Chiral Anomalies and the Nielsen - Ninomiya No-Go Theorem

N. Kadakia[1]

[1] *Department of Physics, University of California at San Diego, La Jolla, CA 92093*

In this letter, a brief introduction to the Nielsen and Ninomiya no-go theorem for putting chiral fermions on a lattice is presented. A simple outline of the phenonenon of fermion doubling is followed by an explanation of its discrepancy with the axial anomaly. The assumptions and consequences of the Nielsen and Ninomiya theorem follow, with a brief explanation of its proof in 1+1 dimensions. The letter concludes with a discussion of a few attempts to evade the fermion doubling problem through a violation of the assumptions of the no-go theorem.

INTRODUCTION

It is well known that scattering, decay, and production amplitudes can be calculated for weakly coupled theories using perturbative techniques around a known solution [1]. Such solutions can be a calculational headache; yet once a method of renormalizing loop diagrams has been decided once and for all, the procedure is entirely systematic. Such a procedure has remarkable success in quantum electrodynamics (QED), where the coupling constant is known to be small for small momenta. But QED is an abelian gauge theory; in non-abelian guage theories, such as quantum chromodynamics (QCD), the situation is reversed. The coupling constant is small for small wavelengths, while for large wavelengths, the perturbative approach is not fruitful [1].

Wilson proposed a method for calculating QCD amplitudes in the regime of strong coupling [2]. The aim was a mechanism which would explain what appeared to be the curious fact that quarks, whose existence explained well the spectrum of hadrons, had never been found experimentally: if they were to exist, they must be strongly confined. His method involved defining the quantum theory on a discrete hypercubic lattice. The lattice model uses an action that inherently preserves gauge symmetry for an arbitrary nonzero lattice spacing, and the lattice spacing serves as a natural momentum cutoff $\Lambda \approx 1/a$. In this method, it is found that quark confinement is unambiguous [3].

The price one pays is that since the continuum has been replaced by a discrete matrix, Lorentz invariance has been sacrificed. In fact, this is not so big of a problem; it can be restored in the limit $a \to 0$. The more severe issues is that of successfully putting Fermions on the lattice.

FERMION DOUBLING

The naive procedure for defining a quantum field theory can be formulated by substituting integrals over spacetime with sums over lattice sites and by substituting derivatives with differences between field values at neighboring sites. Such a procedure for scalar bosons gives the following two point function [4]:

$$< \hat{\phi}_n \hat{\phi}_m >= a^2 \int_{-\pi/a}^{\pi/a} \frac{d^4k}{(2\pi)^4} \frac{e^{ik(n-m)}}{(2/a)^2 \sum_\mu \sin^2(k_\mu a/2) + M^2}$$

The discretization of position space restricts possible k-vectors in momentum space to a single Brillouin zone (BZ), say $-\pi/a$ to π/a. Since the largest values in the integral occur where the denominator is small, the integral is dominated by values where $k_\mu a/2 = n\pi$. This only occurs for one point in the first BZ, $k_\mu \to 0$. Finally taking the $a \to 0$ limit and holding M fixed reproduces the continuum two point correlation function [4].

The crucial difference between this case and the Fermion case is that due to the Dirac action being linear in the derivatives, one finds that the argument of the sine function in the two point function is not $k_\mu a/2$ but $k_\mu a$. Since the integral still spans a single Brillouin zone, the propagator has a well-defined limit both in the continuum $p_\mu = 0$ and at the corners of the Brillouin zone at which p_μ equals $\pi/2$ or 0. At all sixteen of these points, the continuum limit produces an identical propagator: a theory that should have one fermion has been doubled in each dimension to produce 2^d fermions. This is a problem because now the propagator receives contributions from all fermions on the lattice but really do not exist in the original continuum theory. Such a phenomenon would imply, for example, that if one was to define a left-handed neutrino and electron on a lattice, the existence of their right-handed counterparts is unavoidable, contradicting observation [4].

The fermion doubling phenomenon is intimately tied to the axial anomaly, as is discussed next.

RELATION TO AXIAL ANOMALY

The axial anomaly can be demonstrated by calculating, for example, the chiral currents in a $\pi^0 \to \gamma\gamma$ scattering process to one loop order and showing that they are not conserved[1]. The anomaly is in fact true to all orders, and can be shown without explicitly using perturbation theory, a method known as the *Fujikawa method*, or the *Adler-Bardeen Theorem* [5] [6]. In words, the axial anomaly implies that the number of right-handed Fermions minus left-handed ones is not constant. Noether's theorem therefore implies that the axial current is not divergenceless; for example for a $U(1)$ gauge theory, $\partial_\mu J^\mu_5 = -\frac{1}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$ [6]. Lattice regularization provides extra fermions of the correct handedness needed to cancel this anomaly. It seems therefore that if one wishes to prevent fermion doubling, the chilal anomaly cannot be preserved. This statement will be made precise in the next section.

