3.3 θ Angle

In addition to the infinitesimal gauge transformations, one may also perform finite, time-independent gauge transformations, U. These leave the Lagrangian (3.15) and the Hamiltonian (3.18) invariant, and are implemented by the unitary operator $\mathscr{G}_U$.

$$A \to \mathscr{G}_U A \mathscr{G}_U^\dagger = U^{-1} A U - U^{-1} \nabla U \quad ,$$

$$E \to \mathscr{G}_U E \mathscr{G}_U^\dagger = U^{-1} E U \quad ,$$

$$B \to \mathscr{G}_U B \mathscr{G}_U^\dagger = U^{-1} B U \quad ,$$

$$[H, \mathscr{G}_U] = 0 \quad . \tag{3.30}$$

Clearly the effect of $\mathscr{G}_U$ on all states (physical and unphysical) is to gauge transform the argument,

$$\mathscr{G}_U \Psi(A) = \Psi(A^U) \equiv \Psi(U^{-1} A U - U^{-1} \nabla U) \quad . \tag{3.31}$$

Since $\mathscr{G}_U$ and H commute, we may choose $\Psi(A)$ to be an eigenstate of $\mathscr{G}_U$, for a given U, with an eigenvalue which is a phase, since $\mathscr{G}_U$ is unitary:

$$\mathscr{G}_U \Psi(A) = e^{-i\theta_U} \Psi(A) \quad . \tag{3.32}$$

The question now is whether physical states, i.e., those annihilated by G_a, are truly invariant against finite gauge transformations or only phase invariant. Let us observe that no physical principle will be violated if θ_U is nonzero, since the probability density $\Psi^*(A)\Psi(A)$ is gauge invariant.

One might suppose that $\mathscr{G}_U$ could be represented by exponentiating the infinitesimal generator (3.22),

$$\mathscr{G}_U = \exp\left(\frac{i}{\hbar} Q_\theta\right)$$

with $U = e^\Theta$ and $\Theta = \theta^a T^a$, where the exponential operator is defined by its power series. If this were the case, then $\mathscr{G}_U$ would leave physical states invariant, because G_a and hence Q_θ annihilates them. (An integration by parts is needed for G_a to act on the state; we assume no surface terms arise, for otherwise certainly $\mathscr{G}_U$ will not leave the state invariant. A precise statement about large-distance behavior of gauge functions will be given presently.) However, it is easy to see that in three-dimensional space there are finite gauge transformations that are not obtained by operating with the exponential operator.

Consider the following quantity, defined in three-space:

$$W(A) = \frac{-1}{4\pi^2} \int \mathrm{d}r\, \epsilon^{ijk}\, \mathrm{tr}\left(\frac{1}{2} A^i \partial_j A^k - \frac{1}{3} A^i A^j A^k\right)$$

$$= -\frac{1}{16\pi^2} \int \mathrm{d}r \epsilon^{ijk}\, \mathrm{tr}\left(F_{ij} A_k - \frac{2}{3} A_i A_j A_k\right) . \tag{3.33}$$

I shall show that $W(A)$ is not gauge invariant, yet is is invariant against gauge transformations implemented by $\exp(\mathrm{i}Q_\theta/\hbar)$. Observe that $W(A)$ has the property that

$$\frac{\delta W(A)}{\delta A_a^i} = \frac{g^2}{8\pi^2} B_a^i \quad . \tag{3.34}$$

Consequently

$$\mathrm{e}^{\mathrm{i}Q_\theta/\hbar} W(A) \mathrm{e}^{-\mathrm{i}Q_\theta/\hbar}$$

$$= W(A) + \frac{\mathrm{i}}{\hbar} [Q_\theta , W(A)] + \ldots$$

$$= W(A) + \int \mathrm{d}r \frac{\mathrm{i}}{\hbar} [Q_\theta, A_a(r)] \cdot \frac{g^2}{8\pi^2} B_a(r) + \ldots$$

$$= W(A) - \frac{g}{8\pi^2} \int \mathrm{d}r (D\theta)_a \cdot B_a + \ldots$$

$$= W(A) + \frac{g}{8\pi^2} \int \mathrm{d}r \theta_a (D \cdot B)_a + \ldots$$

$$= W(A) \quad , \tag{3.35}$$

since the Biänchi identity states that the covariant divergence of $\boldsymbol{B}$ is zero. (In order to drop surface terms, we have assumed that θ^a goes to a constant at spatial infinity and the $\boldsymbol{B}$ falls faster than $1/r^2$, i.e., there are no magnetic monopoles; see also below.) Thus $W(A)$ is gauge invariant against gauge transformations implemented by the exponential operator $\exp(\mathrm{i}Q_\theta/\hbar)$.

