20

Anomalies in Quantum Field Theory

20.1 The chiral anomaly

In classical field theory, discussed in chapter 3, we learn that symmetries dictate the existence of conservation laws. At the quantum level, conservation laws of continuous symmetries are embodied by Ward identities which dictate the behavior of correlation functions. However we will now see that in quantum field theory there are many instances of symmetries of the classical theory that do not survive at the quantum level due to a quantum *anomaly*. Most examples of anomalies involve Dirac fermions that are massless at the level of the Lagrangian.

Global and gauge symmetries of theories of Dirac fermions involve understanding their currents which are products of Dirac operators at short distances. Such operators require a consistent definition (and a normal-ordering prescription). Hence, some sort of regularization is needed for a proper definition. The problem is that all of the symmetries can simultaneously survive regularization. While in scalar field theories, at least in flat spacetimes, this is not a problem, it turns out to be a significant problem in theories with massless Dirac fermions, for which some formally conserved currents become anomalous.

In the path integral language, a symmetry is anomalous if the action is invariant under the symmetry but the measure of the path integral is not. In this sense, quantum anomalies often arise in the process of regularization of a quantum field theory. We will see, however, that they are closely related to topological considerations as well.

The subject of anomalies in quantum field theory is discussed in many excellent textbooks, e.g. in the book by Peskin and Schroeder (Michael E. Peskin and Daniel V. Schroeder, 1995). It is also a subject that can be fairly technical and mathematically quite sophisticated. For these reasons, we will keep the presentation to be physically transparent and as simple as possible, even at the price of some degree of rigor.

The prototype of a quantum anomaly is the axial (or chiral) anomaly. Classically, and at the free field level, a theory with a single massless Dirac fermion has two natural continuous global $U(1)$ symmetries: gauge invariance and chiral symmetry. Then, as we saw in chapter 7, Noether's theorem implies the existence of two currents: the gauge current $j_\mu = \bar{\psi}\gamma_\mu\psi$, and a axial (chiral) current $j_\mu^5 = \bar{\psi}\gamma_\mu\gamma^5\psi$. Superficially, in the massless theory at the free field level, the gauge and the chiral currents are separately conserved,

$$\partial^\mu j_\mu = 0, \qquad \partial^\mu j_\mu^5 = 0 \tag{20.1}$$

A Dirac mass term is gauge-invariant, but breaks the chiral symmetry explicitly. The chiral current is not conserved in its presence. Indeed, using the Dirac equation one finds that conservation Eq.(20.1) the axial current j_μ^5 is modified to

$$\partial^\mu j_\mu^5 = 2m\, i\bar{\psi}\gamma^5\psi \tag{20.2}$$

where m is the mass.

The anomaly was first discovered in the computation of so-called triangle Feynman diagrams, involved in the study of decay processes of a neutral pion to two photons, $\pi^0 \to 2\gamma$, by Adler (Adler, 1969), and Bell and Jackiw (Bell and Jackiw, 1969). They, and subsequent authors, showed that, in any gauge-invariant regularization of the theory, there is an extra term in the right hand side of Eq.(20.2), even in the massless limit, $m \to 0$. This term is the chiral (or axial) anomaly.

The form of the anomaly term depends on the dimensionality, and turned out to have a topological meaning. Although the anomaly arises from short distance singularities in the quantum field theory, it is a finite and universal term. Here universality means although that its existence depends on how the theory is regularized, its value depends only on the symmetries that are preserved by the regularization and not by its detailed form.

The important physical implication of the anomaly is that in its presence, the anomalous current is not conserved and, consequently, the associated symmetry cannot be gauged. This poses strong constraints on what theories of particle physics are physically sensible, particularly in the weak interaction sector.

20.2 The chiral anomaly in $1 + 1$ dimensions

We will discuss first the chiral anomaly on 1+1-dimensional theories of Dirac fermions. For simplicity, we will consider a theory of a single massless Dirac fermion ψ coupled to a background gauge field A_μ. The Lagrangian is

$$\mathcal{L} = \bar{\psi} i \not{\partial} \psi + e \bar{\psi} \gamma^\mu \psi \, A_\mu \tag{20.3}$$

The standard derivation of the anomaly uses subtle (and important) arguments of regulators and what symmetries they preserve (or break). Although we will do that shortly, it is worthwhile to use first a transparent and physically compelling argument, originally due to Nielsen and Ninomiya (Nielsen and Ninomiya, 1983).

20.2.1 The anomaly as particle-antiparticle pair creation

Recall that in 1+1 dimensions the Dirac fermion is a two-component spinor. We will work in the chiral basis which simplifies the analysis. In the chiral basis the Dirac spinor is $\psi = (\psi_R, \psi_L^t)$ (where the upper index t means transpose). In 1+1 dimensional Minkowski space time, the Dirac matrices are

$$\gamma_0 = \sigma_1, \qquad \gamma_1 = i\sigma_2, \qquad \gamma_5 = \sigma_3 \tag{20.4}$$

where σ_1, σ_2 and σ_3 are the standard 2×2 Pauli matrices. In this basis, the gauge current j_μ and the chiral current j_μ^5 are

$$\begin{aligned} j_0 =& \psi_R^\dagger \psi_R + \psi_L^\dagger \psi_L, \qquad & j_1 =& \ \psi_R^\dagger \psi_R - \psi_L^\dagger \psi_L \\ j_0^5 =& \psi_R^\dagger \psi_R - \psi_L^\dagger \psi_L, \qquad & j_1^5 =& -(\psi_R^\dagger \psi_R + \psi_L^\dagger \psi_L) \end{aligned} \tag{20.5}$$

Hence, j_0 measures the total density of right and left moving fermions, and j_1 measures the difference of the densities of right moving and left moving fermions. Notice that the components chiral current j_μ^5 essentially switch the roles of charge and current. In short, we can write

$$j_\mu^5 = \epsilon^{\mu\nu} j_\nu \tag{20.6}$$

In particular, the total gauge charge $Q = \int dx; j_0(x)$ of a state is the total number of right and left moving particles, $N_R + N_L$, the total chiral charge of a state $Q^5 = \int dx \; j_0^5(x)$, is

$$Q^5 = N_R - N_L \equiv N_R + \bar{N}_L \tag{20.7}$$

where $\bar{N}_L$ is the number of left-moving antiparticles.

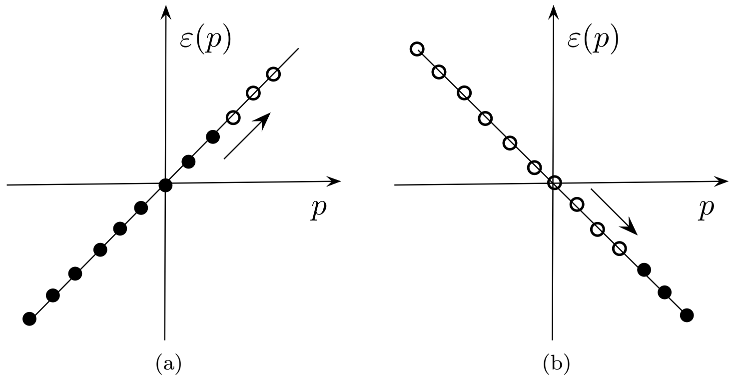

Figure 20.1 Chiral anomaly in 1+1 dimensions: pair creation by an uniform electric field E and the chiral anomaly. a) the Fermi point of the right movers increases with time reflecting particle creation. b) the Fermi point of left-movers decreases reflecting the creation of anti-particles (holes).

In the chiral basis, the Dirac equation in the temporal gauge, $A_0 = 0$,

$$i\partial_0\psi_R(x) = (-i\partial_1 - A^1)\psi_R(x), \qquad i\partial_0\psi_L(x) = (i\partial_1 - A^1)\psi_L(x) \tag{20.8}$$

In the temporal gauge an uniform electric field is $E = \partial_0 A^1$, and A^1 increases monotonically in time. The Dirac equation, Eq.(20.8) states that as A^1 increases, the Fermi momentum $p_F = \varepsilon_F$ (with ε_F being the Fermi energy) increases by the amount

$$\frac{dp_F}{dx_0} = eE \tag{20.9}$$

The total number of right-moving particles N_R. The density of states of a system of length L is $L/(2\pi)$ Then, the rate of change of the number of right moving particles is

$$\frac{dN_R}{dx_0} = \frac{1}{L}\frac{L}{2\pi}\frac{d\varepsilon_F}{dx_0} = \frac{e}{2\pi}E \tag{20.10}$$

We will *assume* that the UV regulator of the theory is such that the total fermion number is conserved, $Q = 0$

$$Q = \int_{-\infty}^{\infty} dx\ j_0(x) = N_R + N_L = 0 \tag{20.11}$$

Therefore if $N_R > 0$ increases, then $N_L < 0$ decreases. Or, what is the same, the number of left-moving *anti-particles* must increase by the same amount that N_R *particles* increase. This leads to the conclusion that the total chiral charge $Q^5 = N_R - N_L = N_R + \bar{N}_L$ must increase at the rate

$$\frac{dQ^5}{dx_0} = \frac{dN_R}{dx_0} + \frac{d\bar{N}_L}{dx_0} = \frac{e}{\pi}E \tag{20.12}$$

But, in this process, the total electric (gauge) charge Q is conserved. Notice that the details of the UV regularization do not affect this result, provided the regularization is gauge-invariant. Hence, the chiral (axial) charge is not conserved if the gauge charge is conserved (and vise versa!).

In covariant notation, the non-conservation of the chiral current j^5_μ is

$$\partial^\mu j^5_\mu = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu} \tag{20.13}$$

This is the chiral (axial) anomaly equation in 1+1 dimensions. Notice that it tells us that the total chiral charge is e times a topological invariant, the total instanton number (the flux) of the gauge field!

20.3 The chiral anomaly and abelian bosonization

We will now reexamine this problem taking care of the short-distance singularities. Let $|0\rangle_D$ be the vacuum state of the theory of free massless Dirac fermions, i.e. the filled Dirac sea. We will normal-order the operators with respect to the Dirac vacuum state. We will need to be careful when treating composite operators such as the currents since they are products of Dirac fields at short distances.

To this effect we will examine carefully the algebra obeyed by the density and current operators. This current algebra leads to the concept of bosonization, introduced by Mattis and Lieb (Mattis and Lieb, 1965), based on results by Schwinger (Schwinger, 1959), and rediscovered (and expanded) by Coleman (Coleman, 1975), Mandelstam (Mandelstam, 1975), Luther and Emery (Luther and Emery, 1974) in the 1970s.

The vacuum state should be charge neutral and have zero total momentum, and zero current. This means that both the charge density j_0 and the charge current j_1 should annihilate the vacuum. As usual, the zero current condition is automatic, while the charge neutrality of the vacuum is insured by a proper subtraction. Since the charge and the current densities are fermion bilinears, they behave as bosonic operators. Naively, one would expect that they should commute with each other. We will now see that this expectation is wrong.

Let us introduce the operators for the right and left moving densities, naively defined as $j_R(x_1) = \psi_R^\dagger(x_1)\psi_R(x_1)$, and $j_L(x_1) = \psi_L^\dagger(x_1)\psi_L(x_1)$, where x_1 is the space coordinate. The propagators of the right and left moving fields are

$$\langle\psi_R^\dagger(x_0,x_1)\psi_R(0,0)\rangle = -\frac{i}{2\pi(x_0-x_1+i\epsilon)}$$
$$\langle\psi_L^\dagger(x_0,x_1)\psi_L(0,0)\rangle = +\frac{i}{2\pi(x_0+x_1+i\epsilon)} \tag{20.14}$$

and diverge at short distances. We will define the normal-ordered densities by a point-splitting procedure (at equal times)

$$j_R(x_1) =: j_R(x_1) : + \lim_{\epsilon\to 0}\langle\psi_R^\dagger(x_1+\epsilon)\psi_R(x_1-\epsilon)\rangle \tag{20.15}$$

and similarly with j_L. Here the expectation value is computed in the free massless Dirac vacuum. From the expressions of the propagators we see that

$$\langle\psi_R^\dagger(x_1+\epsilon)\psi_R(x_1-\epsilon)\rangle = \frac{i}{4\pi\epsilon}$$
$$\langle\psi_L^\dagger(x_1+\epsilon)\psi_L(x_1-\epsilon)\rangle = -\frac{i}{4\pi\epsilon} \tag{20.16}$$

Then, the equal-time commutator of two right-moving currents is found to be (upon taking the limit $\epsilon \to 0$)

$$[j_R(x_1), j_R(x_1')] = -\frac{i}{2\pi}\partial_1\delta(x_1-x_1') \tag{20.17}$$

Similarly, we find

$$[j_L(x_1), j_L(x_1')] = \frac{i}{2\pi}\partial_1\delta(x_1-x_1') \tag{20.18}$$

Clearly the right moving densities (and the left moving densities) do not commute with each other!