NIELSEN AND NINOMIYA THEOREM

The incongruence of these conditions is made explicit in a theorem proven first by Nielsen and Ninomiya in 1981 [7] [8]. The theorem states the following. Under the following assumptions, there appear an equal number of right- and left-handed particles of given quantum numbers in a regularized theory with a chirally invariant action: a) U(1) gauge invariance, b) quadratic Hamiltonian, and c) the correct axial anomaly. The general argument relies on the periodicity of the dispersion relation, $\omega(k)$. For example, in 1+1 dimensions, the dispersion relation is a simple $\omega(k)$ graph. The Fermi energy can be taken as $\omega = 0$; low energy excitations correspond therefore to points about the zeros, with velocities given by $\frac{\partial\omega}{\partial k}$. In one spatial dimension, right-handed particles have positive velocities and left-handed negative velocities. Because a periodic function must, in one cycle, go up through a zero as many times as it goes down ($\frac{\partial\omega}{\partial k} \neq 0$ at the Fermi energy), there are equal numbers of right- and left-handed particles. This follows because the allowed momentum vectors are restricted to one BZ – they have the topology of a circle.

In 3+1 dimensions, the space of **k**-vectors have the topology of a torus; the argument is generalized to deal with curves along which the Fermi energy is zero, and these curves must pass through a given level surface equal numbers of times if they are closed (closed can mean closed within a BZ, or passing through identical points on different surfaces of a BZ). Thus there are equal numbers of opposite-handed fermions [7].

Due to the restrictions implied by the no-go theorem, preventing fermion doubling in the continuum limit of a lattice gauge theory relies on on violating its assumptions. There may be some ways to prevent fermion doubling.

In one scheme, Wilson adds a term to the naive action that explicitly breaks chiral invariance of the Lagrangian [4] [7]. Recall that the lattice action was defined by "naively" extending the continuum action with appropriate re-definitions of integrals and derivatives – a more sophisticated method may prevent its pitfalls. Also recall that while the typical QCD action is chirally invariant, the theory is not: there are indeed anomalies. The result of this explicit breaking of the chiral invariance in the action is that the continuum limit is now dominated by the fermion at the origin only and doubling is prevented. Yet the price one pays is the loss of chirally invariance, which can be reinstated after taking the continuum limit, but only by fine-tuning the mass parameter.

One may also confront the fermion doubling problem by recalling that the issue arose from the fact that the lattice-defined propagator is dominated not only by values of k near the origin, but also at the edges of the BZ. If the Brillouin zone can be reduced to half of its size by doubling the effective lattice spacing, the BZ-edge fermions will not remain in the continuum limit. This method is called "staggered fermions," and involves placing different fermions (or, equivalaently, different combinations of flavors) on adjacent lattice sites so the periodicity of the lattice of spacing a is actually $2a$. The advantage of this method over the Wilson method is that, to a large degree, chiral invariance can be retained. The price one pays is that, due to the technical nature of exactly doubling the lattice spacing, the number of quark flavors must be $2^{d/2}$, whereas in Wilson's formulation it is arbitrary. This scheme is therefore more suited for investigations of chiral symmetry breaking in QCD [4].

Other schemes exist, but as one expects, the advantages do not come without drawbacks. One should of course choose a scheme that retains the features most appropriate to the phenomena worth exploring.

CONCLUSION

While lattice gauge theory is most useful in exploring phenomena particularly in regimes of non-perturbative, strong coupling, one must define the discretized version of the theory with care. Under the typical assumptions of translation invariance and certain restrictions on the form of the action, the Nielsen and Ninomiya no-go thoerem guarantees that one cannot retain chiral invariance in a theory without sacrificing proper fermion counting in the continuum. The theorem is at its heart only a statement about the *topology* of momentum space, and is therefore quite general in its applicability to lattice gauge theories. To sidestep it means certain assumptions, such as chiral invariance of the action, must be violated.

Acknowledgements

The author would like to thank John McGreevy for useful discussions.

[1] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, 1995).

[2] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).

[3] L. H. Karsten and J. Smit, Nucl. Phys. B **183**, 103 (1981).
[4] H. J. Rothe, *Lattice Gauge Theories – An Introduction* (World Scientific, fourth ed., 2012).
[5] K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979).
[6] M. Srednicki, *Quantum Field Theory* (Cambridge Univ. Press, 1st ed., 2007).
[7] H. B. Nielsen and M. Ninomiya, Nucl. Phys. B **193**, 173 (1981).
[8] H. B. Nielsen and M. Ninomiya, Phys. Lett. B **105**, 219 (1981).