & -1/2 & \dots & 0 \end{pmatrix}, \quad (C.4)$$

and further one diagonal generator

$$T_{N^2-1} = \frac{i}{2} \sqrt{\frac{2}{N(N-1)}} \text{diag}(-1, \dots, -1, N-1). \quad (C.5)$$

(For instance, for $SU(3)$ there are two pairs, proportional to the usual λ_4 and λ_5 and λ_6 and λ_7 , and the diagonal hypercharge generator λ_8 .) The idea now is to establish a natural one-to-one correspondence between points in $\mathbf{C}^N$ and points in $\mathbf{R}^{2N}$, namely we write all points $(x^1, \dots, x^{2N})$ in $\mathbf{R}^{2N}$ as points in $\mathbf{C}^N$ as follows: $(ix^1 + x^2, \dots, ix^{2N-1} + x^{2N})$. In

particular the south pole $(0, 0, \dots, 0, 1)$ in $\mathbf{R}^{2N}$ corresponds to the south pole $(0, 0, \dots, 0, 1)$ in $\mathbf{C}^N$ and the sphere $\sum_{i=1}^{2N} (x^i)^2 = 1$ in $\mathbf{R}^{2N}$ corresponds to the hypersphere $\sum_{i=1}^N |z^i|^2 = 1$ in $\mathbf{C}^N$. Points on the sphere S^{2N-1} in $\mathbf{R}^{2N}$ correspond one-to-one to coset elements of $SO(2N)/SO(2N-1)$. The coset generators of $SO(2N)/SO(2N-1)$ are antisymmetric $2N \times 2N$ matrices A_I ($I = 1, \dots, 2N-1$) with the entry $+1/2$ in the last column and $-1/2$ in the last row. The coset element $1 + \delta g = 1 + dt^I A_I$ maps the south pole $s = (0, \dots, 0, 1)$ in $\mathbf{R}^{2N}$ to a point $s + \delta s$ in $\mathbf{R}^{2N}$ where $\delta s = 1/2(dt^1, \dots, dt^{2N-1}, 0)$. We know how points in $\mathbf{C}^N$ correspond to points in $\mathbf{R}^{2N}$, so we can ask which coset element in $SU(N)/SU(N-1)$ maps the south-pole in $\mathbf{C}^N$ to the point in $\mathbf{C}^N$ which corresponds to $s + \delta s$. In $\mathbf{C}^N$ the corresponding point is $s + \delta s$ with $\delta s = 1/2(idt^1 + dt^2, \dots, idt^{2N-1})$. The coset generators of $SU(N)/SU(N-1)$ act in $\mathbf{C}^N$ as follows: $g = 1 + dx^\mu K_\mu$ maps the south-pole s to $s + \delta s$ where now $\delta s = 1/2(idx^1 + dx^2, \dots, i\sqrt{\frac{2(N-1)}{N}}dx^{2N-1})$. We can cover $SO(2N)/SO(2N-1) = S^{2N-1}$ with small patches. Similarly we cover $SU(N)/SU(N-1)$ with small patches. Each patch of S^{2N-1} can be brought by the action of a suitable coset element to the south-pole, and then we can use the inverse of this group element to map this patch back into the manifold $SU(N)/SU(N-1)$. In this way both S^{2N-1} and $SU(N)/SU(N-1)$ are covered by patches which are in a one-to-one correspondence. Each pair of patches has the same ratio of volumes since both patches can be brought to the south pole by the same group element and at the south pole the ratio of their volumes is the same. To find the ratio of the volumes of S^{2N-1} and $SU(N)/SU(N-1)$, it is then sufficient to consider a small patch near the south pole. Near the south pole the vielbeins become unit matrices for coset manifolds, hence the volume of the patches near the south-pole is simply the product of the coordinates of these patches. Consider then a small patch at the south pole of S^{2N-1} with coordinates $(dt^1, \dots, dt^{2N-1})$ and volume $dt^1 \dots dt^{2N-1}$. The same patch at the south pole in $\mathbf{C}^N$ has coordinates dx^μ where $(idt^1 + dt^2, \dots, idt^{2N-1}) = \left(idx^1 + dx^2, \dots, i\sqrt{\frac{2(N-1)}{N}}dx^{2N-1}\right)$. The volume of a patch in $SU(N)/SU(N-1)$ with coordinates $dx^1, \dots, dx^{2N-1}$ is $dx^1 \dots dx^{2N-1}$. It follows that the volume of $SU(N)/SU(N-1)$ equals the volume of S^{2N-1} times $\sqrt{\frac{N}{2(N-1)}}$ ⁵⁸,

$$\text{Vol} \frac{SU(N)}{SU(N-1)} = \sqrt{\frac{N}{2(N-1)}} \text{Vol} S^{2N-1}. \quad (\text{C.6})$$

From here the evaluation of $\text{Vol} SU(N)/H$ is straightforward. Using

$$\text{Vol} S^{2N-1} = \frac{2\pi^N}{(N-1)!} l, \quad (\text{C.7})$$

⁵⁸This result yields the same answer for (6.15) as [52], but it yields $\pi^N/(N N!)$ for the volume of the complex projective space $CP(N) = SU(N+1)/(SU(N) \times U(1))$ which differs from the result $\text{Vol}[U(N+1)/(U(N) \times U(1))] = \text{Vol} S^{2N}$ given in [90].

where $l = 1$ if one uses the normalization $\text{tr } K_m^2 = -2$, but $l = 2^{2N-1}$ with our normalization of $\text{tr } K_m^2 = -\frac{1}{2}$, we obtain

$$\text{Vol } SU(N) = \sqrt{N} \prod_{k=2}^N \frac{\sqrt{2}\pi^k}{(k-1)!} 2^{2k-1} . \quad (\text{C.8})$$

We assumed that $\text{Vol } SU(1) = 1$ which seems a natural value but must be, and will be, justified below. Then

$$\begin{aligned} \text{Vol } H &= \text{Vol } SU(N-2) \text{Vol "U(1)" ,} & \text{Vol "U(1)" } &= 4\pi \sqrt{\frac{N}{N-2}} , \\ \text{Vol } SU(N)/H &= \frac{1}{2} \frac{\pi^{2N-2}}{(N-1)!(N-2)!} 2^{2N-1} 2^{2N-3} . \end{aligned} \quad (\text{C.9})$$

This then produces formula (6.15).

