10

Yang–Mills instantons

The most important examples of instantons are those, first discovered by Belavin, Polyakov, Schwarz, and Tyupkin [227], that arise in non-Abelian gauge theories. In this chapter I will focus on the properties of these and related solutions, and on their relation to the vacuum structure of the gauge theory. Their physical consequences will be discussed in the next chapter, which will also describe the additional instanton-related effects that arise when fermion fields are added to the theory.

10.1 $A_0 = 0$ gauge

Before delving into the details of these solutions, there are some points that must be addressed at the outset. It was shown in the previous chapter how tunneling amplitudes in field theory are related to solutions of the Euclidean field equations. When we turn to gauge theories, two new issues arise. First, the discussion in Chap. 9 was in the context of a theory with a kinetic energy that was purely quadratic in time derivatives. In general, gauge theory kinetic energies are not of this form. Second, we need to understand how the Euclideanization should affect A_0. If we were to simply replace t by $i\tau$, and leave A_0 unchanged, then the F_{0j} components of the field strength would become complex. This would be avoided if A_0 also gained a factor of i, but that would require some justification.

Both of these issues can be clarified by working in a gauge with $A_0 = 0$. From a canonical point of view, this gauge has some unusual features. With A_0 absent, the Lagrangian density takes the form

$$\mathcal{L} = \operatorname{tr} \dot{A}_j^2 - \frac{1}{2} \operatorname{tr} F_{ij}^2. \tag{10.1}$$

This describes a theory with three independent dynamical fields, A_1, A_2, and A_3, whose conjugate momenta are the electric components of the field strength, $\Pi^j = F^{j0}$. Varying with respect to the A_j yields the three equations of motion

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

$$D_\mu F^{\mu j} = 0. \tag{10.2}$$

However, we do not recover the A_0 equation, which is the Gauss's law constraint

$$\mathcal{C}(\mathbf{x}) \equiv D_j F^{j0}(\mathbf{x}) = D_j \Pi^j(\mathbf{x}) = 0. \tag{10.3}$$

It should be kept in mind that $\mathcal{C}$ is an element of the Lie algebra of the gauge group, so this is actually an adjoint multiplet of constraints at each point in space. (Here, and throughout most of this chapter, group indices have been suppressed.)

Although $\mathcal{C}(\mathbf{x})$ does not necessarily vanish, Eq. (10.2) implies that its time derivative does. This vanishing of $\dot{\mathcal{C}}(x)$ can be understood from symmetry considerations. Even after fixing $A_0 = 0$, we still have the freedom to make time-independent gauge transformations. These can be viewed as symmetries of the Lagrangian and, by Noether's theorem, should lead to corresponding conserved charges.

This can be made explicit by considering the infinitesimal gauge transformation corresponding to a gauge function $\Lambda(\mathbf{x})$ that vanishes at spatial infinity. The Noether charge that generates this transformation is

$$\begin{aligned} C_\Lambda &= 2\int d^3x \,\mathrm{tr}\, D_j\Lambda(\mathbf{x})\Pi^j(\mathbf{x}) \\ &= -2\int d^3x \,\mathrm{tr}\, \Lambda(\mathbf{x}) D_j\Pi^j(\mathbf{x}) \\ &= -2\int d^3x \,\mathrm{tr}\, \Lambda(\mathbf{x})\mathcal{C}(\mathbf{x}). \end{aligned} \tag{10.4}$$

(There is no surface term from the integration by parts in the second line because of the requirement that $\Lambda(\mathbf{x})$ vanish at spatial infinity.) Because the gauge transformations are a symmetry of the Lagrangian, C_Λ is conserved and $\dot{C}_\Lambda = 0$ for arbitrary Λ. This is equivalent to the vanishing of $\dot{\mathcal{C}}(\mathbf{x})$ for all $\mathbf{x}$.

At this point, we have a theory with three independent dynamical fields, A_1, A_2, and A_3. This is too many fields, so it is not quite the theory that we want. However, we also have an infinite number of conserved quantities $\mathcal{C}(\mathbf{x})$. By restricting ourselves to the subspace in which $\mathcal{C}(\mathbf{x}) = 0$ for all $\mathbf{x}$, we arrive at the desired Yang–Mills theory. Classically, this corresponds to imposing Gauss's law as a constraint on the initial conditions. Quantum mechanically, it can be implemented by restricting ourselves to the subspace of the Hilbert space in which all components of the $\mathcal{C}(\mathbf{x})$ vanish.[1]

In this gauge the kinetic energy is purely quadratic, and so the arguments relating tunneling to the Euclidean Lagrangian

$$\mathcal{L}_E = \mathrm{tr}\,(\partial_4 A_j)^2 + \frac{1}{2}\mathrm{tr}\, F_{ij}^2 \tag{10.5}$$

[1] Because there are nontrivial commutators between the gauge components of $\mathcal{C}$, one cannot in general find a simultaneous eigenstate of all them. However, the commutators are all proportional to components of $\mathcal{C}$, so eigenstates with zero eigenvalue for all components do exist. This is analogous to the situation with angular momentum, for which one can have states with $L_x = L_y = L_z = 0$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

(with $x_4 \equiv \tau$) go through straightforwardly. In analogy with the Lagrangian of Eq. (10.1), this Lagrangian implies three field equations, but not the Euclidean version of Gauss's law,

$$D_j F^{j4} = 0. \tag{10.6}$$

However, the classical instanton solution must satisfy the boundary conditions that its Euclidean time derivatives $\partial_4 A_j$ vanish at the initial and final values of x_4. This, together with the other field equations, implies that the Gauss's law constraint is satisfied at all intermediate values of x_4.

Although we could just work with the Euclidean theory in this form, we can also give the theory a full Euclidean gauge freedom by introducing a field $A_4(x)$ that essentially acts as a Lagrange multiplier to enforce the Euclidean Gauss's law constraint. We define F_{j4} in the obvious way, and take the Euclidean Lagrangian to be

$$\mathcal{L}_E = \frac{1}{2} \operatorname{tr} F_{rs}^2 \,, \tag{10.7}$$

where r and s run from 1 to 4. Once we have done this, we are free to work in any gauge that we find convenient, with the boundary conditions on the instanton at the initial and final values of x_4 being such that $F_{j4}(\mathbf{x}, x_4^{\text{init}}) = F_{j4}(\mathbf{x}, x_4^{\text{fin}}) = 0$.

In the course of our examination of the Yang–Mills instantons, there will be occasion to refer to both Euclidean and Minkowskian coordinates. I will adopt the convention that Greek indices refer to Minkowskian coordinates that take values from 0 to 3, Roman indices from late in the alphabet (r, s, ...) refer to Euclidean indices that range from 1 to 4, and Roman indices from the middle of the alphabet (i, j, ...) are spatial indices running from 1 to 3, and can occur in either a Minkowskian or a Euclidean context.

10.2 Yang–Mills vacua: $A_0 = 0$ gauge

If instantons are associated with a tunneling process, the natural starting point of the discussion is to specify from where and to where the tunneling is taking place. Somewhat surprisingly, the answer to this question—although not the observable physical consequences—depends on the gauge in which one works. Perhaps the simplest picture is the one associated with the $A_0 = 0$ gauge [233, 234]. As described below, in this gauge there are degenerate classical minima that are separated by energy barriers. Quantum mechanical tunneling through these barriers is described by the instanton.

We have already seen in the previous section that $A_0 = 0$ gauge has some unusual features, and the existence of multiple classical vacua is just one more of these. This suggests that it might be helpful to examine the problem in a more "normal" gauge in which there is a unique vacuum. This will be done in the next section.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

Classically, the vacuum configurations have $F_{\mu\nu} = 0$. In $A_0 = 0$ gauge this means that

$$A_j = \frac{i}{g} G^{-1} \partial_j G \,, \tag{10.8}$$

where $G(\mathbf{x})$ is an element of the gauge group, which for the present we take to be SU(2). Let us assume that G tends to a constant value G_∞ as $r \to \infty$; without loss of generality we can take $G_\infty = I$. $G(\mathbf{x})$ is a map from R^3 into the gauge group. Our assumption that $G(r = \infty)$ is a constant means that we can view spatial infinity as a single point, and space as a three-sphere. The map is then one from S^3 into the group. Such maps, and hence the vacuum configurations, can be classified by elements of $\pi_3(SU(2))$ which, we recall from Eq. (4.46), is Z, the additive group of the integers. Thus, the classical vacua fall into disconnected sets, each corresponding to an element of $\pi_3(SU(2))$. The vacua within a given set are related by gauge transformations that can be built up from a series of infinitesimal gauge transformations with gauge functions Λ that vanish at spatial infinity; we will call these "small" gauge transformations. Vacua in different sets are also gauge-equivalent, but the gauge transformations that connect them are "large" gauge transformations that cannot be obtained from a sequence of infinitesimal gauge transformations.

The correspondence between vacua and homotopy classes can be made explicit by defining a quantity

$$N[G] = \frac{1}{24\pi^2} \epsilon^{ijk} \int d^3x \ \mathrm{tr}\, G^{-1} \partial_i G \, G^{-1} \partial_j G \, G^{-1} \partial_k G. \tag{10.9}$$

The normalization here is such that if G is taken to be in the fundamental representation of SU(2), with the generators $T_a = \sigma_a/2$, $N[G]$ is the winding number that counts how many times SU(2), which is topologically a three-sphere, is mapped onto R^3, which we are also viewing as a three-sphere. To see this, consider the neighborhood of a point where G has some value G_0. Near this point, we can write

$$G(x) = G_0 \exp[i\sigma_a \Lambda_a(x)] \approx G_0[I + i\sigma_a \Lambda_a(x)] \tag{10.10}$$

with Λ_a small. The right-hand side of Eq. (10.9) is then the integral of

$$\begin{aligned} -\frac{i}{24\pi^2} \epsilon^{ijk} (\mathrm{tr}\, \sigma_a \sigma_b \sigma_c) \, \partial_i \Lambda_a \, \partial_j \Lambda_b \, \partial_k \Lambda_c &= \frac{1}{12\pi^2} \epsilon^{ijk} \epsilon^{abc} \, \partial_i \Lambda_a \, \partial_j \Lambda_b \, \partial_k \Lambda_c \\ &= \left(\frac{1}{2\pi^2} \right) \left(\epsilon^{ijk} \, \partial_i \Lambda_1 \, \partial_j \Lambda_2 \, \partial_k \Lambda_3 \right). \end{aligned} \tag{10.11}$$

On the last line the second factor is the Jacobian corresponding to the transformation from the spatial coordinates x_j to the group coordinates Λ_a. This converts the integral over space to one over the group. Dividing by the volume

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

of the unit three-sphere, $2\pi^2$, normalizes this integral so that it gives directly the number of times that SU(2) is covered by G as one goes over all of space.[2]

A check of the normalization can be obtained by explicitly calculating the winding number for

$$g_1(x) = \exp[i\hat{r}_a \sigma_a f(r)]\,, \tag{10.12}$$

where $f(r)$ is any monotonic function with $f(0) = -\pi$ and $f(\infty) = 0$. This gives a one-to-one mapping of the physical space into SU(2), with every element of SU(2) occurring at only a single value of $\mathbf{x}$. One verifies that $N[g_1] = 1$, as required.

Now suppose that the gauge transformation can be written as a product $G = G_1 G_2$. Substitution of this into Eq. (10.9) gives

$$N[G_1 G_2] = N[G_1] + N[G_2] + \frac{1}{8\pi^2}\Delta\,, \tag{10.13}$$

where

$$\begin{aligned}
\Delta &= \epsilon^{ijk} \int d^3x \mathrm{tr}\, \left[G_1^{-1}\partial_i G_1\, \partial_j G_2\, G_2^{-1} \partial_k G_2\, G_2^{-1}\right. \\
&\qquad\qquad \left. + \partial_i G_2\; G_2^{-1} G_1^{-1} \partial_j G_1\, G_1^{-1} \partial_k G_1\right] \\
&= -\epsilon^{ijk} \int d^3x \mathrm{tr}\, \left[G_1^{-1}\partial_i G_1\, \partial_j G_2\, \partial_k G_2^{-1} + \partial_i G_2\, G_2^{-1} \partial_j G_1^{-1}\, \partial_k G_1\right] \\
&= \epsilon^{ijk} \int d^3x \mathrm{tr}\, \left[G_1^{-1}\partial_i G_1\, \partial_j (\partial_k G_2\, G_2^{-1}) - \partial_i G_2\, G_2^{-1} \partial_j (G_1^{-1} \partial_k G_1)\right] \\
&= \epsilon^{ijk} \int d^3x \mathrm{tr}\, \partial_j \left(G_1^{-1}\partial_i G_1 \partial_k G_2\, G_2^{-1}\right) \\
&= 0.
\end{aligned} \tag{10.14}$$

Thus, the winding number of a product of two gauge transformations is the sum of the individual winding numbers,

$$N[G_1 G_2] = N[G_1] + N[G_2]. \tag{10.15}$$

Equation (10.11) shows that for an infinitesimal gauge transformation, $G_\Lambda \approx I + i\Lambda(\mathbf{x}) = I + i\sigma_a \Lambda_a$, with $\Lambda(\mathbf{x})$ vanishing at spatial infinity, the leading contribution to $N[G_\Lambda]$ is the integral of

$$\frac{1}{2\pi^2}\, \epsilon^{ijk}\, \partial_i \Lambda_1\, \partial_j \Lambda_2\, \partial_k \Lambda_3 = \frac{1}{2\pi^2}\, \epsilon^{ijk}\, \partial_i \left(\Lambda_1\, \partial_j \Lambda_2\, \partial_k \Lambda_3\right). \tag{10.16}$$

Since this is the total derivative of a quantity that vanishes at spatial infinity, $N[G_\Lambda] = 0$. It then follows from Eq. (10.15) that any continuous change in G

[2] This is quite analogous to the results in lower dimensions that were given in Eqs. (4.38) and (4.39).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

leaves $N[G]$ invariant, as expected for a winding number. Furthermore, if G_0 is a constant, then G and $G_0 G$ have the same winding number.