The gauge transformation of $W(A)$ may also be computed directly:

$$
\begin{aligned}
\mathscr{G}_U W(A) \mathscr{G}_U^{-1} &= W(A^U) \\
&= W(A) - \frac{1}{8\pi^2} \int \mathrm{d}r \epsilon^{ijk} \partial_i \, tr(\partial_j U U^{-1} A^k) \\
&\quad + \frac{1}{24\pi^2} \int \mathrm{d}r \epsilon^{ijk} \, \mathrm{tr}(U^{-1} \partial_i U U^{-1} \partial_j U U^{-1} \partial_k U) \quad . \qquad (3.36)
\end{aligned}
$$

To evaluate the gauge change in $W(A)$, we must impose a boundary condition on U so that integrals in (3.36) not diverge. We shall assume that U tends to a constant matrix at spatial infinity, and without loss of generality this may be taken to be $\pm I$. Also A is assumed to fall faster than $1/r$. Note that both these statements are consistent with the earlier requirement that θ^a approach a constant and $\boldsymbol{B}$ decrease faster than $1/r^2$. With these, the middle integral in (3.36) vanishes, but the last remains. It occurs only for a non-Abelian theory and makes no reference to the potentials since it depends on only the non-Abelian gauge transformations,

$$
\begin{aligned}
\omega(U) \equiv & \frac{1}{24\pi^2} \int \mathrm{d}r \epsilon^{ijk} \, \mathrm{tr}(U^{-1} \partial_i U U^{-1} \partial_j U U^{-1} \partial_k U) \\
& = W(-U^{-1} \nabla U) \quad . \qquad (3.37)
\end{aligned}
$$

This integration may be performed for a definite U, and one finds a non-vanishing results; see Exercise 3.5.

However, it is not necessary to perform the explicit evaluation. We recognize that the gauge functions U, with large-distance asymptotes $\pm I$, provide a mapping of the three-sphere S_3 (which is equivalent to our three-space once the points at infinity have been identified) into the gauge group. Such mappings fall into disjoint homotopy classes, labeled by the integers, and gauge functions belonging to different classes cannot be deformed continuously into each other. In particular, only those in the zero class are deformable to the identity. This fact is expressed by the mathematical statement that

$$\Pi_3 \text{ (non-Abelian compact gauge group)} = \text{(group of all integers under addition)} \equiv \mathscr{Z} . \tag{3.38}$$

Furthermore, $\omega(U)$ is an analytic expression for the integer which labels U's homotopy class. It is called the "winding number" of the gauge transformation.

Thus we see that $W(A)$ is not gauge invariant against homotopically non-trivial gauge transformations; rather it changes by U's winding number.

$$W(A) \to W(A^U) = W(A) + n_U$$
$$n_U \equiv \omega(U) \tag{3.39}$$

We now recognize that $\mathscr{G}_U$ may be represented by the exponential only when U belongs to the trivial homotopy class and is deformable to the identity. (For these θ^a vanishes at $r = \infty$.) Correspondingly physical states are gauge invariant against these gauge transformations. But homotopically non-trivial gauge transformations that are not deformable to the identity are not implemented by the exponential operator, and physical states are only phase invariant. If $\mathscr{G}_n$ is the unitary operator that implements a representative gauge transformation U_n belonging to the nth homotopy class, then we have

$$\mathscr{G}_n |\text{physical}\rangle = e^{-in\theta} |\text{physical}\rangle ,$$
$$\left.\begin{aligned} \mathscr{G}_0 &= e^{iQ_\theta/\hbar} = I \\ \mathscr{G}_n \mathscr{G}_m &= \mathscr{G}_{n+m} \end{aligned}\right\} \text{ on physical states.} \tag{3.40}$$

[It should be clear that θ in (3.40) is distinct from the gauge parameter θ^a.] The topologically non-trivial gauge transformations are called "large", while the trivial ones are called "small."

This is the origin of the famous vacuum angle in gauge theories [46, 47]; we see that its presence is established without any approximation but rather by carefully following the response of a non-Abelian gauge theory (in four-dimensional space-time) to the large gauge transformations which are topologically richer than anything in electrodynamics [46]. We emphasize that in two spatial dimensions, all gauge transformations are small.