As an application and check of this analysis let us derive a few relations between the volumes of different groups. From now on till the end of this appendix we adopt the normalization $\text{tr}(T_a T_b) = -2\delta_{ab}$ for the generators of all groups involved. Let us check that $\text{Vol } SU(2) = 2\text{Vol } SO(3)$, $\text{Vol } SU(4) = 2\text{Vol } SO(6)$ and $\text{Vol } SO(4) = \frac{1}{2} (\text{Vol } SU(2))^2$ (the latter will follow from $SO(4) = SU(2) \times SU(2)/Z_2$). We begin with the usual formula for the surface of a sphere with unit radius (given already above for odd N)

$$\text{Vol } S^N = \frac{2\pi^{(N+1)/2}}{\Gamma\left(\frac{N+1}{2}\right)} . \quad (\text{C.10})$$

In particular $\text{Vol } S^1 = 2\pi$ and

$$\begin{aligned} \text{Vol } S^2 &= 4\pi , & \text{Vol } S^3 &= 2\pi^2 , & \text{Vol } S^4 &= \frac{8}{3}\pi^2 , \\ \text{Vol } S^5 &= \pi^3 , & \text{Vol } S^6 &= \frac{16}{15}\pi^3 , & \text{Vol } S^7 &= \frac{1}{3}\pi^4 . \end{aligned} \quad (\text{C.11})$$

Furthermore $\text{Vol } SO(2) = 2\pi$ since the $SO(2)$ generator with $\text{tr } T^2 = -2$ is $T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\exp(\theta T)$ is an ordinary rotation $\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$ for which $0 \leq \theta \leq 2\pi$. The vielbein is unity for an abelian group, and thus the Haar measure is⁵⁹ simply $d\theta$.

With $\text{Vol } SO(N) = \text{Vol } S^{N-1} \text{Vol } SO(N-1)$ we obtain $\text{Vol } SO(1) = 1$ and

$$\begin{aligned} \text{Vol } SO(2) &= 2\pi , & \text{Vol } SO(3) &= 8\pi^2 , & \text{Vol } SO(4) &= 16\pi^4 , \\ \text{Vol } SO(5) &= \frac{128}{3}\pi^6 , & \text{Vol } SO(6) &= \frac{128}{3}\pi^9 . \end{aligned} \quad (\text{C.12})$$

Now consider $SU(2)$. In the normalization $T_1 = -i\tau_1$, $T_2 = -i\tau_2$ and $T_3 = -i\tau_3$ (so that $\text{tr } T_a T_b = -2\delta_{ab}$) we find by direct evaluation⁶⁰ using Euler angles $\text{Vol } SU(2) = 2\pi^2$. This

⁵⁹One clearly must specify the normalization of the generators T_a ; for example by choosing $T_a = \begin{pmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{pmatrix}$, the range of θ becomes $0 \leq \theta \leq 4\pi$, but the Haar measure is still $d\theta$.
⁶⁰Parametrize $g = e^{\alpha T_3} e^{\beta T_1} e^{\gamma T_3}$, determine the range of α, β, γ and compute the group vielbeins.

also agrees with (C.6) and (C.8) for $N = 2$, justifying our assumption that $\text{Vol } SU(1) = 1$. For higher N we get

$$\text{Vol } SU(2) = 2\pi^2, \quad \text{Vol } SU(3) = \sqrt{3}\pi^5, \quad \text{Vol } SU(4) = \frac{\sqrt{2}}{3}\pi^9. \quad (\text{C.13})$$

The group elements of $SU(2)$ can also be written as $g = x^4 + i\vec{\tau} \cdot \vec{x}$ with $(x^4)^2 + (\vec{x})^2 = 1$ which defines a sphere S^3 . Since near the unit element $g \approx 1 + i\vec{\tau} \cdot \delta\vec{x}$, the normalization of the generators is as before, and hence for this parametrization $\text{Vol } SU(2) = 2\pi^2$. This is indeed equal to $\text{Vol } S^3$. In the mathematical literature one finds the statement that $\text{Vol } SU(2)$ is twice $\text{Vol } SO(3)$ because $SU(2)$ is the double covering group of $SO(3)$. However, we have just found that $\text{Vol } SU(2) = \frac{1}{4}\text{Vol } SO(3)$. The reason is that in order to compare properties of different groups we should normalize the generators such that the structure constants are the same (the Lie algebras are the same, although the group volumes are not). In other words, we should use the normalization that the **adjoint** representations have the same $\text{tr } T_a T_b$. For $SU(2)$ the generators which lead to the same commutators as the usual $SO(3)$ rotation generators (with entries $+1$ and -1) are $T_a = \{-\frac{i}{2}\tau_1, -\frac{i}{2}\tau_2, -\frac{i}{2}\tau_3\}$. Then $\text{tr } T_a T_b = -\frac{1}{2}\delta_{ab}$. In this normalization, the range of each group coordinate is multiplied by 2, leading to $\text{Vol } SU(2) = 2^3 \cdot 2\pi^2 = 16\pi^2$. Now indeed $\text{Vol } SU(2) = 2\text{Vol } SO(3)$.

For $SU(4)$ the generators with the same Lie algebra as $SO(6)$ are the 15 antihermitean 4×4 matrices $\frac{1}{4}(\gamma_m \gamma_n - \gamma_n \gamma_m)$, $i\gamma_m/2$, $\gamma_m \gamma_5/2$ and $i\gamma_5/2$, where γ_m and γ_5 are the five 4×4 matrices γ_M obeying the Clifford algebra $\{\gamma_M, \gamma_N\} = 2\delta_{MN}$ ⁶¹. Now, $\text{tr } T_a T_b = -\delta_{ab}$ (for example, $\text{tr } \{(\frac{1}{2}\gamma_1 \gamma_2)(\frac{1}{2}\gamma_1 \gamma_2)\} = -1$). Recall that originally we had chosen the normalization $\text{tr } T_a T_b = -2\delta_{ab}$. We must thus multiply the range of each coordinate by a factor $\sqrt{2}$, and hence we must multiply our original result for $\text{Vol } SU(4)$ by a factor $(\sqrt{2})^{15}$. We find then indeed that the relation $\text{Vol } SU(4) = 2\text{Vol } SO(6)$ is fulfilled.

Finally, we consider the relation $SO(4) = SU(2) \times SU(2)/Z_2$. (The vector representation of $SO(4)$ corresponds to the representation $(\frac{1}{2}, \frac{1}{2})$ of $SU(2) \times SU(2)$, but representations like $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ are not representations of $SO(4)$ and hence we must divide by Z_2 . The reasoning is the same as for $SU(2)$ and $SO(3)$, or $SU(4)$ and $SO(6)$.) We choose the generators of $SO(4)$ as follows

$$T_1^{(+)} = \frac{1}{\sqrt{2}}(L_{14} + L_{23}), \quad T_2^{(+)} = \frac{1}{\sqrt{2}}(L_{31} + L_{24}), \quad T_3^{(+)} = \frac{1}{\sqrt{2}}(L_{12} + L_{34}), \quad (\text{C.14})$$

and the same but with minus sign denoted by $T_i^{(-)}$. Here L_{mn} equals $+1$ in the m^{th} column and n^{th} row, and is antisymmetric. Clearly $\text{tr } T_a T_b = -2\delta_{ab}$. The structure constants follow from

$$\left[\frac{1}{\sqrt{2}}(L_{12} + L_{34}), \frac{1}{\sqrt{2}}(L_{14} + L_{23}), \right] = -(L_{31} + L_{24}), \quad (\text{C.15})$$

⁶¹As Dirac matrices in six dimensions we take $\gamma_m \otimes \tau_2, \gamma_5 \otimes \tau_2$ and $I \times \tau_3$.