Now let us define a current

$$\begin{aligned} j_A^\mu &= \frac{g^2}{8\pi^2}\,\epsilon^{\mu\nu\alpha\beta}\,\mathrm{tr}\left(A_\nu\,\partial_\alpha A_\beta - \frac{2ig}{3}A_\nu A_\alpha A_\beta\right) \\ &= \frac{g^2}{16\pi^2}\,\epsilon^{\mu\nu\alpha\beta}\,\mathrm{tr}\left(A_\nu F_{\alpha\beta} + \frac{2ig}{3}A_\nu A_\alpha A_\beta\right). \end{aligned} \tag{10.17}$$

Although this current is not gauge independent, its divergence,[3]

$$\partial_\mu j_A^\mu = \frac{g^2}{16\pi^2}\,\mathrm{tr}\,F_{\mu\nu}\tilde{F}^{\mu\nu}\,, \tag{10.18}$$

is. Here

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} \tag{10.19}$$

(or "F dual") is the dual field strength.

The charge associated with this current is

$$Q_A = \int d^3x\; j_A^0 = \frac{g^2}{16\pi^2}\int d^3x\,\epsilon^{ijk}\,\mathrm{tr}\left(A_i F_{jk} + \frac{2ig}{3}A_i A_j A_k\right). \tag{10.20}$$

(Note that, because j_A is not divergenceless, there is no reason to expect Q_A to be conserved.) For a vacuum configuration, with $F_{ij} = 0$ and $A_j = (i/g)G^{-1}\partial_j G$, we have

$$Q_A = N[G]. \tag{10.21}$$

Now let $A_\mu(\mathbf{x}, t)$ be a sequence of finite energy configurations, parameterized by t, that interpolate between two vacuum configurations with winding numbers N_1 and N_2. The integral

$$\int d^4x\,\partial_\mu j_A^\mu = \int d^4x\left[\partial_0 j_A^0 + \partial_i j_A^i\right] \tag{10.22}$$

can be converted to a sum of integrals over the bounding surface, which is conveniently viewed as a hypercylinder oriented along the t-axis. The first term on the right-hand side contributes only on the initial and final hypersurfaces of constant t, giving $N_2 - N_1$. The second term gives the integral over the surface at $r = \infty$. In $A_0 = 0$ gauge the only contribution to j_A^i comes from the $A_j F_{0k}$ term, but for finite energy configurations this falls fast enough as $r \to \infty$ that the surface integral vanishes. Hence, using Eq. (10.18), we find that

$$\Delta N = N_2 - N_1 = \frac{g^2}{16\pi^2}\int d^4x\;\mathrm{tr}\,F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{10.23}$$

[3] In verifying this equation, one can use the cyclic property of the trace and the fact that $\epsilon^{0123} = -\epsilon^{1230}$ to show that the term quartic in the potential on the right-hand side actually vanishes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

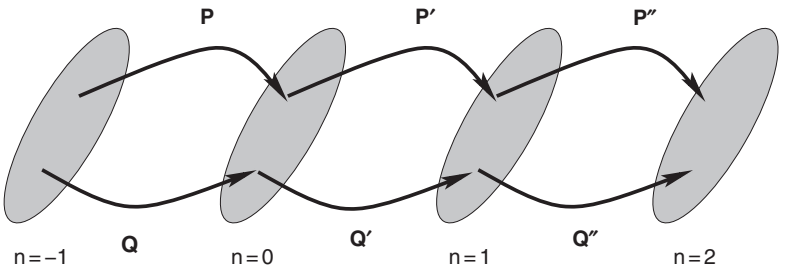


Fig. 10.1. Schematic view of the configuration space in $A_0 = 0$ gauge. The shaded regions correspond to vacuum configurations with various values of the winding number n. These are separated by regions of nonvacuum configurations of nonzero energy. Tunneling paths P, P', and P'' are gauge-equivalent, as are Q, Q', and Q''. Along each of these paths the value of the integral in Eq. (10.24) is determined solely by the difference of the initial and final winding numbers.

A caricature of the vacuum structure in $A_0 = 0$ gauge is shown in Fig. 10.1. Each of the shaded "islands" corresponds to the vacuum configurations of a fixed winding number. These are separated by a "sea" of nonvacuum configurations of higher energy. Each island can be mapped onto the succeeding one by a gauge transformation of unit winding number, such as the $g_1(\mathbf{x})$ of Eq. (10.12). Also shown are a number of paths from one island to the next. These are also gauge equivalent. Paths P' and P'' are obtained by successive applications of g_1 to P, while Q' and Q'' are obtained from Q in a similar manner. Each of these paths connects vacua whose winding numbers differ by one, so integrating along any of these gives

$$\int d^4x \, \mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} = \frac{16\pi^2}{g^2}. \tag{10.24}$$

Note that although Eqs. (10.22)–(10.24) were written with Greek indices, suggesting a Minkowskian interpretation of t as a real time, they are equally valid with a Euclidean interpretation in which x_4 simply parameterizes a sequence of configurations involving only A_j with $j = 1, 2, 3$.

This discussion of vacua has been purely classical up to this point. Let us now turn to the quantum theory. In a theory with a single classical vacuum the wave functional of the quantum ground state is concentrated on field configurations close to the classical vacuum. Here, however, we have many degenerate classical vacua. We are faced with two types of degeneracies. The first, among the states with a given winding number, is handled by imposing the Gauss's law constraint. In general, if a state is an eigenstate of a quantum mechanical charge with zero eigenvalue, then its wavefunction is independent of the conjugate variable; e.g., a one-particle state with zero angular momentum has a wavefunction that is independent of angle. In the present case, Gauss's law restricts us to states for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

which the C_Λ all vanish, implying that their wave functional is uniform (with respect to an appropriate measure) over all configurations that are related by small gauge transformations. If we only had classical vacua with a single winding number, this would resolve the degeneracy, and there would be a unique quantum vacuum.

However, we have to consider vacuum configurations with all possible winding numbers. Perturbative analysis about these would predict one quantum vacuum state for each winding number. Let us denote these "n-vacua" by $|n\rangle$. If it were not for the possibility of tunneling through the potential energy barriers separating vacuum configurations with different winding numbers, these n-vacua would be eigenstates of the Hamiltonian. However, because there is tunneling, they are not. The situation is similar to that of a particle in a periodic potential, and we can proceed in a similar fashion.

To construct the true eigenstates, let us define an operator T that takes a state $|\Psi\rangle$ with wave functional $\Psi[A_j(\mathbf{x})]$ to a state $T|\Psi\rangle = |\Psi'\rangle$ whose wave functional obeys

$$\Psi'[A_j^g(\mathbf{x})] = \Psi[A_j(\mathbf{x})]\,, \tag{10.25}$$

where A_j^g is the result of acting on A_j with the large gauge transformation g_1 defined in Eq. (10.12). In particular, T takes one n-vacuum to the next, so that

$$T|n\rangle = |n+1\rangle. \tag{10.26}$$

The gauge invariance of the theory ensures that T commutes with the Hamiltonian.

We can then define "θ-vacua,"[4]

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{-in\theta}|n\rangle\,, \tag{10.27}$$

that are eigenstates of T with eigenvalue $e^{i\theta}$ [233, 234]. (Note that θ is periodic with period 2π.) These θ-vacua are true eigenstates of the Hamiltonian. To verify this, consider the matrix element of the time evolution operator between two θ-vacua,

$$\begin{aligned}\langle\theta'|e^{-iHt}|\theta\rangle &= \sum_{m,n} e^{im\theta'}e^{-in\theta}\langle m|e^{-iHt}|n\rangle \\ &= \sum_{m,n} e^{im(\theta'-\theta)}e^{i(m-n)\theta}\langle m|e^{-iHt}|n\rangle.\end{aligned} \tag{10.28}$$

[4] Although an intuitive specification of the n-vacua is fairly clear, a precise definition is a bit more difficult, since their wave functionals have tails extending out to configurations, far from any classical vacuum, for which the winding number does not have even an approximate meaning. One way to proceed is to define the θ-vacua as the lowest energy states with given values of θ, and to then define the n-vacua by inverting the Fourier transform in Eq. (10.27).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

Because H is gauge-invariant, the matrix element on the right-hand side can only depend on the difference between the winding numbers, $m - n$. Hence,

$$\begin{aligned}\langle\theta'|e^{-iHt}|\theta\rangle &= \sum_{m,k} e^{im(\theta'-\theta)} e^{ik\theta}\langle k|e^{-iHt}|0\rangle \\ &= 2\pi\delta(\theta-\theta')\sum_k e^{ik\theta}\langle k|e^{-iHt}|0\rangle. \end{aligned} \tag{10.29}$$

Thus, one θ-vacuum cannot evolve into a different one. This argument can clearly be extended to the one- and many-particle states built upon these vacua.[5] If we assume that we live in a state of definite θ, then states with other values of θ are inaccessible to us, and so θ is effectively a constant of nature.

The explicit factor of $e^{ik\theta}$ that appears in the sum in Eq. (10.29) can be absorbed by modifying the Lagrangian. Writing the matrix element of the time evolution operator as a path integral, we have

$$\langle k|e^{-iHt}|0\rangle = \int [dA_\mu] e^{i\int d^4x \mathcal{L}}, \tag{10.30}$$

where the integral is over paths that run from a configuration with zero winding number at $t = -\infty$ to one with winding number k at $t = \infty$. Equation (10.23) tells us that over any such path

$$k = \frac{g^2}{16\pi^2}\int d^4x \,\mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}, \tag{10.31}$$

so the factor of $e^{ik\theta}$ is equivalent to adding to the Lagrangian density a term of the form [234, 235]

$$\Delta\mathcal{L} = \frac{\theta g^2}{16\pi^2}\,\mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{10.32}$$

Recalling Eq. (10.18), we see that this additional term is a total divergence. Classically, such a term would have no effect, since adding a total divergence to the Lagrangian density leaves the Euler–Lagrange equations unchanged. Even quantum mechanically, this has no effect in perturbation theory: If we think of this term as giving rise to a new type of vertex in Feynman graphs, the overall derivative gives a factor of the net momentum entering the vertex which is, of course, zero. This is why it is usually omitted in introductory discussions of Yang–Mills theories, and why the importance, or even the possibility, of such a term was not appreciated until the discovery of the Yang–Mills instanton solutions. This term will be discussed in more detail in Sec. 11.5.

[5] Given an arbitrary state $|\Psi\rangle$, a state of definite θ can be defined by $|\Psi,\theta\rangle = \sum_n e^{-in\theta}T^n|\Psi\rangle$. T is clearly invertible, so its negative powers are well defined.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

10.3 Yang–Mills vacuum: axial gauge

The discussion in the previous section depended crucially on the fact that we were working in a gauge that admitted multiple classical vacua. One more commonly imagines working in a "physical" gauge where the potential is uniquely determined by the field strength. In such a gauge there is only one classical vacuum. In electrodynamics, Coulomb gauge is a simple example of a gauge that gives a one-to-one mapping between field strength and potential. This is not the case in the non-Abelian theory, where the Gribov ambiguity shows that there are configurations of field strengths that allow several distinct choices of Coulomb gauge potentials, as well as other configurations that cannot be brought into Coulomb gauge at all [85, 236]. These issues are often swept under the rug when working in perturbation theory, but cannot be ignored when considering potentials of order $1/g$, for which the commutator term in the field strength is comparable in size to the derivative terms. As suggested by Eq. (10.8), it is precisely such potentials with which we will be concerned.

A choice that does fix the potentials uniquely is the axial gauge, defined by

$$A_3(x,y,z,t) = 0 \tag{10.33}$$

and the subsidiary conditions

$$\begin{aligned} A_2(x,y,0,t) &= 0\,, \\ A_1(x,0,0,t) &= 0\,, \\ A_0(0,0,0,t) &= 0. \end{aligned} \tag{10.34}$$

To show that this gauge can always be achieved, let us assume that we are given A_μ everywhere in space, but not necessarily in axial gauge. Our task is to find a gauge transformation $U(x)$ that transforms A_μ to an axial gauge potential A'_μ via

$$A'_\mu = U^{-1} A_\mu U + \frac{i}{g} U^{-1} \partial_\mu U. \tag{10.35}$$

Requiring that $A'_3 = 0$ gives

$$\partial_3 U(x,y,z,t) = igA_3(x,y,z,t)U(x,y,z,t)\,, \tag{10.36}$$

whose solution is

$$U(x,y,z,t) = P \exp\left[ig \int_0^z dz'\, A_x(x,y,z',t)\right] U(x,y,0,t)\,, \tag{10.37}$$

where P denotes path ordering. Requiring that A_2 vanish on the hyperplane $z = 0$ then gives

$$\partial_2 U(x,y,0,t) = -igA_2(x,y,0,t)U(x,y,0,t)\,, \tag{10.38}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

whose solution can again be written in terms of a path-ordered exponential of an integral along the y-direction on the hyperplane $z = 0$. By proceeding in a similar fashion, the remaining two subsidiary conditions can also be achieved.

If a potential is already in axial gauge, then Eq. (10.33) allows only z-independent gauge transformations. The three subsidiary conditions then rule out, successively, y-dependent, x-dependent, and t-dependent gauge transformations, leaving only the freedom to make global gauge transformations.

In fact, the axial gauge potential is given in terms of the field strength by a rather simple expression. Because $F_{3\mu} = \partial_3 A_\mu$,

$$A_\mu(x, y, z, t) = \int_0^z dz'\, F_{3\mu}(x, y, z', t) + C_\mu(x, y, t). \tag{10.39}$$

The subsidiary conditions determine C_μ uniquely, giving

$$\begin{aligned} A_\mu(x, y, z, t) = &\int_0^z dz'\, F_{3\mu}(x, y, z', t) + \int_0^y dy'\, F_{2\mu}(x, y', 0, t) \\ &+ \int_0^x dx'\, F_{1\mu}(x', 0, 0, t) + \int_0^t dt'\, F_{0\mu}(0, 0, 0, t'). \end{aligned} \tag{10.40}$$

In particular, in the classical vacuum with $F_{\mu\nu} = 0$ the gauge potential vanishes everywhere.

Physical results should not depend on the choice of gauge. Let us see how we can recover the results of the previous section [235]. With the multiple vacua of the $A_0 = 0$ gauge, there were families of configurations, such as those lying along the path P of Fig. 10.1b, that connected one vacuum to another. Transforming these configurations into axial gauge gives a path $\tilde{P}$ that starts and ends at the same point. In contrast with the $A_0 = 0$ gauge instantons that describe tunneling through the energy barrier separating distinct vacua, axial gauge instantons correspond to a somewhat peculiar process in which one tunnels into an energy barrier merely to tunnel back out and return to one's starting point.