These results imply that the equal-time commutators of properly regularized charge density and current operators j_0 and j_1 are

$$[j_0(x_1), j_1(x_1')] = -\frac{i}{\pi}\partial_1\delta(x_1-x_1'), \quad [j_0(x_1), j_0(x_1')] = [j_1(x_1), j_1(x_1')] = 0 \tag{20.19}$$

In other words, the currents of a theory of free massless Dirac fermions (which has a global $U(1)$ symmetry) obey the algebra of Eqs. (20.17), (20.18), and (20.19). This is known as the $U(1)$ Kac-Moody current algebra.

The singular results of the commutators of currents are known as Schwinger terms. We should recall here that we obtained a similar result in section 10.11 where we discussed the local conservation laws of a system of non-relativistic fermions at finite density. This is not an accident since in one space dimension, at low energy, a system of non-relativistic fermions at finite density is equivalent to a theory of a massless Dirac field (where the speed of light is identified with the Fermi velocity).

The $U(1)$ current algebra of Eq.(20.19) is reminiscent of the equal-time canonical commutation relations of a scalar field $\phi(x)$ and its canonical momentum $\Pi(x)$,

$$[\phi(x_1), \Pi(x_1')] = i\delta(x_1 - x_1') \tag{20.20}$$

Indeed, we can identify the normal-ordered density $j_0(x_1)$ with

$$j_0(x) = \frac{1}{\sqrt{\pi}}\partial_1\phi(x) \tag{20.21}$$

and the normal-ordered current $j_1(x)$ with

$$j_1(x) = -\frac{1}{\sqrt{\pi}}\Pi(x) = -\frac{1}{\sqrt{\pi}}\partial_0\phi(x) \tag{20.22}$$

With these operator identifications, the canonical commutation relations of the scalar field $\phi(x)$, Eq(20.20), imply

$$\frac{1}{\pi}[\partial_1\phi(x_1), \Pi(x_1')] = \frac{i}{\pi}\partial_1\delta(x_1 - x_1') \tag{20.23}$$

which reproduces the Schwinger term of the $U(1)$ Kac-Moody current algebra of Eq.(20.19).

The operator identifications of Eqs.(20.21) and (20.22) can be written in the Lorentz covariant form

$$j_\mu = \frac{1}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu\phi \tag{20.24}$$

which satisfies the local conservation condition

$$\partial^\mu j_\mu = 0 \tag{20.25}$$

Thus, the regularization we adopted is consistent with the conservation of the $U(1)$ current. The identification (or mapping) of the $U(1)$ fermionic current of Eq.(20.24) shows that there can be a mapping between operators of the theory of massless Dirac fermions to operators of a theory of scalar fields, which are bosons. Such mappings are called *bosonization*.

However, is it compatible with the conservation of the chiral current j_μ^5? In Eq.(20.6) we showed that the chiral and the $U(1)$ currents are related to each other by $j_\mu^5 = \epsilon_{\mu\nu}j^\nu$. Therefore, the divergence of the chiral current is identified with

$$\partial^\mu j_\mu^5(x) = \epsilon^{\mu\nu}j_\nu(x) = \frac{1}{\sqrt{\pi}}\epsilon^{\mu\nu}\epsilon_{\nu\lambda}\partial^\lambda\phi(x) = \frac{1}{\sqrt{\pi}}\partial^2\phi \tag{20.26}$$

Therefore,

$$\partial^\mu j^5_\mu(x) = 0 \Longleftrightarrow \partial^2\phi = 0 \tag{20.27}$$

and the chiral (or axial) current is conserved if and only if the scalar free is free and massless, which has the Lagrangian $\mathcal{L}_B$,

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\phi)^2 \tag{20.28}$$

We also see that the Hamiltonian density of the Dirac theory

$$\mathcal{H}_D = -(\psi_R^\dagger i\partial_1\psi_R - \psi_L^\dagger i\partial_1\psi_L) \tag{20.29}$$

must be identified with hamiltonian density of the mass less scalar field

$$\mathcal{H}_B = \frac{1}{2}\left(\Pi^2 + (\partial_1\phi)^2\right) \tag{20.30}$$

which, after normal-ordering, can be expressed in terms of the $U(1)$ density and current in the (Sugawara) form

$$\mathcal{H} = \frac{\pi}{2}\left(j_0^2 + j_1^2\right) \tag{20.31}$$

To see how this is related to the chiral anomaly, we will couple the Dirac theory to an external $U(1)$ gauge field A_μ. The coupling term in the Dirac Lagrangian is

$$\mathcal{L}_{\text{int}} = eA^\mu j_\mu \tag{20.32}$$

where j_μ is the $U(1)$ Dirac current. In the theory of the massless scalar field we identify this term with

$$\mathcal{L}_{\text{int}} = \frac{e}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu\phi\; A^\mu \tag{20.33}$$

which states that the coupling of the fermions to the gauge field is equivalent to the coupling the scalar field to a source

$$J(x) = \frac{e}{2\sqrt{\pi}}\epsilon_{\mu\nu}F^{\mu\nu} \equiv \frac{e}{2\sqrt{\pi}}F^* \tag{20.34}$$

where F^* is the dual of the field strength $F^{\mu\nu}$.

The equation of motion of the scalar field now is

$$\partial^2\phi = \frac{e}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\mu A^\nu \tag{20.35}$$

Therefore, we find that the divergence of the axial current j^5_μ does not vanish and is given by

$$\partial^\mu j^5_\mu = \frac{1}{\sqrt{\pi}}\partial^2\phi = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu} \tag{20.36}$$

which reproduces the chiral (axial) anomaly of Eq.(20.13)!

Let us consider a system with periodic boundary conditions in space, which we will be take to be a circle of circumference L. The total charge is

$$Q = e\int_0^L dx_1 \; j_0(x_1) \tag{20.37}$$

which is measured relative to the vacuum state, which neutral, $Q_{\text{vacuum}} = 0$. It then follows that the charge must be quantized and is an *integer* multiple of the electric charge, $Q = Ne$. On the other hand, using the identification of the current, Eq.(20.24), we obtain the condition

$$Q = \frac{e}{\sqrt{\pi}}\int_0^L dx_1 \; \partial_1\phi(x_1) = \frac{e}{\sqrt{\pi}}\Delta\phi \tag{20.38}$$

where we defined

$$\Delta\phi = \phi(x_1 + L) - \phi(x_1) \tag{20.39}$$

Therefore, if the fermions are in the sector of the Hilbert space with $N \in \mathbb{Z}$ fermions, the scalar field must obey the generalized periodic boundary condition (here $x = x_1$)

$$\phi(x + L) = \phi(x) + 2\pi N R_\phi \tag{20.40}$$

where

$$R_\phi = \frac{1}{2\sqrt{\pi}} \tag{20.41}$$

Hence, the scalar field is *compactified* and R_ϕ is the compactification radius.

Therefore the target space of this scalar field are not the real numbers, $\mathbb{R}$, but the circle S^1, whose radius is R_ϕ. This fact, which is a consequence of the charge quantization of the Dirac fermions, restricts the allowed observables of the bosonized side of the theory to be *invariant* under shifts $\phi \to \phi + 2\pi n R_\phi$, with $n \in \mathbb{Z}$. For example, the *vertex operators* V_α,

$$V_\alpha(x) = \exp(i\alpha\phi(x)) \tag{20.42}$$

are allowed only for the values $\alpha = 2\pi n R_\phi = n\sqrt{\pi}$.

Using methods similar to those we have discussed here, Mandelstam showed that the bosonic counterpart of the Dirac fermion operator, ψ_R and ψ_L, can be identified as

$$\begin{aligned}\psi_R(x) =& \frac{1}{\sqrt{2\pi a}} : \exp(-\sqrt{\pi}\int_{-\infty}^{x_1} dx_1'\ \Pi(x_0,x_1') + i\sqrt{\pi}\phi(x)) : \\ \equiv& \frac{1}{\sqrt{2\pi a}} : \exp(2i\sqrt{\pi}\phi_R(x)) : \\ \psi_L(x) =& \frac{1}{\sqrt{2\pi a}} : \exp(-\sqrt{\pi}\int_{-\infty}^{x_1} dx_1'\ \Pi(x_0,x_1') - i\sqrt{\pi}\phi(x)) : \\ \equiv& \frac{1}{\sqrt{2\pi a}} : \exp(-2i\sqrt{\pi}\phi_L(x)) :\end{aligned} \tag{20.43}$$

where a is a short-distance cutoff, and $: \mathcal{O} :$ denotes normal ordering. Notice that in terms of the scalar field, the Mandelstam operators are, essentially, a product of an operator that creates a kink (a soliton), which shifts the field by $\sqrt{\pi}$, and a vertex operator that measures the charge. After some algebra one can show that these operators create states that carry the unit of charge.

In Eq.(20.43) we introduced the fields ϕ_R and ϕ_L, the right and left moving components of the scalar field, respectively,

$$\phi = \phi_R + \phi_L, \qquad \vartheta = -\phi_R + \phi_L \tag{20.44}$$

where

$$\vartheta(x) = \int_{-\infty}^{x_1} dx_1'\ \Pi(x_0,x_1') \tag{20.45}$$

is called the dual field. The fields ϕ and ϑ satisfy the Cauchy-Riemann equations

$$\partial_\mu\phi = \epsilon_{\mu\nu}\partial^\nu\vartheta \tag{20.46}$$

We also record here the identification of the Dirac mass operator $\bar\psi\psi$ and of the chiral mass operator $i\bar\psi\gamma^5\psi$,

$$\bar\psi\psi = \frac{1}{2\pi a} : \cos(2\sqrt{\pi}\ \phi) :, \qquad i\bar\psi\gamma^5\psi = \sin(2\sqrt{\pi}\ \phi) \tag{20.47}$$

These identifications can be derived using the Operator Product Expansion.

In particular, these operator identifications imply that the Lagrangian of the free massive Dirac theory

$$\mathcal{L} = \bar\psi i\partial\!\!\!/\psi - m\bar\psi\psi \tag{20.48}$$

maps to the sine-Gordon Lagrangian

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\phi)^2 - g : \cos(2\sqrt{\pi}\ \phi) : \tag{20.49}$$

with $g = m/(2\pi a)$.

20.4 Solitons and fractional charge

We will now see that in theories of scalar fields coupled to Dirac fermions have states with fractional charge if the scalar fields have solitons. For simplicity we will focus on 1+1-dimensional theories.

20.4.1 The sine-Gordon soliton

We already saw a hint of this in the derivation of the connection between the sine-Gordon theory and a massive Dirac fermion. The Lagrangian of the sine-Gordon theory is

$$\mathcal{L}_{\rm SG} = \frac{1}{2}(\partial_\mu \phi)^2 + g\cos(2\sqrt{\pi}\,\phi) \tag{20.50}$$

This theory is invariant under the discrete shifts of the field, $\phi(x) \to \phi(x) + 2\pi n/\beta$, where $n \in \mathbb{Z}$. The Hamiltonian density

$$\mathcal{H}_{\rm SG} = \frac{1}{2}\Pi^2 + \frac{1}{2}(\partial_x \phi)^2 - g\cos(2\sqrt{\pi}\,\phi) \tag{20.51}$$

is, of course, invariant under the same global symmetry.

The classical vacua of this theory are $\phi_c = 2\pi N/\beta$. In addition to these uniform classical states, this theory has solitons that connect the uniform states. Let $\phi(x)$ be a static, i.e. time-independent, field configuration (here x is the space coordinate). The total energy of such a configuration is

$$E[\phi] = \int_{-\infty}^{\infty} dx \left[\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 - g\cos(2\sqrt{\pi}\,\phi(x))\right] \tag{20.52}$$

We are interested in non-uniform configurations $\phi(x)$ that extremize $E[\phi]$ that interpolate between two classical vacuum states. These classical solitons are solutions the Euler-Lagrange equation

$$\frac{d^2\phi}{dx^2} = 2g\sqrt{\pi}\,\sin(2\sqrt{\pi}\,\phi) \tag{20.53}$$

with the boundary conditions

$$\lim_{x\to-\infty}\phi(x) = 0, \qquad \lim_{x\to+\infty}\phi(x) = \sqrt{\pi} \tag{20.54}$$

The classical equation for the static soliton is the same as the equation for a physical pendulum of coordinate ϕ as a function of time. This is a similar problem to what our discussion of quantum tunneling and instantons in Chapter 19. The classical energy of the pendulum is the negative of the sine-Gordon potential, and the classical vacuum states correspond to the top of the pendulum potential. The soliton is the solution that rolls down

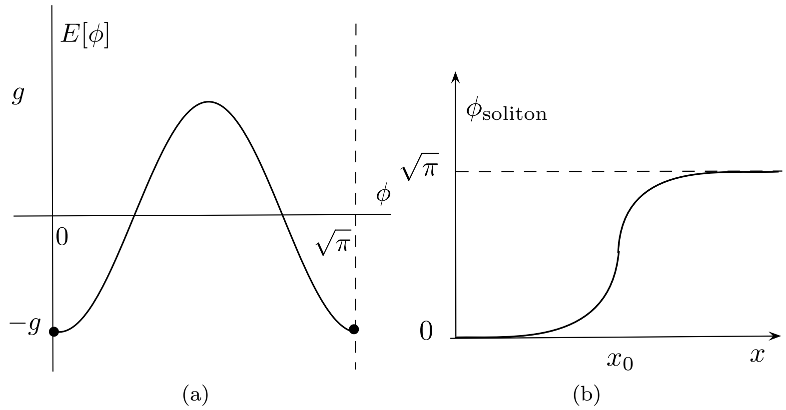

Figure 20.2 a) The sine-Gordon potential energy if a periodic function of ϕ of period $\sqrt{\pi}$; the black dots are the classical vacua at $\phi = 0, \sqrt{\pi}$. b) sketch of the sine-Gordon soliton.

the hill from the peak at $\phi = 0$ to the valley at $\phi = \sqrt{\pi}/2$ and then climbs to the summit of the next hill at $\phi = \sqrt{\pi}$. The soliton solution is

$$\phi_{\text{soliton}}(x) = \frac{1}{\sqrt{\pi}} \arccos\left[\tanh(2\sqrt{\pi}\,\sqrt{g}\,(x - x_0))\right] \tag{20.55}$$

where x_0 is a zero mode of the soliton. The total energy of the soliton (measured from the energy of the uniform classical solution) is finite, $E_{\text{soliton}} = 8\sqrt{g}/\beta$. In this soliton solution, the total change of the angle is $\Delta\phi_{\text{soliton}} = \sqrt{\pi}$ which, using Eq.(20.38), we see corresponds to a fermion charge $Q = e\frac{\Delta\phi}{\sqrt{\pi}} = e$. Thus, wee see that the fermion of the free massive Dirac theory is the soliton of the sine-Gordon theory.

