Chiral Symmetry and Lattice Fermions: Lecture 1

David B. Kaplan

Institute for Nuclear Theory, University of Washington, Seattle, WA 98195-1550

Lectures delivered at the "Frontiers in Lattice QCD" Summer School
Peking University, June 24-July 12, 2019

Acknowledgements

This work is supported in part by U.S. DOE grant No. DE-FG02-00ER41132.

Contents

1

Chiral symmetry and anomalies in d=1+1.

1.1 Introduction

Chiral symmetries play an important role in the spectrum and phenomenology of both the standard model and various theories for physics beyond the standard model. In many cases chiral symmetry is associated with nonperturbative physics which can only be quantitatively explored in full on a lattice. It is therefore important to implement chiral symmetry on the lattice, which turns out to be less than straightforward. In these lectures I discuss what chiral symmetry is, why it is important, how it is broken, and ways to implement it on the lattice. There have been many hundreds of papers on the subject and this is not an exhaustive review; the limited choice of topics I cover reflects on the scope of my own understanding and not the value of the omitted work.

1.2 Dirac fermions in 1+1 dimensions

1.2.1 γ-matrices

Consider a free, massive Dirac fermion in $d = 1 + 1$ dimensions, whose coordinates we will call $x^0 = t$, $x^1 = x$. The Lagrange density is

$$\mathcal{L} = \bar{\psi}\left(i\partial_\mu\gamma^\mu - m\right)\psi \ , \tag{1.1}$$

where ψ is a 2-component spinor, and the γ-matrices satisfy

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \ , \qquad \eta^{\mu\nu} = \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \ . \tag{1.2}$$

A convenient representation for the γ-matrices in terms of the Pauli matrices is

$$\gamma^0 = i\sigma_1 \ , \qquad \gamma^1 = \sigma_2 \ . \tag{1.3}$$

Lorentz transformation of ψ takes the form

$$\psi(x) \to e^{i\frac{1}{2}\omega_{\mu\nu}\sigma^{\mu\nu}}\psi(\Lambda^{-1}x) \ , \qquad \sigma^{\mu\nu} = \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right] \ , \tag{1.4}$$

where $\Lambda_\mu{}^\nu(\omega)$ is the corresponding Lorentz transformation matrix for a 2-vector. This is a bit heavy-handed: life in $d = 1 + 1$ dimensions is life on a wire, and the only Lorentz transformations are boosts in the x direction; for

$$\omega_{01} = -\omega_{10} = \theta \tag{1.5}$$

we have

$$\Lambda_\mu^{\ \nu} = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} , \qquad \sigma^{01} = -i\frac{\sigma_3}{2} . \tag{1.6}$$

Note that we can define a third γ-matrix I will denote Γ which is Hermitian and which anticommutes with both γ^μ, the analogue of γ_5 in $d = 3 + 1$:

$$\Gamma = \Gamma^\dagger , \qquad \Gamma^2 = 1 , \qquad \{\Gamma, \gamma^\mu\} = 0 \quad \Longrightarrow \quad [\Gamma, \sigma^{\mu\nu}] = 0 . \tag{1.7}$$

In the above basis, we can take $\Gamma = \sigma_3$. Since Γ commutes with Lorentz transformations, we conclude that we have a *reducible* representation of the Lorentz group. We can define projection operators

$$P_\pm = \frac{1 \pm \Gamma}{2} , \qquad P_+^2 = P_+ , \qquad P_-^2 = P_- , \qquad P_+ + P_- = 1 . \tag{1.8}$$

(where "1" means the 2×2 unit matrix). Then we define

$$\psi_R = P_+\psi , \quad \psi_L = P_-\psi , \quad \bar\psi_L = \psi_L^\dagger\gamma^0 = \bar\psi P_+ , \quad \bar\psi_R = \psi_R^\dagger\gamma^0 = \bar\psi P_- , \tag{1.9}$$

and we know that $\psi_{L,R}$ will not mix under Lorentz transformations. With the above definitions, writing $\psi = \psi_L + \psi_R$ and plugging back into our Lagrange density in eqn. (1.1) we find we can rewrite it as

$$\mathcal{L} = \bar\psi_L i\partial\!\!\!/\psi_L + \bar\psi_R i\partial\!\!\!/\psi_R - m\left(\bar\psi_L\psi_R + \bar\psi_R\psi_L\right) = \bar\psi_L i\partial\!\!\!/\psi_L + \bar\psi_R i\partial\!\!\!/\psi_R - \left(m\bar\psi_L\psi_R + \text{h.c.}\right) \tag{1.10}$$

1.2.2 The massless case and chiral symmetry

Let's set the fermion mass to zero. We see that the above Lagrange density has two $U(1)$ symmetries: we can rotate the fermions with the two independent phases α and β as

$$\begin{aligned} \psi_L &\to e^{i\alpha}\psi_L , \qquad & \bar\psi_L &\to e^{-i\alpha}\bar\psi_L , \\ \psi_R &\to e^{i\beta}\psi_R , \qquad & \bar\psi_R &\to e^{-i\beta}\bar\psi_R , \end{aligned} \tag{1.11}$$

without affecting $\mathcal{L}$ in eqn. (1.10), provided that $m = 0$. Especially interesting is that the symmetry persists even if we add gauge interactions, replacing ∂_μ by a gauge covariant derivative D_μ. Therefore, even with gauge interactions, there are apparently two conserved currents

$$j_R^\mu = \bar\psi\gamma^\mu P_+\psi , \qquad j_L^\mu = \bar\psi\gamma^\mu P_-\psi . \tag{1.12}$$