Because $\mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}$ is gauge-invariant, its integrals over the paths P and $\tilde{P}$ must be equal. However, when these are converted to surface integrals of the gauge-variant j_A^μ the contributions arise from different parts of the surface. We saw previously that for the $A_0 = 0$ gauge path P the only contributions were from the initial and final surfaces, giving the integral in terms of a difference of winding numbers. In axial gauge, where the vacuum winding number clearly vanishes, the surface integral from path $\tilde{P}$ comes entirely from the surfaces at $z = \pm\infty$.

A particularly striking consequence of the multiple classical vacua in $A_0 = 0$ gauge was the existence of multiple quantum vacua labeled by a continuous periodic parameter θ. This naturally led us to the inclusion of the new, θ-dependent, term in the Lagrangian that was given in Eq. (10.32). The axial gauge analysis doesn't lead so compellingly to this term, but once we realize that it can appear, there is no reason not to include it [235]. There is the difference that θ is a fixed

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

coupling constant in axial gauge, but a continuous variable in $A_0 = 0$ gauge. However, as was already noted, the fact that transitions between states of different θ are forbidden means that θ can be viewed as a constant of nature even in the latter gauge.

An analogy that may be instructive is given by the single-particle Lagrangian

$$L = \frac{1}{2}B\dot{\alpha}^2 - K(1 - \cos\alpha). \tag{10.41}$$

If α is viewed as ranging over all real numbers, this describes a particle of mass B moving in a periodic potential with minima analogous to the $A_0 = 0$ gauge n-vacua. As is well known, the states of this system fall into bands, with states within a band labeled by a periodic variable analogous to the gauge theory θ. Alternatively, α can be taken to be periodic with period 2π, and the system interpreted as a pendulum with moment of inertia B swinging in a constant gravitational field. There is now just a single classical ground state. Tunneling between adjacent minima of the periodic potential corresponds to a process in which the pendulum makes a full 2π rotation even though its energy is too low for this to be classically possible. The analogue of the gauge theory $\theta F\tilde{F}$ term is the total time derivative term

$$\Delta L = \frac{\theta}{2\pi}\,\dot{\alpha}. \tag{10.42}$$

We will see, in Sec. 11.5, that such a term can have measurable effects.

10.4 Some topology

In the course of our discussion of classical vacua in $A_0 = 0$ gauge we found that the values of the volume integral of $\mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}$ were quantized. One might guess that this has a topological interpretation. This is indeed the case. We have already encountered homotopy classes in the course of studying topological solitons. What we are concerned with now is a different type of topological object, one associated with a gauge field on a manifold.

Although we are actually interested in fields on four-dimensional Euclidean space, R^4, it is perhaps best to begin by discussing fields on a four-sphere, S^4. Let us therefore suppose that we are given a non-Abelian gauge field A_p on S^4. This field is not assumed to satisfy any particular field equations. We could require that it be everywhere nonsingular, but our experience with the Dirac monopole suggests that this might exclude some physically interesting cases. Instead, let us allow singularities, but only those that can be removed by a gauge transformation. (This would allow singularities similar to those along a Dirac string, but not a singularity such as that at the center of a Dirac monopole.) If this is the case, we can cover the four-sphere by multiple gauge patches. In each patch the corresponding gauge potential is nonsingular, while in the overlap between

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

two patches the corresponding potentials are related by a nonsingular gauge transformation. Thus, in the overlap of Regions I and II there is a nonsingular gauge function U such that

$$A_p^{\mathrm{II}} = U A_p^{\mathrm{I}} U^{-1} - \frac{i}{g}(\partial_p U) U^{-1}. \tag{10.43}$$

This is quite analogous to the two-patch treatment of the Dirac monopole field that was introduced in Sec. 5.1. With this in mind, consider the case of an Abelian gauge field on a two-sphere. (This could be thought of as a two-sphere enclosing one or more monopoles in ordinary three-dimensional space, although that is not necessary.) The field strength F_{pq} is assumed to be nonsingular everywhere on the sphere. The potentials A_p are nonsingular on gauge patches, and (because the theory is Abelian) satisfy equations of the form

$$A_p^{\mathrm{II}} - A_p^{\mathrm{I}} = -\partial_p \Lambda \tag{10.44}$$

on the overlaps between regions.

Now let us define

$$I_1 = \frac{1}{4\pi} \int d^2x \, \epsilon^{pq} F_{pq} \,, \tag{10.45}$$

where the x^p are coordinates on the two-sphere. (There will be the usual spherical coordinate singularities unless one uses several coordinate charts, but these have no effect on our results.)

This quantity is manifestly gauge-invariant. It has two other important properties. First, it is a topological invariant, in the sense that it is unchanged under a smooth deformation

$$\begin{aligned} A_p^{\mathrm{I}} &\to A_p^{\mathrm{I}} + \delta A_p^{\mathrm{I}} \\ A_p^{\mathrm{II}} &\to A_p^{\mathrm{II}} + \delta A_p^{\mathrm{II}} \end{aligned} \tag{10.46}$$

of the gauge potentials. To see this, note first that Eq. (10.44) implies that δA_p is continuous from one patch to the next, with

$$\delta A_p^{\mathrm{II}} = \delta A_p^{\mathrm{I}} \,. \tag{10.47}$$

Next, substituting Eq. (10.46) into Eq. (10.45) gives

$$\delta I_1 = \frac{1}{4\pi} \int d^2x \, \epsilon^{pq} \delta F_{pq} = \frac{1}{2\pi} \int d^2x \, \partial_p (\epsilon^{pq} \delta A_q). \tag{10.48}$$

The integrand in the last expression is a total derivative of a quantity that varies continuously from patch to patch. Integrating this over the entire manifold gives a surface integral over the boundary of the manifold. Since the two-sphere has no boundary, the integral vanishes and I_1 is invariant, as promised.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

The second property relates I_1 to a homotopy class of the gauge group. To obtain this result, we write

$$I_1 = \frac{1}{2\pi}\int d^2x\,\partial_p(\epsilon^{pq}A_q) \equiv \int d^2x\,\partial_p j^p. \tag{10.49}$$

As in Eq. (10.48), we have the integral of a total divergence. In contrast with the previous case, it is the divergence of a current that varies from patch to patch, so more care is need in evaluating the integral. Let us assume, for simplicity, that there are only two coordinate patches, and let Σ be a circle in the overlap region. The integral can then be evaluated by using $A_p^{\rm I}$ in the portion of Region I bounded by Σ, and $A_p^{\rm II}$ in the complementary portion of Region II, also bounded by Σ. In both cases the integral can be expressed as a line integral over Σ (traversed in opposite directions for the two cases), and we obtain

$$I_1 = \frac{1}{2\pi}\int_\Sigma dl_p\,\epsilon^{pq}(A_q^{\rm I} - A_q^{\rm II}) = \frac{1}{2\pi}\int_\Sigma dl_p\,\epsilon^{pq}\partial_q\Lambda. \tag{10.50}$$

The last expression is just the winding number that counts how many times the U(1) is traversed as one goes around Σ. It thus relates I_1 to an element of $\pi_1(U(1))$.

The integrand in Eq. (10.45) for I_1 is known as the first Chern form, and the integral itself as the first Chern number. To see the generalization to a higher dimension, let us now return to our non-Abelian theory on a four-sphere and define the gauge-invariant second Chern number[6]

$$I_2 = \frac{g^2}{32\pi^2}\int d^4x\,\epsilon^{pqrs}\,\mathrm{tr}\,F_{pq}F_{rs}. \tag{10.51}$$

We first want to show that I_2 is a topological invariant that is unchanged under deformation of the gauge potentials. In contrast with the Abelian case, δA is not continuous across patches, but instead transforms covariantly, with

$$\delta A_p^{\rm II} = U\delta A_p^{\rm I}U^{-1}. \tag{10.52}$$

The change in I_2 resulting from this deformation is

$$\delta I_2 = \frac{g^2}{16\pi^2}\int d^4x\,\epsilon^{pqrs}\,\mathrm{tr}\,F_{pq}\,\delta F_{rs} = \frac{g^2}{8\pi^2}\int d^4x\,\epsilon^{pqrs}\,\mathrm{tr}\,F_{pq}D_r\delta A_s\,, \tag{10.53}$$

where the covariant derivative in the last expression is taken with respect to the unperturbed potential. However,

$$\epsilon^{pqrs}\,\mathrm{tr}\,F_{pq}D_r\delta A_s = \epsilon^{pqrs}\,\partial_r(\mathrm{tr}\,F_{pq}\delta A_s) - \epsilon^{pqrs}\,\mathrm{tr}\,(D_rF_{pq})\delta A_s. \tag{10.54}$$

[6] Notice that this expression, like that for I_1 above, does not involve the metric. This can be understood by noting that the gauge potential and field strength are one- and two-forms, respectively, and that the integrand here is a four-form, whose integral over a four-manifold is defined independently of the metric.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

The second term on the right-hand side vanishes by the Bianchi identity, and we have

$$\delta I_2 = \frac{g^2}{8\pi^2}\int d^4x\, \partial_r(\epsilon^{pqrs}\,\mathrm{tr}\, F_{pq}\delta A_s). \tag{10.55}$$

The quantity in parentheses on the right-hand side is a gauge-invariant quantity. Integrating the divergence of this quantity over the four-sphere, which has no boundary, gives zero. Hence, I_2 is invariant under smooth deformations of the gauge potential, as was claimed.

Next, we want to relate I_2 to the topology of the gauge group. From the discussion in Sec. 10.2 we know that

$$\begin{aligned}\frac{g^2}{32\pi^2}\epsilon^{pqrs}\,\mathrm{tr}\, F_{pq}F_{rs} &= \frac{g^2}{16\pi^2}\,\partial_p\left[\epsilon^{pqrs}\,\mathrm{tr}\,\left(A_qF_{rs} + \frac{2ig}{3}A_qA_rA_s\right)\right]\\ &\equiv \partial_p j_A^p.\end{aligned} \tag{10.56}$$

Thus I_2 is itself the integral of a total divergence over the four-sphere. As with I_1, the current involved is gauge-variant, so the integration over the manifold reduces to a sum of surface integrals over the boundaries separating the gauge patches. In analogy with our previous calculation, we can divide the four-sphere into two parts separated by a three-sphere hypersurface Σ that lies in the overlap of the two gauge patches. The integral then becomes

$$I_2 = \int_\Sigma dS_p(j_A^{\mathrm{I}p} - j_A^{\mathrm{II}p}). \tag{10.57}$$

Using Eq. (10.43), we find that

$$\begin{aligned}j_A^{\mathrm{II}p} - j_A^{\mathrm{I}p} = &-\frac{1}{24\pi^2}\,\epsilon^{pqrs}\,\mathrm{tr}\, U^{-1}\partial_qU\, U^{-1}\partial_rU\, U^{-1}\partial_sU\\ &+\frac{ig}{8\pi^2}\,\epsilon^{pqrs}\,\partial_q(\mathrm{tr}\, U^{-1}\partial_rU\, A_s).\end{aligned} \tag{10.58}$$

In the overlap region, and on Σ in particular, all the fields are nonsingular, so when this expression is substituted into Eq. (10.57) the last term leads to an integral of a total divergence over the closed manifold Σ, which vanishes. This leaves us with

$$I_2 = \frac{1}{24\pi^2}\,\epsilon^{pqrs}\int_\Sigma dS_p\ \mathrm{tr}\, U^{-1}\partial_qU\, U^{-1}\partial_rU\, U^{-1}\partial_sU. \tag{10.59}$$

If the gauge group is SU(2), comparison with Eq. (10.9) shows that I_2 is just the winding number that counts the number of times that SU(2) (which is topologically a three-sphere) is covered as one moves over Σ (which is also a three-sphere). It thus corresponds to an element of $\pi_3(G)$. I_2 is also quantized for larger simple groups. In particular, for $G = \mathrm{SU}(N)$, with the matrices in the fundamental representation, the normalization used here makes $I_2 = 1$ for the unit instanton.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

One can also work with a single gauge patch, at the cost of introducing a singularity in the gauge field. This can be viewed as taking the limit of the multipatch formalism in which all but one of the patches is shrunk to a point and patch I covers the entire sphere, with the boundaries between integration regions becoming infinitesimal hypersurfaces enclosing the singularities. The volume integrals over the shrunken regions clearly vanish, while that over region I can be converted into a sum of surface integrals over the hypersurfaces enclosing the singularities.

Our real interest is in a gauge field on four-dimensional Euclidean space, R^4, not on a four-sphere. However, if the field strength on R^4 falls rapidly enough at spatial infinity, we can view the field on R^4 as being on S^4. This can be made more precise by projecting the four-sphere (covered by five-dimensional coordinates r_a with $\sum r_a^2 = 1$) onto R^4 (with coordinates x_p). The mapping is given by

$$\begin{aligned} r_p &= \frac{2x_p}{1+x^2}\,, \quad p = 1,2,3,4, \\ r_5 &= \frac{1-x^2}{1+x^2}\,, \end{aligned} \tag{10.60}$$

so the "south pole", $(0,0,0,0,-1)$, is mapped to spatial infinity on R^4. The R^4 potential A_p is related to the S^4 potential $\hat{A}_a$ (which satisfies $\sum r_a \hat{A}_a = 0$) by

$$\frac{1+x^2}{2} A_p = \hat{A}_p - x_p \hat{A}_5\,, \tag{10.61}$$

so that if the S^4 potential is nonsingular at the south pole, then the R^4 potential falls at least as fast as $1/x^2$ at large distance. This mapping between the two spaces applies whether or not the gauge fields obey the field equations. If one of them happens to be a solution, then the conformal invariance of the classical Yang–Mills theory implies that the other is also a solution [237].