20.4.2 Fractionally charged solitons

We will now consider a theory, again in 1+1 dimensions, with two real scalar fields, φ_1 and φ_2, coupled to a massless Dirac fermion. We will assume that the scalar fields have a non-vanishing vacuum expectation value. The Lagrangian is

$$\mathcal{L} = \bar{\psi} i \not{\partial} \psi + g\bar{\psi}(\varphi_1 + i\gamma^5\varphi_2)\psi \tag{20.56}$$

We recognize that a constant value of $\langle\varphi_1\rangle$ is the same as a Dirac mass $m = g\langle\varphi_1\rangle$, and that a constant value of $\langle\varphi_2\rangle$ is the same as a chiral mass $m_5 = g\langle\varphi_2\rangle$. In Section 17.4 we discussed the chiral and the non-chiral Gross-Neveu models (which we solved in their large-N limit) where these mass terms arise from spontaneous breaking of the chiral symmetry,

$$\psi \to e^{i\eta\gamma_5}\psi \tag{20.57}$$

where η was an arbitrary angle for the chiral model, and equal to $\pi/2$ for the model with a discrete chiral symmetry

We can make the relation with chiral symmetry more apparent we write the scalar fields as

$$\varphi_1 = |\varphi|\cos\theta, \qquad \phi_2 = |\varphi|\sin\theta \tag{20.58}$$

where $|\varphi| = (\varphi_1^2 + \varphi_2^2)^{1/2}$, and the Lagrangian now is

$$\mathcal{L} = \bar{\psi} i \partial\!\!\!/ \psi + g|\varphi|\bar{\psi}e^{i\theta\gamma_5}\psi \tag{20.59}$$

Thus, a global chiral transformation with angle θ, Eq.(20.57), changes the phase field $\theta \to \theta + 2\eta$.

We will now consider the case in which the scalar field (φ_1, φ_2) has a soliton, such that the phase θ winds by $\Delta\theta$ as x goes from $-\infty$ to $+\infty$. We will assume that the soliton is static so that this winding is adiabatic. The question that we will consider is what is the charge of the soliton.

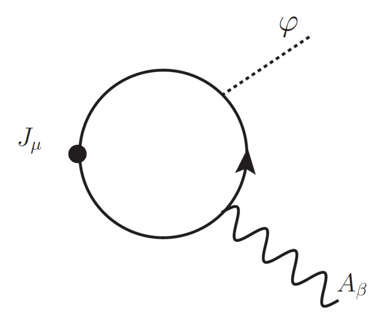

Figure 20.3 Induced current $\langle J_\mu \rangle$ by a small and adiabatic change in the phase of the complex scalar field φ.

Goldstone and Wilczek used a perturbative calculation (the Feynman diagram of Fig.20.3) in the regime where $|\varphi|$ is large compared to the gradients of φ_1 and φ_2, and showed there is an induced charge current $\langle j_\mu \rangle$ given by (Goldstone and Wilczek, 1981)

$$\langle j_\mu \rangle = \frac{1}{2\pi}\epsilon_{\mu\nu}\epsilon^{ab}\frac{\varphi_a \partial_\nu \varphi_b}{|\varphi|^2} = \frac{1}{2\pi}\epsilon_{\mu\nu}\partial^\nu \tan^{-1}(\varphi_2/\varphi_1) \tag{20.60}$$

This equation implies that the total charge Q is

$$Q = \int_{-\infty}^{\infty} dx\, j_0(x) = \frac{1}{2\pi}\Delta \tan^{-1}(\varphi_2/\varphi_1) \tag{20.61}$$

The same result can be found using the bosonization identities. Indeed, using Eqs.(20.47) we can readily see that the bosonized expression of the Lagrangian of Eq.(20.56) is

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\phi)^2 + \frac{g}{2\pi a}\cos(2\sqrt{\pi}\phi - \theta) \tag{20.62}$$

The minimum of the energy of this sine-Gordon theory occurs for $\phi = \theta/\sqrt{4\pi}$.

A soliton can be represented by a fermion with Dirac mass m, and hence $\varphi_1 = m/g$, and a φ_2 field that approaches the values $\pm v$ as $x \to \pm\infty$. In this case the phase changes by $\Delta\theta = 2\tan^{-1}(gv/m)$. Therefore, to minimize the energy, the sine-Gordon field ϕ winds by $\Delta\phi = \Delta\theta/\sqrt{4\pi}$. Then the total charge Q induced by the soliton is

$$Q_{\text{soliton}} = e\frac{\Delta\phi}{\sqrt{\pi}} = e\frac{\Delta\theta}{2\pi} = \frac{e}{\pi}\tan^{-1}(\frac{gv}{m}) \tag{20.63}$$

Of particular interest is the limiting case $m \to 0$, for which the soliton charge approaches the half-integer value

$$Q_{\text{soliton}} = \frac{e}{2}, \quad \text{and} \quad \Delta\theta = \pi \tag{20.64}$$

This fractionally charged soliton was known to exist in ϕ^4 theory (see section 19.2) coupled to Dirac fermions (Jackiw and Rebbi, 1976), and in the one-dimensional conductor polyacetylene (Su et al., 1979). For example, let us consider a theory with one massless Dirac fermion coupled to a ϕ^4 field in its $\mathbb{Z}_2$ broken symmetry state with vacuum expectation value $\langle\phi\rangle = \phi_0$. The Lagrangian is

$$\mathcal{L} = \bar{\psi}i\partial\!\!\!/\psi - g\phi\bar{\psi}\psi + \frac{1}{2}(\partial_\mu\phi)^2 - \lambda(\phi^2 - \phi_0^2)^2 \tag{20.65}$$

This theory has a discrete chiral symmetry under which $\bar{\psi}\psi \mapsto -\bar{\psi}\psi$ and $\phi \mapsto -\phi$. We will now consider a static soliton state of the ϕ field whose configuration is $\phi(x) = \pm\phi_0\tanh\left(\frac{(x-x_0)}{\sqrt{2}\xi}\right)$ (see Eq.(19.55)). The arguments given above predict that the soliton has fractional charge $+1/2$ (or, more properly, fractional fermion number).

How does the presence of the soliton affect the fermionic spectrum? The stationary states of the Dirac Hamiltonian for the two-component spinor can be written as

$$H_D[\phi]\psi = \alpha i\partial_1\psi + \beta g\phi(x_1)\psi \tag{20.66}$$

where α and β are two Pauli matrices which can always be chosen to be real, say σ_1 and σ_3, respectively. The Dirac Hamiltonian anti-commutes with the the third Pauli matrix, σ_2. This implies that if ψ is an eigenstate of energy E, then $\sigma_2\psi$ is an eigenstate with energy $-E$. Therefore, the spectrum on states with non-zero eigenvalue is symmetric and there is a one-to-one correspondence between positive, and negative energy states.

However, if the field $\phi(x)$ has a soliton (or an anti-soliton) there is a state whose energy is exactly $E = 0$, a zero mode. To construct the state, we rewrite the Dirac equation as

$$i\beta\alpha\partial_1\psi(x_1) + g\phi(x_1)\psi(x_1) = 0 \tag{20.67}$$

We recognize that $\beta\alpha = \gamma_1$ which is an anti-hermitian matrix, $\gamma_1^\dagger = -\gamma_1$, and hence its eigenvalues are $\pm i$. We can choose a basis in which the zero modes are eigenstates of γ_1, and write them as $\psi_\pm(x_1)\chi_\pm$, where $\chi_\pm = (1,0)^\dagger$ or $(0,1)^\dagger$, respectively. The functions $\psi_\pm(x_1)$ satisfy the first order differential equation

$$\partial_1\psi_\pm(x_1) = \pm g\phi(x_1)\psi_\pm(x_1) \tag{20.68}$$

whose physically sensible solution must be square integrable,

$$\int_{-\infty}^{\infty} dx_1 \, |\psi_\pm(x_1)|^2 = 1 \tag{20.69}$$

For a soliton, which asymptotically satisfies $\lim_{x\to\pm\infty}\phi(x_1) = \pm\phi_0$, the square integrable solution the eigenstate has spinor

$$\chi_- = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{20.70}$$

whose γ_1 eigenvalue is $-i$, and the wave function is

$$\psi_-(x_1) = \psi_-(0)\ \exp\left(-\int_0^{x_1} dx_1' \ \phi(x_1') \right) \tag{20.71}$$

This wave function is even under $x_1 \to -x_1$ since since the soliton $\phi(x_1)$ changes sign under this operation. For the soliton profile of Eq.(19.55), at long distances we can approximate $\psi_-(x_1) \simeq \psi_-(0)\ \exp(-\phi_0|x_1|)$, and find an exponential decay (as expected for a bound state). For the anti-soliton, the situation is reversed: now the scalar field obeys $\lim_{x_1\pm\infty}\phi(x_1) = \mp\phi_0$, and we must choose the spinor with γ_1 eigenvalue $+i$. However, the normalizable wave function is still the same as in Eq.(20.71).

20.5 The axial anomaly in 3+1 dimensions

The axial anomaly was discovered in four dimensions in the computation of triangle Feynman diagrams with fermionic loops. Here too, the axial anomaly is the non-conservation of the axial (chiral) current j^5_μ in a theory with a gauge-invariant regularization. If the fermions are massless the chiral anomaly for a $U(1)$ theory is the identity

$$\partial^\mu j^5_\mu = -\frac{e^2}{8\pi^2} F^{\mu\nu} F^*_{\mu\nu} \tag{20.72}$$

where $F^*_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}F^{\lambda\rho}$ is the dual field strength tensor. This result is known to hold in QED to all orders in perturbation theory. Just as in the 1 + 1-dimensional case, the right hand side is a total divergence which has a topological meaning.