thus

$$\left[T_i^{(+)}, T_j^{(+)}\right] = -\sqrt{2}\epsilon_{ijk}T_k^{(+)} , \quad \left[T_i^{(-)}, T_j^{(-)}\right] = -\sqrt{2}\epsilon_{ijk}T_k^{(-)} , \quad \left[T_i^{(+)}, T_j^{(-)}\right] = 0 . \quad (\text{C.16})$$

We choose for the generators of $SU(2) \times SU(2)$ the representation

$$T_i^{(+)} = \frac{i\tau_i}{\sqrt{2}} \otimes \mathbb{1} , \quad T_i^{(-)} = \mathbb{1} \otimes \frac{i\tau_i}{\sqrt{2}} . \quad (\text{C.17})$$

Then we get the same commutation relations as for $SO(4)$ generators (C.16); however, the generators are normalized differently, namely $\text{tr } T_a T_b = -2\delta_{ab}$ for $SO(4)$ but $\text{tr } T_a T_b = -\delta_{ab}$ for $SU(2)$. With the normalization $\text{tr } T_a T_b = -2\delta_{ab}$ we found $\text{Vol } SU(2) = 2\pi^2$. In the present normalization we find $\text{Vol } SU(2) = 2\pi^2 (\sqrt{2})^3$. The relation $\text{Vol } SO(4) = \frac{1}{2} (\text{Vol } SU(2))^2$ is now indeed satisfied

$$\text{Vol } SO(4) = 16\pi^4 = \frac{1}{2} (\text{Vol } SU(2))^2 = \frac{1}{2} \left(2\pi^2 (\sqrt{2})^3 \right)^2 . \quad (\text{C.18})$$

D Zero modes and conformal symmetries

The bosonic collective coordinates obtained for gauge group $SU(2)$ and the one-instanton solution could all be identified with rigid symmetries of the action: a_μ with translations, ρ with scale transformations and θ^a with rigid gauge symmetries. Similarly, the fermionic collective coordinates for $SU(2)(\xi^\alpha$ and $\bar{\eta}_{\dot{\alpha}}$ with $\alpha, \dot{\alpha} = 1, 2)$ could be identified with ordinary supersymmetry and conformal supersymmetry. However, the full conformal algebra in 4 Euclidean dimensions is $SO(5, 1)$, and its generators are $P_\mu, K_\mu, D, M_{\mu\nu}$, so one might expect that the conformal boost transformations K_μ and the Lorentz rotations $M_{\mu\nu}$ produce further collective coordinates. As we now show, the transformations due to these symmetries can be undone by suitably chosen gauge transformations with constant gauge parameters [29]. So there are no further bosonic collective coordinates, as we already know from the index theorem discussed in the main text.

Consider first rigid Lorentz transformations. Here one should not forget that in addition to a spin part which acts on the indices of a field they also contain an orbital part that acts on the coordinates: $M_{\mu\nu} = \Sigma_{\mu\nu} + L_{\mu\nu}$. For example, for a spinor one has $\delta(\lambda_{mn})\psi = \frac{1}{4}\lambda_{mn}\gamma_{mn}\psi + (\lambda_{mn}x_m\partial_n)\psi$. One may check that only with this orbital part present the Dirac action is Lorentz invariant. In fact, starting with only the spin part or the orbital part, one can find the other part by requiring invariance of the action. We begin by considering the field strength $F_{\mu\nu} = 2\bar{\sigma}_{\mu\nu}\rho^2/(x^2 + \rho^2)^2$ for an instanton with $k = 1$ in the regular gauge. Under a Lorentz transformation with parameter $\lambda_{\mu\nu} = -\lambda_{\nu\mu}$ one has $\delta_M A_\mu = \lambda_{\mu\nu}A_\nu + \lambda_{mn}x_m\partial_n A_\nu$. (Note that coordinates transform opposite to fields: $\delta x^m = -\lambda_{mn}x^n$). One may check this transformation rule by showing that the Maxwell action is Lorentz

invariant (use the Bianchi identities⁶²), or just by writing down the transformation law for a covariant vector in general relativity. The Lagrangian transforms into $\partial_\mu(\xi^\mu \mathcal{L})$, where $\xi^\mu = \lambda^{\rho\mu} x_\rho$. The field strength of the instanton transforms as follows

$$\delta_M F_{\mu\nu} = \lambda_{\mu\rho} F_{\rho\nu} + \lambda_{\nu\rho} F_{\mu\rho} \quad (\text{D.1})$$

There is no contribution from the orbital part because x^2 is Lorentz invariant. On the other hand, under a gauge transformation with parameter $\Lambda_{\rho\sigma}$ we obtain⁶³

$$\delta_{\text{gauge}} F_{\mu\nu} = [\bar{\sigma}_{\mu\nu}, \frac{1}{4} \Lambda_{\rho\sigma} \bar{\sigma}_{\rho\sigma}] (2\rho^2 / (x^2 + \rho^2)^2) = \Lambda_{\nu\sigma} F_{\mu\sigma} - \Lambda_{\mu\sigma} F_{\nu\sigma} \quad (\text{D.2})$$

Thus $F_{\mu\nu}$ is invariant under combined Lorentz and gauge transformations with opposite parameters, $\Lambda_{\rho\sigma} = -\lambda_{\rho\sigma}$. Using $\bar{\sigma}_{\rho\sigma} = i\eta_{a\rho\sigma}\tau_a$, it is clear that the $SU(2)$ gauge parameter Λ_a is proportional to $\eta_{a\rho\sigma}\lambda_{\rho\sigma}$. Only the selfdual part of $\lambda_{\rho\sigma}$ contributes. For an anti-instanton we would have needed the anti-selfdual part of $\lambda_{\rho\sigma}$. So we have only proven that $F_{\mu\nu}$ is invariant under combined Lorentz and gauge transformation if the Lorentz parameter is self dual. However, the anti-self dual part of $\lambda_{\rho\sigma}$ leaves $F_{\mu\nu}$ separately invariant, without the need to add compensating gauge transformations. One can prove this directly, using that $\lambda_{\mu\rho} F_{\rho\nu} = -(*\lambda_{\mu\rho})(*F_{\rho\nu})$ and then working out the product of two ϵ -tensors and finally antisymmetrizing in $\mu\nu$, but it is already clear from the index structure: $F_{\mu\nu}$ is proportional to $(\bar{\sigma}_{\mu\nu})_{\alpha'\beta'}$ while an anti-selfdual $\lambda_{\rho\sigma}$ has in spinor notation only undotted indices.

Let us now repeat this exercise for the gauge field A_μ . One finds for the combined Lorentz and gauge transformation

$$\delta A_\mu = \lambda_{\mu\nu} A_\nu + \lambda_{\rho\sigma} x_\rho \partial_\sigma A_\mu + [A_\mu, \frac{1}{4} \Lambda_{\rho\sigma} \bar{\sigma}_{\rho\sigma}] \quad (\text{D.3})$$

The instanton field A_μ for $k = 1$ in the regular gauge is given by $A_\mu = (-\bar{\sigma}_{\mu\nu} x^\nu) / (x^2 + \rho^2)$. The orbital part with $\lambda_{\rho\sigma}$ now contributes, but there is no term $\partial_\mu \Lambda^a$ in the gauge transformation of A_μ since Λ^a is constant. One obtains

$$\delta A_\mu = \lambda_{\mu\nu} A_\nu + \frac{\bar{\sigma}_{\mu\nu}(\lambda_{\nu\rho} x_\rho)}{x^2 + \rho^2} - \frac{(\Lambda_{\nu\sigma} \bar{\sigma}_{\mu\sigma} - \Lambda_{\mu\sigma} \bar{\sigma}_{\nu\sigma}) x^\nu}{x^2 + \rho^2} \quad (\text{D.4})$$

For $\Lambda_{\mu\nu} = -\lambda_{\mu\nu}$ all terms again cancel. Hence, Lorentz symmetry does not yield further zero modes.