Exercise 3.5. Consider the SU(2) gauge group, and evaluate (3.37) with

$$U = e^{\theta^a T^a} \quad , \qquad T^a = \sigma^a/2i \quad . \tag{E3.8}$$

Show that the integrand in (3.37) may be written as a total divergence, so that the volume integral for $\omega(U)$ may be recast as an integral over the surface at infinity, provided θ^a is regular in the interior.

$$\omega(U) = \frac{1}{16\pi^2}\int dS^i \epsilon^{ijk} \epsilon_{abc}\, \hat{\theta}^a \partial_j \hat{\theta}^b \partial_k \hat{\theta}^c (\sin|\theta| - |\theta|) \quad ,$$

$$|\theta| = \sqrt{\theta^a\theta^a} \quad , \qquad \hat{\theta}^a = \theta^a/|\theta| \quad . \tag{E3.9}$$

Exercise 3.6. For an arbitrary SU(2) gauge transformation parametrized as in (E3.8), the condition $U \xrightarrow[r\to\infty]{} \pm I$ sets the requirement that $|\theta| \xrightarrow[r\to\infty]{} 2\pi n$. Thus the general formula (E3.9) reduces to

$$\omega(U) = -\frac{n}{8\pi}\int dS^i \epsilon^{ijk} \epsilon_{abc}\, \hat{\theta}^a \partial_j \hat{\theta}^b \partial_k \hat{\theta}^c \quad . \tag{E3.10}$$

By parametrizing the unit three-vector $\hat{\theta}^a$ as

$$\hat{\theta}^1 = \sin\psi\cos\phi \quad , \qquad \hat{\theta}^2 = \sin\psi\sin\phi \quad , \qquad \hat{\theta}^3 = \cos\psi \quad , \tag{E3.11}$$

and the surface of the two-sphere at $r = \infty$ by angles α and β on which ψ and ϕ depend, show that

$$\omega(U) = -\frac{n}{4\pi}\int_0^{2\pi} d\beta \int_0^{\pi} d\alpha \sin\psi \left(\frac{\partial\psi}{\partial\alpha}\frac{\partial\phi}{\partial\beta} - \frac{\partial\psi}{\partial\beta}\frac{\partial\phi}{\partial\alpha}\right) \quad . \tag{E3.12}$$

The quantity in parentheses is the Jacobian, apart from sign, of the transformation from (α, β) to (ψ, ϕ). Hence the above integral is the integer that counts the (signed) number of times (ψ, ϕ) range over their two-sphere as (α, β) range over theirs. This shows quite generally, that $\omega(U)$ is an integer.

Verify the above analysis by considering the gauge function

$$\theta^a = \hat{r}^a f(r), \qquad f(0) = 0 \quad , \qquad f(\infty) = 2\pi n \quad . \tag{E3.13}$$

Evaluate (3.37), (E3.9) and (E3.12). (It is important to appreciate that here one is not compactifying R_3 to S_3. Even though the points at infinity can be identified as far as the gauge function U is concerned — by hypothesis U tends to a uniform angle-independent limit — the Lie-algebra valued quantity Θ does not possess a uniform limit; $\Theta \xrightarrow[r\to\infty]{} -i\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\pi n$. Thus with the asymptote (E3.13), the Lie-algebra cannot be defined on a compact manifold, and this is why a total derivative expression for the integrand $\omega(U)$ can be given. On a compact manifold no such formula is available. The generalization of (E3.9) to an arbitrary group on R_3 is given in ref. [48]).

Exercise 3.7. Prove that $\exp[\pm(8\pi^2/\hbar g^2)\,W(A)]$ solves the non-Abelian functional Schrödinger equation (3.29) with zero eigenvalue. Regrettably this remarkable solution diverges for large A and does not seem to have physical meaning [49].

3.4 *Functional Integral Formulation*

The Yang-Mills quantum theory can be transcribed into a functional integral formulation, which has the advantage of exhibiting the vacuum angle in an unambiguously gauge- and Lorentz-invariant fashion. In order to derive the functional integral, we prefer to work with states which are invariant even against large gauge transformations. Thus we inquire whether we can modify our physical wave functionals so that they are gauge invariant. This is easy to do. Recall from Eq. (3.39) that $W(A)$ shifts by n under a gauge transformation in homotopy class n. Hence

$$\Phi \equiv e^{iW(A)\theta}\,\Psi_{\mathrm{phys}} \,, \tag{3.41}$$

continues to be annihilated by G_a, and is also gauge invariant against all gauge transformations, large and small. However, Φ satisfies a Schrödinger equation more complicated than (3.29), which follows from (3.34) and (3.41).

$$\int dr \left[\frac{1}{2}\left(\frac{\hbar}{i}\frac{\delta}{\delta A_a} - \frac{\hbar\theta g^2}{8\pi^2}B_a\right)^2 + \frac{1}{2}B_a^2\right]\Phi = E\Phi \,. \tag{3.42}$$

The path integral for this Hamiltonian is given by analogy with the constrained quantum mechanical system; see (3.11).