We will, in fact, only be concerned with fields whose field strengths fall rapidly enough at large distance that they can be mapped onto S^4. For such fields, I_2 becomes the integer instanton number

$$k = \frac{g^2}{16\pi^2} \int d^4x \,\mathrm{tr}\, F_{pq} \tilde{F}^{pq} = \frac{g^2}{32\pi^2} \int d^4x \, F^a_{pq} \tilde{F}^{apq}. \tag{10.62}$$

10.5 't Hooft symbols

Before embarking on the study of explicit instanton solutions, let us pause to develop some notation that proves to be quite useful. Recall that SO(4) and SU(2)×SU(2) share the same Lie algebra. To explore this further, let us define two sets of matrices e_p and $e_p^\dagger$ by

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

$$e_p = \begin{cases} i\tau_p\,, & p = 1,2,3, \\ I\,, & p = 4, \end{cases} \tag{10.63}$$

$$e_p^\dagger = \begin{cases} -i\tau_p\,, & p = 1,2,3, \\ I\,, & p = 4, \end{cases} \tag{10.64}$$

where τ_1, τ_2, and τ_3 are the usual Pauli matrices.

An SO(4) vector V_p can be represented as a 2×2 matrix

$$V = e_p^\dagger V_p = \begin{pmatrix} V_4 - iV_3 & -V_2 - iV_1 \\ V_2 - iV_1 & V_4 + iV_3 \end{pmatrix} \tag{10.65}$$

with determinant

$$\det V = V_p V_p. \tag{10.66}$$

Note that $(\det V)^{-1/2} V$ is a unitary matrix. Rotations are linear transformations on V_p that preserve its length. These can be implemented via matrix multiplications of the form

$$V \to U_L^{-1} V U_R\,, \tag{10.67}$$

where U_L and U_R are unitary matrices (to preserve the form of V as a constant times a unitary matrix) with unit determinant (to preserve $\det V$, and hence the length of V_p). They are thus elements of SU(2), and we have

$$\mathrm{SO}(4) = [\mathrm{SU}(2) \times \mathrm{SU}(2)]/Z_2\,, \tag{10.68}$$

with the division by Z_2 arising because $U_L = U_R = -I$ leaves V unchanged.

The effect of a rotation on the components of V_p can be obtained by writing

$$U_L = e^{-i\boldsymbol{\omega}_L \cdot \boldsymbol{\tau}}\,, \qquad U_R = e^{-i\boldsymbol{\omega}_R \cdot \boldsymbol{\tau}}. \tag{10.69}$$

For an infinitesimal transformation, the change in V is

$$\delta V = i(\boldsymbol{\omega}_L \cdot \boldsymbol{\tau})V - iV(\boldsymbol{\omega}_R \cdot \boldsymbol{\tau}). \tag{10.70}$$

Using the identity

$$V_p = \frac{1}{2} \mathrm{tr}\, V e_p\,, \tag{10.71}$$

we then obtain

$$\delta V_p = \left(\omega_L^j \bar{\eta}_{pq}^j + \omega_R^j \eta_{pq}^j\right) V_q\,, \tag{10.72}$$

where the 2×2 matrices

$$\eta_{pq} = \eta_{pq}^j \tau_j\,, \qquad \bar{\eta}_{pq} = \bar{\eta}_{pq}^j \tau_j \tag{10.73}$$

are defined by [238]

$$\begin{aligned} \eta_{pq} &= -i\left(e_p e_q^\dagger - \delta_{pq} I\right), \\ \bar{\eta}_{pq} &= -i\left(e_p^\dagger e_q - \delta_{pq} I\right). \end{aligned} \tag{10.74}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

They obey $\eta_{pq} = -\eta_{qp}$ and $\bar{\eta}_{pq} = -\bar{\eta}_{qp}$, with the nonzero η^j_{pq} and $\bar{\eta}^j_{pq}$ fixed by

$$\eta^k_{ij} = \bar{\eta}^k_{ij} = \epsilon_{ijk}\,,$$

$$\eta^k_{i4} = -\bar{\eta}^k_{i4} = \delta_{ik}. \tag{10.75}$$

It follows from the above equations that η_{pq} is self-dual,

$$\eta_{pq} = \frac{1}{2}\epsilon_{pqrs}\,\eta_{rs}\,, \tag{10.76}$$

while $\bar{\eta}_{pq}$ is anti-self-dual, with

$$\bar{\eta}_{pq} = -\frac{1}{2}\epsilon_{pqrs}\,\bar{\eta}_{rs}. \tag{10.77}$$

An antisymmetric rank two SO(4) tensor transforms as the $(1,0)\oplus(0,1)$ representation. The two terms, corresponding to the self-dual and anti-self-dual parts, respectively, are each left unaltered by one of the two SU(2) factors of the group. This correspondence is clarified by comparing Eqs. (10.76) and (10.77) with the roles played by η_{pq} and $\bar{\eta}_{pq}$ in Eq. (10.72).

Two useful identities are

$$\eta_{pq}\eta_{qr} = -2i\eta_{pr} - 3\delta_{pr}I\,,$$

$$\bar{\eta}_{pq}\bar{\eta}_{qr} = -2i\bar{\eta}_{pr} - 3\delta_{pr}I. \tag{10.78}$$

Further identities can be found in [238].

10.6 The unit instanton

We are now ready to look for instanton solutions. We could do this by directly attacking the second-order Euclidean Yang–Mills equations. However, an easier path is to adopt the method that we used to find the BPS solutions in Sec. 8.2. Because $F^2_{rs} = \tilde{F}^2_{rs}$, the Euclidean action can be written as

$$\begin{aligned}
\int d^4x\,\frac{1}{2}\,\mathrm{tr}\,F^2_{rs} &= \int d^4x\,\frac{1}{4}\,\mathrm{tr}\,(F^2_{rs} + \tilde{F}^2_{rs}) \\
&= \int d^4x\,\frac{1}{4}\,\mathrm{tr}\,(F_{rs} \mp \tilde{F}_{rs})^2 \pm \int d^4x\,\frac{1}{2}\,\mathrm{tr}\,F_{rs}\tilde{F}_{rs} \\
&= \pm\left(\frac{8\pi^2}{g^2}\right)k + \int d^4x\,\frac{1}{4}\,\mathrm{tr}\,(F_{rs} \mp \tilde{F}_{rs})^2 \\
&\geq \left(\frac{8\pi^2}{g^2}\right)|k|\,,
\end{aligned} \tag{10.79}$$

where k is the instanton number. For fixed k, a minimum of the Euclidean action, and hence a solution to the field equations, is obtained if the field strength is either self-dual, with

$$F_{rs} = \tilde{F}_{rs}\,, \tag{10.80}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

or anti-self-dual, with

$$F_{rs} = -\tilde{F}_{rs}\,, \tag{10.81}$$

according to whether k is positive or negative, respectively. The standard terminology is to call only the self-dual solutions instantons, with the anti-self-dual solutions being termed anti-instantons.

Let us now seek a solution of the self-duality equation (10.80). It is simplest to not impose any particular gauge condition at the outset; once a solution has been found, it can be transformed into whatever gauge is desired. As we have often done before, we start by adopting an ansatz that exploits the symmetries of the theory. Specifically, we require that the solution be invariant under the combination of a global SU(2) gauge transformation and a transformation by one of the SU(2) factors of SO(4). This suggests trying the ansatz

$$A_p = \eta_{pq}\, x_q\, f(x^2). \tag{10.82}$$

The corresponding field strength is

$$F_{pq} = 2\eta_{qp} f + 2(\eta_{qr} x_r x_p - \eta_{pr} x_r x_q) f' - ig\,[\eta_{pr}, \eta_{qs}]\, x_r x_s f^2\,, \tag{10.83}$$

where $f' \equiv df/d(x^2)$.

Requiring this to be self-dual imposes constraints on $f(x^2)$. Since our ansatz is rotationally invariant, it is sufficient to examine the fields along the positive x_4-axis. Along this axis we have

$$\begin{aligned} F_{ij} &= -2\eta_{ij} f - ig[\eta_{i4}, \eta_{j4}] x^2 f^2\,, \\ F_{k4} &= -2\eta_{k4} f - 2\eta_{k4} x^2 f'. \end{aligned} \tag{10.84}$$

The self-duality condition $F_{ij} = \epsilon_{ijk} F_{k4}$ then reduces to

$$\frac{df}{d(x^2)} = -g f^2. \tag{10.85}$$

This is solved by

$$f = \frac{1}{g}\,\frac{1}{x^2 + \lambda^2}\,, \tag{10.86}$$

where λ^2 is an integration constant that must be positive for the solution to be nonsingular.

We thus have [227]

$$A_p = \frac{1}{g}\,\frac{\eta_{pq} x_q}{x^2 + \lambda^2} \tag{10.87}$$

or, in terms of SU(2) components,

$$A^j_p = \frac{2}{g}\,\frac{\eta^j_{pq} x_q}{x^2 + \lambda^2}. \tag{10.88}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

It can be seen that λ gives a characteristic size for the instanton. The fact that all values of λ are allowed is a consequence of the scale invariance of the classical Yang–Mills theory. The anti-self-dual anti-instanton solution has exactly the same form, except that η_{pq} is replaced by $\bar{\eta}_{pq}$.

The field strength that follows from the instanton potential is

$$F_{pq} = -\frac{2}{g}\,\frac{\eta_{pq}\lambda^2}{(x^2+\lambda^2)^2}. \tag{10.89}$$

Since A_p falls as $1/|x|$ at large distance, one might have expected F_{pq} to fall as $1/x^2$. Its actual $1/x^4$ falloff reflects the fact that, as we will see shortly, the leading large-distance component of A_p is a pure gauge potential.

Because of the spherical symmetry of the solution, it is trivial to evaluate the instanton number by integrating $\operatorname{tr} F_{pq}\tilde{F}_{pq} = \operatorname{tr} F_{pq}F_{pq}$ over all space. One finds that $k = 1$. Alternatively, one could use Eq. (10.18) and calculate the instanton number as an integral of the normal component of j^p_A over the surface at spatial infinity. Because $F_{pq} \sim 1/x^4$, the contribution comes entirely from the final, cubic, term in Eq. (10.17). Here again, the calculation can be simplified by using spherical symmetry and evaluating the integrand along the positive z-axis.

Now consider the unitary matrix

$$S = \frac{e^\dagger_p x_p}{|x|}. \tag{10.90}$$

It is easily verified that

$$\frac{i}{g}S^{-1}\partial_p S = \frac{1}{g}\,\frac{\eta_{pq}x_q}{x^2}. \tag{10.91}$$

Hence the instanton solution of Eq. (10.87) can be written as

$$A_p = \frac{i}{g}S^{-1}\partial_p S\left(\frac{x^2}{x^2+\lambda^2}\right), \tag{10.92}$$

which shows that it approaches a pure gauge form at large distance. Let us now gauge-transform this potential to

$$\begin{aligned} A'_p &= SA_pS^{-1} - \frac{i}{g}\,\partial_p S\,S^{-1} \\ &= \frac{i}{g}\partial_p S\,S^{-1}\left(\frac{x^2}{x^2+\lambda^2} - 1\right) \\ &= \frac{1}{g}\,\frac{\bar{\eta}_{pq}x_q\lambda^2}{x^2(x^2+\lambda^2)}. \end{aligned} \tag{10.93}$$

[Note that the self-dual η_{rs} of Eq. (10.87) has been replaced by the anti-self-dual $\bar{\eta}_{rs}$.] In this form the potential falls rapidly enough at spatial infinity that the surface integral of j^p_A at infinity vanishes. The instanton number is still recovered,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

because the singularity of the potential at $x = 0$ requires that when we evaluate the instanton number as a surface integral we include an infinitesimal surface enclosing the origin. This surface gives a unit contribution to k, as required.

The relation between these two forms of the instanton solution is particularly simple when they are mapped onto the four-sphere. Both correspond to using a single gauge patch, implying that the potential has a singularity somewhere on the sphere. For the solution of Eq. (10.87) that singularity is at the south pole, while for that of Eq. (10.93) it is at the north pole.

10.7 Multi-instanton solutions

The latter form of the one-instanton solution suggests a rather powerful ansatz, first found by 't Hooft [239], that yields multi-instanton solutions. Consider a potential of the form

$$A_p = \bar{\eta}_{pr} f_r(x)\,, \tag{10.94}$$

where the $f_r(x)$ are arbitrary functions of x_p, with no particular symmetry assumed. The field strength is

$$F_{pq} = \bar{\eta}_{qr}\partial_p f_r - \bar{\eta}_{pr}\partial_q f_r - ig[\bar{\eta}_{pr}, \bar{\eta}_{qs}] f_r f_s. \tag{10.95}$$

Requiring this to be self-dual is equivalent to requiring that its product with the anti-self-dual matrix $\bar{\eta}_{pq}$ vanish. Using the second identity in Eq. (10.78), we obtain the condition

$$\begin{aligned} 0 &= \bar{\eta}_{pq} F_{pq} \\ &= -4i\bar{\eta}_{pq}\partial_p f_q - 6\partial_p f_p + 12 g f_p f_p. \end{aligned} \tag{10.96}$$

The coefficient of $\bar{\eta}_{pq}$ in the last line vanishes if f_p is taken to be the gradient of a scalar function. The remaining terms then cancel if we write

$$f_p = -\frac{1}{2g}\,\partial_p \ln \rho, \tag{10.97}$$

where ρ obeys

$$0 = \frac{\Box \rho}{\rho}. \tag{10.98}$$

This is solved by

$$\rho = 1 + \sum_{a=1}^{k} \frac{\lambda_a^2}{(x - w_a)^2}. \tag{10.99}$$

For $k = 1$, with $w_1 = 0$, this reproduces the one-instanton solution in the form given by Eq. (10.93). For $k > 1$, the instanton number receives a unit contribution from a surface integral about each of the poles, and none from the surface at

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

infinity. This is therefore a k-instanton solution, with a four-vector position w_a and a scale λ_a for each of the component instantons.[7]

Jackiw, Nohl, and Rebbi [240] pointed out a curious generalization of this solution. Equation (10.98) is still satisfied if we leave out the first term in Eq. (10.99) and write

$$\rho = \sum_{a=1}^{k+1} \frac{\lambda_a^2}{(x - w_a)^2}. \tag{10.101}$$

Again, surface integrals enclosing each of the poles make unit contributions to the instanton number. However, with the initial 1 omitted, A_p falls as $1/|x|$, and so there is also a contribution from the surface at infinity. With the η_{pq} replaced by an $\bar{\eta}_{pq}$, the latter surface gives a contribution of -1, so the total instanton number is one less than the number of poles.