We will now show how this result arises using the four-dimensional version of the the Nielsen-Ninomiya argument that we discussed in 1+1 dimensions. We will consider a theory with a single, right-handed, Weyl fermion in an uniform magnetic field B along the direction x_3, whose vector potential is $A_2 = Bx_1$, and all other components are zero. The Dirac equation for the right-handed Weyl (i.e. two-component) spinor ψ_R is

$$[i\partial_0 - (\boldsymbol{p} - e\boldsymbol{A}) \cdot \boldsymbol{\sigma}]\psi_R = 0 \tag{20.73}$$

where $\boldsymbol{p} = -i\boldsymbol{\partial}$ is the momentum operator, and $\boldsymbol{\sigma}$ are the three 2×2 Pauli matrices. As usual, the solutions to this equation are expressed in terms of the Weyl spinor Φ,

$$\psi_R = [i\partial_0 + (\boldsymbol{p} - e\boldsymbol{A}) \cdot \boldsymbol{\sigma}]\Phi = 0 \tag{20.74}$$

leading to

$$[i\partial_0 - (\boldsymbol{p} - e\boldsymbol{A}) \cdot \boldsymbol{\sigma}][i\partial_0 + (\boldsymbol{p} - e\boldsymbol{A}) \cdot \boldsymbol{\sigma}]\Phi = 0 \tag{20.75}$$

The spinor Φ can be chosen to be an eigenstate of p_2 and p_3. Then, Φ is the solution of the harmonic oscillator equation

$$[-\partial_1^2 + (eB)^2(x^1 + \frac{p_2}{eB}) + p_3^2 + eB\sigma_3]\Phi = \omega^2\Phi \tag{20.76}$$

The eigenvalues are the Landau levels

$$\omega(n, p_3, \sigma_3) = \pm\Big[2eB(n + \frac{1}{2}) + p_3^2 + eB\sigma_3\Big]^{1/2} \tag{20.77}$$

with $n = 0, 1, 2, \ldots$, except for the mode $n = 0$ and $\sigma_3 = -1$ for which

$$\omega(n = 0, \sigma_3 = -1, p_3) = \pm p_3 \tag{20.78}$$

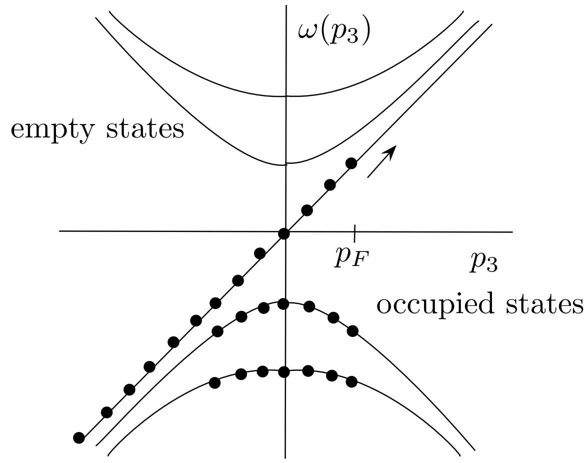

Figure 20.4 The axial anomaly in 3+1 dimensions: spectrum of right handed Weyl fermions in parallel B and E uniform magnetic and electric fields along the direction x_3. The arrow shows that right handed fermions are being created, and the Fermi momentum p_F increases with time.

It is straightforward to find the eigenfunctions. They are, with the exception of the zero modes, the usual linear oscillator wavefunctions time plane waves (for each spinor eigenstates, $(1,0)$ and $(0,1)$). The wave function ψ_R for the zero modes with $n = 0$ and $\sigma_3 = -1$ and eigenvalue $\omega = -p_3$ vanishes, and it is non zero for the zero mode with $n = 0$, $\sigma_3 = -1$ and for $\omega = p_3$. The spectrum is shown in Fig.20.4.

We now turn on an uniform electric field E parallel to the magnetic field B. The levels with $n \neq 0$ are either totally occupied or or empty. In fact, there is a one-to-one correspondence between these levels, and their contributions cancel out to the quantity of our interest. Only the zero mode $n = 0$, $\sigma_3 = -1$ and $\omega = p_3$ matters. Except for a density of states factor, $LeB/(4\pi^2)$ (where L is the linear size of the system), its contribution is the same as in the 1+1-dimensional case. Thus, we find that the rate of creation of right-handed Weyl fermions is

$$\frac{dN_R}{dx_0} = \frac{1}{L}\frac{LeB}{4\pi^2}\frac{dp_F}{dx_0} = \frac{e^2}{4\pi^2}EB = \frac{dQ_R}{dx_0} \tag{20.79}$$

Similarly, for left-handed Weyl fermions, ψ_L, we find that the rate is

$$\frac{dN_L}{dx_0} = -\frac{1}{L}\frac{LeB}{4\pi^2}\frac{dp_F}{dx_0} = -\frac{e^2}{4\pi^2}EB = \frac{dQ_L}{dx_0} \tag{20.80}$$

Therefore, the rate of creation or right handed particles and of left handed antiparticles is

$$\frac{dN_R}{dx_0} + \frac{d\bar{N}_L}{dx_0} = \frac{e^2}{2\pi^2} EB = \frac{dQ_5}{dx_0} \tag{20.81}$$

which is the axial anomaly in 3+1 dimensions.

20.6 Fermion path integrals, the chiral anomaly, and the index theorem

At the beginning of this chapter we noted that in the context of the path integral the axial (or chiral) anomaly arises due to a non-invariance of the fermionic measure under chiral transformations (Fujikawa, 1979). Thus, although the action is invariant, the partition function is not.

We will see how this takes place in a simple theory, a free massless Dirac fermion coupled to a $U(1)$ gauge field. In an even D-dimensional Euclidean space-time it is

$$\mathcal{L} = \bar{\psi} \not{D} \psi \tag{20.82}$$

where, as usual, $\not{D} = \gamma^\mu(\partial_\mu + iA_\mu)$, and $\not{D}$ is a hermitian operator.

In the definition of the path integral one needs to specify a complete basis of field configurations that will specify the integration measure. It is natural to use the eigenstates of the operator $\not{D}$. Let $\{\psi_n\}$ be a complete set of spinor eigenstates of $\not{D}$,

$$\not{D}\psi_n(x) = \lambda_n \psi_n(x) \tag{20.83}$$

where the eigenvalues are $\lambda_n \in \mathbb{R}$, and the eigenstates are complete and orthonormal,

$$\int d^dx \; \psi_n^\dagger(x)\psi_m(x) = \delta_{n,m}, \quad \sum_n \psi_n^\dagger(x)\psi_n(y) = \delta(x-y) \tag{20.84}$$

where the eigenstates are functions of the gauge field (since $\not{D}$ depends on the gauge field).

Then, we expand the fields in this basis,

$$\psi(x) = \sum_n a_n \psi_n(x), \qquad \bar{\psi}(x) = \sum_n b_n \psi_n^\dagger(x) \tag{20.85}$$

where the coefficients $\{a_n\}$ and $\{b_n\}$ are are independent Grassmann numbers. The fermionic integration measure is defined as

$$\mathcal{D}\bar{\psi}\mathcal{D}\psi \equiv \prod_n da_n db_n \tag{20.86}$$

Under a local chiral transformation

$$\psi(x) \mapsto \psi'_n(x) = e^{-i\alpha(x)\gamma_5}\psi(x), \quad \bar{\psi}(x) \mapsto \bar{\psi}'_n(x) = \bar{\psi}(x)e^{-i\alpha(x)\gamma_5} \tag{20.87}$$

a linear transformation is induced on the coefficients $\{a_n\}$ and $\{b_n\}$,

$$a_n \mapsto a'_n = \sum_m C_{nm} a_m, \quad b_n \mapsto b'_n = \sum_m C_{nm} b_m \tag{20.88}$$

where

$$C_{nm} = \int d^dx\, \psi_n^\dagger(x) e^{-i\alpha(x)\gamma_5}\psi_m(x) \tag{20.89}$$

are numbers.

Since $\{\gamma_5, \not{D}\} = 0$ we can choose the states $\psi_n(x)$ to be eigenstates of γ_5 and have definite chirality. Then, for each eigenvalue $\lambda_n \neq 0$ there are two linearly independent states with opposite chirality. Then, in this representation, the measure of the coefficients $\{a_n\}$ and $\{b_n\}$ changes by a Jacobian factor

$$\prod_n da'_n = (\det C)^{-1} \prod_n da_n \tag{20.90}$$

and similarly for the b_n coefficients. For an infinitesimal chiral transformation, the Jacobian is

$$\begin{aligned}(\det C)^{-1} &= \prod_n \left[1 - i\int d^dx\, \alpha(x) \sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x)\right] \\ &= \exp(i\int d^dx \alpha(x) \sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x))\end{aligned} \tag{20.91}$$

Therefore, under an infinitesimal local chiral transformation, the fermionic measure changes as

$$\mathcal{D}\bar{\psi}\mathcal{D}\psi \mapsto \mathcal{D}\bar{\psi}'\mathcal{D}\psi' = \exp(2iN_f \sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x)) \tag{20.92}$$

where we allowed for the possibility of having N_f flavors of fermions. If it is not equal to 1, the exponential prefactor is the non-invariance of the measure.

The problem now reduces to the computation of the sums in the exponent of Eq.(20.92). However this sum is ambiguous and its definition requires a regulator. We will define the sum to be

$$\sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x) = \lim_{\varepsilon\to 0^+} \sum_n e^{-\lambda_n^2\epsilon}\psi_n^\dagger(x)\gamma_5\psi_n(x) \tag{20.93}$$

where we introduced a factor in each term of the sum that damps out the contribution of the eigenstates of $\not{D}$ with large eigenvalues. Notice that here we made a choice to regularize the sum using the spectrum of the *gauge-invariant operator* $\exp(-\not{D}^2\epsilon)$. This is the heat kernel method used by Fujikawa. The detailed computation of the fermion determinants using the ζ-function approach is found in (Gamboa Saraví et al., 1984).

Atiyah and Singer proved a theorem, the *Atiyah-Singer Theorem*, which relates the computation of this sum to an *index* of the Dirac operator (Atiyah and Singer, 1968). We will not go through the details of this proof, but we will quote the result of the sum in $D=2$ and $D=4$ dimensions,

$$
\begin{aligned}
\sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x) &= \frac{e}{4\pi}\epsilon_{\mu\nu}F^{\mu\nu}, && \text{for } U(1) \text{ in } d=2\\
\sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x) &= \frac{1}{8\pi^2}F^{\mu\nu}F^*_{\mu\nu}, && \text{for } U(1) \text{ in } d=2\\
\sum_n \psi_n^\dagger(x)\gamma_5\psi_n(x) &= \frac{e^2}{16\pi^2}\mathrm{tr}(F^{\mu\nu}F^*_{\mu\nu}), && \text{for } SU(N) \text{ in } d=4
\end{aligned}
\tag{20.94}
$$

where, once again, we recognize that the right hand side of these equations involves the topological density of the gauge field.

Let us return to Eqs. (20.94) and integrate both sides over all Euclidean space time. The key observation now is that for the states with eigenvalues $\lambda_n \neq 0$ the contributions of the two degenerate chiral eigenstates cancel each other, and, consequently, only the states with zero eigenvalue, $\lambda_n = 0$ (the "zero modes") contribute to the sum. Then, the result of the sum is finite. Let $n_\pm$ be the number of zero modes with chirality ±1. On the other hand, the integral of the right hand side yields the topological charge Q of the gauge field.

The Atiyah-Singer Theorem states that if $\mathcal{D}$ is an (elliptic) differential operator and $\mathcal{D}^\dagger$ is its adjoint, the index $I_\mathcal{D}$ is

$$
I_\mathcal{D} = \dim\ker(\mathcal{D}) - \dim\ker(\mathcal{D}^\dagger) \tag{20.95}
$$

where $\ker(\mathcal{D})$ denotes the subspace spanned by the kernel of the operator $\mathcal{D}$. Consider now the quantity

$$
I_\mathcal{D}(M^2) = \mathrm{Tr}\left(\frac{M^2}{\mathcal{D}^\dagger\mathcal{D}+M^2}\right) - \mathrm{Tr}\left(\frac{M^2}{\mathcal{D}\mathcal{D}^\dagger+M^2}\right) \tag{20.96}
$$

As $M^2 \to 0$ only the contribution of the zero eigenvalues of $\mathcal{D}^\dagger\mathcal{D}$ survives. The normalizable zero eigenvalues of $\mathcal{D}^\dagger\mathcal{D}$ (which are the same as those of $\mathcal{D}$) each contribute with 1 to this expression. Likewise, the normalizable zero eigenvalues of $\mathcal{D}\mathcal{D}^\dagger$ each contribute with -1 (which are those of $\mathcal{D}^\dagger$) . Therefore, we can write index of $\mathcal{D}$ as

$$I_{\mathcal{D}} = \lim_{M^2 \to 0} I_{\mathcal{D}}(M^2) \tag{20.97}$$

Then, it follows that the topological charge Q is equal to the index

$$n_+ - n_- = Q \tag{20.98}$$

This result is actually general and also holds for non-abelian gauge theories. It implies that if the gauge field has an instanton, say, with $Q = 1$, then there must be at least one zero mode.

20.7 The parity anomaly and Chern-Simons gauge theory

So far we considered on even-dimensional space-times. On a space-time of even dimension $D = 2n$, the Dirac spinors have $2n$ components. This in 1+1 dimensions they are two-spinors, in 3+1 dimensions they are four-spinors, etc. Similarly, in 1+1 dimensions the Dirac operator is a 2×2 matrix-valued differential operator and the Dirac gamma matrices are 2×2 matrices that can be chosen to be real (in the Euclidean signature). Likewise in 3+1 dimensions the Dirac operator is a 4×4 matrix and the Dirac gamma matrices can also be chosen to be real (again, in the Euclidean signature). In both cases, there is a special matrix, γ_5, which anti-commutes with the Dirac operator. If γ_5 does not enter in the Dirac operator, then the theory is time-reversal (or CP) invariant.