$$Z = \int \mathscr{D}E_a\,\mathscr{D}A_a\,\delta(G_a)\,\delta(\chi_b)\det\{G_a,\chi_b\}$$
$$\times \exp\frac{-i}{\hbar}\int dx\left\{E_a\cdot\dot{A}_a + \frac{1}{2}\left(E_a + \frac{\hbar\theta g^2}{8\pi^2}B_a\right)^2 + \frac{1}{2}B_a^2\right\}. \tag{3.43a}$$

Here χ_b is an arbitrary gauge choice, taken to depend on A_a. Since G_a generates gauge transformations, the Poisson bracket is just an infinitesimal gauge transformation of χ_b, with "parameter" θ^a stripped off; we represent it by $\delta_a \chi_b$. The constraint delta function may be written as a (functional) phase integral, with an integration variable which we call A_a^0. Thus (3.43a) becomes

$$Z = \int \mathscr{D}A_\mu^a \, \mathscr{D}E_a \, \delta(\chi_b) \det(\delta_a \chi_b) \exp \frac{-i}{\hbar} \times \int dx \left\{ E_a \cdot \dot{A}_a - A_a^0 (D \cdot E)_a + \frac{1}{2}\left(E_a + \frac{\hbar \theta g^2}{8\pi^2} B_a\right)^2 + \frac{1}{2} B_a^2 \right\} . \tag{3.43b}$$

Next the Gaussian E_a integral is performed, leaving an expression that may be written in invariant notation.

$$Z = \int \mathscr{D}A_\mu^a \, \delta(\chi_b) \det(\delta_a \chi_b) \exp \frac{i}{\hbar} \int dx \mathscr{L},$$

$$\mathscr{L} \equiv \frac{1}{2g^2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} - \frac{\hbar\theta}{16\pi^2} \operatorname{tr} {*F^{\mu\nu}} F_{\mu\nu} . \tag{3.43c}$$

Note that even though in our derivation the gauge condition χ depends only on A, the result holds for arbitrary χ; hence we may allow χ to depend also on A^0 provided the determinant is correspondingly adjusted [50].

Aside from the familiar gauge fixing delta function and gauge compensating determinant, we have arrived at the functional integral formulated in terms of the gauge- and Lorentz-invariant Yang-Mills action, but with an additional term contributing to the Lagrangian as a consequence of the angle θ. This gauge invariant term does not contribute to the equations of motion because it is a total divergence (of a gauge variant quantity).

$$\mathscr{P} \equiv - \frac{1}{16\pi^2} \operatorname{tr} {*F^{\mu\nu}} F_{\mu\nu} = \partial_\mu \mathscr{C}^\mu , \tag{3.44a}$$

$$\mathscr{C}^\mu = -\frac{1}{16\pi^2}\epsilon^{\mu\alpha\beta\gamma}\operatorname{tr}(F_{\alpha\beta}A_\gamma - \tfrac{2}{3}A_\alpha A_\beta A_\gamma) \tag{3.44b}$$

Also it does not contribute to the energy-momentum tensor; $\theta^{\mu\nu}$ in (2.33) remains conserved. However, tr $^*F^{\mu\nu}F_{\mu\nu}$ does affect the canonical formalism because it depends on time derivatives of the canonical variables. Observe that $W(A)$ is given by the spatial integral of $\mathscr{C}^0$, with the time dependence suppressed,

$$W(A) = \int d\mathbf{r}\,\mathscr{C}^0(A) \quad . \tag{3.45}$$

Exercise 3.8. Derive the energy-momentum tensor (2.33) for the Lagrangian in (3.43c) by using Noether's theorem and an appropriate improvement.

The form of the θ term in the Lagrangian shows that θ is Lorentz and gauge invariant. However, since tr $^*F^{\mu\nu}F_{\mu\nu}$ is odd under P and T reflection symmetries, the θ term is P and T [or CP] violating. Moreover, we see now that an independent argument for the existence of a θ-parameter in four-dimensional Yang-Mills theory is the fact that tr $^*F^{\mu\nu}F_{\mu\nu}$ could have been added to the original Lagrangian without affecting classical dynamics, which is entirely determined by equations of motion. However, such a term modifies quantum dynamics which depends on the Lagrangian and on the action, as is seen, for example, from the functional integral formulation.

Exercise 3.9. Starting with the Lagrangian of (3.43c) in the Weyl gauge ($A^0 = 0$) show that the Hamiltonian is conventional,

$$H = \frac{1}{2}\int d\mathbf{r}\,(\mathbf{E}_a^2 + \mathbf{B}_a^2) \quad . \tag{E3.14}$$

Derive the canonical form for the Hamiltonian in terms of canonical variables and compare with (3.42).