Evidently these are both symmetry currents only for massless fermions, since the mass term in eqn. (1.10) couples ψ_L to ψ_R, and the Lagrange density is not invariant under independent phase rotations. It is useful, therefore, to consider two different linear combinations of these currents, referred to as the vector and axial currents,

$$j^\mu = \bar\psi\gamma^\mu\psi \ , \qquad j^\mu_A = \bar\psi\gamma^\mu\Gamma\psi \ . \tag{1.13}$$

These two currents correspond to the two independent transformations

$$\psi \to e^{i\theta}\psi \ , \qquad \psi \to e^{i\omega\Gamma}\psi \tag{1.14}$$

respectively. The first conserved quantity is just fermion number, the second, which counts right minus left number, is called axial charge. The fact that $U(1)_A$ axial transformations are a symmetry of the kinetic operator is a consequence of the property

$$\{\Gamma, \not{D}\} = 0 \ . \tag{1.15}$$

Later we will see that on the lattice, even for massless fermions it is not possible to define a kinetic operator analogous to $\not{D}$ which anti-commutes with Γ, but that the above equation will have to be modified.

The reason why both $j^\mu_{L,R}$ currents are conserved for massless fermions is easy to see if we look at the equation of motion for the free massless fermion:

$$0 = i\not\partial\psi = -\begin{pmatrix} 0 & \partial_t - \partial_x \\ \partial_t + \partial_x & 0 \end{pmatrix}\psi \ , \tag{1.16}$$

which has plane wave solutions

$$\psi_R = e^{-ik(t-x)}\begin{pmatrix}1\\0\end{pmatrix} \qquad \psi_L = e^{-ik(t+x)}\begin{pmatrix}0\\1\end{pmatrix} \ , \tag{1.17}$$

with $P_+\psi_R = \psi_R$ and $P_-\psi_L = \psi_L$. We see that ψ_R corresponds to fermions moving at the speed of light to the right (positive x-direction) and ψ_L corresponds to particles moving to the left. Clearly, Lorentz boosts cannot change the number of either, which is therefore conserved quantities. Thus we expect to be able to write

$$\partial_\mu j^\mu_{L,R} = \partial_\mu j^\mu = \partial_\mu j^\mu_A = 0 \ . \tag{1.18}$$

We will show below, however, that this is not the case, and that quantum effects spoil some of these conservation laws through "anomalies".

Because ψ_L and ψ_R transform under independent irreducible representations of the Lorentz group, we can consider a theory with just one of them, such as

$$\mathcal{L} = \bar\psi_R i\not{D}\psi_R \ , \tag{1.19}$$

which, if gauged as in the above example, would be called an example of a "chiral gauge theory". Note that ψ_R is a 2-component spinor, where the lower component equals zero. So we could have written this as a 1-component fermion, but it is convenient often to write it as a Dirac spinor, with one component projected out by P_+. This theory looks like it should make sense because the gauge field appears to be coupled to a conserved current, but again, anomalies will spoil this assumption.

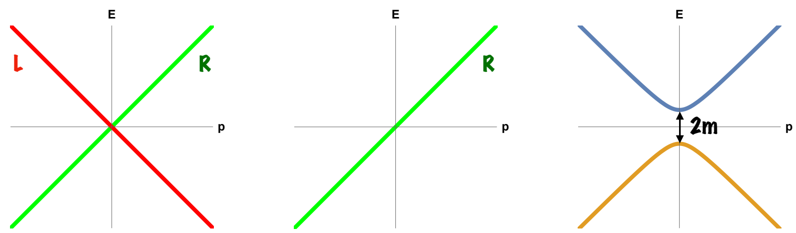

Fig. 1.1 *The spectrum of (i) a free massless Dirac fermion in $d = 1+1$, (ii) a free massless RH chiral fermion, (iii) a free massive Dirac fermion. In the first case both LH and RH fermion numbers are independently conserved — or equivalently, both fermion number and axial charge are conserved; in the second case there is a conserved RH fermion number; for the massive case there is only a conserved fermion number.*

1.2.3 The massive Dirac fermion

In contrast to the massless case, if the fermion is massive we can always boost between frames where a left-mover in one frame is a right-mover in the other, and so the number of either cannot be individually conserved. Figure 1.1 shows the spectrum for of free fermions for the cases we have discussed. If you take at the Lagrange density in eqn. (1.10) and perform the axial transformation in eqn. (1.14), you find

$$\mathcal{L} \to \bar{\psi}_L i\partial\!\!\!/ \psi_L + \bar{\psi}_R i\partial\!\!\!/ \psi_R - \left(m e^{2i\omega} \bar{\psi}_L \psi_R + \text{h.c.} \right) \ , \tag{1.20}$$

which is equivalent to rotating the mass term by a phase, $m \to m e^{2i\omega}$. (This shows that the phase of the fermion mass has no physical meaning in a theory where it is the only source of axial symmetry violation, since we can change the phase at will with a change of variables.) Noether's theorem then tells us that for the massive case we have

$$\partial_\mu j^\mu = 0, \qquad \partial_\mu j^\mu_A = 2im\bar{\psi}\Gamma\psi \ . \tag{1.21}$$