In spinor notation these results are almost obvious. In general the selfdual part of a curvature reads in spinor notation

$$(F_{\mu\nu})^u{}_v (\bar{\sigma}_{\mu\nu})_{\alpha'\beta'} \quad (\text{D.5})$$

⁶²One has $\delta_M \frac{1}{4} F_{\mu\nu}^2 = F_{\mu\nu} \partial_\mu (\lambda_{\nu\rho} A_\rho) + F_{\mu\nu} \partial_\mu (\lambda_{mn} x_m \partial_n A_\nu) = F_{\mu\nu} \lambda_{mn} x_m \partial_\mu \partial_n A_\nu$. Replacing $\partial_\mu \partial_n A_\nu$ by $-\partial_n \partial_\nu A_\mu - \partial_\nu \partial_\mu A_n$ yields $\partial_\mu(\xi^\mu \mathcal{L})$.

⁶³The usual form of an $SU(2)$ gauge transformation is $\delta F_{\mu\nu} = [F_{\mu\nu}, \Lambda^a(x) \frac{\tau_a}{2i}]$, but using $\eta_{a\mu\nu} \eta_{b\mu\nu} = 4\delta_{ab}$ and $\bar{\sigma}_{\rho\sigma} = i\eta_{a\rho\sigma} \tau_a$, this can be rewritten as $\delta F_{\mu\nu} = [F_{\mu\nu}, \frac{1}{4} \Lambda^{\rho\sigma} \bar{\sigma}_{\rho\sigma}]$ where $\Lambda_{\rho\sigma} = -\frac{1}{2} \eta_{a\rho\sigma} \Lambda^a$.

where u, v are the indices of $(\tau^a)^u_v$, and α', β' are the spinor indices. If we raise/lower indices by ϵ tensors, we get for the instanton solution

$$(F_{\mu\nu})^{uv}(\bar{\sigma}_{\mu\nu})_{\alpha'\beta'} \equiv F_{\alpha'\beta'}^{uv} \sim \delta_{\alpha'}^u \delta_{\beta'}^v + \delta_{\alpha'}^v \delta_{\beta'}^u \quad (\text{D.6})$$

It is then clear that $F_{\mu\nu}$ is invariant under diagonal transformations of $SU(2)_R$ and $SU(2)_{\text{gauge}}$, and separately invariant under $SU(2)_L$. For an anti-instanton, the roles of $SU(2)_L$ and $SU(2)_R$ are interchanged.

We come now to the more complicated problem of conformal transformations. A conformal transformation of a field φ with constant parameter a^m is given by⁶⁴

$$\begin{aligned} \delta(a^m K_m)\varphi &= (2a \cdot x x^m - a^m x^2)\partial_m \varphi + \delta(2a \cdot x D^{(\text{spin})})\varphi \\ &+ \delta(2a_m x_n M_{mn}^{(\text{spin})})\varphi \end{aligned} \quad (\text{D.7})$$

where D^{spin} and M_{mn}^{spin} act only on $\varphi(0)$ and $\delta(a^m K_m)\varphi(x)$ is by definition $[\varphi(x), a^m K_m]$. As the notation indicates, only the spin parts of the dilatational generator D and the Lorentz generators contribute. For example

$$\delta(D^{\text{spin}})A_\mu = [A_\mu, D^{(\text{spin})}] = A_\mu, \quad \delta(\tfrac{1}{2}\lambda_{mn}M_{mn}^{(\text{spin})})A_\mu = \lambda_{\mu\nu}A_\nu. \quad (\text{D.8})$$

Consider first $F_{\mu\nu}$. We obtain

$$\begin{aligned} \delta(a^m K_m)F_{\mu\nu} &= (2a \cdot x x^m - a^m x^2)\partial_m F_{\mu\nu} + 4a \cdot x F_{\mu\nu} \\ &+ 4\delta(\tfrac{1}{2}a_m x_n M_{mn}^{(\text{spin})})F_{\mu\nu} \text{ with } F_{\mu\nu} = \frac{2\bar{\sigma}_{\mu\nu}\rho^2}{(x^2 + \rho^2)^2}. \end{aligned} \quad (\text{D.9})$$

We already know that the last term can be canceled by a suitable gauge transformation (there are no contributions from $M_{mn}^{(\text{orb})}$ because x^2 is Lorentz invariant). The first term gives $-4\frac{a \cdot x x^2}{x^2 + \rho^2}F_{\mu\nu}$. The first and second term together produce then $\frac{4a \cdot x \rho^2}{x^2 + \rho^2}F_{\mu\nu}$. But this is the opposite of a translation with parameter $a^m \rho^2$, namely

$$\delta(a^m \rho^2 P_m)F_{\mu\nu} = \frac{-4a \cdot x \rho^2}{x^2 + \rho^2}F_{\mu\nu}; \delta(P_m)\varphi = \partial_m \varphi. \quad (\text{D.10})$$

⁶⁴This formula follows from $\delta(a^m K_m)\varphi(x) = [\varphi(x), a^m K_m]$, and $\varphi(x) = e^{-P \cdot x}\varphi(0)e^{P \cdot x}$ with $[\varphi(0), P_\mu] = \partial_\mu \varphi(0)$. One may then use $e^{P \cdot x}K_m = (e^{P \cdot x}K_m e^{-P \cdot x})e^{P \cdot x}$ and $[K_m, P_n] = -2\delta_{mn}D - 2M_{mn}$; $[P_m, D] = P_m$; $[P_m, M_{rs}] = \delta_{mr}P_s - \delta_{ms}P_r$ and this yields (D.7). In the same way one may derive the Lorentz transformation rule for a spinor $\psi(x)$, with both spin and orbital parts, by using that the spin part is given by $[\psi(0), \tfrac{1}{2}\lambda_{mn}M_{mn}] = \tfrac{1}{4}\lambda_{mn}\gamma_{mn}\psi(0)$. One finds then the correct result: $\delta(\tfrac{1}{2}\lambda_{mn}M_{mn})\psi(x) = \tfrac{1}{4}\lambda_{mn}\gamma_{mn}\psi(x) + \lambda_{mn}x_m \partial_n \psi(x)$. Given the spin part of the transformation rule of the field at the origin, one derives in this way the orbital part. In this way one finds that the generators of the conformal algebra act as follows on the coordinates: $\delta(P_m)x^n = \delta_m^n$, $\delta(D)x^n = x^n$, $\delta(M_{st})x^m = x_s \delta_t^m - x_t \delta_s^m$ and $\delta(K_m)x^n = 2x_m x^n - x^2 \delta_m^n$. Note that coordinates transform contragradiently to fields. For example, whereas $[\delta(K_m), \delta(P_n)]\varphi = -\delta([K_m, P_n])\varphi$ (by definition), one finds $[\delta(K_m), \delta(P_n)]x^s = \delta([K_m, P_n])x^s$.