The conclusion therefore is that a four-dimensional Yang-Mills quantum theory is characterized not only by its coupling constant g, but also by a hidden parameter θ, which enters on the quantum level and involves tr$^*F^{\mu\nu}F_{\mu\nu}$. (This effect does not occur in three space-time dimensions, but we shall see later that there too an unexpected parameter characterizes the Yang-Mills theory.)

The novel addition to the action has a well-established place in mathematics: $\mathscr{P} \equiv -(1/16\pi^2)$ tr $^*F^{\mu\nu}F_{\mu\nu}$ is called the "Pontryagin density" and $\mathscr{C}^\mu$ — the vector whose divergence equals the Pontryagin density — is the "Chern-Simons secondary characteristic class".

For well-defined classical potentials, the Pontryagin index

$$q \equiv \int \mathrm{d}x \, \mathscr{P} \tag{3.46}$$

is a topological invariant; it does not change under local variations of the potentials.

$$\delta q = -\frac{1}{8\pi^2} \int \mathrm{d}x \, \mathrm{tr}\, {}^*F^{\mu\nu} \, \delta F_{\mu\nu}$$

$$= -\frac{1}{4\pi^2} \int \mathrm{d}x \, \mathrm{tr}\, {}^*F^{\mu\nu} D_\mu \, \delta A_\nu = 0 \; . \tag{3.47}$$

No surface terms arise in the integration by parts, which together with the Bianchi identity is needed to pass from the second to the third equalities, since δA_ν is arbitrary, and therefore may be taken to be localized. According to (3.44) q is given by the large-distance properties of the gauge potential — another hallmark of a topological quantity.

$$q = \int \mathrm{d}S_\mu \, \mathscr{C}^\mu \; . \tag{3.48}$$

If we consider gauge potentials that tend to a pure gauge at large distances in all four directions,

$$A_\mu \xrightarrow[x \to \infty]{} U^{-1} \partial_\mu U \; ; \tag{3.49a}$$

then q may be represented in terms of U, by substituting into the expression (3.44b) for $\mathscr{C}^\mu$ the asymptotic form of A_μ, (3.49a).

$$q = \frac{1}{24\pi^2} \int \mathrm{d}S_\mu \, \epsilon^{\mu\alpha\beta\gamma} \, \mathrm{tr}\,(U^{-1} \partial_\alpha U U^{-1} \partial_\beta U U^{-1} \partial_\gamma U) \; . \tag{3.49b}$$

Finally, note also that q is a geometrical invariant; even in curved space-time no factors of the metric are needed to make (3.46) a world scalar [51].

These topological remarks apply to well-behaved *classical* potentials which are all that mathematicians are concerned with. They are not directly relevant to quantum operator fields, nor to functional integrals, where the integration ranges over irregular field configurations. Most frequently the

the Pontryagin index is used for classical potentials defined on Euclidean four-space, and we shall meet it again when we discuss the semi-classical picture for the θ-angle. However, even on spaces with Minkowski signature q plays a role; for non-Abelian gauge fields that describe a 't Hooft-Polyakov monopole, q coincides with the monopole strength [52].

In Euclidean space q is an integer for regular potentials. This is seen by compactifying R_4 to S_4, and recognizing that the limit (3.49a) defines a gauge function U on S_3, the boundary of S_4. Again because Π_3 (gauge group) $= \mathscr{Z}$, the U's fall into integer-labeled homotopy classes and (3.49b) is an analytic expression for the winding number. Alternatively, one may work on R_4 (or even in Minkowski space) and present the limit (3.49a) in the following way.

Assume that in three directions, A_μ goes rapidly to zero, faster than $1/r$, where r is the modulus of the three-vector. In the fourth direction (time or Euclidean time) at negative infinity take A_μ to vanish, but at positive infinity to tend to a pure gauge. From (3.44) and (3.46) we have

$$q = \int_{-\infty}^{\infty} \mathrm{d}x^0 \int \mathrm{d}\boldsymbol{r}\, \partial_0 \mathscr{C}^0 + \int_{-\infty}^{\infty} \mathrm{d}x^0 \int \mathrm{d}\boldsymbol{r}\, \nabla \cdot \boldsymbol{\mathscr{C}} \quad . \tag{3.50a}$$

The last term is converted to a three-surface integral and may be dropped; the x^0 integration in the remaining integral is trivial to perform, with the contribution from negative infinity vanishing. Thus

$$q = \int \mathrm{d}\boldsymbol{r}\, \mathscr{C}^0 \big|_{x^0 = \infty} \quad . \tag{3.50b}$$

However, as already noted in (3.45), the integral (3.50b) coincides with $W(A)$, hence at $x^0 = \infty$ it is the winding number of the gauge transformation to which A tends.