We will now consider the Dirac theory in 2+1 dimensions. The Dirac spinors are still two-component, as they are in 1+1 dimensions. In the case of a single massive Dirac field, the Lagrangian is

$$\mathcal{L} = \bar{\psi}(i\not{D} - m)\psi \tag{20.99}$$

where D_μ is the covariant derivative. Although this Lagrangian has the same form as in even dimensional space times, there is the fundamental difference that it involves all three γ_μ Dirac gamma matrices (with $\mu = 0, 1, 2$). In particular, the Dirac Hamiltonian

$$H_D = \boldsymbol{\alpha} \cdot \boldsymbol{p} + \beta m \tag{20.100}$$

involves the three Pauli matrices. However the Hamiltonian is not real since, for instance if we chose a representation such that $\alpha_1 = \sigma_1$, $\alpha_2 = \sigma_3$, then the mass term involves the complex matrix $\beta = \sigma_2$. This means that the Hamiltonian is not invariant under time-reversal which will map it onto the complex conjugate and flip the *sign* of the mass term. Likewise, we could have chosen a representation in which $\alpha_2 = \sigma_2$. In that representation, complex conjugation is equivalent to parity, defined as $x_1 \to x_1$ and $x_2 \to -x_2$. In addition, there is no natural definition of a γ_5 matrix and there is no chirality.

These simple observations suggest that in 2+1 dimensions parity (or time-reversal) maybe necessarily broken. We will now see that this is a subtle question and that the physics depends on the regularization. Here too, there is a choice of either making the theory gauge-invariant or parity (and time-reversal) invariant.

This phenomenon is known as the parity anomaly (Redlich, 1984a,b). It arises in the computation of the effective action of the gauge field A_μ coupled to a theory of massive Dirac fermions. If gauge invariance is preserved by the regularization, then the effective action at low energies, low compered to the mass of the fermion, must be a sum of locally gauge invariant operators. In even-dimensional spacetimes the operator with lowest dimensions is the Maxwell term (or the Yang-Mills term in the non-abelian case). If we define the covariant derivative as $D_\mu = \partial_\mu + iA_\mu$, then the gauge field carries units of momentum (in all dimensions!). This means that the Maxwell term (and the Yang-Mills term) is an operator of dimension 4. If this operator appears from integrating out massive fermions, then the prefactor of this effective action in $D = 3$ spacetime dimensions must be proportional to $1/m$, where m is the mass of the Dirac field.

However, in 2+1 dimensions there is a locally gauge-invariant term with dimension 3 that one can write. It is the Chern-Simons term (Deser et al., 1982a,b).

$$\mathcal{L}_{\text{CS}} = \frac{k}{4\pi}\epsilon_{\mu\nu\lambda}A^\mu\partial^\nu A^\lambda \equiv \frac{1}{4\pi}A \wedge dA \tag{20.101}$$

in the abelian theory, and

$$\mathcal{L}_{\text{CS}} = \frac{k}{8\pi}\text{tr}\left(\epsilon_{\mu\nu\lambda}A^\mu\partial^\nu A^\lambda + \frac{2}{3}\epsilon^{\mu\nu\lambda}A_\mu A_\nu A_\lambda\right) \equiv \frac{k}{8\pi}\text{tr}\left(A \wedge dA + \frac{2}{3}A \wedge A \wedge A\right) \tag{20.102}$$

in the non-abelian theory. Here A_μ is in the algebra of a non-abelian simply connected Lie group G. Here we wrote the Lagrangians in a simpler form using the notation of differential forms. (The factor of $1/2$ difference in the prefactors of Eqs. (20.101) and (20.102) is due to the normalization of the traces in Eq.(20.102).)

The Chern-Simons gauge theory has deep connections with topology, particularly the theory of knots (Witten, 1989), and has many applications in physics (e.g. in the physics of the quantum Hall effects, see (Fradkin, 2013)).

We will discuss the Chern-Simons theory in Chapter 22. Since it is first order in derivatives, and involves the Levi-Civita symbol, the Chern-Simons action is locally gauge invariant, and it is *odd* under parity and time reversal. In this sense, it is a natural term to consider as an effective action for a theory with broken parity and time reversal.

However, the Chern-Simons action is not gauge-invariant if the space-time manifold is open (has an edge). But, even if the space-time manifold is closed (e.g. a three-sphere S^3, a three-torus, etc.), the action (or, rather, the weight in the path integral) is not invariant under *large* gauge transformations, which wrap around the space manifold, unless the parameter k is quantized: $k \in \mathbb{Z}$.

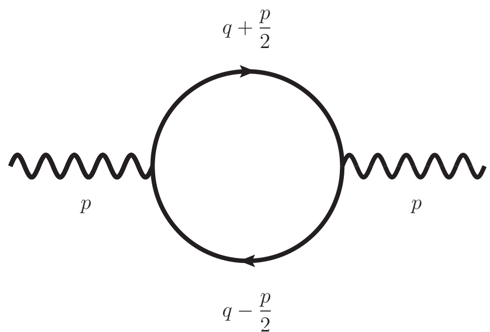

Figure 20.5 One-loop fermion diagram.

Let us do a one-loop perturbative calculation of the effective action of the gauge fields (Redlich, 1984b). In the abelian theory, this requires the computation a Feynman diagram with a single fermion internal loop and two gauge fields in the external legs (a polarization bubble diagram), shown in Fig.20.5. In the non-abelian theory there is, in addition, a triangle diagram with three external gauge fields and a single internal fermion loop.

Both the bubble and the triangle diagrams are UV divergent and require regularization. Since we are after an anomalous contribution, we cannot use dimensional regularization. Anomalies are dimension-specific and dimensional regularization yields ambiguous results, although ad hoc procedures that have been devised to solve this problem. The most commonly used regularization is Pauli-Villars. As we already saw before, Pauli-Villars amounts to the introduction of a set of fields (fermions in this case) with very large mass. This procedure will keep the theory gauge invariant but there will be a finite term that breaks parity (and time reversal), the parity anomaly.

We will only quote the result for the abelian bubble diagram, which we will denote by $\Pi_{\mu\nu}(p)$. Here p_μ is the external momentum (of the gauge fields). The expression to be computed is

$$\Pi^{\mu\nu}(p) = \int \frac{d^3q}{(2\pi)^3} \text{tr}\left[S(q+\frac{p}{2})\gamma^\mu S(q-\frac{p}{2})\gamma^\nu \right] \tag{20.103}$$

where $S(q)$ is the propagator of a Dirac fermion of mass m.

By power counting, this diagram is superficially linearly divergent as the momentum cutoff $\Lambda \to \infty$. Lorentz invariance yields an expression of the form (again in Euclidean space time),

$$\Pi_{\mu\nu}(p) = \Pi_0(p^2)g_{\mu\nu} - i\epsilon_{\mu\nu\lambda}p^\lambda \Pi_A(p^2) + (p^2 g^{\mu\nu} - p^\mu p^\nu)\Pi_E(p^2) \tag{20.104}$$

where $\Pi_0(p^2)$ contains the linearly divergent contribution, and it is not gauge-invariant. The other two remaining terms are finite and gauge-invariant.

If we assume that the only effect of the Pauli-Villars regulators is to subtract the linearly-divergent (and parity-even) term, the low-energy limit, or what is the same, for large fermion mass the gauge-invariant contributions yield an effective action for the abelian gauge field A_μ of the form (again in Euclidean spacetime)

$$S_{\text{eff}}[A] = i\frac{1}{8\pi}\text{sign}(m) \int d^3x\, \epsilon_{\mu\nu\lambda}A^\mu \partial^\nu A^\lambda + \frac{1}{4g^2}\int d^3x\, F_{\mu\nu}^2 + \dots \tag{20.105}$$

where $g^2 = 1/(\pi|m|)$. Thus, we obtain a parity and time-reversal odd contribution, the Chern-Simons term, and a parity and time-reversal even Maxwell term. Clearly these are the first two terms of an expansion in powers of $1/m$.

However, this well established result, upon closer examination, has a problem. We see that the coefficient of the Chern-Simons term is (up to a sign) $k = 1/2$. This is a problem since on a closed manifold invariance under large gauge transformations require that it should be an integer. Furthermore, this calculation predicts that it should be present even in the massless limit, where the theory should be time-reversal invariant, and with this term it is not.

The solution of this apparent conundrum lies in the fact that the Pauli-Villars regulator is a heavy Dirac fermion, which has two contributions to the effective action. One is the cancellation of the linearly divergent and non-gauge-invariant term in the polarization tensor of Eq.(20.104). However, it has also finite contributions. One of them is a Chern-Simons term, with the same form as in Eq.(20.105), also with coefficient $1/2$, but also with a sign ambiguity. This means that the total contribution to the coefficient of the Chern-Simons term is either $k = 0$ or $k = \pm 1$. In the first case, time-reversal invariance is preserved (and there is no parity anomaly). In the second case, the coefficient is correctly quantized (and time-reversal symmetry is broken), but the sign of k depends no just on the sign of the fermion mass but also on the sign of the regulator mass.

More formally, we write the fermion determinant as

$$\det(i\not{D}[A]) = |\det(i\not{D}[A])| \exp(-i\frac{\pi}{2}\eta[A]) \tag{20.106}$$

where $\eta[A]$ is the Atiyah-Patodi-Singer η-invariant,

$$\eta[A] = \sum_{\lambda_k > 0} 1 - \sum_{\lambda_k < 0} 1 \tag{20.107}$$

is the regularized and gauge-invariant spectral asymmetry of the Dirac operator. In this language, this contribution is often expressed as the $U(1)_{-1/2}$ Chern-Simons action (Alvarez-Gaumé et al., 1985).

A physically more intuitive way to reach the same result is to consider a theory in which the Dirac fermions are defined on a spatial lattice. This is the setting appropriate to investigate the quantum anomalous Hall effect. There is a theorem by Nielsen and Ninomiya which states that it is not possible to have a local theory or chiral fermions on a (any) lattice (and in any dimensions). The upshot of this “fermion doubling” theorem is that the number of Dirac fermions must be even. This theorem has been a long standing obstacle to non-perturbative studies of theories of weak interactions using lattice gauge theory methods (Nielsen and Ninomiya, 1981).

In 2+1 dimensions, the Nielsen-Ninomiya theorem implies that, at a minimum, there should be two Dirac fermions. So, we see that if the mass of one fermion is much larger that the mass of the other, then the “doubler” plays the role of the Pauli-Villars regulator. The other relevant consideration is that the coefficient of the Chern-Simons term is the same as the value of the Hall conductivity of the theory of fermions. In units of e^2/h, the hall conductivity is given by a topological invariant of the occupied band of states of the lattice model. This topological invariant is equal to an integer, the first Chern number. This topological invariant is the non-trivial winding number of the Berry phase of the single-particle states on the Brillouin zone.

20.8 Anomaly inflow

In the previous sections we discussed chiral anomalies in 1+1 and 3+1 dimensions and the parity anomaly in 2+1 dimensions. We will now see that these anomalies are related. The relation involves considering systems on different space dimensions, where the lower dimensional system is defined on a topological defect of the higher dimensional partner. These topological defects are vortices and domain walls. The connection between anomalies in different dimensions in theories with topological defects is known as the anomaly inflow.

We will see that, in the presence of topological defects, the anomalies of the theory defined on the defect, precisely match the anomalies of the theory in the larger system (the "bulk"). Here we will follow the general construction of the anomaly inflow by Callan and Harvey (Callan and Harvey, 1985).

20.8.1 Axion strings

For the sake of concreteness we will consider a theory in even ($D = 4$) dimensions with a γ^5 coupling to a string-like topological defect, a vortex (or "axion string"). We will assume that the Dirac fermions are coupled through mass terms to a complex scalar field $\varphi = \varphi_1 + i\varphi_2$, that has a vortex topological defect running along the x_3 axis. The Lagrangian in the vortex background is

$$\mathcal{L} = \bar{\psi} i \partial\!\!\!/ \psi + \bar{\psi}(\varphi_1 + i\gamma^5 \varphi_2)\psi \tag{20.108}$$

The vortex has the usual form $\varphi = f(\rho)\exp(i\vartheta)$, where ϑ is the phase of the complex scalar field. Here we will use cylindrical coordinates, with $x_1 = \rho\cos\phi$, and $x_2 = \rho\sin\phi$. The vortex solution behaves at infinity as $\lim_{\rho\to\infty} f(\rho) = \varphi_0$, and vanishes as $\rho \to 0$, where the vortex has a singularity. In the asymptotic regime, the phase of a single vortex with winding number $N = +1$, is $\vartheta = \phi$. The vortex singularity is an "axion string."