Thus the following combination of symmetry transformations leaves $F_{\mu\nu}$ invariant

$$a^m K_m + \rho^2 a^m P_m + \delta_{\text{gauge}}(\Lambda_{mn} = -2a_m x_n + 2x_m a_n) \quad (\text{D.11})$$

Let us now check that also A_μ itself is invariant under this combination of symmetries. We find by direct evaluation, using $A_\mu = (-\bar{\sigma}_{\mu\nu} x^\nu)/(x^2 + \rho^2)$ and (D.9) and (D.2)

$$\begin{aligned} \delta A_\mu &= \left(\frac{-2a \cdot x x^2}{x^2 + \rho^2} A_\mu - \bar{\sigma}_{\mu\nu} \frac{(2a \cdot x x_\nu - a_\nu x^2)}{x^2 + \rho^2} \right) + 2a \cdot x A_\mu \\ &+ (2a_\mu x_\nu A_\nu - 2x_\mu a_\nu A_\nu) + \left(-\frac{\rho^2 2a \cdot x}{x^2 + \rho^2} A_\mu - \frac{\bar{\sigma}_{\mu\nu} a_\nu \rho^2}{x^2 + \rho^2} \right) \\ &+ \partial_\mu (-a_\rho x_\sigma \bar{\sigma}_{\rho\sigma}) + [A_\mu, -a_\rho x_\sigma \bar{\sigma}_{\rho\sigma}] . \end{aligned} \quad (\text{D.12})$$

As in the case of $F_{\mu\nu}$, the sum of the first, third and sixth term cancels. This takes care of the dilatation term and the denominator of A_μ . We are left with terms from the numerator, and Lorentz and gauge terms

$$\begin{aligned} &(2a \cdot x A_\mu + (\bar{\sigma}_{\mu\rho} a_\rho) \left(\frac{x^2}{x^2 + \rho^2} \right) + (0 - 2x_\mu a_\nu A_\nu) + \\ &\left(\frac{-\bar{\sigma}_{\mu\nu} a_\nu \rho^2}{x^2 + \rho^2} \right) + (\bar{\sigma}_{\mu\rho} a_\rho) + \frac{2x^\nu}{x^2 + \rho^2} \begin{pmatrix} a_\nu x_\sigma \bar{\sigma}_{\mu\sigma} - a_\rho x_\nu \bar{\sigma}_{\mu\rho} \\ +0 + a_\rho x_\mu \bar{\sigma}_{\nu\rho} \end{pmatrix} \end{aligned} \quad (\text{D.13})$$

The terms denoted by “0” vanish due to $x_\nu A_\nu = 0$. All other terms cancel in the following combinations

- (i) the second, fourth, fifth, and seventh nonvanishing contributions sum up to zero. These are the terms with $\bar{\sigma}_{\mu\rho} a_\rho$.
- (ii) the first and third-but-last nonvanishing term cancel each other. Here conformal boosts cancel a gauge term.
- (iii) the remaining Lorentz term $-2x_\mu a_\nu A_\nu$ cancels the remaining gauge term $2x^\nu (x^2 + \rho^2)^{-1} (a_\rho x_\mu \bar{\sigma}_{\nu\rho})$.

Hence, conformal boosts do not lead to further zero modes either.

References

- [1] A.V. Belitsky, S. Vandoren, P. van Nieuwenhuizen, *Yang-Mills and D-instantons*, *Class. Quant. Grav.* **17** (2000) 3521, hep-th/0004186.
- [2] R. Rajaraman, *Solitons and Instantons*, North-Holland, (Amsterdam, 1982);
S. Coleman, *The uses of instantons and Classical lumps and their quantum descendants*, in Proc. Int. School of Subnuclear Physics, Erice (1977), reprinted in *Aspects of Symmetry*, Cambridge University Press, (Cambridge, 1985) 265;

- A. Vainshtein, V. Zakharov, V. Novikov and M. Shifman, *ABC of Instantons*, Sov. Phys. Usp. 25 (1982) 195; *Instantons in Gauge Theories*, M. Shifman, World Scientific, (Singapore, 1994);
- M. Shifman, A. Vainshtein, *Instanton versus Supersymmetry: Fifteen Years Later*, ITEP Lectures, Edt by M. Shifman, World Scientific, Singapore 1999, Vol. 2, 485, hep-th/9902018;
- N. Dorey, T.J. Hollowood, V.V. Khoze and M.P. Mattis, *The calculus of many instantons*, Phys. Rept. **371** (2002) 231, hep-th/0206063;
- A.S. Goldhaber, A. Rebhan, P. van Nieuwenhuizen and R. Wimmer, *Quantum corrections to mass and central charge of supersymmetric solitons*, Phys. Rept. **398** (2004) 179, hep-th/0401152;
- D. Tong, *TASI lectures on solitons: Instantons, monopoles, vortices and kinks*, hep-th/0509216.
- [3] A. Belavin, A. Polyakov, A. Schwartz, Y. Tyupkin, *Pseudo-particle solutions of the Yang-Mills equations*, Phys. Lett. B **59** (1975) 85.
- [4] G. 't Hooft, *Computation of the quantum effects due to a four-dimensional pseudoparticle*, Phys. Rev. D **14** (1976) 3432.
- [5] G. 't Hooft, *Symmetry breaking through Bell-Jackiw anomalies*, Phys. Rev. Lett. **37** (1976) 8.
- [6] E. Gildener and A. Patrascioiu, *Instanton contributions to the energy spectrum of a one-dimensional system*, Phys. Rev. D **16** (1977) 423.
- [7] G. Gibbons, M. Green, M. Perry, *Instantons and seven-branes in type IIB superstring theory*, Phys. Lett. B **370** (1996) 37, hep-th/9511080.
- [8] Instantons can also be defined in dimensions higher than four. In terms of complex coordinates z^α and $\bar{z}^\alpha$ in even dimensions the equations are $F_{\alpha\beta} = F_{\bar{\alpha}\bar{\beta}} = 0$ and $\sum_\alpha F_{\alpha\bar{\alpha}} = 0$. In 8 dimensions one can define instantons by $F_{\alpha\beta} = \frac{1}{2} \sum_{\gamma,\delta} \epsilon_{\alpha\beta\gamma\delta} F_{\bar{\gamma}\bar{\delta}}$ and $\sum_\alpha F_{\alpha\bar{\alpha}} = 0$. One can rewrite the first relation as $F_{8a} = \frac{1}{2} f_{abc} F_{bc}$ where f_{abc} are the octonionic structure constants. See: E. Corrigan, C. Devchand, D.B. Fairlie and J. Nuyts, *First order equations for gauge fields in spaces of dimension greater than four*, Nucl. Phys. B **214** (1983) 452;
- S. Fubini and H. Nicolai, *The octonionic instanton*, Phys. Lett. B **155** (1985) 369;
- J. Harvey and A. Strominger, *Octonionic superstring solitons*, Phys. Rev. Lett. **66** (1991) 549.