The above discussion of the Pontryagin density is restricted to four-dimensional space-time, which is where the θ-angle of Yang-Mills theory appears. However, as a topological/mathematical object it can be defined in any even-dimensional space, and always it is a divergence of a vectorial Chern-Simons secondary characteristic. For example, in two dimensions, for an Abelian gauge theory,

$$\mathscr{P}_2 = \frac{-\mathrm{i}}{2\pi}\, {}^*F = \frac{-\mathrm{i}}{4\pi}\, \epsilon^{\mu\nu} F_{\mu\nu} \quad , \qquad \mathscr{C}_2^\mu = \frac{-\mathrm{i}}{2\pi}\, \epsilon^{\mu\nu} A_\nu \quad , \tag{3.51}$$

while in six dimensions

$$\mathscr{P}_6 = \frac{i}{384\pi^3} \epsilon^{\alpha\beta\gamma\delta\epsilon\phi} \operatorname{tr} F_{\alpha\beta} F_{\gamma\delta} F_{\epsilon\phi} \quad ,$$

$$\mathscr{C}_6^\mu = \frac{i}{192\pi^3} \epsilon^{\mu\alpha\beta\gamma\delta\epsilon} \operatorname{tr} (F_{\alpha\beta} F_{\gamma\delta} A_\epsilon - F_{\alpha\beta} A_\gamma A_\delta A_\epsilon + \tfrac{2}{5} A_\alpha A_\beta A_\gamma A_\delta A_\epsilon) \quad . \qquad (3.52)$$

We shall discuss some physical consequences of the two-dimensional Pontryain density below, while the six-dimensional quantities will appear in our discussion of chiral anomalies.

3.5 *Semi-Classical Picture for the Vacuum Angle*

The emergence of a phase in the response of a physical state to a finite symmetry transformation is reminiscent of the quantum mechanical Bloch momentum associated with the wave function of a particle in a periodic potential $V(q)$, of the type pictured in Fig. 1. Even though the Hamiltonian $H = p^2/2m + V(q)$ is invariant under the shift $q \to q + a$, the wave function acquires a phase: $\psi(q+a) = e^{i\theta}\,\psi(q)$, where θ is proportional to the Bloch momentum. Moreover, even though the classical zero-energy configuration is infinitely degenerate ($q^{(n)} = na$; $n = 0, \pm 1, \pm 2, \ldots$), quantum mechanical tunnelling between classical minima lifts the degeneracy and produces a band spectrum $E(\theta)$. This is the physical situation in a crystal.

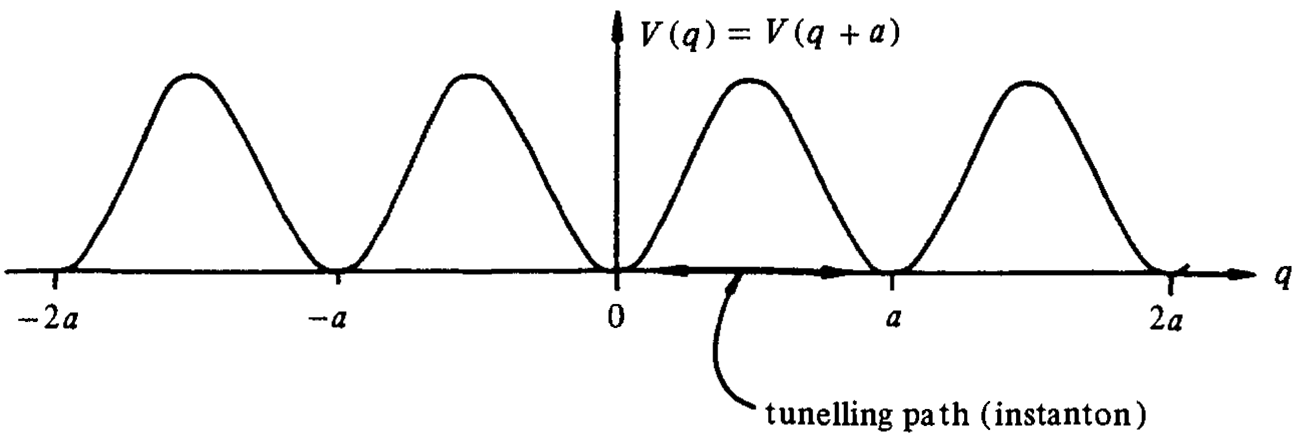

Fig. 1. Periodic potential whose quantum mechanics is analogous to the quantized Yang-Mills theory, with its gauge freedom with respect to infinitesimal gauge transformations removed. Zero-energy configurations $q^{(n)} = an$, $n = 0, \pm 1, \pm 2, \ldots$, are analogous to $A^{(n)} = -U_n^{-1} \nabla U_n$ and shifting by a is analogous to gauge transforming by U_1. Tunneling is discovered by finding an imaginary-time classical solution (instanton) that follows the tuneling path.