We will first show that, in the presence of the vortex, the Dirac operator has zero modes. Let us rewrite $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$ as $\gamma^5 = \Gamma^{\text{int}}\Gamma^{\text{ext}}$, where $\Gamma^{\text{int}} = i\gamma^0\gamma^1$, and $\Gamma^{\text{ext}} = \gamma^2\gamma^3$. Here, Γ^{int} is the γ^5 matrix of the 1+1 dimensional world, and measures the chirality of the modes of the Dirac equation on the string. Let $\psi_\pm$ be eigenfunctions of γ^5 with eigenvalue ± 1. In this basis, the Dirac equation is

$$\begin{aligned} i\partial\!\!\!/^{\text{int}}\psi_- + i\gamma^2(\cos\varphi + i\Gamma^{\text{ext}}\sin\varphi)\partial_\rho\psi_- &= f(\rho)e^{i\vartheta}\psi_+ \\ i\partial\!\!\!/^{\text{int}}\psi_+ + i\gamma^2(\cos\varphi + i\Gamma^{\text{ext}}\sin\varphi)\partial_\rho\psi_+ &= f(\rho)e^{i\vartheta}\psi_- \end{aligned} \tag{20.109}$$

The zero mode solution is

$$\psi_- = \eta(x^{\text{int}})\exp\left(-\int_0^\rho d\rho'\, f(\rho')\right), \qquad \text{where } \psi_+ = -i\gamma^2\psi_- \tag{20.110}$$

where the spinor η satisfies

$$i\partial\!\!\!/^{\text{int}}\eta = 0, \qquad \Gamma^{\text{int}}\eta = -\eta \tag{20.111}$$

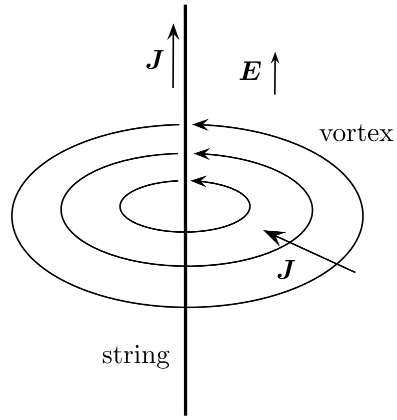

Figure 20.6 The anomaly inflow in the axion string, the singularity of a vortex. In the presence of an uniform electric field $\boldsymbol{E}$ parallel to the axion string, in the 'bulk' three-dimensional world there is a Hall-like current flowing towards the string, and flows the string, parallel to $\boldsymbol{E}$. The direction of the current on the string is determined by the sign of the vorticity.

For the antivortex, with winding number $N = -1$, the phase field is $\vartheta = -\phi$, and the fermion zero mode on the axion string is left moving.

Hence, there is a massless chiral fermion traveling along the string, and the direction of propagation is determined by the axial charge of the fermion. Therefore, on the string there is a chiral fermion. Now, we see that in the presence of an external electromagnetic field, A_μ, the chiral fermion has a gauge anomaly,

$$\partial^\mu J_\mu = \frac{1}{2\pi}\epsilon_{\mu\nu}\partial^\mu A^\nu \tag{20.112}$$

where now $\mu, \nu = 0, 1$. However, we saw that the anomaly of a chiral fermion, Eq.(20.10), means that in the presence of an electric field E parallel to the string, the gauge charge is not conserved. Where is this charge coming from?

To see where this charge may be coming from, we need to consider the rest of the spectrum of the Dirac operator (i.e. the non-zero modes). this means that we need to examine what happens in the higher dimensional bulk. This can be done using the method of Goldstone and Wilczek, see Fig.20.3. In 3+1 dimensions the results is

$$\langle J_\mu \rangle = -i\frac{e}{16\pi^2}\epsilon_{\mu\nu\lambda\rho}\frac{(\varphi^*\partial^\nu\varphi - \varphi\partial^\nu\varphi^*)}{|\varphi|^2}F^{\lambda\rho} \tag{20.113}$$

far for the vortex singularity we can approximate $\varphi \simeq \phi_0 \exp(i\vartheta(x))$ to find

$$\langle J_\mu \rangle = \frac{e}{8\pi^2}\epsilon_{\mu\nu\lambda\rho}\partial^\nu \vartheta(x) F^{\lambda\rho} \tag{20.114}$$

For a vortex running along the x_3 axis, far from the singularity we get $\vartheta(x) = \phi$. If the background field is an electric field of magnitude E along the direction x_3, the induced current J_μ has only a radial component,

$$J_\rho = -\frac{e}{4\pi^2 \rho} E \tag{20.115}$$

and it is pointing inwards, towards the axion string, and is perpendicular to the applied electric field. It is this inwards component of the current that compensates the anomaly on the string. Moreover, using the singularity of the phase, $(\partial_1\partial_2 - \partial_2\partial_1)\vartheta(x_1, x_2) = 2\pi\delta(x_1)\delta(x_2)$, we obtain

$$\partial^\mu \langle J_\mu \rangle = \frac{e}{4\pi}\epsilon^{ab} F_{ab}\delta(x_1)\delta(x_2) \tag{20.116}$$

Hence, the current is conserved away from the singularity, and the non-conservation at the string singularity is matched by the gauge anomaly on the string!

20.8.2 Domain wall in 2+1 dimensions

We will now consider the effects of domain walls. We will consider first a theory in 2+1 dimensions with the Lagrangian

$$\mathcal{L} = \bar{\psi} i \partial\!\!\!/ \psi + \frac{1}{2}(\partial_\mu \phi)^2 - g\phi\bar{\psi}\psi - U(\phi) \tag{20.117}$$

where ψ is a two-component Dirac fermion. Here ϕ is a real scalar field in its broken symmetry state, and we write the interaction in the form $U(\phi) = \frac{\lambda}{4!}(\phi_0^2 - \phi^2)^2$. We will consider a static domain wall, i.e. a time-independent classical solution of the field ϕ that has a kink (soliton) along the direction x_2 and is constant along x_1. In the sequel $\phi_W(x_2)$ denotes the domain wall.

As in the preceding subsection, we set $\Gamma^{\rm int} = \gamma^0\gamma^1$, and $\Gamma^{\rm ext} = \gamma^2$, and the Dirac equation becomes

$$i\partial\!\!\!/^{\rm int}\psi - \Gamma^{\rm int}\partial_2\psi = g\phi_W(x_2)\psi \tag{20.118}$$

The zero mode is

$$\psi = \eta(x_{\rm int}) \exp\left(- \int_0^{x_2} dx_2'\, g\phi_W(x_2)\right) \tag{20.119}$$

with

$$i\not{\partial}^{\text{int}}\eta = 0, \qquad \text{and } \Gamma^{\text{int}}\eta = \eta \tag{20.120}$$

This is a right-moving Dirac fermion along the wall.

Since in the broken symmetry state of the scalar field the fermions are massive and their mass is $\phi_0 \text{sign}(x_2)$, depending on the side of the wall. The fermion contribution of the fermions to the effective action of a gauge field A_μ, far from the location of the wall is

$$S_{\text{eff}} = \int d^3x \frac{(1 + \text{sign}(x_2))}{8\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \tag{20.121}$$

where the integral extends over all the 2+1-dimensional spacetime, except from an infinitesimal region surrounding the domain wall.

Under a gauge transformation, $A_\mu \to A_\mu + \partial_\mu \Phi$, this action changes as

$$\begin{aligned}\Delta S_{\text{eff}} &= \int d^3x \ \frac{1}{8\pi}(1 + \text{sign}(x_2))\epsilon^{\mu\nu\lambda}\partial_\mu \Phi \ \partial_\nu A_\lambda \\ &= -\frac{1}{2\pi}\int d^2x \ \Phi(x)\epsilon_{\mu\nu}F^{\mu\nu}\end{aligned} \tag{20.122}$$

where we performed an integration by parts. This result cancels the gauge anomaly of the chiral fermion in 1+1 dimensions, carried by the fermion zero modes on the wall. In the theory of the quantum Hall effects, these are the chiral edge states.

It is straightforward to see that a uniform electric field parallel to the wall induces a charge current that flows towards the wall. On the other hand, the electric field will induce a current along the wall, by the pair creation mechanism we discussed. In each sector, bulk and boundary, there is a gauge anomaly, but these anomalies cancel each other, as we just saw.

20.8.3 Domain wall in 3+1 dimensions

We will now discuss, briefly, the problem Dirac fermions coupled to a domain wall in 3+1 dimensions. Physically, this problem arises in the behavior of three-dimensional topological insulators (Qi et al., 2008; Boyanovsky et al., 1987), and as a technique in lattice gauge theory to study theories of chiral fermions (Kaplan, 1992). The Lagrangian is

$$\mathcal{L} = \bar{\psi} i \not{\partial} \psi + g\varphi(x)\bar{\psi}\psi \tag{20.123}$$

Since we are in 3+1 dimensions the fermions are the usual Dirac four component spinors. Here, $\varphi(x)$ is a real scalar field. We will assume that $\varphi(x)$ is a static classical configuration with a domain wall, i.e. $\varphi(x)$ depends only on x_3: $\varphi(x) = \phi_0\, f(x_3)$, where $f(x_3) \to \pm 1$ as $x_3 \to \pm\infty$; ϕ_0 is the vacuum expectation value of $\varphi(x)$. Clearly, $g\varphi(x)$ plays the role of a Dirac mass, $m(x_3)$, that varies form $g\phi_0$ to $-g\phi_0$.

The Dirac Hamiltonian has the standard form,

$$H = -i\boldsymbol{\alpha} \cdot \nabla + m(x_3)\beta \tag{20.124}$$

where $\boldsymbol{\alpha}$ and β are the Dirac matrices. The geometry of this problem suggests that we split the Hamiltonian into on and off the wall components, $H = H_{\text{wall}} + H_\perp$, where

$$H_{\text{wall}} = -i\alpha_1\partial_1 - i\alpha_2\partial_2, \qquad H_\perp = -i\alpha_3\partial_3 + m(x_3)\beta \tag{20.125}$$

Let $\psi_\pm$ be an eigenstate of the anti-hermitian Dirac matrix $\gamma_3 = \beta\alpha_3$ with eigenvalues $\pm i$, respectively,

$$\gamma_3\psi_\pm = \pm i\psi_\pm \tag{20.126}$$

We will seek a zero-mode solution $H_\perp\psi_\pm = 0$. Hence,

$$\pm\partial_3\psi_\pm + m(x_3)\psi_\pm = 0 \tag{20.127}$$

and a solution of

$$(i\gamma_0 - i\gamma_1\partial_1 - i\gamma_2\partial_2)\psi_\pm = 0 \tag{20.128}$$

In other words, is a solution of a massless Dirac equation in 2+1 dimensions. The requirement that $\psi_\pm$ be an eigenstate of γ_3, reduces the number of spinor components from four to two.

The solution of these equations has the form

$$\psi_\pm = \eta_\pm(x_0, x_1, x_2)\; F_\pm(x_3), \qquad \text{where } \pm\,\partial_3 F_\pm(x_3) = -m(x_3)F_\pm(x_3) \tag{20.129}$$

The solution $F_\pm(x_3)$ is

$$F_\pm(x_3) = F(0)\exp\left(\mp\int_0^{x_3} dx_3'\; m(x_3')\right) \tag{20.130}$$

If the domain wall behaves as $m(x_3) \to g\phi_0$ as $x_3 \to \infty$, then $F_+(x_3)$ is the normalizable solution. This also implies that the spinor $\eta_\pm(x_0, x_1, x_2)$ should be an eigenstate of γ_3 with eigenvalue $+i$. It then follows that the Dirac fermions on the 2+1-dimensional wall are massless, Eq.(20.128).

We conclude that the Dirac equation in 3+1 dimensions coupled to a domain wall defect has zero modes that the behave as massless Dirac fermions in 2+1 dimensions confined to the wall. Note that, if the Dirac fermions have, in addition to the Dirac mass m, a γ^5 mass m_5, the fermions of the

2+1 dimensional world of the domain wall now have a mass equal to m_5. Therefore, we expect that the fermions on the wall should exhibit the parity anomaly determined by the sign of m_5!

This question can be answered using, again, an extension of the Callan-Harvey argument. Indeed, instead of coupling the fermions to a real scalar field, they will now be coupled to a complex scalar field $\varphi = \varphi_1 + i\varphi_2$, with the same coupling to the fermions as in the case of the axion string, Eq.(20.108), except that now we will have a domain wall, with $g\varphi_1(x_3) = m(x_3)$ and $g\varphi_2 = m_5$. Once again, we can use the callan-Harvey result of Eq.(20.113), and infer that there is a current J_μ induced by the coupling to the electromagnetic gauge field A_μ. Indeed, since the θ angle is a slowly varying function of x_3, we find that an electric field parallel to the wall induces a current also parallel to the wall and the the electric field, a result reminiscent of the Hall effect. Likewise, a magnetic field normal to the wall induces a charge on the wall.

In addition, since there is only one Dirac spinor of the wall, the effective action of the gauge field on the wall is the same as predicted from the parity anomaly, with a coefficient 1/2 the quantized value. So, we seemingly find again the same difficulties. However, the bulk effective action supplies, in this case, we requisite missing contribution. Hence, here too, the anomaly cancellation is non-local and involves the whole edge and the bulk contributions.

Even in the abelian theory, whose gauge field sector does not have instantons, it has implications if the three dimensional space of the 3 + 1-dimensional spacetime Ω has a boundary $\partial\omega$. Using the divergence theorem, one readily finds that the θ term of the action integrates to the boundary, where it becomes (for $\theta = \pi$)

$$S_{\text{boundary}} = \frac{1}{8\pi}\int_{\partial\Omega\times\mathbb{R}} d^3x\, \epsilon_{\mu\nu\lambda}A^\mu\partial^\nu A^\lambda \tag{20.131}$$

This is a Chern-Simons action with level $k = 1/2$. On the other hand, the effective action Dirac fermion at the boundary is also a Chern-Simons gauge their with level $k = 1/2$. We discussed above that this result is incompatible with the requirement that the level k of the Chern-Simons theory be an integer. We see now, that if the 2+1 dimensional theory is the boundary of a 3+1-dimensional theory, the contribution of the bulk to the effective action matches the parity anomaly of the boundary leading to a consistent theory. This result has important implications in the theory of topological insulators in 3+1 dimensions.