- [9] S.W. Hawking, *Gravitational instantons*, Phys. Lett. A **60** (1977) 81;
 T. Eguchi and A.J. Hanson, *Asymptotically flat selfdual solutions to euclidean gravity*, Phys. Lett. B **74** (1978) 249;
 G.W. Gibbons and S.W. Hawking, *Classification of gravitational instanton symmetries*, Commun. Math. Phys. **66** (1979) 291;
 G. 't Hooft, *A physical interpretation of gravitational instantons*, Nucl. Phys. B **315** (1989) 517;
 G.M. Gibbons and S.W. Hawking, *Euclidean Quantum Gravity*, World Scientific, Singapore, 1993.
- [10] A.A. Abrikosov, *Instantons and multi-instantons in curvilinear coordinates*, Nucl. Phys. B **586** (2000) 589.
- [11] C. Callan Jr., R. Dashen, D. Gross, *The structure of the gauge theory vacuum*, Phys. Lett. B **63** (1976) 334; *Toward a theory of the strong interactions*, Phys. Rev. D **17** (1978) 2717.
- [12] R. Jackiw, C. Rebbi, *Vacuum periodicity in a Yang-Mills quantum theory*, Phys. Rev. Lett. **37** (1976) 172.
- [13] A. Polyakov, *Quark confinement and topology of gauge groups*, Nucl. Phys. B **120** (1977) 429.
- [14] A. D'Adda and P. Di Vecchia, *Supersymmetry and Instantons*, Phys. Lett. B **73** (1978) 162.
- [15] V. Novikov, M. Shifman, A. Vainshtein and V. Zakharov, *Supersymmetric instanton calculus (gauge theories with matter)*, Nucl. Phys. B **260** (1985) 157; and *Instanton Effects In Supersymmetric Theories*, Nucl. Phys. B **229** (1983) 407.
- [16] I. Affleck, M. Dine, N. Seiberg, *Supersymmetry breaking by instantons*, Phys. Rev. Lett. **51** (1983) 1026; *Dynamical supersymmetry breaking in supersymmetric QCD*, Nucl. Phys. B **241** (1984) 493 and *Dynamical supersymmetry breaking in four-dimensions and its phenomenological implications*, Nucl. Phys. B **256** (1985) 557;
 D. Amati, G. Rossi and G. Veneziano, *Instanton effects in supersymmetric gauge theories*, Nucl. Phys. B **249** (1985) 1.
- [17] D. Amati, K. Konishi, Y. Meurice, G. Rossi, G. Veneziano, *Nonperturbative effects in supersymmetric gauge theories*, Phys. Rep. **162** (1988) 169.

- [18] N. Seiberg, E. Witten, *Electric-magnetic duality, monopole condensation, and confinement in $N=2$ supersymmetric Yang-Mills theory*, Nucl. Phys. B **426** (1994) 19, (Erratum: Nucl. Phys. B **430** (1994) 485); *Monopoles, duality and chiral symmetry breaking in $N=2$ supersymmetric QCD*, Nucl. Phys. **431** (1994) 484.
- [19] D. Finnell, P. Pouliot, *Instanton calculations versus exact results in four-dimensional susy gauge theories*, Nucl. Phys. B **453** (1995) 225, hep-th/9503115;
K. Ito, N. Sasakura, *One instanton calculations in $N=2$ supersymmetric $SU(N(C))$ Yang-Mills theory*, Phys. Lett. B **382** (1996) 95, hep-th/9602073; *Exact and microscopic one instanton calculations in $N=2$ supersymmetric Yang-Mills theories*, Nucl. Phys. B **484** (1997) 141, hep-th/9608054.
- [20] N. Dorey, V.V. Khoze, M. P. Mattis, *Multi-instanton calculus in $N=2$ supersymmetric gauge theory*, Phys. Rev. D **54** (1996) 2921, hep-th/9603136; *Multi-instanton calculus in $N=2$ supersymmetric gauge theory. 2. Coupling to matter*, Phys. Rev. D **54** (1996) 7832, hep-th/9607202;
F. Fucito, G. Travaglini, *Instanton calculus and nonperturbative relations in $N = 2$ supersymmetric gauge theories*, Phys. Rev. D **55** (1997) 1099;
H. Aoyama, T. Harano, M. Sato, S. Wada, *Multi-instanton calculus in $N = 2$ supersymmetric QCD*, Phys. Lett. B **338** (1996) 331, hep-th/9607076.
- [21] N. Nekrasov, *Seiberg-Witten prepotential from instanton counting*, Adv. Theor. Math **7** (2004) 831, hep-th/0206161; N. Nekrasov and S. Shadchin, *The ABCD of instantons*, Commun. Math. Phys. **252** (2004) 359, hep-th/0404225.
- [22] J. Maldacena, *The large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998) 231, hep-th/9711200;
S. Gubser, I. Klebanov, A. Polyakov, *Gauge theory correlators from non-critical string theory*, Phys. Lett. B **428** (1998) 105, hep-th/9802109;
E. Witten, *Anti-de Sitter space and holography*, Adv. Theor. Math. Phys. **2** (1998) 253, hep-th/9802150.
- [23] M. Bianchi, M. Green, S. Kovacs, G. Rossi, *Instantons in supersymmetric Yang-Mills and D-instantons in IIB superstring theory*, JHEP **9808** (1998) 013, hep-th/9807033.
- [24] N. Dorey, V.V. Khoze, M.P. Mattis, S. Vandoren, *Yang-Mills instantons in the large- N limit and the AdS/CFT correspondence*, Phys. Lett. B **442** (1998) 145, hep-th/9808157.
- [25] N. Dorey, T. Hollowood, V.V. Khoze, M.P. Mattis, S. Vandoren, *Multi-instantons and Maldacena's conjecture*, JHEP **9906** (1999) 023, hep-th/9810243; *Multi-instanton*