This appears to give a k-instanton solution that depends on $5k+4$ parameters. (An overall scale can clearly be factored out of ρ without affecting A_p.) This is more than is needed to specify the positions and scales. For $k = 1$ and 2 some of these extra parameters can be changed by gauge transformations, while maintaining the form of the ansatz. For $k = 1$, four of the apparent parameters can be absorbed in this manner, so there is no truly new solution here. [However, this does give the curious result that the positions of the poles in Eq. (10.101) are not special locations, and are not simply related to the center of the instanton.] For $k = 2$, there is one degree of gauge freedom, and thus 13, rather than the apparent 14 parameters. For $k \geq 3$, the poles can only be moved by a gauge transformation if they all lie on a common circle, so the general solution does indeed depend on $5k + 4$ parameters.

In fact, this is still not the most general multi-instanton solution. This can be seen by using index theory methods to determine the number of parameters needed to specify an arbitrary solution with instanton number k.

10.8 Counting parameters with an index theorem

Counting the number of parameters needed to specify an instanton solution is equivalent to counting the number of linearly independent perturbations of the solution that preserve its self-duality, with the proviso that perturbations that are simply gauge transformations should not be counted. As we will see shortly, the perturbations that we are considering satisfy an equation of the form

[7] The instanton scale is modified by the presence of the other instantons. For the case of widely separated instantons, the quantity comparable to the λ that appears in the one-instanton solution is actually

$$\lambda_a' = \lambda_a \left[1 + \sum_{b \neq a} \frac{\lambda_b}{(w_b - w_a)^2} \right]. \tag{10.100}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

$$\mathcal{D}\chi = 0\,, \tag{10.102}$$

where $\mathcal{D}$ is a matrix differential operator. The kernel of an operator is defined as the subspace that it annihilates, so the number we want is the dimension of the kernel of $\mathcal{D}$. Several approaches were used to determine this number for the SU(2) instantons [241–244]; the discussion in this section will be closest to that of [244].

What we will actually calculate is the index of $\mathcal{D}$,

$$\mathcal{I}(\mathcal{D}) = \dim\ker(\mathcal{D}) - \dim\ker(\mathcal{D}^\dagger)\,, \tag{10.103}$$

which is the difference in the dimensions of the kernels of $\mathcal{D}$ and of its adjoint.[8] As we will see, there is a "vanishing theorem" that shows that $\mathcal{D}^\dagger$ has no zero eigenvalues, and hence no kernel, so that $\mathcal{I}(\mathcal{D})$ is just the number we want.

The motivation for this somewhat roundabout approach requires some explanation. First, note that the zero modes of $\mathcal{D}$ (i.e., its normalizable eigenfunctions with eigenvalue zero) are the same as those of $\mathcal{D}^\dagger\mathcal{D}$. It is obvious that if $\mathcal{D}\chi = 0$ then $\mathcal{D}^\dagger\mathcal{D}\chi = 0$. To show the converse, let us assume that $\mathcal{D}^\dagger\mathcal{D}\chi = 0$ and consider the inner product with $\chi^\dagger$. We have

$$\begin{aligned} 0 &= \int d^4x\, \chi^\dagger(\mathcal{D}^\dagger\mathcal{D}\chi) \\ &= \int d^4x\, (\mathcal{D}\chi)^\dagger(\mathcal{D}\chi)\,, \end{aligned} \tag{10.104}$$

from which it follows that $\mathcal{D}\chi = 0$. Second, the nonzero eigenvalues of $\mathcal{D}^\dagger\mathcal{D}$ and $\mathcal{D}\mathcal{D}^\dagger$ are the same, since if

$$\mathcal{D}^\dagger\mathcal{D}\chi = \lambda\chi \tag{10.105}$$

then

$$\mathcal{D}\mathcal{D}^\dagger(\mathcal{D}\chi) = \lambda(\mathcal{D}\chi). \tag{10.106}$$

The first result tells us that counting the zero eigenvalues of $\mathcal{D}^\dagger\mathcal{D}$ is equivalent to counting those of $\mathcal{D}$, and similarly for $\mathcal{D}\mathcal{D}^\dagger$ and $\mathcal{D}^\dagger$. The second gives us important information about the behavior of these eigenvalues under smooth variations of $\mathcal{D}$. Let us suppose that $\mathcal{D}$ depends upon some external field—in our case, $A_p(x)$. If $A_p(x)$ is smoothly varied, then the eigenvalues of $\mathcal{D}^\dagger\mathcal{D}$ will presumably vary continuously. In particular, some zero eigenvalues might become nonzero, or some nonzero eigenvalues might become zero. However, the same must happen to the eigenvalues of $\mathcal{D}\mathcal{D}^\dagger$, with the zero eigenvalues of the two operators appearing or disappearing simultaneously. As a result, the index of $\mathcal{D}$ is not affected by such smooth variations. One might therefore expect—as indeed turns out to be the case—that the index can be expressed in terms of topological invariants of the fields entering $\mathcal{D}$.

[8] For further discussion of index calculations, see Appendix B.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

These results lead us to consider the quantity

$$\mathcal{I}(\mathcal{D}, M^2) = \mathrm{Tr}\left(\frac{M^2}{\mathcal{D}^\dagger\mathcal{D} + M^2} - \frac{M^2}{\mathcal{D}\mathcal{D}^\dagger + M^2}\right), \tag{10.107}$$

with the trace being understood in a functional sense and M an arbitrary real number. Let us take the limit $M^2 \to 0$, performing the evaluation in a basis where the operators are diagonal. The terms corresponding to nonzero eigenvalues vanish in this limit, while those from the zero eigenvalues each contribute 1 or -1, depending on the term in which they appear. Hence, we conclude that

$$\lim_{M^2\to 0} \mathcal{I}(\mathcal{D}, M^2) = \mathcal{I}(\mathcal{D}). \tag{10.108}$$

In fact, for any value of M^2 the terms from the nonzero eigenvalues cancel pairwise between the two expressions in Eq. (10.107), while the zero eigenvalues still contribute 1 or -1. Thus, despite appearances, $\mathcal{I}(\mathcal{D}, M^2)$ is actually independent of M^2, which means that we can evaluate it using any value for M^2 that we find convenient; it will turn out to be simplest to evaluate it in the limit $M^2 \to \infty$.

A word of caution is needed here. The discussion above was phrased as if the spectra of the various operators were discrete. This would indeed be the case if we were working with differential operators on a closed manifold, such as S^4. Instead, we are working on infinite four-dimensional Euclidean space, R^4, and these spectra include continuous portions. As a result, the formal arguments above may lead us into error.

One possibility is that the densities of continuum eigenvalues near zero could be sufficiently singular as to give a nonzero contribution to the left-hand side of Eq. (10.108). This portion of the spectrum is sensitive to the large-distance behavior of the instanton field. Because the part of A_p that is not pure gauge falls as $1/x^3$, there is in fact no problem, and there is no anomalous continuum contribution.[9] This is essentially equivalent to the statement that the falloff of the instanton solutions is fast enough that they can be mapped onto solutions on S^4, as discussed at the end of Sec. 10.4.

A second potential problem is that the pairwise cancellation of terms in Eq. (10.107) is ill-defined for continuous spectra, bringing into question the M^2-independence of $\mathcal{I}(\mathcal{D}, M^2)$. However, we will see presently that this M^2-independence can be verified by an explicit calculation.

Let us now proceed to determine $\mathcal{D}$. We assume that we are given a gauge potential $A_p^a(x)$ that gives rise to a self-dual field strength, and consider a perturbation

$$A_p^a \to A_p^a + \delta A_p^a. \tag{10.109}$$

[9] A continuum contribution does arise in other contexts, as discussed in Appendix B.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

The change in the field strength is then

$$\begin{aligned}\delta F^a_{pq} &= \partial_p \delta A^a_q - \partial_q \delta A^a_p + g f_{abc} A^b_p \delta A^c_q - g f_{abc} A^b_q \delta A^c_p \\ &= (D_p \delta A_q)^a - (D_q \delta A_p)^a,\end{aligned} \tag{10.110}$$

where the covariant derivative of δA_p is defined by

$$(D_p \delta A_q)^a = \partial_p \delta A^a_q - i g A^b_p (T^b)_{ac} \delta A^c_q \,, \tag{10.111}$$

with the $(T^b)_{ac} = i f_{abc}$ being the generators of the group in the adjoint representation. If δF_{pq} is self-dual, then contraction with the anti-self-dual matrix $\bar{\eta}_{pq}$ gives

$$0 = \bar{\eta}_{pq} D_p \delta A_q. \tag{10.112}$$

Many of the solutions of this equation are simply gauge transformations of the original solution and can be written in the form

$$\delta_G A_p = D_p \Lambda. \tag{10.113}$$

To exclude these physically uninteresting modes, we require that δA_p be orthogonal to all such modes, so that

$$\begin{aligned}0 &= \int d^4x \; \delta_G A_p \, \delta A_p \\ &= \int d^4x \, D_p \Lambda \, \delta A_p \\ &= -\int d^4x \, \Lambda \, D_p \delta A_p.\end{aligned} \tag{10.114}$$

Requiring that this hold for all $\Lambda(x)$ gives the background gauge condition

$$0 = D_p \delta A_p. \tag{10.115}$$

Note, however, that the integration by parts in the last line of Eq. (10.114) assumes that Λ falls sufficiently rapidly at large distance. Consequently, although we have removed modes arising from local gauge transformations, those corresponding to global gauge transformations still remain and have to be handled separately.

By recalling Eq. (10.74), we can combine Eqs. (10.112) and (10.115) into the single equation

$$0 = e^\dagger_p e_q D_p \delta A_q. \tag{10.116}$$

If we define

$$\Psi = -i e_q \delta A_q \,, \tag{10.117}$$

Eq. (10.116) can be recast as the Dirac-type equation

$$0 = -e^\dagger_p D_p \Psi. \tag{10.118}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

Thus, our immediate task is to determine the index of the operator

$$\mathcal{D} \equiv -e_p^\dagger D_p = -e_p^\dagger(\partial_p - igA_p^b T^b). \tag{10.119}$$

For the application to counting instanton parameters, the generator T^b in this equation must be in the adjoint representation. However, it will also be of interest to know the index of $\mathcal{D}$ for arbitrary representations, so we will leave the representation unspecified until the end of the calculation. For the moment, we just note that the generators can be chosen to satisfy

$$\operatorname{tr} T^a T^b = \delta_{ab}\, T(R)\,, \tag{10.120}$$

where $T(R)$ depends on the representation. For SU(2), $T(R)$ is most easily calculated by evaluating $\operatorname{tr} T_3^2$, which in the representation with "isospin" t gives

$$T(R) = \sum_{t_z=-t}^{t} t_z^2 = \frac{1}{3}t(t+1)(2t+1). \tag{10.121}$$

In particular, $T = 2$ for the adjoint representation and $T = 1/2$ for the fundamental doublet representation.

There is an important point to note concerning the relation between Ψ and δA_p. If $\Psi(x)$ is a solution of Eq. (10.118), then so is $i\Psi(x)$. As Dirac solutions these are viewed as being linearly dependent, and so count as a single solution when determining the index. However, when δA_p is extracted from Ψ via

$$\delta A_p = \frac{i}{2}\operatorname{tr} e_p^\dagger \Psi\,, \tag{10.122}$$

Ψ and $i\Psi$ produce linearly independent sets of δA_p. Hence, the index $\mathcal{I}_b$ for the operator in the bosonic Eq. (10.116) and the index $\mathcal{I}_f$ of the fermionic operator in Eq. (10.118) are related by

$$\mathcal{I}_b = 2\mathcal{I}_f. \tag{10.123}$$

Hence, the bosonic index must be an even number. In fact, a stronger result holds. If Ψ is a solution, then not only is $i\Psi$ a solution but so is ΨU, where U is any 2×2 unitary matrix. The result is that the number of bosonic zero modes must be a multiple of four.

Let us now proceed to our calculation. A set of Euclidean gamma matrices obeying

$$\{\gamma_p, \gamma_q\} = 2\delta_{pq} \tag{10.124}$$

is given by

$$\gamma_p = \begin{pmatrix} 0 & -ie_p \\ ie_p^\dagger & 0 \end{pmatrix}, \tag{10.125}$$

with

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}. \tag{10.126}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

Using these, we can write

$$i\gamma \cdot D \equiv i\gamma_p D_p = \begin{pmatrix} 0 & \mathcal{D}^\dagger \\ \mathcal{D} & 0 \end{pmatrix} \tag{10.127}$$

and

$$(i\gamma \cdot D)^2 = \begin{pmatrix} \mathcal{D}^\dagger \mathcal{D} & 0 \\ 0 & \mathcal{D}\mathcal{D}^\dagger \end{pmatrix}, \tag{10.128}$$

while Eq. (10.107) for the index gives

$$\mathcal{I}_f = \mathrm{Tr}\left[-\gamma_5 \frac{M^2}{-(\gamma \cdot D)^2 + M^2}\right]. \tag{10.129}$$

It must be remembered that the trace in this expression encompasses traces over Dirac indices and group indices, as well as a functional trace.