20.9 θ vacua

In this section will investigate the role of the topological charge in several models of interest. We will do this by weighing different topological sector of the configuration space by their topological charge. In a theory defined in Euclidean space time, since the topological charge Q is an integer, we will weigh the configurations by a factor $\exp(i\theta Q)$, where θ is an angle defined in $[0, 2\pi)$. Thus the partition function now depends on the parameter θ

$$Z(\theta) = \int \mathcal{D}\text{fields}\, e^{-S}\, e^{i\theta Q} \tag{20.132}$$

The dependence of the partition function on the angle θ provides a direct way to assess the role of instantons in the physical vacuum of the theory (Callan et al., 1976; Callan, Jr. et al., 1978).

20.9.1 The Schwinger model and θ vacua

The Schwinger model is the theory of quantum electrodynamics in 1+1 dimensions (Schwinger, 1962). The Lagrangian of the Schwinger model is

$$\mathcal{L} = \bar{\psi} i \not{\partial} \psi - m\bar{\psi}\psi - \bar{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4e^2} F_{\mu\nu}^2 - \frac{\theta}{4\pi}\epsilon_{\mu\nu}F^{\mu\nu} \tag{20.133}$$

Here we included the last term, proportional to the topological charge, and parametrized by the θ angle. In this theory, this term is equivalent to a Wilson loop of charge $\frac{\theta}{2\pi}$ on a closed contour at the boundary of the two-dimensional space time. Therefore, it can be interpreted as the effect of the uniform electric field E generated by two charges $\pm\frac{\theta}{2\pi}$ at $\pm\infty$.

In 1962 Schwinger considered the massless theory (with $\theta = 0$) and showed that the spectrum is exhausted by a free massive scalar field, a boson of mass $m_{\text{Schwinger}} = \frac{e}{\sqrt{\pi}}$. Recall that in 1+1 dimensions the QED coupling constant, the electric charge e, has units of mass. The massless fermions of the Lagrangian are not present in the spectrum of physical, gauge-invariant states. He further showed that the gauge fields became massive and that its mass is the mass of the boson.

The world of 1+1 dimensions is quite restrictive, and simple. Indeed, in 1+1 dimensions, for obvious reasons, the gauge field does not have transverse components. When coupled to a charged matters field, it acquires a longitudinal component which is the same as the local charge fluctuations of the matter field. In particular, if the matter field is absent, the Coulomb gauge condition $\partial_1 A_1 = 0$ is solved by $A_1 = \text{const.}$. The *confinement* of charge matter fields is easily seen even classically. Indeed, in one spatial dimension the electric field created by a pair of static charges is necessarily uniform, and the Coulomb interaction grows linearly with the separation R between the charges, $V(R) = -qR$. Therefore, at least at the classical level, the static sources are confined.

We will now look at the quantum version of the theory. The simplest way analyze the quantum theory is to use the method of bosonization, already discussed in Section 20.3. There we used bosonization to identify fermionic current (Eq.(20.24)), the fermion operators (Eq.(20.43)), and the mass terms (Eq.(20.47)). The bosonized version of the Lagrangian of Eq.(20.133) is

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - g\cos(\sqrt{4\pi}\phi) + \frac{1}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu \phi A^\mu - \frac{1}{4e^2}F_{\mu\nu}^2 - \frac{\theta}{4\pi}\epsilon_{\mu\nu}F^{\mu\nu} \quad (20.134)$$

where $g \propto m$. Upon gauge fixing, e.g. the Feynman-'t Hooft gauges, we can integrate-out the gauge field and obtain the effective Lagrangian (Kogut and Susskind, 1975b; Coleman et al., 1975)

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{e^2}{2\pi}\phi^2 - g\cos(\sqrt{4\pi}\phi + \theta) \quad (20.135)$$

The parameter θ was originally introduced in the operator solution of the massless Schwinger model to insure consistency of the boundary conditions (Lowenstein and Swieca, 1971). It was later given a physical interpretation in terms of the global degrees of freedom of the gauge theory, i.e. as a background electric field (Kogut and Susskind, 1975b).

Let us discuss the massless case first (Lowenstein and Swieca, 1971; Casher et al., 1974; Kogut and Susskind, 1975b). If the fermion mass is zero, $m = 0$, the boson ϕ is massive and its mass squared is $m^2_{\text{Schwinger}} = e^2/\pi$. This is Schwinger's result. After some simple calculations one can also show that the fermion propagator (in the Coulomb gauge) is zero. This means that there are no fermions in the spectrum, i.e. the fermions are confined. On the other hand, since the boson is massive, then we see that the Dirac mass operator has a finite expectation value, $\langle \bar{\psi}\psi \rangle = \langle \cos(\sqrt{4\pi}\phi \rangle \neq 0$, since, up to weak gaussian fluctuations, the boson ϕ is pinned to the classical value $\phi_c = 0$. Therefore, the chiral symmetry is spontaneously broken. In the massless case the parameter θ is irrelevant in the sense that the vacuum energy is independent of the value of θ. Vacua with different values of θ rotate into each other under the action of the global chiral charge, Q_5, without any energy cost.

Things are different if the fermions are massive, $m \neq 0$. In this case the cosine term in the action of Eq.(20.135) is now present. While here too the fermions are confined, and the scalar field ϕ is massive, there is a tension between the Schwinger mass term and the cosine term. Although this theory is no longer trivially solvable, we see that the parameter θ will change the vacuum energy. To lowest order in g, the vacuum energy density is a periodic function of θ, $\varepsilon(\theta) = g\cos\theta$. The physical interpretation is that, since the background electric field is quantized in multiples of 2π, it will change abruptly at the value $\theta = \pi$ (mod 2π).

20.9.2 θ vacua and the $\mathbb{CP}^{N-1}$ model

Since the $\mathbb{CP}^{N-1}$ model can be solved in the large-N limit (see Section 17.3), and has instantons for all values of N, we may ask what role do instantons play in the large-N limit. Following the approach used in Section 17.3, we rescale the coupling constant $g^2 \to g_0^2/N$. In this limit, the bound for the Euclidean action is $S \geq 2\pi N|Q|/g_0^2$, and one would expect that the instanton contribution should be $O(\exp(-cN))$.

We will address this problem by modifying the action of the of the $\mathbb{CP}^{N-1}$ model by adding a term proportional to the topological charge Q of the field configuration. In Euclidean spacetime, the partition function becomes (Witten, 1979b)

$$Z[\theta] = \int \mathcal{D}\boldsymbol{z}\mathcal{D}\boldsymbol{z}^*\mathcal{D}A_\mu \, \exp\left(-\frac{1}{g^2}\int d^2x \, |(\partial_\mu + iA_\mu)z_a|^2 + i\theta\, Q\right) \quad (20.136)$$

The contribution to the weight of the path integral of the term involving the angle θ can be expressed as

$$\exp(i\theta Q) = \exp\left(i\frac{\theta}{2\pi}\int_\Sigma d^2x \, \epsilon_{\mu\nu}\partial^\mu A^\nu\right) = \exp\left(i\frac{\theta}{2\pi}\oint_{\partial\Sigma} dx_\mu A^\mu\right) \quad (20.137)$$

or, what is the same, we can write

$$\frac{Z[\theta]}{Z[0]} = \left\langle \exp\left(i\frac{\theta}{2\pi}\oint_{\partial\Sigma} dx_\mu A^\mu\right)\right\rangle_{\theta=0} \quad (20.138)$$

Therefore, the addition of the term with the θ angle is equivalent to the computation of the expectation value of the Wilson loop operator with a test charge of charge $q = \theta/(2\pi)$ on a contour $\partial\Sigma$ that encloses the full spacetime Σ. Thus, the θ term con be interpreted as the response to a background electric field $E = \theta/(2\pi)$.

The presence of the θ term does not change the asymptotically free character of the theory in the weak coupling regime. Indeed, its beta function predicts that under the RG the theory flows to strong coupling, a regime inaccessible to the perturbative RG. So other approaches must be used to study this regime. One non-perturbative approach is large-N, and the $1/N$ expansion.

Since $Q \in \mathbb{Z}$, the parameter $\theta \in [0, 2\pi)$, and one expects the partition function (and the free energy) to be a periodic function of the angle θ. Notice that since the weight of a configuration is now a complex number, the theory is not invariant under time-reversal, or equivalent under CP, unless $\theta = 0, \pi \bmod (2\pi)$. Terms of this type often arise in theories with Dirac fermions due to the effects of quantum anomalies.

Since the topological charge Q is the integral of a total derivative, the saddle-point equations of this theory are the same as in the theory with $\theta = 0$. Then, the θ dependence on physical quantities comes from quantum fluctuations. To leading order in the $1/N$ expansion, the θ dependence of the ground state energy density, $\varepsilon_0(\theta)$, i.e. the response to the background electric field E, is computed by

$$\varepsilon_0(\theta) = -\lim_{L,T\to\infty} \frac{1}{LT} \ln\left(\frac{Z[\theta]}{Z[0]}\right) \tag{20.139}$$

The result turns out to be a simple periodic function of θ (shown in Fig.20.7),

$$\varepsilon_0(\theta) = \frac{1}{2}\left(\frac{e_{\text{eff}}\theta}{2\pi}\right)^2 + O(1/N^2), \qquad \text{for} |\theta| < \pi \tag{20.140}$$

and periodically defined in other intervals. here we defined

$$e_{\text{eff}}^2 = 12\pi^2 \frac{M^2}{N} \tag{20.141}$$

where M^2 is the dynamically generated mass of the $\mathbb{CP}^{N-1}$ model,

$$M = \mu\, e^{-2\pi/g_R^2} \tag{20.142}$$

The fact that $0 < \varepsilon_0[\theta] < \infty$ implies that the test particles of (fractional) charge $\theta/(2\pi)$ are confined.

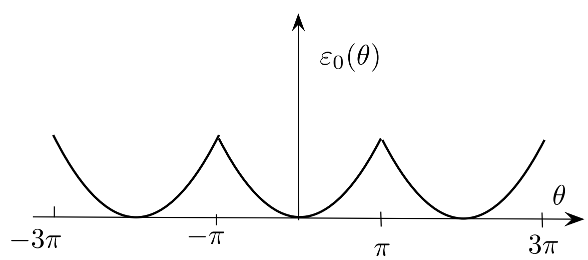

Figure 20.7 Ground state energy density of the $\mathbb{CP}^{N-1}$ model in the large N limit as a function of the angle θ. It has cusps for θ an odd integer multiple of π, where there is a first order transition and CP is spontaneously broken.

The function $\varepsilon_0(\theta)$ is a periodic quadratic function of θ with cusp singularities at odd multiples of π. It has a smooth $1/N$ expansion, in clear contradiction with the expected N-dependence from a naive instanton argument which predicts a $\cos\theta$ dependence. This cusp implies that there is a discontinuity (a jump) in the topological density, $q = \langle Q \rangle = \frac{\partial \varepsilon_0}{\partial \theta}$ for θ an odd multiple of π. Since $\langle Q \rangle \neq 0$ violates CP (or T) invariance, this symmetry is spontaneously broken, at least in the large N limit.

20.9.3 The $O(3)$ non-linear sigma model and θ vacua.

The $\mathbb{CP}^1$ model is equivalent to the $O(3)$ non-linear sigma model. Its topological charge is given by Eq.(19.105), and the modified Euclidean action now is

$$S = \frac{1}{2g^2} \int d^2x \, (\partial_\mu \boldsymbol{n})^2 + i\frac{\theta}{8\pi} \int_{S^2} d^2x \, \epsilon_{\mu\nu} \boldsymbol{n}(x) \cdot \partial_\mu \boldsymbol{n}(x) \times \partial_\nu \boldsymbol{n}(x) \quad (20.143)$$

In particular, the effective action of Eq.(20.143) is the effective field theory of one-dimensional spin-s quantum Heisenberg antiferromagnets, with $\theta = 2\pi s$, and $g^2 \propto 1/s$. Hence, for large values of s the theory is weakly coupled, and for small values of s is strongly coupled, e.g. for $s = 1/2$. Since s can only be an integer of a half-integer, the value of θ is either 0 (mod 2π) for $s \in \mathbb{Z}$, or π (mod 2π) for $s = 1/2$ (mod $\mathbb{Z}$).

The perturbative RG shows that the theory is asymptotically free at weak coupling regardless the value of θ. For integer s, where the θ term is absent, the coupling constant g^2 flows under the RG to strong coupling. But, for half-integer values of s it has been shown that the RG flows to a fixed point with a finite value of g^2 and $\theta = \pi$. We will now see how this works (Haldane, 1983; Affleck and Haldane, 1987).