For our gauge field theory, we recognize that gauge transforming by a gauge function U_1, belonging to the first homotopy class, is analogous to shifting by a, and the analog of the infinite number of classical zero-energy configurations are the pure gauge potentials,

$$A^{(n)} = -U_n^{-1} \nabla U_n \quad , \tag{3.53}$$

for which $\boldsymbol{E}$ and $\boldsymbol{B}$ are zero, and hence the energy vanishes. (It is understood that the gauge freedom associated with homotopically trivial gauge transformations is completely fixed by the imposition of Gauss' law.)

How can we recognize if there is tunnelling in the gauge theory, which would close the analogy with the crystal and produce a band spectrum? It is well known to chemists and condensed-matter physicists that semi-classical evidence for tunnelling is obtained by solving the classical equations of motion, not in real time, but in imaginary time $\tau = \mathrm{i}t$, and by finding a solution which interpolates as τ passes from $-\infty$ to $+\infty$ between adjacent classical minima. For the ground state, these should have vanishing (imaginary time) energy, $-\frac{1}{2}m\dot{q}^2 + V(q) = 0$. Moreover, a semi-classical formula for the tunnelling probability amplitude Γ is gotten by dominating the functional integral, continued to imaginary time, with the (imaginary time) solution, which these days is called an "instanton"; i.e., $\Gamma \propto \mathrm{e}^{-I/\hbar}$ where I is the classical (imaginary time) action $I = \int_{-\infty}^{\infty} \mathrm{d}\tau [\frac{1}{2}m\dot{q}^2 + V(q)]$ evaluate for the instanton. The zero-energy instanton satisfies $\dot{q} = \pm(2V(q)/m)^{1/2}$, hence

$$I = \int_{-\infty}^{\infty} \mathrm{d}\tau 2V(q) = \int \mathrm{d}q (\mathrm{d}\tau/\mathrm{d}q)\, 2V(q) = \int \mathrm{d}q \sqrt{2mV(q)} \quad ,$$

and $\mathrm{e}^{-I/\hbar}$ is recognized as the WKB approximation to the (zero-energy) tunnelling amplitude. (Higher corrections involve computing the quadratic fluctuations around the instanton and performing a Gaussian functional integral.) Evidently the action must be finite to get a non-vanishing result [53]. [Let me emphasize that in the exact functional integral, one integrates over configurations, not merely solutions, and that finite-action solutions and configurations give a vanishingly small contribution. The infinite-action configurations are so much more numerous, that even though each one gives zero for $\mathrm{e}^{-I/\hbar}$, their large number (entropy) ensures a finite answer. It is only in the semi-classical approximation that finite action solutions are relevant.]

These ideas carry over to the Yang-Mills theory [46]. The imaginary time theory becomes defined in Euclidean space; the imaginary time energy is $\frac{1}{2}\int \mathrm{d}\boldsymbol{r}(-\boldsymbol{E}_a^2 + \boldsymbol{B}_a^2)$; and the Euclidean action becomes

$$I = \int \mathrm{d}x \left(-\frac{1}{2g^2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} + \frac{\mathrm{i}\hbar\theta}{16\pi^2} \operatorname{tr} {}^*F^{\mu\nu} F_{\mu\nu} \right) . \tag{3.54}$$

(The topological term retains its factor of i in the continuation to Euclidean space, since it is a world scalar.) Zero energy is assured if $\boldsymbol{E} = \pm \boldsymbol{B}$ or in covariant notation

$${}^*F^{\mu\nu} = \pm F^{\mu\nu} . \tag{3.55}$$

As mentioned earlier, solutions to (3.55) automatically satisfy the Euclidean-space Yang-Mills equation, by virtue of the Bianchi identity.

Since the action must be finite, we may expect the potentials to be sufficiently regular so that Pontryagin index is an integer, $q = N \neq 0$, and the solution is called $|N|$ instanton solution. The action then becomes

$$I_N = \frac{8\pi^2 |N|}{g^2} - \mathrm{i}N\hbar\theta . \tag{3.56}$$

For the physical application we need the smallest action, hence $|N| = 1$ and the two cases $N = \pm 1$ add coherently. Thus the semi-classical tunnelling amplitude is

$$\Gamma \propto \exp(-8\pi^2/\hbar g^2)\cos\theta . \tag{3.57}$$

(Instantons with $|N| \geqslant 2$ give exponentially subdominant contributions.) Also the quadratic fluctuations have been evaluated [54].