1.3 The $U(1)_A$ anomaly in $d = 1+1$

So far we have blithely assumed that symmetries of the Lagrange density imply symmetries of the quantum theory. However, one of the fascinating features of chiral symmetry is that sometimes it is not a symmetry of the quantum field theory even when it is a symmetry of the Lagrangian. In particular, Noether's theorem can be modified in a theory with an infinite number of degrees of freedom; the modification is called "an anomaly". Anomalies turn out to be very relevant both for phenomenology, and central for understanding the challenges for implementing chiral symmetry in lattice field theory. The reason anomalies affect chiral symmetries is that regularization requires a cut-off on the infinite number of modes above some mass scale, while chiral symmetry is incompatible with fermion masses[1].

[1] Dimensional regularization is not a loophole, since chiral symmetry cannot be analytically continued away from odd space dimensions.

A simple way to derive anomalies (and in some ways, overly simple) is to look at what happens to the ground state of a theory with a single flavor of massless Dirac fermion in $(1+1)$ dimensions in the presence of an electric field. Suppose one adiabatically turns on a constant positive electric field $E(t)$, then later turns it off; the equation of motion for the fermion is [2] $\frac{dp}{dt} = eE(t)$ and the total change in momentum is

$$\Delta p = e \int E(t)\, dt \ . \tag{1.22}$$

Thus the momenta of both left- and right-moving modes increase; if one starts in the ground state of the theory with filled Dirac sea, after the electric field has turned off, both the right-moving and left-moving sea levels have shifted to the right as in Fig. 1.2. The the final state differs from the original by the creation of particle- antiparticle pairs: right moving particles and left moving antiparticles. Thus while there is a fermion current in the final state, fermion number has not changed. This is what one would expect from conservation of the $U(1)$ current:

$$\partial_\mu j^\mu = 0 \ , \tag{1.23}$$

However, recall that right-moving and left-moving particles have positive and negative chirality respectively; therefore the final state in Fig. 1.2 has net axial charge, even though the initial state did not. This is peculiar, since the coupling of the electromagnetic field in the Lagrangian does not violate chirality. We can quantify the effect: if we place the system in a box of size L with periodic boundary conditions, momenta are quantized as $p_n = 2\pi n/L$. The change in axial charge is then

$$\Delta Q_A = 2\frac{\Delta p}{2\pi/L} = \frac{e}{\pi}\int d^2x\, E(t) = \frac{e}{2\pi}\int d^2x\, \epsilon_{\mu\nu}F^{\mu\nu} \ , \tag{1.24}$$

where I expressed the electric field in terms of the field strength F, where $F^{01} = -F^{10} = E$. This can be converted into the local equation using $\Delta Q_A = \int d^2x\, \partial_\mu j_A^\mu$, a modification of eqn. (1.18):

$$\partial_\mu j_A^\mu = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu} \ , \tag{1.25}$$

where in the above equation I have included the classical violation due to a mass term as well. The second term is the axial anomaly in $1+1$ dimensions.

Exercise 1.1 Use the above arguments to derive the anomaly that results if one gauges the axial current instead of the vector current.

[2] While in much of these lectures I will normalize gauge fields so that $D_\mu = \partial_\mu + iA_\mu$, in this section I need to put the gauge coupling back in. If you want to return to the nicer normalization, rescale the gauge field by e so that there is no coupling constant in the covariant derivative and a $1/e^2$ factor appears in front of the gauge action.

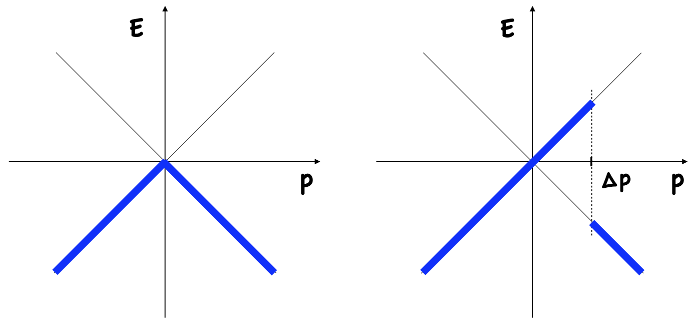

Fig. 1.2 *On the left: the ground state for a theory of a single massless Dirac fermion in* $(1+1)$ *dimensions; on the right: the theory after application of an adiabatic electric field with all states shifted to the right by* Δp*, given in eqn.* (1.22)*. Filled states are indicated by the heavier blue lines.*

So how did an electric field end up violating chiral charge? Note that this analysis relied on the Dirac sea being infinitely deep. If there had been a finite number of negative energy states, then they would have shifted to higher momentum, but there would have been no change in the axial charge. With an infinite number of degrees of freedom, though, one can have a “Hilbert Hotel”: the infinite hotel which can always accommodate another visitor, even when full, by moving each guest to the next room and thereby opening up a room for the newcomer. This should tell you that it will not be straightforward to represent chiral anomalies on the lattice: a lattice field theory approximates quantum field theory with a finite number of degrees of freedom — the lattice may be a big hotel, but it is quite conventional. In such a hotel there can be no anomaly, since there is no ambiguity about how many occupants it has.