- calculus and the AdS/CFT correspondence in $N = 4$ superconformal field theory*, Nucl. Phys. B **552** (1999) 88, hep-th/9901128.
- [26] M. Green and S. Kovacs, *Instanton induced Yang-Mills correlation functions at large N and their $AdS_5 \times S^5$ duals*, JHEP **0304:058** (2003), hep-th/0212332.
 - [27] S. Kovacs, *On instanton contributions to anomalous dimensions in $N = 4$ supersymmetric Yang-Mills theory*, Nucl. Phys. B **684** (2004) 3, hep-th/0310193.
 - [28] M. Bianchi, S. Kovacs and G. Rossi, *Instantons and supersymmetry*, in 'String Theory and Fundamental Interactions'. Edited by M. Gasperini and J. Maharana. Lecture Notes in Physics, Springer, 2007, hep-th/0703142.
 - [29] R. Jackiw, C. Rebbi, *Conformal properties of a Yang-Mills pseudoparticle*, Phys. Rev. D **14** (1976) 517; R. Jackiw, C. Nohl and C. Rebbi, *Conformal properties of pseudoparticle configurations*, Phys. Rev. D **15** (1977) 1642.
 - [30] F. Wilczek, *Inequivalent embeddings of $SU(2)$ and instanton interactions*, Phys. Lett. B **65** (1976) 160.
 - [31] M. Atiyah, V. Drinfeld, N. Hitchin, Yu. Manin, *Construction of instantons*, Phys. Lett. A **65** (1978) 185. In this paper the general multi-instanton solution is given in terms of a set of complicated quadratic equations for quaternions. However, for the 3-instanton case, this system is equivalent to a set of linear equations, see [35].
 - [32] N. Dorey, T. Hollowood, V. Khoze and M. Matthis, *The calculus of many instantons*, Phys. Rept. **371** (2002) 231, hep-th/0206063. The 2-instanton solution with all moduli is also explicitly known but one needs the ADHM formalism.
 - [33] H. Osborn, *Semiclassical Functional Integrals For Selfdual Gauge Fields*, Annals Phys. **135** (1981) 373.
 - [34] N. H. Christ, E. J. Weinberg and N. K. Stanton, *General Self-Dual Yang-Mills Solutions*, Phys. Rev. D **18** (1978) 2013.
 - [35] V. Korepin and S. Shatashvili, *Rational parametrization of the three instanton solutions of the Yang-Mills equations*, Math. USSR Izversiya **24** (1985) 307.
 - [36] C. Bernard, N. Christ, A. Guth, E. Weinberg, *Pseudoparticle parameters for arbitrary gauge groups*, Phys. Rev. D **16** (1977) 2967.

- [37] S. Vandoren, *Instantons and quaternions*, in the proceedings of the 4th Annual European TMR Conference on Integrability, Nonperturbative Effects and Symmetry in Quantum Field Theory, Paris, France, 7-13 Sep. 2000, hep-th/0009150.
- [38] J.P. Bourguignon, H.B. Lawson and J. Simons, *Stability and gap phenomena for Yang-Mills fields*, Proc. Natl. Acad. Sci. USA **76** (1979), 1550.
- [39] J.P. Bourguignon and H.B. Lawson, *Stability and isolation phenomena for Yang-Mills fields*, Commun. Math. Phys. **79** (1981) 189.
- [40] C.H. Taubes, *Stability in Yang-Mills theories*, Commun. Math. Phys. **91** (1983) 235.
- [41] L.M. Sibner, R.J. Sibner and K. Uhlenbeck, Proc. Natl. Acad. Sci. USA **86** (1989), 8610.
- [42] L. Sagun and J. Segert, *Non-self-dual Yang-Mills connections with nonzero Chern number*, Bull. Amer. Math. Soc. **24** (1991) 163; *Stationary points of the Yang-Mills action*, Comm. Pure and Appl. Math. **45** (1992) 461.
- [43] J. Burzlaff, *Non-self-dual solutions of $SU(3)$ Yang-Mills theory and a two-dimensional Abelian Higgs model*, Phys. Rev. D **24** (1981) 546.
- [44] J. Schiff, *Hyperbolic vortices and some non-self-dual classical solutions of $SU(3)$ gauge theory*, Phys. Rev. D **44** (1991) 528.
- [45] E. Bergshoeff, A. Collinucci, A. Ploegh, A. Van Riet and S. Vandoren, *Non-extremal D-instantons and the AdS/CFT correspondence*, JHEP **0601:061** (2006), hep-th/0510048.
- [46] L. Brown, R. Carlitz, C. Lee, *Massless excitations in pseudoparticle fields*, Phys. Rev. D **16** (1977) 417.
- [47] A. Rebhan, P. van Nieuwenhuizen and R. Wimmer, *Nonvanishing quantum corrections to the mass and central charge of the $N = 2$ vortex and BPS saturation*, Nucl. Phys. B **679** (2004) 382.
- [48] A. Rebhan, P. van Nieuwenhuizen and R. Wimmer, *The anomaly in the central charge of the supersymmetric kink from dimensional regularization and reduction*, Nucl. Phys. B **648** (2003) 174, hep-th/0207051.
- [49] E. J. Weinberg, *Parameter Counting For Multi - Monopole Solutions*, Phys. Rev. D **20** (1979) 936.

- [50] E. J. Weinberg, *Index Calculations For The Fermion - Vortex System*, Phys. Rev. D **24** (1981) 2669.
- [51] A. Rebhan, P. van Nieuwenhuizen and R. Wimmer, *Quantum mass and central charge of supersymmetric monopoles: Anomalies, current renormalization, and surface terms*, hep-th/0601029.
- [52] C. Bernard, *Gauge zero modes, instanton determinants, and QCD calculations*, Phys. Rev. D **19** (1979) 3013.
- [53] E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, 2nd ed., Gauthier-Villars, Paris.
- [54] P. van Nieuwenhuizen and P.C. West, *Principles of supersymmetry and supergravity*, Cambridge University Press, to appear.
- [55] E. Corrigan, P. Goddard, S. Templeton, *Instanton Green's functions and tensor products*, Nucl. Phys. B **151** (1979) 93;
E. Corrigan, D. Fairlie, P. Goddard, S. Templeton, *A Green's function for the general selfdual gauge field*, Nucl. Phys. B **140** (1978) 31.
- [56] J.L. Gervais, B. Sakita, *Extended particles in quantum field theories*, Phys. Rev. D **11** (1975) 2943;
E. Tomboulis, *Canonical quantization of nonlinear waves*, Phys. Rev. D **12** (1975) 1678.
- [57] A.A. Aleinikov, E.V. Shuryak, *Instantons in quantum mechanics. Two loop effects*, Sov. J. Nucl. Phys. **46** (1987) 76;
S. Olejnik, *Do nongaussian effects decrease tunneling probabilities ? Three loop instanton density for the double well potential*, Phys. Lett. B **221** (1989) 372;
C.F. Wöhler, E.V. Shuryak, *Two loop correction to the instanton density for the double well potential*, Phys. Lett. B **333** (1994) 467, hep-ph/9402287.
- [58] D. Amati and A. Rouet, *Renormalization Of Yang-Mills Theory Developed Around An Instanton*, Nuovo Cim. A **50** (1979) 265.;
L.G. Yaffe, *Quantizing Gauge Theories: Nonclassical Field Configurations, Broken Symmetries, And Gauge Copies*, Nucl. Phys. B **151** (1979) 247;
T.R. Morris, D.A. Ross, C.T. Sachrajda, *Higher Order Quantum Corrections In The Presence Of An Instanton Background Field*, Nucl. Phys. B **255** (1985) 115.
- [59] F. Ore, *How to compute determinants compactly*, Phys. Rev. D **16** (1977) 2577.

