Chiral Fermions on the Lattice

Rajamani Narayanan and Herbert Neuberger

Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855-0849
(Received 12 August 1993)

An expression for the lattice effective action induced by chiral fermions in any even dimensions in terms of an overlap of two states is shown to have promising properties in two dimensions: The correct Abelian anomaly is reproduced and instantons are suppressed.

PACS numbers: 11.15.Ha, 11.30.Rd

The fundamental building blocks of matter, as known today, are chiral. If there are no anomalies, relativistic field theory can consistently describe chiral matter interacting via gauge boson exchange to any order in the Feynman diagram expansion. If matter were vectorlike and all couplings were asymptotically free, work based on lattice formulations shows that one can sum all diagrams unambiguously. On the other hand, difficulties encountered in regularizing chiral fermions (even in asymptotically free situations) have raised doubts about the existence of chiral gauge theories at the nonperturbative level. This is a central problem of particle-physics-oriented field theory and its resolution may open up new possibilities in the endeavor to unify more of the observed forces in nature.

We focus on the lattice formalism because one can then regularize the bosonic degrees of freedom in a gauge invariant manner. The difficulties arising upon the addition of chiral lattice fermions are well known [1]. If we do not break gauge invariance in the fermionic sector, uninvited "doublers" are produced, making the particle content vectorlike and protecting the theory from ever having any anomalies. If we break gauge invariance, recovering it in the continuum limit for anomaly free representations is, at best, a matter of delicate fine tuning that would be very difficult to implement [2].

The modern approach [3] to the quantization of chiral gauge theories is a two-step process: (a) regularization of the chiral fermionic determinant in an arbitrary smooth gauge background and (b) subsequent integration over all gauge fields. If there are no anomalies, the answer to the first step is gauge invariant and the subsequent functional integration over the gauge fields can be attempted. The problem of regularizing chiral fermions on the lattice is encountered at the first step and we concern ourselves mainly with this step in this Letter. However, our formulation accepts arbitrary lattice gauge backgrounds. Some readers may object to the strategy which led us to the two steps in the first place, because fermionic and bosonic fields are quantized sequentially rather than simultaneously. We cannot overrule this objection; however, the progress in our understanding of anomalies in the mid 1980's [3] and the discovery of the important instanton effects in the 1970's [4] were made within this two-step framework.

An easy way of presenting a new approach due to Kaplan [5] is as follows [6]. We start with a generic vectorlike gauge theory that is easy to regularize:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_g(A) ,$$
$$\mathcal{L}_1 = i\bar{\psi}(\not{\partial} - ig\not{A})\psi + \bar{\psi}(MP_+ + M^\dagger P_-)\psi . \tag{1}$$

$P_\pm = (1 \pm \gamma_5)/2$ and $A_\mu = A_\mu^a T^a$ where T^a's are generators in the appropriate representation of the gauge group. $\mathcal{L}_g(A)$ is the pure gauge part, whose explicit form is irrelevant to us here. To achieve chiral symmetries, we need massless fermions. M is a square matrix and therefore M and $M^\dagger$ will have equal rank, implying that massless fermions will occur in pairs of opposite chirality. We can overcome this only if we make M infinite dimensional and endow M with a nonzero analytical index. The infinite-dimensional "flavor" space, whose sole role is to deal with the specific problem of chiral massless fermions, adds an additional source of indeterminacy to our problem. We can therefore regularize the ultraviolet infinities in any way we wish (since they are "released" from the "duty" to also induce anomalies), and we choose the lattice. The specific problems having to do with chirality are relegated to the step when we try to control the infinity of the internal (flavor) space.

The simplest way to realize a nontrivial index for M is to choose the internal space as the real line and choose $M = \partial_s + f(s)$ with $f(s)$ asymptotically approaching constants of different signs when $s \to \pm \infty$. This produces the wall system discussed by Callan and Harvey [7]. s space can also be discretized and the entire Callan-Harvey analysis can be transferred to the lattice. This observation is due to Kaplan [5] who also showed that on the lattice one may work with a defect strictly local in s.