Finally, one may represent schematically the ground state wave functional by analogy with the tight-binding approximation of crystal physics. The quantum state should be a superposition of wave functionals, $\Psi_n(A)$, each peaked in A-function space near the classical zero-energy configuration $-U_n^{-1}\nabla U_n$. They must be gauge invariant against homotopically trivial gauge transformations, but the non-trivial ones shift n.

$$\mathscr{G}_n \Psi_{n'} = \Psi_{n'+n'} . \tag{3.58}$$

The ground state band functional then is

$$\Psi(A) = \sum_n e^{in\theta}\, \psi_n(A) \quad , \tag{3.59}$$

and the band energy (density) has the form $\alpha + \beta\cos\theta$ [46, 47].

Let me emphasize that the existence of the vacuum angle does not rely on instantons; it is an exact statement. Instantons provide an approximate method for calculating the consequences of the angle. This is just as with the periodic potential in quantum mechanics: Bloch-Floquet theory makes exact statements, and the tight-binding approximation provides approximate analysis.

I shall describe the physical importance of these results when I consider a more realistic model: Yang-Mills theory coupled to fermions.

Explicit Euclidean-space instantons have been found. For the SU(2) theory the self-dual one-instanton potential is [55]

$$A^\mu = \frac{-2i}{x^2 + \lambda^2}\, \alpha^{\mu\nu} x_\nu \quad , \tag{3.60}$$

where the self-dual 2×2 matrices $\alpha^{\mu\nu}$ are defined by

$$\alpha^{\mu\nu} = {}^*\alpha^{\mu\nu} = \frac{1}{4i}(\bar\alpha^\mu \alpha^\nu - \bar\alpha^\nu \alpha^\mu) \quad , \qquad \alpha^\mu = (-i\boldsymbol{\sigma}, I) \quad ,$$

$$\bar\alpha^\mu = (\alpha^\mu)^\dagger = (i\boldsymbol{\sigma}, I) \quad , \tag{3.61}$$

and the ensuing field strength is

$$F^{\mu\nu} = \frac{4i\lambda^2 \alpha^{\mu\nu}}{(\lambda^2 + x^2)^2} \quad . \tag{3.62}$$

The configurations depend on five parameters: the instanton "size" λ, and four parameters specifying the location, here set at zero. The solution is invariant against SO(5) rotations [in the sense (2.69)] which form the maximal compact subgroup of the Euclidean space SO(5, 1) conformal group — the symmetry group for classical Euclidean-space Yang-Mills theory [56]. The anti-self-dual solution can be gotten from the above by replacing $\alpha^{\mu\nu}$ by the anti-self-dual matrices $\bar\alpha^{\mu\nu}$.

$$\bar{\alpha}^{\mu\nu} = -{}^*\bar{\alpha}^{\mu\nu} = \frac{1}{4\mathrm{i}}(\alpha^\mu\bar{\alpha}^\nu - \alpha^\nu\bar{\alpha}^\mu) \quad . \tag{3.63}$$

Although only the $|N| = 1$ solutions are physically important, both physicists and mathematicians have been fascinated by the multi-instanton configurations. It has been shown that for SU(2) the $|N|$-instanton solution depends on $8|N| - 3$ parameters, which can be interpreted as $5|N|$ size and position parameters, and $3|N| - 3$ parameters specifying relative orientations in the 3-parameter SU(2) group space [57]. The most general instanton solution which can be simply written has been found by physicists. The self-dual solution is [58]

$$A^\mu = \mathrm{i}\bar{\alpha}^{\mu\nu}\partial_\nu \ln\rho \quad , \qquad \rho = \sum_{i=1}^{N+1} \frac{\lambda_i^2}{(x-y_i)^2} \quad . \tag{3.64}$$

It is regular — the singularities are gauge artifacts — closed under the conformal group, and depends on $5N + 4$ gauge invariant parameters for $N \geqslant 3$, on thirteen parameters for $N = 2$, and on five for $N = 1$. [Some of the parameters in (3.64) are gauge artifacts.] While the field strength is complicated, the Pontryagin density is given by the elegant formula

$$-\frac{1}{16\pi^2}\,\mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu} = \Box\Box\ln\rho \quad , \tag{3.65}$$

and the Pontryagin index is N [59]. Evidently, for $N \geqslant 3$ (3.64) is not the most general solution. Mathematicians have given a constructive procedure, which involves (many!) finite steps, for constructing any of the most general $8|N| - 3$ parameter solutions [60]; however, no closed expression analogous to (3.64) is available.

Exercise 3.10. Show that the (Euclidean space version of) of the energy-momentum tensor (2.33) vanishes on self-dual and anti-self-dual field strengths.

4. Fermion Interactions and Anomalies

4.1 Quantized Fermions

To be realistic, a Yang-Mills theory must be supplemented by other fields, coupling to it in a gauge invariant manner. While scalar fields are used to