This method of deriving the anomaly gives the correct answer, but is a bit too simplistic. For one thing, there is no need to assume that the gauge field must change adiabatically. For another, it doesn’t help one figure out what happens in the case where there is a fermion mass and a gap, where the correct answer is that one just adds together the anomalous and explicit symmetry violation, modifying eqn. (1.21) to read

$$\partial_\mu j^\mu_A = 2im\overline{\psi}\Gamma\psi + \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu} \ , \tag{1.26}$$

We can derive the anomaly in other ways, such as by computing the anomaly diagram Fig. 1.3, or by following Fujikawa (Fujikawa, 1979; Fujikawa, 1980) and carefully accounting for the Jacobian from the measure of the path integral when performing a chiral transformation. It is particularly instructive for our later discussion of lattice fermions to compute the anomaly in perturbation theory using Pauli-Villars regulators of mass M. Consider the fermion determinant obtained from the path integral in Euclidian spacetime:

$$\det\left(\not{D} + m\right) \ . \tag{1.27}$$

Here we assume hermitian γ-matrices satisfying $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$ and so $\not{D}$ is an anti-hermitian operator with unbounded imaginary eigenvalues. The determinant is formally given by

$$\det\left(\not{D} + m\right) = \prod_i (i\lambda_i + m) \ , \tag{1.28}$$

but this is ill defined. To better define it, we consider instead a regulated version,

$$\lim_{M\to\infty} \frac{\det\left(\not{D} + m\right)}{\det\left(\not{D} + M\right)} = \lim_{M\to\infty} \prod_i \frac{(i\lambda_i + m)}{(i\lambda_i + M)} \ . \tag{1.29}$$

Note that at fixed M, for $\lambda_i \gg M \gg m$ the contributions to the regulated determinant all go to factors of 1, so the effect of the regulator is to cancel off contributions from those states. Of course, in the end we take $M \to \infty$ and recover the theory we are interested in. The Feynman rules for this regulated determinant are simple: we just add a heavy "Pauli-Villars" fermion Φ with a Dirac action with mass M, but instead of having a factor of -1 from each closed loop, we get a $+1$ in order to obtain the inverse determinant, as we would get from a boson field. It is important that the Φ have all the same couplings as the fermion, including to external sources. As a result, we should consider the divergence of the regulated axial current

$$j^\mu_{A,\mathrm{reg}} = \overline{\psi}\gamma^\mu\Gamma\psi + \overline{\Phi}\gamma^\mu\Gamma\Phi \ , \tag{1.30}$$

where it follows from Noether's theorem that

$$\partial_\mu j^\mu_{A,\mathrm{reg}} = 2im\overline{\psi}\Gamma\psi + 2iM\overline{\Phi}\Gamma\Phi \ . \tag{1.31}$$

note that Φ contributes a new contribution to axial symmetry breaking proportional to M. However, *this* current does not have any additional anomalous divergence, because we have essentially removed all high λ_i states from consideration and have a "conventional hotel". Therefore, if we are to recover the anaomaly, it must come the Pauli-Villars contribution somehow. As we are interested in matrix elements of $j^\mu_{A,\mathrm{reg}}$ in a background gauge field between states that do not contain any Pauli-Villars particles, we need to evaluate the expectation value $\langle 2iM\overline{\Phi}\Gamma\Phi\rangle$ in a background gauge field and take the limit $M \to \infty$, in order to see if $\partial_\mu j^\mu_{A,\mathrm{reg}}$ picks up any anomalous contributions that do not decouple as we remove the cutoff $M \to \infty$.

To compute $\langle 2iM\overline{\Phi}\Gamma\Phi\rangle$ we need to consider all Feynman diagrams with a Pauli-Villars loop, and insertion of the $\overline{\Phi}\Gamma\Phi$ operator, and any number of external $U(1)$ gauge fields. By gauge invariance, a graph with n external photon lines will contribute n powers of the field strength tensor $F^{\mu\nu}$. For power counting, it is convenient that we normalize the gauge field so that the covariant derivative is $D_\mu = (\partial_\mu + iA_\mu)$; then the gauge field has mass dimension 1, and $F^{\mu\nu}$ has dimension 2. In $(1+1)$ dimensions $\langle 2iM\overline{\Phi}\Gamma\Phi\rangle$ has dimension 2, and so simple dimensional analysis implies that the graph with n photon lines must make a contribution proportional to $(F^{\mu\nu})^n/M^{2(n-1)}$. Therefore only the graph in Fig. 1.3 with one photon insertion can make a contribution that survives the $M \to \infty$ limit (the graph with zero photons vanishes). Calculation of this diagram yields the same result for the divergence of the regulated axial current as we found in eqn. (1.26); to show this is an exercise.