This is a good point to take a moment to verify that $\mathcal{I}$ really is independent of M^2. Differentiating the right-hand side of Eq. (10.129) gives

$$\frac{\partial \mathcal{I}_f}{\partial M^2} = \mathrm{Tr}\left\{\gamma_5 \frac{(\gamma \cdot D)^2}{[-(\gamma \cdot D)^2 + M^2]^2}\right\}. \tag{10.130}$$

Next, we bring a factor of $\gamma \cdot D$ from the numerator to the left of the γ_5, giving a minus sign from the anticommutator, and then use the cyclic property of the trace to bring it back to its original position. This shows that the right-hand side of the equation is equal to minus itself, and hence that

$$\frac{\partial \mathcal{I}_f}{\partial M^2} = 0. \tag{10.131}$$

Let us now return to the evaluation of $\mathcal{I}_f$. By separating symmetric and antisymmetric terms, we can write

$$\begin{aligned}(\gamma \cdot D)^2 &= \frac{1}{4}\{\gamma_p, \gamma_q\}\{D_p, D_q\} + \frac{1}{4}[\gamma_p, \gamma_q][D_p, D_q] \\ &= D^2 - \frac{ig}{2}\gamma_p \gamma_q F_{pq} \\ &\equiv D^2 + \Delta.\end{aligned} \tag{10.132}$$

We can then expand the right-hand side of Eq. (10.129) in a power series, obtaining

$$\begin{aligned}\mathcal{I}_f = -\mathrm{Tr}\Bigg\{\gamma_5 M^2 \Bigg[&\frac{1}{(-D^2 + M^2)} + \frac{1}{(-D^2 + M^2)}\Delta\frac{1}{(-D^2 + M^2)} \\ &+ \frac{1}{(-D^2 + M^2)}\Delta\frac{1}{(-D^2 + M^2)}\Delta\frac{1}{(-D^2 + M^2)} + \cdots\Bigg]\Bigg\}.\end{aligned} \tag{10.133}$$

Let us consider the terms in this expansion one by one. The first term is proportional to the trace of γ_5, and therefore vanishes. The second term contains

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

the trace of $\gamma_5\gamma_p\gamma_q$, and so also vanishes. The terms represented by the ellipsis vanish when we take the limit $M^2 \to \infty$. (This can be seen by dimensional analysis, but in any case will be clear after the third term has been evaluated.) Hence, in the $M^2 \to \infty$ limit we only need to consider the third term, and can write

$$\mathcal{I}_f = \lim_{M^2\to\infty} \left(\frac{g}{2}\right)^2 (\mathrm{tr}\,\gamma_5\gamma_p\gamma_q\gamma_r\gamma_s)(\mathrm{tr}\,T^aT^b) \times \int d^4x \left\langle x \left| \frac{M^2}{(-D^2+M^2)} F^a_{pq} \frac{1}{(-D^2+M^2)} F^b_{rs} \frac{1}{(-D^2+M^2)} \right| x \right\rangle . \quad (10.134)$$

The traces over Dirac and group indices give

$$(\mathrm{tr}\,\gamma_5\gamma_p\gamma_q\gamma_r\gamma_s)(\mathrm{tr}\,T^aT^b) = 4\epsilon_{pqrs}\,\delta_{ab}\,T(R). \quad (10.135)$$

This leaves us with the functional trace over the Green's function in the second line. We can move the F_{pq} past the factor of $(-D^2+M^2)^{-1}$, because the commutator gives a term, containing the derivative of the field strength, that is suppressed by an additional factor of M^2. For similar reasons, in the large M^2 limit we can replace D^2 by ∂^2. Making these changes, and using Eq. (10.135), we have

$$\mathcal{I}_f = \lim_{M^2\to\infty} 2g^2T(R) \int d^4x \left\langle x \left| \frac{M^2}{(-\partial^2+M^2)^3} \right| x \right\rangle F^a_{pq}\tilde{F}^a_{pq}. \quad (10.136)$$

The first factor in the integral is actually independent of x, and given by

$$\left\langle x \left| \frac{M^2}{(-\partial^2+M^2)^3} \right| x \right\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{M^2}{(p^2+M^2)^3} = \frac{1}{32\pi^2}. \quad (10.137)$$

Using Eq. (10.62) to relate the integral of $F^a_{pq}\tilde{F}^a_{pq}$ to the instanton number, we obtain

$$\mathcal{I}_f = 2T(R)k. \quad (10.138)$$

For counting instanton parameters, the generators are in the $t = 1$ adjoint representation, for which $T(R) = 2$. We get an additional factor of two in going from the fermionic to the bosonic operator, giving

$$\mathcal{I}_b = 8k. \quad (10.139)$$

We are not quite done, however. First, the index only gives us the difference between the number of zero eigenvalues of $\mathcal{D}$ and the corresponding number for $\mathcal{D}^\dagger$, whereas what we actually want is the former quantity. A calculation essentially identical to that which yielded Eq. (10.132) shows that

$$\mathcal{D}\mathcal{D}^\dagger = -D^2 - \frac{g}{2}\bar{\eta}_{pq}F_{pq} = -D^2\,, \quad (10.140)$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

where the last equality follows because $\bar{\eta}_{pq}$ is anti-self-dual while F_{pq} is self-dual. Because $-D^2$ is a positive definite operator, $\mathcal{D}^\dagger$ has no zero eigenvalues, and so $\mathcal{D}$ has $8k$.

Second, we need to worry about the global gauge modes that were not excluded by the background gauge condition. With a gauge group SU(2), of dimension three, there are three such modes, and so the number of physical parameters entering a solution with instanton number k is

$$n = 8k - 3. \tag{10.141}$$

This result suggests that any such solution should be interpreted as being composed of k unit instantons, with no essentially new objects appearing at higher values of k. For each of the component instantons there are one size and four position parameters, just as in the 't Hooft solution. In addition, each has three parameters that specify its SU(2) orientation. Finally, three parameters must be subtracted, because only the relative SU(2) orientations are physically meaningful. Although this simple picture implicitly assumes that the separation between instantons is large compared to their sizes, the parameter counting is correct for all self-dual solutions.

As a final point, recall that there is a correspondence between the instanton solutions on Euclidean R^4 and those on the four-sphere, S^4. Counting of parameters by index methods should therefore yield the same answer for both cases. On S^4, where one does not need to worry about a surface term in the analogue of Eq. (10.114), the background gauge condition eliminates all gauge modes, and so there are no global gauge modes to be subtracted at the end. However, when the methods of this section are generalized to the sphere, the metric enters the differential operators. This leads to an extra contribution of -3, so the result of Eq. (10.141) is recovered [241]. This result on the four-sphere is a manifestation of the Atiyah–Singer index theorem [245], which states that the index on a compact manifold can be expressed solely in terms of topological invariants.

10.9 Larger gauge groups

The results of the preceding section can be generalized to a larger gauge group G, which I will assume to be a compact connected simple group. Because $\Pi_3(G) = Z$ for any such group, the topological arguments for the existence of instanton solutions again apply. In fact, it is quite easy to explicitly exhibit some such solutions.

For any choice of G there are two or more inequivalent SU(2) subgroups. Using one of these subgroups to embed an SU(2) solution into G clearly gives an instanton solution in the larger group. To be explicit, let $\mathcal{T}^a$ be the generators of one such subgroup, normalized to give the standard commutation relations, and satisfying

$$\operatorname{tr} \mathcal{T}^a \mathcal{T}^b = \frac{c}{2}\, \delta^{ab}. \tag{10.142}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

A solution is then given by

$$A_p = A_p^a \mathcal{T}^a \,, \tag{10.143}$$

where the A_p^a are the components of the SU(2) solution. If the embedding solution had instanton number $k_{\mathrm{SU}(2)}$, the new solution has instanton number

$$\begin{aligned} k_G &= \frac{g^2}{16\pi^2} \int d^4x \, \mathrm{tr}\, F_{pq} \tilde{F}_{pq} \\ &= \left(\mathrm{tr}\, \mathcal{T}^a \mathcal{T}^b\right) \frac{g^2}{16\pi^2} \int d^4x F_{pq}^a \tilde{F}_{pq}^b \\ &= c\, k_{\mathrm{SU}(2)}. \end{aligned} \tag{10.144}$$

For example, if $G = \mathrm{SU}(3)$, viewed as a group of 3×3 matrices, there is an obvious SU(2) subgroup corresponding to the upper left 2×2 block. This has generators $\lambda_1/2$, $\lambda_2/2$, and $\lambda_3/2$, with $c = 1$. Embedding via this subgroup leaves the instanton number unchanged. There is also an SU(2) [or, more properly, an SO(3) subgroup] with generators λ_2, λ_5, and λ_7. For these, $c = 4$, so embedding via this subgroup quadruples the topological charge. In particular, embedding the unit SU(2) instanton gives a spherically symmetric instanton with $k = 4$. At first sight, one might think that there were now two different species of fundamental instantons, one with $k = 1$ and one with $k = 4$. This turns out not to be the case. The index theory results given below show that this $k = 4$ solution has precisely the number of zero modes that would be expected for a four-instanton solution, and so should be viewed as just a particularly symmetric configuration of overlapping unit instantons. For SU(3), and for all larger G, there is only one fundamental instanton, with $k = 1$.

For any G, the raising and lowering operators corresponding to any one of the roots can be used to define an SU(2) subgroup. [Explicit formulas for the generators are given in Eq. (A.7).] If the roots of G are all of equal length, embedding via any such subgroup yields a unit instanton; when there are two root lengths, a long root must be used. For the classical groups these subgroups are easily visualized. If $G = \mathrm{SU}(N)$, it is a subgroup defined by any 2×2 block of the $N \times N$ matrices. For $\mathrm{SO}(N)$, with $N \geq 5$, the relevant subgroup is obtained by taking an $\mathrm{SO}(4) = \mathrm{SU}(2) \times \mathrm{SU}(2)$ subgroup of $\mathrm{SO}(N)$ in the obvious way and then taking one of the SU(2) factors. For $\mathrm{Sp}(2N)$, viewed as a group of $N \times N$ quaternionic matrices, it is the $\mathrm{SU}(2) = \mathrm{Sp}(2)$ corresponding to any diagonal element.

Extending the index calculation of Sec. 10.8 to an arbitrary gauge group is relatively straightforward [254]. The steps leading up to Eq. (10.138) do not depend on the gauge group. Taking that result, and choosing R to be the adjoint representation of G, we have[10]

$$\mathcal{I}_b = 4kT(\mathrm{adjoint}(G)). \tag{10.145}$$

[10] The normalization here is such that the long roots of the group have unit length. In this case $T(\mathrm{adjoint}(G))$ is equal to the dual Coxeter number of the group.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

We might therefore expect that after subtracting the global gauge zero modes the number of parameters would be

$$n = 4kT(\text{adjoint}(G)) - \dim(G). \tag{10.146}$$

While this is correct for large values of k, it clearly can't be right for small k. For SU(N), we have $T(\text{adjoint}(G)) = N$ and $\dim(G) = N^2 - 1$, so we would have $n = 4Nk - N^2 + 1$, which for low values of k and $N \geq 5$ predicts a negative number of parameters. The explanation is simple. Consider the one-instanton solution in SU(N). This can always be brought into the upper left 2×2 block by a gauge transformation. Clearly, the U($N-2$) corresponding to the lower right $(N-2) \times (N-2)$ block leaves the instanton invariant, so the total number of global gauge modes is not equal to the dimension of SU(N), but rather

$$\dim[\mathrm{SU}(N)] - \dim[\mathrm{U}(N-2)] = 4N - 5. \tag{10.147}$$

Subtracting this from $\mathcal{I}_b = 4Nk = 4N$ gives five parameters, which are just the position and scale parameters inherited from the SU(2) solution. A generic two-instanton solution cannot be put into a 2×2 block, but it can be rotated into a 4×4 one and so will be left invariant by a U($N-4$) subgroup. It is only when $k \geq N/2$ that all of the elements of G act nontrivially on a generic solution. Extending the above analysis to arbitrary $k < N/2$, one finds that for SU(N) the number of parameters is

$$n = \begin{cases} 4Nk - (N^2 - 1)\,, & k \geq N/2, \\ 4k^2 + 1\,, & k < N/2. \end{cases} \tag{10.148}$$

The $4N$ that appears as the coefficient of k is just the sum of four position, one scale, and $4N - 5$ group orientation parameters for each component unit instanton.

The situation is similar for the other simple groups. For Sp($2N$), with $N \geq 2$,

$$n = \begin{cases} 4(N+1)k - N(2N+1)\,, & k \geq N, \\ 2k^2 + 3k\,, & k < N, \end{cases} \tag{10.149}$$

while for SO(N), with $N \geq 5$,

$$n = \begin{cases} 4(N-2)k - \frac{N(N-1)}{2}\,, & k \geq (N-1)/4, \\ 8k^2 - 6k\,, & 2 \leq k < (N-1)/4, \\ 5\,, & k = 1. \end{cases} \tag{10.150}$$

[Recall that SO(3) = Sp(2) = SU(2), and that SO(4) is not simple.]

For the exceptional groups there are five parameters when $k = 1$, just as with the classical groups. For larger values of k we have

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

$$
\begin{aligned}
G_2: &\qquad n = 16k - 14\,, \qquad k \geq 2\,,\\
F_4: &\qquad n = 36k - 52\,, \qquad k \geq 2\,,\\
E_6: &\qquad n = \begin{cases} 20\,, & k = 2,\\ 48k - 78\,, & k \geq 3,\end{cases}\\
E_7: &\qquad n = \begin{cases} 20\,, & k = 2,\\ 72k - 133\,, & k \geq 3,\end{cases}\\
E_8: &\qquad n = \begin{cases} 20\,, & k = 2,\\ 120k - 248\,, & k \geq 3.\end{cases}
\end{aligned}
\tag{10.151}
$$

10.10 The Atiyah–Drinfeld–Hitchin–Manin construction

A very powerful construction that transforms the problem of solving the self-duality differential equation (10.80) to one of solving a set of algebraic matrix equations is due to Atiyah, Drinfeld, Hitchin, and Manin [246, 247]. In this section I will present this ADHM construction for the SU(2) theory, leaving the extension to larger groups to the next section.

The construction is most easily expressed in terms of quaternions. Recall that quaternions are generalizations of complex numbers that involve three imaginary units i, j, and k that anticommute and obey

$$i^2 = j^2 = k^2 = -1\,, \qquad ijk = -1. \tag{10.152}$$

An arbitrary quaternion can be written as

$$q = a + bi + cj + dk \tag{10.153}$$

where a, b, c, and d are real numbers. If only a is nonzero, q is said to be real; if $a = 0$, then q is pure imaginary. The conjugate of a quaternion generalizes complex conjugation, so that the conjugate of the above quaternion is defined to be

$$q^* = a - bi - cj - dk\,. \tag{10.154}$$

Finally, a unit quaternion is one obeying $q^*q = 1$.

For our purposes here it will be convenient to define

$$\hat{e}_1 = -i\,, \quad \hat{e}_2 = -j\,, \quad \hat{e}_3 = -k\,, \quad \hat{e}_4 = 1 \tag{10.155}$$

and to expand an arbitrary quaternion as

$$q = q_r \hat{e}_r. \tag{10.156}$$

These conventions have been chosen so that the combination

$$\chi_{rs} = \hat{e}_r \hat{e}_s^* - \hat{e}_s \hat{e}_r^* \tag{10.157}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

is self-dual. [An easy way to see this is to note that the $\hat{e}_r$ can be represented by the matrices e_r, and then use Eq. (10.74).]