- [60] S. Chadha, P. Di Vecchia and F. Nicodemi, *Zeta function regularization of the quantum fluctuations around the Yang-Mills pseudoparticle*, Phys. Lett. B **72** (1977) 103.
- [61] B. Zumino, *Supersymmetry and the vacuum*, Phys. Lett. B **89** (1975) 535.
- [62] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Exact Gell-Mann-Low function of supersymmetric Yang-Mills theories from instanton calculus*, Nucl. Phys. B **229** (1983) 381, and *Beta function in supersymmetric gauge theories, instantons versus traditional approach*, Phys. Lett. B **166** (1986) 329.
- [63] V. A. Novikov, M. A. Shifman, A. I. Vainshtein, M. B. Voloshin and V. I. Zakharov, *Supersymmetry transformations of instantons*, Nucl. Phys. B **229** (1983) 394;
- [64] D.R.T. Jones, *Asymptotic behavior of supersymmetric Yang-Mills theories in the two-loop approximation*, Nucl. Phys. B **87** (1975) 127, eq. (19);
D.R.T. Jones, *Scheme dependence and the NSVZ beta function*, Nucl. Phys. B **486** (1997) 479, eq. (2.7).
- [65] F. Gliozzi, J. Scherk, D. Olive, *Supersymmetry, supergravity and the dual spinor model*, Nucl. Phys. B **122** (1977) 253;
L. Brink, J. Schwarz, J. Scherk, *Supersymmetric Yang-Mills theories*, Nucl. Phys. B **121** (1977) 77.
- [66] B. Zumino, *Euclidean supersymmetry and the many-instanton problem*, Phys. Lett. B **69** (1977) 369.
- [67] H. Nicolai, *A possible constructive approach to (SUPER Φ^3) in four dimensions. 1. Euclidean formulation of the model*, Nucl. Phys. B **140** (1978) 294.
- [68] P. van Nieuwenhuizen and A. Waldron, *On Euclidean spinors and Wick rotations*, Phys. Lett. B **389** (1996) 29, hep-th/9608174.
- [69] M. Blau, G. Thompson, *Euclidean SYM theories by time reduction and special holonomy manifolds*, Phys. Lett. B **415** (1997) 242, hep-th/9706225;
B. Acharya, J. Figueroa-O'Farrill, M. O'Loughlin, B. Spence, *Euclidean D-branes and higher dimensional gauge theory*, Nucl. Phys. B **514** (1998) 583, hep-th/9707118.
- [70] A.V. Belitsky, S. Vandoren, P. van Nieuwenhuizen, *Instantons, Euclidean supersymmetry and Wick rotations*, Phys. Lett. B **477** (2000) 335, hep-th/0001010.

- [71] P. van Nieuwenhuizen, *Phys. Rept. C* **68** (1981) 189, appendix; *An introduction to simple supergravity and the Kaluza-Klein program*, in Proceedings of Les Houches Summer School on Theoretical Physics: *Relativity, Groups and Topology II*, ed. B.S. DeWitt and R. Stora, North-Holland (Amsterdam, 1984) 823.
- [72] I. Affleck, *On constrained instantons*, Nucl. Phys. B **191** (1981) 429.
- [73] M. Nielsen and N.K. Nielsen, *Explicit construction of constrained instantons*, Ann. of Phys. **321** (1996) 331, hep-th/9912006;
P.M. Glerfoss, J. Hysberg, and N.K. Nielsen, *Instanton constraints in supersymmetric gauge theories. I: Supersymmetric QCD*, hep-th/0503119;
N.K. Nielsen, *Instanton constraints in supersymmetric gauge theories. II: $N = 2$ Yang-Mills theory*, hep-th/0503120;
P.M. Glerfoss and N.K. Nielsen, *Instanton constraints and renormalization*, Annals Phys. **321** (2006) 331, hep-th/0504178.
- [74] G.H. Derrick, *Comments on nonlinear wave equations as models for elementary particles*, J. Math. Phys. **5** (1964) 1252;
R. Hobart, *Proc. Royal. Soc. London* **82** (1963) 201.
- [75] D. Bellisai, F. Fucito, A. Tanzini, G. Travaglini, *Instanton calculus, topological field theories and $N = 2$ super Yang-Mills theories*, JHEP **0007:017** (2000), hep-th/0003272.
- [76] K.M. Bitar and S-J. Chang, *Vacuum tunneling of gauge theory in Minkowski space*, Phys. Rev. D **17** (1978) 486.
- [77] T.I. Banks and C.M. Bender, *Anharmonic oscillator with polynomial self-interaction*, Journ. Math. Phys. **13** (1972) 1320;
C.M. Bender and T.T. Wu, *Anharmonic oscillator. 2. A study in perturbation theory in large order*, Phys. Rev. D **7** (1973) 1620;
T.I. Banks, C.M. Bender and T.T. Wu, *Coupled anharmonic oscillators. 1. Equal mass case*, Phys. Rev. D **8** (1973) 3346;
T.I. Banks and C.M. Bender, *Coupled anharmonic oscillators. 2. Unequal mass case*, Phys. Rev. D **8** (1973) 3366.
- [78] E. Brezin and J.L. Gervais, *Nonperturbative aspects in quantum field theory*, Phys. Rept. **49** (1979) 131, and references therein on page 141.
- [79] S. Coleman, *The fate of the false vacuum. 1. Semiclassical theory*, Phys. Rev. D **15** (1977) 2929, Erratum-ibid. D **16** (1977) 1248.

- [80] C.G. Callen and S. Coleman, *The fate of the false vacuum. 2. First quantum corrections*, Phys. Rev. D **16** (1977) 1762.
- [81] H.J. Katz, *Lifetime of metastable vacuum states*, Phys. Rev. D **17** (1978) 1056.
- [82] G. 't Hooft, *How instantons solve the $U(1)$ problem*, Phys. Rep. **142** (1986) 357. For an alternative point of view, see R. Crewther, *Chirality selection rules and the $U(1)$ problem*, Phys. Lett. B **70** (1977) 349, and G.A. Christos, *Chiral symmetry and the $U(1)$ problem*, Phys. Rept. **116** (1984) 251.
- [83] S. Weinberg, *The $U(1)$ problem*, Phys. Rev. D **11** (1975) 3583.
- [84] S. Weinberg, *The quantum theory of fields*, volume 2, Cambridge University Press, 2000.
- [85] Chang and Li, *Gauge theory of elementary particle physics*, Oxford University Press, 1984.
- [86] V.V. Khoze, M. Mattis, M.J. Slater, *The instanton Hunter's guide to supersymmetric $SU(N)$ gauge theories*, Nucl. Phys. B **536** (1998) 69, hep-th/9804009.
- [87] T. Hollowood, V.V. Khoze, M. Mattis, *Summing the instanton series in $N = 2$ superconformal large N QCD*, JHEP **9910** (1999) 019, hep-th/9905209;
E. Gava, K.S. Narain, M.H. Sarmadi, *Instantons in $N = 2$ $Sp(N)$ superconformal gauge theories and the AdS/CFT correspondence*, Nucl. Phys. B **569** (2000) 183, hep-th/9908125.
- [88] T. Hollowood, V.V. Khoze, W. Lee, M. Mattis, *Breakdown of cluster decomposition in instanton calculations of the gluino condensate*, Nucl. Phys. B **570** (2000) 241, hep-th/9904116.
- [89] T. Schaefer and E.V. Shuryak, *Instantons in QCD*, Rev. Mod. Phys. **70** (1998) 323, hep-ph/9610451.
- [90] R. Gilmore, *Lie groups, Lie algebras, and some of the applications*, Wiley 1974, chapter 5, section 6.