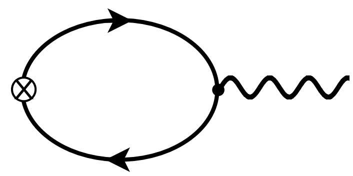

Fig. 1.3 *The anomaly diagram in 1+1 dimensions, with one Pauli-Villars loop and an insertion of $2iM\overline{\Phi}\Gamma\Phi$ at the X.*

Exercise 1.2 Compute the diagram in Fig. 1.3 using the conventional normalization of the gauge field $D_\mu = (\partial_\mu + ieA_\mu)$ and verify that $2iM\langle\overline{\Phi}\Gamma\Phi\rangle = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu}$ when $M \to \infty$. Note that you are looking for a contribution proportional to $i\epsilon_{\mu\nu}k^\nu$, where k^ν is the momentum of the external gauge boson and μ is its polarization.

Note that in this description of the anomaly we (i) effectively rendered the number of degrees of freedom finite by introducing the regulator; (ii) the regulator explicitly broke the chiral symmetry; (iii) as the regulator was removed, the symmetry breaking effects of the regulator never decoupled, indicating that the anomaly arises when the two vertices in Fig. 1.3 sit at the same spacetime point. While we used a Pauli-Villars regulator here, the use of a lattice regulator will have qualitatively similar features, with the inverse lattice spacing playing the role of the Pauli-Villars mass, and we can turn these observations around: A lattice theory will not correctly reproduce anomalous symmetry currents in the continuum limit, unless that symmetry is broken explicitly by the lattice regulator. This means we would be foolish to expect that a continuum field theory with anomalies could ever be represented by a lattice theory with exact chiral symmetry.

1.4 A lattice Hamiltonian for d=1+1 fermions

1.4.1 Doubling of as chiral fermion

Let's reconsider the theory of a single gauged right-handed fermion, as in eqn. (1.19). We now know that the current will have an anomalous divergence, which means that the theory is not gauge invariant! Such theories are known to be sick, so it should not be possible to give them a definition on the lattice. To keep things simple, I will first consider a latticized version of the Hamiltonian for the free theory, which only involves discretizing space, not time. The continuum Hamiltonian in our γ-matrix basis dfor the free fermion is simply

$$H = -i\partial_x \tag{1.32}$$

with naive discretization

$$H = -i\frac{1}{2a}\sum_n c_n^\dagger\left(c_{n+1} - c_{n-1}\right) \tag{1.33}$$

where a is the lattice spacing, and the $c_n, c_n^\dagger$ are fermionic ladder operators at site n:

$$\{c_m, c_n\} = 0 \ , \qquad \{c_m, c_n^\dagger\} = \delta_{mn} \ . \tag{1.34}$$

This theory has an exact $U(1)$ symmetry, which is fermion number:

$$Q = \sum_n c_n^\dagger c_n \ , \qquad [Q, H] = 0 \ . \tag{1.35}$$

This is the symmetry we can gauge. The single-particle eigenstates of H are

$$|p\rangle = \sum_n e^{iapn} c_n^\dagger |0\rangle \tag{1.36}$$

with energy eigenvalue

$$H|p\rangle = E_p|p\rangle \ , \qquad E_p = \frac{\sin ap}{a} \ , \qquad -\frac{\pi}{a} \le p \le \frac{\pi}{a} \ . \tag{1.37}$$

Note from the construction of the state $|p\rangle$ that shifting $p \to p + 2\pi a$ gives back the same state, so p-space is a circle and so taking the above range for p (the Brillouin zone) accounts for all states.

What is the continuum limit of this theory? Naively, the continuum limit $ap \to 0$ gives $E_p = p$, the desired continuum result corresponding to a single right-mover, shown in Fig. 1.1. However, if we rewrite $p = \pi - k$, then the $ak \to 0$ limit gives $E_k = -k$, a left-mover! We see that the continuum theory describes a single massless Dirac fermion in the continuum, with both right and left modes, not a single right-mover. That is because the dispersion relation $E_p = \sin ap/a$ crosses the p-axis in two places, $p = 0$ and $p = \pm\pi$, so there will always be two low energy modes, even as $a \to 0$. Furthermore, the exact $U(1)$ symmetry of the lattice is just fermion number in the continuum theory, so if we gauge it, the result looks like QED in $d = 1 + 1$, a sensible theory with a conserved gauge current, unlike the chiral gauge theory in eqn. (1.19).

Can we add a local interaction that will gap the spectrum at $p = \pm\pi/a$, to get rid of the continuum left-mover? Obviously not: the function E_p will be a continuous function of p and therefore must be periodic; a periodic function of p cannot cross the p-axis an odd number of times.