This non-perturbative result is known form a series of mappings between different models. The spin-1/2 one-dimensional Heisenberg quantum antiferromagnet is an integrable system wit a global $SU(2)$ symmetry. It is solvable by the the Bethe ansatz, and the entire spectrum if known, as are (most of) its correlation functions. In particular the scaling dimensions of all the local operators are known. At the isotropic point the theory is massless and and conformally invariant at low energies. The massless excitations are fermionic solitons. It is a conformal field theory, a subject that we will discuss in chapter 21, with central charge $c = 1/2$. In particular, the correlation functions or the (normalized) chiral $SU(2)$ (spin) currents $J^{\pm}(x)$ are known to decay as a power law as

$$\langle J^{\pm}(x) J^{\pm}(y) \rangle = \frac{k/2}{|x-y|^2} \tag{20.144}$$

In the case at hand, $k = 1$. The exponent of this power law implies that $J^{\pm}$ has scaling dimension 1.

The WZW model is an $SU(N)$ non-linear sigma model in 1+1 dimensions with field $g(x) \in SU(N)$ (and $g^{-1} = g^{\dagger}$), i.e. its target space is the group manifold $G = SU(N)$. The Minkowski space action is

$$S[g] = \frac{1}{4\lambda^2} \int_{S^2_{\text{base}}} d^2x \, \text{tr}\left(\partial^{\mu} g \partial_{\mu} g^{-1}\right) + \frac{k}{12\pi} \int_B \epsilon^{\mu\nu\lambda} \text{tr}\left(g^{-1}\partial_{\mu} g \, g^{-1}\partial_{\nu} g \, g^{-1}\partial_{\lambda} g\right) \tag{20.145}$$

The second term of this action is the Wess-Zumino-Witten term. We will see in section 21.6.4 that the level k must be an integer. Much as with other non-linear sigma models, the WZW model is asymptotically free at weak coupling, and the coupling constant runs to strong coupling under the perturbative RG. However, by explicit computation it has been shown that, if $k \neq 0$, the beta function has a zero, and the RG has an IR fixed point, at a finite value of the coupling constant,

$$\lambda_c^2 = \frac{4\pi}{k} \tag{20.146}$$

At this IR fixed point the theory is conformally invariant. This conformal field theory known as the $SU(2)_k$ Wess-Zumino-Witten (WZW) model with level k. We will discuss this CFT in the next chapter. The case of interest in for the $O(3)$ non-linear sigma model is the $SU(2)_1$ WZW theory.

The $SU(2)_1$ WZW CFT has only one CP invariant relevant perturbation, the operator $\mu(\text{tr}g)^2$. For $\mu > 0$, the theory flows in the IR to a fixed point with $\text{tr}g = 0$, where $g \to i\boldsymbol{\sigma} \cdot \boldsymbol{n}$, with $\boldsymbol{n}^2 = 1$. In this limit, the WZW term of the action becomes $\pi Q[\boldsymbol{n}]$, where $Q[\boldsymbol{n}]$ is the topological charge of the non-linear sigma model. Hence, in the IR the theory flows to the $O(3)$ non-linear sigma model of Eq.(20.143) with $\theta = \pi$. These arguments then imply that the $O(3)$ non-linear sigma model with $\theta = \pi$ is at a non-trivial fixed point described by the $SU(2)_1$ WZW CFT.

On the other hand, for $\theta = 2\pi n$ (with $n \in \mathbb{Z}$) the topological term non-linear sigma model action of Eq.(20.143) has no effect since, for these values of θ, the topological term of the action is equal to an integer multiple of 2π. Hence, for $\theta = 2\pi n$ the theory flows to the massive phase (and CP invariant) phase that we described above. This is also the behavior found in spin chains where s is an integer. Results on extensive, and highly accurate,

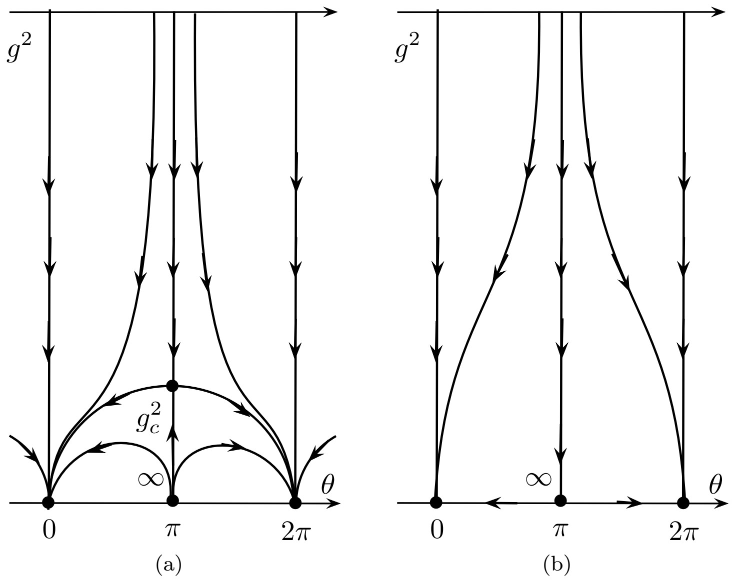

Figure 20.8 Conjectured RG flows of the $\mathbb{CP}^{N-1}$ model as a function of the coupling constant g^2 and of the angle θ: a) the RG flow for $N = 2$ (the $O(3)$ non-linear sigma model) has a non-trivial finite IR fixed point at $(g^2, \theta) = (g_c^2, \pi)$, and b) the RG flow in the large-N limit describes a first order transition at $\theta = \pi$ (mod 2π) and a spontaneous breaking of CP invariance.

numerical simulations of different quantum spin chains (using the density matrix renormalization group approach) are consistent with this picture.

In spite of much analytic and numerical work, what happens for general values of N and θ remains a matter of (educated) speculation. The currently accepted as the '"best guess" for the global RG flow is shown is Fig. 20.8. Fig.20.8a is the conjectured RG flow for $N = 2$ and shows a continuous phase transition at $\theta = \pi$ controlled by the finite fixed point at (g_c^2, π). Fig.20.8b shows the (also conjectured) RG flow for $N \geq 2$. It depicts the case of a first order transition controlled by a fixed point at (∞, π) where the correlation length vanishes, and the mass diverges (a "discontinuity" fixed point).

Analytically, these flows have been derived by saturating the partition function of the non-linear sigma model with the one instanton (and anti-instanton) contribution (Pruisken, 1985). As we noted above, this is problematic since the integration over instanton sizes is infrared divergent. The divergence is suppressed by means of an IR cutoff which breaks the classical scale invariance of the theory. In turn, this leads to an instanton contribution to the RG. The result is suggestive, and predicts an RG flow of the type shown in Fig.20.8. However, this procedure is problematic in many ways. In particular, since the θ term of the action is the total instanton number and, in the absence of the IR regulator, it is insensitive to the effects of local fluctuations. Thus, a priori one would have not expected that θ could be renormalized.

There is, however, numerical (Monte Carlo) evidence of a lattice version of the $\mathbb{CP}^{N-1}$ model that shows that, for $N \geq 3$ as $\theta \to \pi$, the topological density remains finite but has a jump at $\theta = \pi$. On the other hand, for the $\mathbb{CP}^1$ model the topological susceptibility appears to be divergent as $\theta \to \pi$, which is consistent with a continuous phase transition. However, since the Gibbs weight of a generic configuration is not a positive real number, the Monte Carlo method converges poorly at low temperatures, leading to significant error bars in the results, these results cannot be regarded as definitive.

20.9.4 Instantons in gauge theory in 3+1 dimensions and their θ vacua

Returning to the problem of the Dirac fermions coupled to a background gauge field, whose Lagrangian is given in Eq.(20.82), we see that, at least at the formal level, we can integrate out the fermions. The partition function is

$$Z[A] = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi \exp\left(- \int d^4x \ \bar{\psi}(i\not{D}[A] + m)\psi \right) = \det(i\not{D}[A] + m) \tag{20.147}$$

The chiral anomaly implies that the low-energy effective action of the gauge field, in an expansion in powers of $1/m$, in addition to the Maxwell term must contain an extra term

$$S_{\text{eff}}[A] = -\text{tr} \ln(i\not{D}[A]+m) = i\frac{\theta}{32\pi^2} \int d^4x \ \epsilon_{\mu\nu\lambda\rho} F^{\mu\nu} F^{\nu\lambda} + \frac{1}{4e^2} \int d^4x \ F_{\mu\nu}F^{\mu\nu} \tag{20.148}$$

where $\theta = \pi$ for a single flavor of fermions. A similar result is found in the non-abelian case.

Although in four dimensional gauge theory the θ term is a total derivative, and as such it does not affect the equations of motion, it has profound implications. In non-abelian gauge theories, such as a Yang-Mills, the axial anomaly has far reaching consequences.

The θ term has strong implications in non-abelian theories, since these theories have instantons. 't Hooft has shown that in an $SU(2)$ Yang-Mill theory with N_f doublets of massless Dirac fermions, the chiral anomaly of the $U(1)$ current j_μ^5 is the analogous expression

$$\partial^\mu j_\mu^5 = -\frac{N_f g^2}{8\pi^2} \mathrm{tr}(F^{\mu\nu} F_{\mu\nu}^*) \tag{20.149}$$

However, the integral of this expression on the right hand side over all (Euclidean) four-dimensional spacetime is proportional to the instanton number Q (the Pontryagin index of Eq.(19.183)) of the configuration of gauge fields. The total change of the axial charge, ΔQ^5, due to the instanton is ('t Hooft, 1976)

$$\Delta Q^5 = 2N_f Q \tag{20.150}$$

This result implies that in a theory of this type the axial symmetry is broken explicitly, and non-perturbatively, by instantons.

Even the abelian $U(1)$ theory is interesting. The Dirac quantization condition requires that the flux $F_{\mu\nu}$ an an arbitrary closed surface of four-dimensional Euclidean space time be an integer multiple of 2π. This condition, in turn, requires that the weight of the path integral be a periodic function of θ with period 2π. Let us define the complex quantity

$$\tau = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2} \tag{20.151}$$

The partition function is a function of τ. Periodicity in $\theta \to \theta + 2\pi$ means that the partition function (on a manifold with Euclidean signature) is invariant under the transformation $\mathcal{T} : \tau \to \tau + 1$. Witten has shown that in addition, and quite generally, there is an S *duality* transformation, $\mathcal{S} : \tau \to -1/\tau$, under which the partition function transforms simply. More concretely, the transformations $\mathcal{S}$ and $\mathcal{T}$ do not commute with each other, and generate the infinite group of modular transformations $SL(2,\mathbb{Z})$. More precisely, the 2×2 integer-valued matrices are elements of a subgroup $\Gamma \subseteq SL(2,\mathbb{Z})$,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \tag{20.152}$$

such that, under a modular transformation in Γ, a function of the complex variable $F(\tau)$ transforms as

$$F\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^u (c\bar{\tau} + d)^v F(\tau) \tag{20.153}$$

where $\bar{\tau}$ is the complex conjugate of τ, and the exponents u and v depend on topological invariants of the manifold (such as the Euler characteristic). Depending on the manifold, the partition function may be invariant under $SL(2,\mathbb{Z})$ or transform irreducibly as in the above transformation law.

Another example is a theory of two bosons ϕ_I (with $I = 1,2$) in two dimensions, each satisfying the compactification conditions $\phi_I \simeq \phi_I + 2\pi$. Therefore the field (ϕ_1, ϕ_2) is a mapping of the base manifold, which we will take it to be the two-torus T^2, to its target manifold, which is also a two torus T^2. These mappings are classified by the homotopy group $\pi_2(T^2) = \mathbb{Z}$. We now consider the (Euclidean) action

$$S = \int d^2x \frac{1}{4\pi g^2}[(\partial_\mu\phi_1)^2 + (\partial_\mu\phi_2)^2] + i\frac{B}{4\pi}\int d^2x \epsilon_{IJ}\epsilon_{\mu\nu}\partial_\mu\phi_I\partial_\nu\phi_J \quad (20.154)$$

Actions of this type are often considered in toroidal compactifications of (closed) string theory. It is easy to see that the second term is the analog of the θ angle term we discussed above. Indeed, this term is equal to $2\pi iBQ$, where $Q \in \mathbb{Z}$ is the topological charge of the field configuration (ϕ_1, ϕ_2). As a result, the partition function is a periodic function of B with period 1. It turns out that it is also invariant under the same duality transformation as the Maxwell theory and, hence, are invariant under the extended $SL(2,\mathbb{Z})$ modular symmetry (Cardy and Rabinovici, 1982; Cardy, 1982; Shapere and Wilczek, 1989). As we will see in a problem, this structure leads to a complex phase diagram in which electric-magnetic composites condense. In the four-dimensional case, the composite objects are magnetic monopoles with electric charges, called *dyons* (Witten, 1979a).