With these preliminaries completed, let us proceed with the construction. We begin with a $(k+1) \times k$ quaternionic matrix of the form

$$M = B - Cx\,, \tag{10.158}$$

where B and C are constant matrices and x is the quaternionic representation of a point in Euclidean space via

$$x = x_r \hat{e}_r\,. \tag{10.159}$$

We require that the $k \times k$ matrix

$$R = M^\dagger M \tag{10.160}$$

(where the dagger denotes the combination of the matrix transpose and quaternionic conjugation) be real and invertible.

The assertion is that if N is a $(k+1)$-component column vector obeying

$$N^\dagger M = 0 \tag{10.161}$$

and

$$N^\dagger N = I\,, \tag{10.162}$$

then a self-dual SU(2) gauge field with instanton number k is given by the quaternion

$$\mathcal{A}_p = \frac{1}{g} N^\dagger \partial_p N = -\frac{1}{g} (\partial_p N^\dagger) N\,, \tag{10.163}$$

with the second equality following from the normalization condition on N.

To see that $\mathcal{A}_p$ defines an SU(2) field, note first that the two expressions for $\mathcal{A}_p$ are minus the conjugates of each other, showing that $\mathcal{A}_p$ is a purely imaginary quaternion and so can be expanded as

$$\mathcal{A}_p = -\sum_{a=1}^{3} \frac{\hat{e}_a}{2} A_p^a. \tag{10.164}$$

The A_p^a are then the components of the potential in the standard SU(2) notation.

To show that the field is self-dual, we write the field strength

$$\begin{aligned}
\mathcal{F}_{pq} &= -\sum_{a=1}^{3} \frac{\hat{e}_a}{2} F_{pq}^a \\
&= \partial_p \mathcal{A}_q + g \mathcal{A}_p \mathcal{A}_q - (p \leftrightarrow q) \\
&= \frac{1}{g} \left[\partial_p N^\dagger \partial_q N - \partial_p N^\dagger N N^\dagger \partial_q N - (p \leftrightarrow q) \right] \\
&= \frac{1}{g} \left[\partial_p N^\dagger \left(I - N N^\dagger \right) \partial_q N - (p \leftrightarrow q) \right].
\end{aligned} \tag{10.165}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

The quantity in parentheses on the last line is a projection operator that projects onto the subspace orthogonal to N, which is the same as the subspace spanned by M. The identity of these two subspaces implies that

$$I - NN^\dagger = M(M^\dagger M)^{-1}M^\dagger = MR^{-1}M^\dagger. \tag{10.166}$$

Substituting this into the last line of Eq. (10.165) and using the orthogonality of N and M gives

$$\begin{aligned}\mathcal{F}_{pq} &= \frac{1}{g}\left[\partial_p N^\dagger M R^{-1} M^\dagger \partial_q N - (p \leftrightarrow q)\right] \\ &= \frac{1}{g}\left[N^\dagger\, \partial_p M\, R^{-1} \partial_q M^\dagger\, N - (p \leftrightarrow q)\right] \\ &= \frac{1}{g}\left[N^\dagger C \hat{e}_p R^{-1} \hat{e}_q^* C^\dagger N - (p \leftrightarrow q)\right].\end{aligned} \tag{10.167}$$

It is at this point that we use the requirement that $R = M^\dagger M$ be real. Because all of its components are real, it commutes with all of the $\hat{e}_p$, allowing us to write

$$\mathcal{F}_{pq} = \frac{1}{g} N^\dagger C R^{-1}(e_p \hat{e}_q^* - e_q \hat{e}_p^*) C^\dagger N. \tag{10.168}$$

We saw above that the quantity in parentheses was self-dual, thus showing that the field strength is self-dual, as claimed.

Directly demonstrating that the instanton number is equal to k is a less trivial matter. However, we will see presently that the 't Hooft solution with instanton number k can be obtained in this manner. Because the instanton number takes only integer values, any matrices M that satisfy the reality condition and are continuously connected to the M that yields the 't Hooft solution must also have instanton number k. Assuming that the solutions of the reality condition with a given k are all continuously connected then yields the desired result.

The matrix M is not uniquely determined by the gauge field. The gauge field A_p is unchanged if we make the replacement

$$\begin{aligned} M &\to SMT\,, \\ N &\to SN\,, \end{aligned} \tag{10.169}$$

where S and T are x-independent matrices with S a quaternionic matrix obeying $S^\dagger S = I$ and T any invertible real matrix. We can use this freedom to put the matrix C into a canonical form [248, 249]

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{10.170}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

in which the first row vanishes and the remainder form a $k \times k$ unit matrix. This does not eliminate all the freedom, but still allows transformations with the $(k+1) \times (k+1)$ matrix S in the block diagonal form

$$S = \begin{pmatrix} u & 0 \\ 0 & T^{-1} \end{pmatrix}, \tag{10.171}$$

where u is a unit quaternion and T is a $k \times k$ orthogonal matrix.

With C in this canonical form, it is convenient to decompose B as

$$B = \begin{pmatrix} v \\ b \end{pmatrix} \tag{10.172}$$

where v is a k-component row vector of quaternions and b is a $k \times k$ quaternionic matrix.

The reality condition is that

$$\begin{aligned} M^\dagger M &= B^\dagger B - x^* C^\dagger B - B^\dagger C x + x^* C^\dagger C x \\ &= \left(v^\dagger v + b^\dagger b\right) - \left(x^* b + b^\dagger x\right) + x^* x I \end{aligned} \tag{10.173}$$

be real. The last term is manifestly real. The terms linear in x are of the form

$$x^* b_{mn} + (b^\dagger)_{mn} x = x^* b_{mn} + (b_{nm})^* x \tag{10.174}$$

and are real for arbitrary x if and only if b is symmetric. With b symmetric, the reality condition then reduces to the requirement that

$$B^\dagger B = v^\dagger v + b^\dagger b \tag{10.175}$$

be real.

At this point let us stop and count the number of free parameters. The vector v has $4k$ real components, and the symmetric matrix b has $4[k(k+1)/2]$. The diagonal elements of $B^\dagger B$ are automatically real, while the elements in the upper right are the conjugates of those in the lower left. Thus, the vanishing of the imaginary components of $B^\dagger B$ gives $3[k(k-1)/2]$ constraints on the parameters in v and b, leaving $k(k+15)/2$ independent components. Of these, $[k(k-1)/2]+3$ represent the freedom to make $\mathrm{SO}(k) \times \mathrm{SU}(2)$ gauge transformations while still preserving the canonical form of C. Subtracting these leaves precisely $8k-3$ physical parameters. This is just the number required for the general instanton number k solution, strongly suggesting that the ADHM construction yields all instanton solutions. An explicit confirmation that it does is given by an inverse construction [250]. Here one starts with a self-dual gauge field with instanton number k. Equation (10.138) shows that the massless Dirac equation for a fermion in the fundamental doublet representation in this background has k zero modes. The vector v and the matrix b can be written in terms of these zero modes, and one can then show that they obey the requisite reality condition, thus completing the demonstration of completeness.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

As an explicit example of the ADHM construction, consider the case where the components of v are all real and b is diagonal. We then have

$$B = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & w_k \end{pmatrix}, \tag{10.176}$$

with the λ_j being real and the w_j being quaternions that can be interpreted as positions as in Eq. (10.159). This clearly satisfies the reality constraint.

We now need to find a quaternionic vector N that satisfies Eq. (10.161). If we label the first row by 0, this equation can be written as

$$0 = (N^\dagger M)_j = N_0^\dagger \lambda_j + N_j^\dagger (w_j - x). \tag{10.177}$$

Without any loss of generality, we can write the first component of N as

$$N_0 = \frac{u(x)}{\sqrt{\rho(x)}}, \tag{10.178}$$

where u is a unit quaternion satisfying $u^* u = 1$ and ρ is a real normalization factor to be determined shortly. Equation (10.177) then gives

$$N_j = \frac{\lambda_j (x - w_j)}{\sqrt{\rho}|x - w_j|^2}\, u(x) \tag{10.179}$$

and the normalization condition, Eq. (10.162), becomes

$$\rho = u^* u + \sum_j \frac{\lambda_j^2}{(x - w_j)^2} = 1 + \sum_j \frac{\lambda_j^2}{(x - w_j)^2}. \tag{10.180}$$

Setting $u(x) = 1$ yields the 't Hooft solution of Eqs. (10.94) and (10.97). Making a different choice for $u(x)$ gives a gauge transformation of this solution, since the transformation $N \to NW$ gives

$$\mathcal{A}_p = \frac{1}{g} N^\dagger \partial_p N \to W^\dagger \mathcal{A}_p W + \frac{1}{g} W^\dagger \partial_p W. \tag{10.181}$$

If we allow the components of v to be arbitrary quaternions, which can be written as $\lambda_j s_j$, with λ_j a positive real number and s_j a unit quaternion, b must have nonzero off-diagonal elements, and solving Eq. (10.175) becomes a rather nontrivial matter. However, if the λ_j are all small compared to the separations $|w_i - w_j|$, one can solve perturbatively for the off-diagonal elements of b [248]. This gives a solution that corresponds to k well-separated instantons. As before, the λ_j and w_j are scale and position parameters for the individual instantons. The $3k$ angles in the s_j are SU(2) orientation angles, of which three are redundant because multiplying all of the s_j by a common unit quaternion has no effect on the Euclidean space gauge fields.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

10.11 The ADHM construction for larger gauge groups

The ADHM formalism can be generalized from SU(2) to the classical groups SO(n), SU(n), and Sp($2n$) [246–249]. As with SU(2), one begins with a matrix

$$M = B - Cx \tag{10.182}$$

and then finds a matrix N obeying

$$N^\dagger M = 0\,, \qquad N^\dagger N = I. \tag{10.183}$$

The gauge potential is again given by

$$A_p = \frac{1}{g}\, N^\dagger \partial_p N. \tag{10.184}$$

In each case we want A_p to be an $n \times n$ matrix. Its elements, and therefore those of N, should be quaternions, complex numbers, or real numbers, for Sp($2n$), SU(n), and SO(n), respectively. Satisfying these requirements imposes conditions on the size and properties of M.

As one might guess from the quaternionic character of the ADHM construction, the simplest case is that of Sp($2n$). For SU(2) = Sp(2), M is an $(k+1) \times k$ matrix satisfying the requirement that $M^\dagger M$ be real and invertible. The latter requirement ensures that M has rank k. Hence, the k columns of M together with the vector N span the $(k+1)$-dimensional quaternionic vector space. The extension to Sp($2n$) is obtained by taking M to be a $(k+n) \times k$ matrix, with $M^\dagger M$ again real and invertible. N is then $(k+n) \times n$, so that A_p is $n \times n$, as required. Taken together, the columns of M and N span the $(k+n)$-dimensional quaternionic vector space.

As with the SU(2) case, it is convenient to put B and C into a canonical form, which in this case is

$$C = \begin{pmatrix} 0_{n\times k} \\ \\ I_k \end{pmatrix} \qquad\qquad B = \begin{pmatrix} v \\ \\ b \end{pmatrix}\,, \tag{10.185}$$

where v is $n \times k$ and $B^\dagger B$ is real and invertible.

An embedding of the SU(2) 't Hooft solutions is obtained by taking b to be diagonal and the columns of v to be real multiples of a fixed quaternionic vector. The nonzero elements of b then give the instanton positions. The magnitudes of the columns of v give the instanton scales, while their common direction as quaternionic vectors gives the gauge orientation parameters that define an Sp(2) subgroup.

Let us now turn to SO(n). To obtain an SO(n) gauge field, we need N to be a real matrix with n columns. Although M is still naturally viewed as a quaternionic matrix, it can also be decomposed into four real matrices M_p, with

$$M = M_p \hat{e}_p. \tag{10.186}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

One is led to take M to be a $(4k+n)\times k$ quaternionic matrix. Its columns define k quaternionic vectors or, by separating out the coefficients of the $\hat{e}_p$, $4k$ real $(4k+n)$-component vectors. For the construction to work, we need that these, together with the n columns of N, span the $(4k+n)$-dimensional vector space. This requires that we impose additional conditions on M, since its columns can be linearly independent as quaternionic vectors without their components all being linearly independent as real vectors. A hint comes from noting that requiring that a quaternionic vector v be such that $v^\dagger \hat{e}_p v$ is real for all four values of p is equivalent to requiring that the four real vectors derived from v be mutually orthogonal and of equal length. Indeed, the proper condition on M turns out to be that

$$M^\dagger \hat{e}_q M \equiv 4R_q \tag{10.187}$$

be real for all four values of q and that

$$R = R_q \hat{e}_q \tag{10.188}$$

be invertible. These conditions imply that

$$M_p^t M_q = \delta_{pq} R_4 + \eta^j_{pq} R^j . \tag{10.189}$$

The canonical forms for C and B are

$$C = \begin{pmatrix} 0_{n\times k} \\ I_k \\ \hat{e}_1 I_k \\ \hat{e}_2 I_k \\ \hat{e}_3 I_k \end{pmatrix} \qquad B = \begin{pmatrix} v \\ b_0 + b_1 + b_2 + b_3 \\ \hat{e}_1(b_0 - b_1) \\ \hat{e}_2(b_0 - b_2) \\ \hat{e}_3(b_0 - b_3) \end{pmatrix} . \tag{10.190}$$

Here v is an $n \times k$ quaternionic matrix, while the b_r are $k \times k$ quaternionic matrices with b_0 symmetric and the other three antisymmetric.

To obtain the 't Hooft solutions we set $b_1 = b_2 = b_3 = 0$ and take b_0 to be diagonal, with its elements specifying the instanton positions. The columns of v must then be real multiples of a common quaternionic vector u for which the $u^\dagger e_p u$ are all real. As noted above, this implies that the four component vectors of u are mutually orthogonal. These then define an SO(4) subgroup of SO(n), and the construction associates the 't Hooft solution with one of the SU(2) factors of this SO(4).