And what about the anomalous $U(1)_A$ global symmetry in $d = 1 + 1$ QED? How does the lattice model realize the anomaly? The answer is that the lattice theory does not have a second $U(1)$ symmetry that we can identify with $U(1)_A$ in the continuum...that would require rotating states with $p \sim 0$ with the opposite phase from states near $p \sim \pm\pi/a$, which is not a symmetry of H. Consider an eigenstate of H at finite lattice spacing which we will call the vacuum when $a \to 0$, with every negative energy state occupied and every positive energy state empty, as shown on the left in Fig. 1.4. Now consider what happens when we turn on an electric field for some time in the x direction: all states will move to the right (increasing p) and we end up with the state shown on the right in that figure. In the continuum theory that corresponds to a state that still has no net fermion number, but a nonzero axial charge. The lattice correctly reproduces the axial anomaly by having no exact axial symmetry.

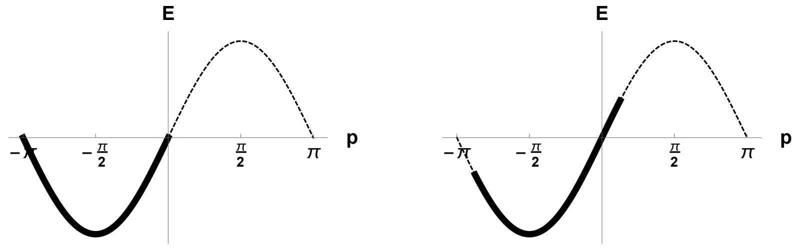

Fig. 1.4 *The ground state of a lattice Dirac fermion (left) and how it has evolved after application of an electric field in the x direction (right). The solid line denotes occupied 1-particle states, and the dashed line vacant states. In the continuum limit, it will appear as if an anomaly has violated axial charge, giving rise to right moving particles and left moving anti-particles.*

Note that it would have been wrong to consider the lattice theory to be similar to a continuum Dirac theory with a momentum cutoff at $p \sim 1/a$: such a theory would have an exact $U(1)_A$ symmetry (since a momentum cutoff is a regulator that does not violate axial rotations) but could not be gauged (since a momentum cutoff violates gauge symmetry). It behaves much more like a continuum theory with a Pauli-Villars regulator with mass $M \sim 1/a$, a regulator that preserves gauge symmetry while breaking axial symmetry.

It seems we should be happy that the lattice was "smart enough" to not give us a sick theory in the continuum, a gauge theory whose gauge symmetry was broken by an anomaly. However there are problems we can see even with very simple variants of this model. The first has to do with the role chiral symmetry plays in the Standard Model, protecting fermion masses from additive mass renormalization. The second has to do with creating a lattice regulator for chiral gauge theories that are not sick....like the Standard Model itself.

1.4.2 Problems for chiral gauge theories

It is possible to construct a theory with several chiral fermions that has an anomaly-free $U(1)$ symmetry that can be gauged. If the fermion representation is such that one cannot write down gauge-invariant mass terms for the fermions, then the theory is called a chiral gauge theory. From our discussion of the anomaly we see that an example of a chiral $U(1)$ gauge theory in $d = 1 + 1$ dimensions is the $3 - 4 - 5$ model which consists of three fermions, right-movers $\psi_{3,4}$ with electric charges 3 and 4, and a left-mover χ_5 with charge 5,

$$\mathcal{L} = \bar{\psi}_3 i\not{D} P_+ \psi_3 + \bar{\psi}_4 i\not{D} P_+ \psi_4 + \bar{\chi}_5 i\not{D} P_- \chi_5 \ . \tag{1.38}$$

Note that unlike in QED, it is impossible to write a gauge invariant fermion mass. That requires both a left- and a right-moving particle...however the two right movers we have in the theory have charges 3 and 4, while the only left-mover has charge 5, and neither $\bar{\chi}_5 P_+ \psi_3$ nor $\bar{\chi}_5 P_+ \psi_4$ operators are invariant under the gauge symmetry.

It appears that there are three conserved currents in this theory, one for each type of fermion number:

$$j_3^\mu = \bar{\psi}_3 \gamma^\mu P_+ \psi_3 \ , \qquad j_4^\mu = \bar{\psi}_4 \gamma^\mu P_+ \psi_4 \ , \qquad j_5^\mu = \bar{\chi}_5 \gamma^\mu P_- \chi_5 \ , \tag{1.39}$$

Note that because $P_\pm = (1 \pm \Gamma)/2$ these current can be written as a sum or difference of a vector and an axial current, with a factor of $1/2$. We have seen that the vector currents are conserved in the presence of a background electric field, while the axial currents are anomalous, so that we get:

$$\partial_\mu j_3^\mu = 3 \frac{e}{4\pi} \epsilon_{\mu\nu} F^{\mu\nu} \ , \qquad \partial_\mu j_4^\mu = 4 \frac{e}{4\pi} \epsilon_{\mu\nu} F^{\mu\nu} \ , \qquad \partial_\mu j_5^\mu = -5 \frac{e}{4\pi} \epsilon_{\mu\nu} F^{\mu\nu} \ . \tag{1.40}$$

However, the current that the gauge field couples to is divergenceless, by construction:

$$J^\mu = 3 e j_3^\mu + 4 e j_4^\mu + 5 e j_5^\mu \ , \qquad \partial_\mu J^\mu = \frac{e^2}{4\pi} \left(3^2 + 4^2 - 5^2\right) \epsilon_{\mu\nu} F^{\mu\nu} = 0 \ . \tag{1.41}$$

Therefore we do not expect this to be a sick theory and would like to study it on a computer.

What happens when we try to construct this theory on the lattice, using a copy of H from eqn. (1.33) for each fermion and adding gauge fields with appropriate charges? We get a sensible theory in the continuum limit, but not the one we wanted: a theory of three Dirac fermions with Lagrangian

$$\mathcal{L} = \bar{\psi}_3 i \not{D} \psi_3 + \bar{\psi}_4 i \not{D} \psi_4 + \bar{\chi}_5 i \not{D} \chi_5 \ . \tag{1.42}$$

Note that this theory has no chiral projection operators in the kinetic terms, and that it is possible to write down gauge invariant Dirac mass terms for each field...this is not a chiral gauge theory. Obviously there are an infinite number of healthy chiral gauge theories and we do not seem to have a way to regulate them on the lattice. This is not just an academic problem because the Standard Model is a chiral gauge theory in $d = 3 + 1$. There have been ideas on how to construct lattice gauge theories which I will discuss later, but it remains an open problem.

1.5 Doubling of a Dirac fermion and the need for fine tuning

Some operators in a Lagrangian suffer from additive renormalizations, such as the unit operator (cosmological constant) and scalar mass terms, such as the Higgs mass in the Standard Model, $|H|^2$. Therefore, the mass scales associated with such operators will naturally be somewhere near the UV cutoff of the theory, unless the bare couplings of the theory are fine-tuned to cancel radiative corrections. Such fine tuning problems have obsessed particle theorists since the work of Wilson and 't Hooft on renormalization and naturalness in the 1970s. However, such intemperate behavior will not occur for operators which violate a symmetry respected by the rest of the theory: if the bare couplings for such operators were set to zero, the symmetry would ensure they could not be generated radiatively in perturbation theory. Fermion mass operators generally fall into this benign category.

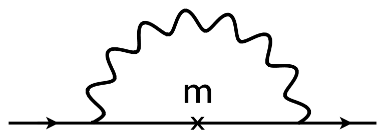

Fig. 1.5 *One-loop renormalization of the electron mass in QED due to photon exchange. A mass operator flips chirality, while gauge interactions do not. A contribution to the electron mass requires an odd number of chirality flips, and so there has to be at least one insertion of the electron mass in the diagram: the electron mass is multiplicatively renormalized. A scalar interaction flips chirality when the scalar is emitted, and flips it back when the scalar is absorbed, so replacing the photon with a scalar in the above graph again requires a fermion mass insertion to contribute to mass renormalization.*

Consider the following toy model: QED with a charge-neutral complex scalar field coupled to the electron:

$$\mathcal{L}\overline{\Psi}(i\not{D} - m)\Psi + |\partial\phi|^2 - \mu^2|\phi|^2 - g|\phi|^4 + y\left(\overline{\Psi}_R\phi\Psi_L + \overline{\Psi}_L\phi^*\Psi_R\right) \ . \tag{1.43}$$

Note that in the limit $m \to 0$ this Lagrangian respects a chiral symmetry $\Psi \to e^{i\alpha\gamma_5}\Psi$, $\phi \to e^{-2i\alpha}\phi$. The symmetry ensures that if $m = 0$, a mass term for the fermion would not be generated radiatively in perturbation theory. With $m \neq 0$, this means that any renormalization of m must be proportional to m itself (i.e. m is "multiplicatively renormalized"). This is evident if one traces chirality through the Feynman diagrams; see Fig. 1.5. Multiplicative renormalization implies that the fermion mass can at most depend logarithmically on the cutoff (by dimensional analysis): $\delta m \sim (\alpha/4\pi)m \ln m/\Lambda$. Try plugging in some numbers here: with $\alpha = 1/137$ and $\Lambda = M_{\text{Planck}} = 10^{19}$ GeV we get a radiative correction to the electron mass δm which is about 3% of the electron mass, not a shift that requires fine tuning.

In contrast, the scalar mass operator $|\phi|^2$ does not violate any symmetry and therefore suffers from additive renormalizations, such as through the graph in Fig. 1.6. By dimensional analysis, the scalar mass operator can have a coefficient that scales quadratically with the cutoff: $\delta\mu^2 \sim (y^2/16\pi^2)\Lambda^2$. This is called an additive renormalization, since $\delta\mu^2$ is not proportional to μ^2. It is only possible in general to have a scalar in the spectrum of this theory with mass much lighter than $y\Lambda/4\pi$ if the bare couplings are finely tuned to cause large radiative corrections to cancel. For $\Lambda = M_{\text{Planck}}$, we require a bare mass term to cancel this one-loop radiative contribution to one part in 10^{30} to get a 100 GeV Higgs. When referring to the Higgs mass in the Standard Model, this is called the hierarchy problem.