It is now easy to see how the construction for SU(n) works. We take M to be a $(2k+n)\times k$ quaternionic matrix such that

$$M^\dagger M = R_4 , \qquad M^\dagger \hat{e}_1 M = R_1 \tag{10.191}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

are both real, with $R_4 + \hat{e}_1 R_1$ invertible. N is a $(2k+n) \times n$ complex matrix (i.e., one containing $\hat{e}_1$ and $\hat{e}_4$, but not $\hat{e}_2$ or $\hat{e}_3$). The canonical form is

$$C = \begin{pmatrix} 0_{n\times k} \\ I_k \\ \hat{e}_2 I_k \end{pmatrix} \qquad B = \begin{pmatrix} v \\ b_0 + b_2 \\ \hat{e}_2(b_0 - b_2) \end{pmatrix}, \tag{10.192}$$

with b_0 and b_2 as in the SO(n) construction. The 't Hooft solutions are obtained by setting $b_2 = 0$, taking b_0 to be diagonal, and choosing the columns of v to be real multiples of a quaternionic vector u with $u^\dagger \hat{e}_1 u$ real.

The general multi-instanton solutions are obtained by allowing the columns of v to be independently chosen quaternionic vectors. For well-separated instantons in Sp($2n$) the diagonal elements of b give the instanton positions and the remaining elements of B can be determined perturbatively, as in the SU(2) case. The input data for this consist of $n+1$ quaternions (one for the position and n for the components of the column of v) for each instanton. This gives a total of $4(n+1)$ real parameters per instanton, which agrees with the coefficient of k in Eq. (10.149). The remaining terms in that equation correspond to the freedom to perform linear transformations on the rows of v.

When this approach is extended to the case of well-separated instantons in SO(n) and SU(n), the columns of v must be chosen so that they each obey the appropriate reality conditions. For SO(n), this can be done by defining an n-component vector u whose first four elements are 1, $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$, with the remaining $n-4$ elements vanishing. An acceptable column for v is then obtained by acting on u with an arbitrary SO(n) matrix and then multiplying by a real scale parameter. The procedure for SU(n) is similar, except that one starts with a vector u whose first components are 1 and $\hat{e}_2$, with the remainder vanishing. One can check that these prescriptions again give the correct number of parameters.

There is an equivalent, although superficially somewhat different, formulation of the ADHM construction for SU(n). We start by regrouping the eight real matrices that make up the quaternionic matrices b_0 and b_2 into four Hermitian complex matrices Z_p. To do this, we write

$$\begin{aligned} b_0 &= \hat{e}_p X_p \\ b_2 &= \hat{e}_1 \hat{e}_p Y_p \end{aligned} \tag{10.193}$$

and define

$$Z_p = X_p + i Y_p. \tag{10.194}$$

Similarly, we trade the quaternionic $n \times k$ matrix v for a complex $n \times 2k$ matrix

$$\omega_a^r = \begin{pmatrix} \omega_{1a}^r \\ \omega_{2a}^r \end{pmatrix}, \tag{10.195}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

where r runs from 1 to k and a from 1 to n. These matrices can then be combined to give the $(n+2k)\times 2k$ complex matrix

$$\Delta = \begin{pmatrix} \omega^t \\ (Z_p - x_p I_k)\otimes e_p \end{pmatrix}. \tag{10.196}$$

We then proceed as before; i.e., we look for an $(n+2k)\times n$ complex matrix N satisfying

$$\begin{aligned} \Delta^\dagger N &= 0 \\ N^\dagger N &= I \end{aligned} \tag{10.197}$$

and again define the gauge field to be

$$A_p = \frac{1}{g} N^\dagger \partial_p N. \tag{10.198}$$

In this language, the self-duality conditions that $B^\dagger B$ and $B^\dagger \hat{e}_1 B$ be real become the triplet of constraints

$$0 = i[Z_p, Z_q]\bar{\eta}^j_{pq} - \sum_{a=1}^{n} \omega^\dagger_{a\alpha}(\sigma^i)_{\alpha\beta}\omega_{a\beta}. \tag{10.199}$$

[For clarity, I have suppressed the SU(k) indices on Z_p and ω.] Interestingly, this latter formulation of the construction with the self-duality condition as given in Eq. (10.199) can be naturally obtained as the condition for minimizing the energy of a configuration of k D0-branes in the presence of n D4-branes [251, 252].

10.12 One-loop corrections

The classical instanton solution gives the dominant exponential factor associated with tunneling. The leading contribution to the pre-exponential factor, corresponding to the quantity K in Eq. (9.39), arises from the fluctuations about the instanton configuration. It was first obtained by 't Hooft [238] in the context of an SU(2) gauge theory; the extension to larger groups is straightforward. The full details of his calculation are too lengthy to reproduce here. Instead, I will give an overview of the calculation, explaining how the various contributions arise. An alternative approach, using the O(5) invariance that follows by projecting the solution onto a four-sphere and using the conformal invariance of the theory, is given in [253].

The path integral derivation of this "one-loop" correction about an instanton was described in Sec. 9.3. That discussion was phrased in terms of a scalar field. When extending this to the Yang–Mills case, it is necessary to impose a gauge condition. This is most conveniently done by adding to the Lagrangian a gauge-fixing term of the form $-\frac{1}{2}(D_p A^{\rm qu}_p)^2$, where $A^{\rm qu}_p$ denotes the fluctuation

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

about the instanton solution and the covariant derivative is defined using only the background instanton potential. This then determines the form of the ghost Lagrangian.

To start, let us consider the contribution due to a single instanton with fixed position and scale size λ, written in the form given in Eq. (10.87). The relevant parts of the action can then be written schematically as

$$S = S_{\rm cl} - \frac{1}{2} A^{\rm qu} \mathcal{M}^A A^{\rm qu} - \bar{c} \mathcal{M}^{\rm gh} c + \cdots . \tag{10.200}$$

Here the terms for the gauge and ghost fields are shown explicitly, with integration over the spatial coordinates and traces over the implied indices understood, while the contributions from any matter fields that might be present are represented by an ellipsis. Ignoring these matter fields for the present, we wish to calculate

$$W(\lambda) = e^{-S_{\rm cl}} \left[\frac{\det' \mathcal{M}^A}{\det \mathcal{M}_0^A} \right]^{-1/2} \left[\frac{\det \mathcal{M}^{\rm gh}}{\det \mathcal{M}_0^{\rm gh}} \right] \prod_a J_a . \tag{10.201}$$

The prime indicates that the zero modes are to be omitted in the calculation of the determinant of $\mathcal{M}_A$. Instead, they are handled by introducing collective coordinates, which entails a Jacobian factor J_a for each zero mode. A subscript 0 denotes the corresponding determinant calculated about the vacuum configuration $A_p = 0$.

The determinants can be understood formally as products of eigenvalues. One way to realize this calculationally is to put the system in a box of finite size, so that the spectra of the various operators are discrete. The resulting infinite products are divergent, even after the division by the vacuum determinants, and so a regularization scheme is needed. Dimensional regularization cannot be used, because the background instanton solution is inherently four-dimensional. An alternative is to use a Pauli–Villars-type regulator, which leads to regularized determinant ratios of the form[11]

$$\frac{\det \mathcal{M}}{\det \mathcal{M}_0} = \prod_{j=1}^{\infty} \frac{\omega_j \left(\omega_{j0} + M^2\right)}{\omega_{j0} \left(\omega_j + M^2\right)} , \tag{10.202}$$

where the ω_j are the eigenvalues of $\mathcal{M}$ and M is the regulator mass.

Because of the spherical symmetry of the instanton, the modes can be chosen to be eigenstates of the spin and orbital $\mathrm{SO}(4) = \mathrm{SU}(2) \times \mathrm{SU}(2)$ angular momenta and the SU(2) gauge symmetry. The determination of the eigenvalues is then reduced to the solution of a radial Schrödinger-type equation that can be converted to a hypergeometric equation whose eigenvalues can be found. After

[11] I am oversimplifying somewhat. For the full details of the regularization, see [238].

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

taking the size of the spatial box to infinity, the net result for the gauge and ghost field determinants is

$$\left[\frac{\det' \mathcal{M}^A}{\det \mathcal{M}_0^A}\right]^{-1/2} \left[\frac{\det \mathcal{M}^{\rm gh}}{\det \mathcal{M}_0^{\rm gh}}\right] = C_1 (M\lambda)^{-2/3}, \tag{10.203}$$

where C_1 is a numerical constant. Note that this ratio of determinants is dimensionless, despite the omission of the zero modes in the numerator of the first factor, because of the way the regularized determinants have been defined in Eq. (10.202).

We must next consider the zero modes. Each of these corresponds to the variation of a collective coordinate z_a and so is of the form

$$a_p^a = \mathcal{N}_a^{-1/2} \left(\frac{\partial A_p^{\rm inst}}{\partial z_a} + D_p \Lambda^a \right), \tag{10.204}$$

where $\mathcal{N}_a$ is a normalization constant and Λ^a is the gauge transformation that brings the mode into background gauge. The Jacobian factor in Eq. (10.201) is related to the normalization constant by

$$J_a = \sqrt{\frac{\mathcal{N}_a}{2\pi}}\, M. \tag{10.205}$$

The factor of M arises from the regulator mode associated with the zero mode, while the $\sqrt{2\pi}$ appears because the integration over the collective coordinate replaces a Gaussian integral over a mode coefficient.

There are four translational zero modes. The derivatives with respect to instanton position are, up to a sign, the same as the spatial derivatives of the instanton solution. These modes therefore fall as $1/x^2$ at large distance and would lead to non-normalizable modes were it not for the transformation to background gauge. In background gauge, these modes are proportional to the field strength, fall as $1/x^4$, and are normalizable.

A fifth zero mode is associated with the scale parameter. It is given in background gauge simply by the derivative of the instanton with respect to λ, with no gauge transformation needed

Finally, there are three gauge zero modes that are not eliminated by the gauge-fixing term in the Lagrangian. These are of the form $a_p = D_p \Lambda(x)$, where the gauge function Λ is nonvanishing at spatial infinity. They can be viewed as corresponding to global gauge rotations.

Before normalization, each of these eight modes inherits a factor of $1/g$ from the instanton solution, leading to a factor of $1/g^2$ in the corresponding $\mathcal{N}_a$. The spatial integral of the square of each mode should equal unity, so the a_p^a must have dimensions of [mass]^2. The $\mathcal{N}_a$ for the translation and scale modes are therefore dimensionless, while those for the three global gauge modes have dimensions of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

$[\text{mass}]^{-2} \sim [\text{length}]^2$. Since λ is the only scale in the problem, these last three $\mathcal{N}_a$ must be proportional to λ^2. Hence, we see that

$$\prod_a J_a = C_2\, g^{-8} M^8 \lambda^3\,, \tag{10.206}$$

where C_2 is another numerical constant.

Substituting our results into Eq. (10.201) gives

$$W(\lambda) = C_1 C_2\, g^{-8}\, \lambda^{-5} \exp\left[-\frac{8\pi^2}{g^2} + \frac{22}{3}\ln(M\lambda)\right]. \tag{10.207}$$

Let us examine the exponent in this expression. Up to now, we have been simply writing the gauge coupling g without taking into account that it runs with energy and so is not well-defined until a scale is specified. Because we have not explicitly included any counterterms in our Lagrangian, the g we are using must be the bare coupling, or rather, since we have introduced a regulator to make everything finite, $g(M)$. The one-loop beta function for SU(2) gives the running of the coupling as

$$\frac{8\pi^2}{g^2(\mu)} = \frac{8\pi^2}{g^2(M)} - \frac{22}{3}\ln\left(\frac{M}{\mu}\right), \tag{10.208}$$

so the effect of the regulator contribution in Eq. (10.207) is to specify that the gauge coupling appearing in the exponent should be evaluated at an energy $1/\lambda$ or, equivalently, at a distance scale λ. This is not a surprise, but it is nice to see how it naturally comes about.[12] It also clear that a calculation to the next order would ensure that the factors of g in the prefactor would also be evaluated at a scale λ.

We now need to integrate over the gauge orientation and scale collective coordinates. After performing the integral over the SU(2) gauge orientation parameters χ^a and evaluating the numerical constants, we find that the weighting factor for instantons of size λ is

$$\int d^3\chi\, W(\lambda) = 2^{14}\pi^6 e^A \frac{1}{\lambda^5}\left[\frac{1}{g(\lambda)}\right]^8 \exp\left[-\frac{8\pi^2}{g(\lambda)^2}\right], \tag{10.209}$$

where $A = 6.9984$ is a constant for which an analytic expression is given in [238]. (Note that product of the numerical prefactors is 1.7×10^{10}.)

We now have to consider the integral over λ. At short distance and high energy the gauge coupling tends to zero, the instanton action tends to infinity, and W, which contains the exponential of a large negative number, tends to zero fast enough that this integral converges at the short-distance end. For large instantons

[12] With a gauge group larger than SU(2), the contributions to Eq. (10.203) from the additional global gauge modes give precisely the extra factors of $M\lambda$ that are needed to give the correct beta function, and thus the proper running coupling, for the larger group. The extra modes also give additional factors of g^{-1} in the prefactor.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011

we find the opposite behavior. With increasing λ the gauge coupling grows and the instanton action decreases, so that our semiclassical expansion eventually becomes unreliable. Unless one is considering a process that puts a natural upper cutoff on λ, the integration over instanton size diverges at large λ.

Furthermore, the dilute-gas approximation, which underlies this calculation, breaks down for large instantons. This approximation assumes that the density of instantons is small relative to the instanton size, which requires that the integral

$$\int d\lambda \, \lambda^4 W(\lambda) \ll 1. \tag{10.210}$$

With $W(\lambda)$ given by Eq. (10.209), this integral diverges at large λ, and the required inequality is clearly violated.

In some cases, such as the electroweak instantons that will be discussed in Sec. 11.4, other considerations restrict the instanton size to a narrow range. The integration over λ is then finite, and for weak coupling the instanton effects are small, just as expected for a tunneling process. For many QCD applications, on the other hand, there is no such cutoff on the instanton size and large instantons, with a strong effective coupling, are important. The net result is that instanton effects can be expected to be large, although they can no longer be reliably calculated with precision. Nevertheless, important qualitative conclusions can be drawn.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 17:57:32, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.011