Let's return to the problem of lattice Hamiltonians for fermions in $d = 1 + 1$. Suppose we want a lattice model to describe a massive Dirac fermion in $d = 1 + 1$. The continuum Hamiltonian is

$$H = \gamma^0(i\gamma^1\partial_x + m) \ , \tag{1.44}$$

and so we again naively write down a Hamiltonian for a free fermion, this time of the form

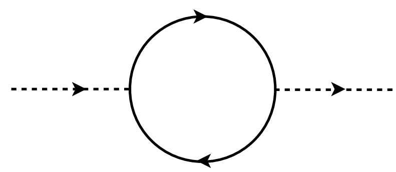

Fig. 1.6 *One-loop additive renormalization of the scalar mass due to a quadratically divergent fermion loop.*

$$H = \sum_n \bar{\psi}_n \left[-\frac{i}{2a} \sigma_3 \left(\psi_{n+1} - \psi_n \right) + m\psi_n \right] \tag{1.45}$$

where ψ_n is a 2-component fermion ladder operator at site n. The eigenvalues of H are the eigenvalues of the matrix

$$\left[\sigma_3 \frac{\sin ap}{a} + m\sigma_1 \right] , \tag{1.46}$$

or

$$E_p = \pm \sqrt{\left(\frac{\sin ap}{a} \right)^2 + m^2} \tag{1.47}$$

Expanding about $a = 0$ we get $E_p \simeq \pm\sqrt{p^2 + m^2}$ for $p = O(1)$, while writing $p = -\pi/a + k$ and expanding about $a = 0$ we get $E_k = \pm\sqrt{k^2 + m^2}$... so we find *two* Dirac fermions in the continuum. This is the same doubling of the spectrum we saw when we tried to construct a lattice model for just a right moving 1-component fermion. In thaty case the doubling kep us from creating a sick theory....here it is just annoying, because a single massive Dirac fermion is a perfectly fine theory, and the one we want.

Can we make the mode near $p \sim \pi/a$ very heavy and get rid of it? Yes, now the spectrum never crosses the p axis and there is no reason a periodic function could exhibit a small gap $\sim m$ at $p = 0$ and a large gap $\sim 1/a$ at $p = \pi/a$. We can do that by adding to H a term of the form

$$-a\bar{\psi}\partial_x^2\psi \tag{1.48}$$

which looks like a mass term that will only affect modes with wavenumber $p \sim 1/a$ in the $a \to 0$ limit. The lattice realization of this operator is

$$H_w = -\frac{ra}{2} \sum_n \frac{1}{a^2} \bar{\psi}_n \left(\psi_{n+1} - 2\psi_n + \psi_{n-1} \right) , \tag{1.49}$$

where the $r/2$ factor is a parameter we can adjust. Now the energy is given by the eigenvalues of

$$\left[-\frac{i}{2a} \sigma_3 2i \sin ap + \left(m + r \left(1 - \cos ap \right) / a^2 \right) \sigma_1 \right] , \tag{1.50}$$

or

$$E_p = \pm\sqrt{\left(\frac{\sin ap}{a}\right)^2 + \left(m + r\,(1-\cos ap)/a^2\right)^2} \tag{1.51}$$

Now if we expand about $a = 0$ for $p = O(1)$ we get $E_p = \pm\sqrt{p^2 + m^2}$ as before, but for $p = -\pi/a + k$ we get $E_k = 2r/a + O(1)$. This is Wilson's solution for getting rid of the unwanted "doubler".

There has been a cost, however. Now we have two terms violating chiral symmetry in the Lagrangian: the $m\bar\psi\psi$ term, and the $a\bar\psi\nabla^2\psi$ term. Thus when we add interactions (e.g. by gauging fermion number) the Wilson interaction term $\bar\psi D^2\psi$ will renormalize the mass term through gauge boson loops, and we expect radiative corrections of size

$$\delta m \sim r\frac{\alpha}{a} \ , \tag{1.52}$$

which looks very similar to the Higgs mass fine tuning problem: the bare mass will have to be fine tuned to one part in $\sim ma/\alpha$ in order to obtain a physical mass m, which becomes harder to do the smaller a becomes. In getting rid of our doubler fermion we lost our approximate chiral symmetry, and have to fine tune parameters to recover it in the continuum limit. And the problem seemed to arise from the need for the lattice to properly account for anomalies.

Exercise 1.3 Compute the spectrum of the Wilson-Dirac Hamiltonian, $H + H_w$ and plot it for various values of m, r with $a = 1$.

1.6 What we have found

- Chiral symmetries appear in theories of massless fermions
- Chiral symmetries can be broken in the continuum by quantum effects called anomalies
- Anomalies cannot exist in a finite system; lattice theories of fermions break potentially anomalous symmetries explicitly
- ...or else double the fermions so that the exact lattice symmetries are vector symmetries in the continuum.
- Lattice doubling of the spectrum makes it problematic to construct sensible chiral gauge theories in the continuum limit
- Eliminating doubling in vector-like gauge theories eliminates global chiral symmetry and requires fine tuningin order to find light fermions in the continuum limit.

What we would like is a lattice fermion formulation which at the very least correctly accounts for anomalies, while protecting fermion masses from additive renormalization, like in the continuum. Better: we would like to know how to construct lattice models for chiral gauge theories.

References

Fujikawa, Kazuo (1979). *Phys. Rev. Lett.*, **42**, 1195.
Fujikawa, Kazuo (1980). *Phys. Rev.*, **D21**, 2848.