We deal with the additional infinity by viewing internal space as the Euclidean time axis of an s-independent Hamiltonian $\hat{H}$ that is subjected to a mass shock at $s = 0$ [8]. The chiral determinant is replaced by the overlap of two ground states, corresponding to $\hat{H}_\pm$ for $s > 0$ and $s < 0$, respectively. A straightforward derivation [8] gives

$$\hat{H}_\pm = -\hat{a}^\dagger \mathbf{H}_\pm \hat{a},$$

$$e^{\mathbf{H}_\pm} = \begin{pmatrix} 1/\mathbf{B}^\pm & (1/\mathbf{B}^\pm)\mathbf{C} \\ \mathbf{C}^\dagger(1/\mathbf{B}^\pm) & \mathbf{C}^\dagger(1/\mathbf{B}^\pm)\mathbf{C} + \mathbf{B}^\pm \end{pmatrix} . \tag{2}$$

0031-9007/93/71(20)/3251(4)$06.00   3251
© 1993 The American Physical Society

$\hat{a}_{nAi}$ and $\hat{a}_{nAi}^{\dagger}$ are fermion operators satisfying canonical anticommutation relations: $\{\hat{a}_{nAi}, \hat{a}_{mBj}^{\dagger}\} = \delta_{nm}\delta_{AB}\delta_{ij}$, $\{\hat{a}_{nAi}, \hat{a}_{mBj}\} = 0$, $\{\hat{a}_{nAi}^{\dagger}, \hat{a}_{mBj}^{\dagger}\} = 0$. The indices A and B run over $2^{d/2}$ spinor components and i and j over group indices. n is the d-dimensional space index. $\mathbf{B}^{\pm}$ and $\mathbf{C}$ are matrices depending on the gauge fields. Explicitly,

$$\mathbf{B}_{n\bar{a}i,m\beta j}^{\pm} = (d + 1 \mp m)\delta_{nm}\delta_{\alpha\beta}\delta_{ij}$$
$$- \frac{1}{2}\delta_{\alpha\beta}\sum_{\mu}[\delta_{m,n+\hat{\mu}}U_{n,\mu}^{ij} + \delta_{n,m+\hat{\mu}}U_{m,\mu}^{ji*}] ,$$

$$\mathbf{C}_{n\alpha i,m\beta j} = \frac{1}{2}\sum_{\mu}[\delta_{m,n+\hat{\mu}}U_{n,\mu}^{ij} - \delta_{n,m+\hat{\mu}}U_{m,\mu}^{ji*}]\sigma_{\mu}^{\alpha\beta} . \qquad (3)$$

$U_{n,\mu}$ is a matrix in the appropriate representation of the gauge group and is associated with the gauge field on the link at site n in the positive μ direction. We have used the following representation for the γ matrices:

$$\gamma_{\mu} = \begin{bmatrix} 0 & \sigma_{\mu} \\ \sigma_{\mu}^{\dagger} & 0 \end{bmatrix}, \quad \gamma_{5} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

where $\sigma_0 = i$ and σ_j; $j = 1, \ldots, d-1$ are the $2^{d/2-1} \times 2^{d/2-1}$ generalization of Pauli matrices to d (even) dimensions. $0 < m < 1$ is a mass parameter. The two Hamiltonians in (2) are Hermitian due to the hermiticity of $\mathbf{B}^{\pm}$ in (3). Both $\mathbf{B}^{\pm}$ are positive definite for all values of the gauge field and therefore $\mathbf{H}_{\pm}$ are well defined for all gauge fields. One can compute the determinants of $e^{\mathbf{H}_{\pm}}$ and prove that both $\mathbf{H}_{\pm}$ are traceless [8]. This indicates the presence of both positive and negative eigenvalues. For any choice of gauge fields, $\mathbf{H}_{-}$ has no eigenvalues in the interval $I = (-\ln(1+m), \ln(1+m))$. The proof of the above statement is as follows. Let $\binom{u}{v}$ be an eigenvector of $e^{\mathbf{H}_{-}}$ with eigenvalue λ. From (2), we have $(\lambda\mathbf{B}^{-} - 1)u = \mathbf{C}v$ and $[1 - (1/\lambda)\mathbf{B}^{-}]v = \mathbf{C}^{\dagger}u$, implying $u^{\dagger}(\lambda\mathbf{B}^{-} - 1)u + v^{\dagger}[(1/\lambda)\mathbf{B}^{-} - 1]v = 0$. Since $\|\mathbf{B}^{-}\| \geq 1 + m$, we have the desired result, namely, $\lambda \notin I$. Therefore, the eigenvalues of $\mathbf{H}_{-}$ as a function of gauge fields can never cross zero and $\mathbf{H}_{-}$ has an equal number of positive and negative eigenvalues for all gauge fields. The same is not true of $\mathbf{H}_{+}$.

The gauge action induced by the integration over fermions is then given by the following formula:

$$e^{S_i(U)} = \frac{_U\langle 0 - |0+\rangle_U}{_1\langle 0 - |0+\rangle_1} e^{i[\Phi_+(U) - \Phi_-(U)]} . \qquad (4)$$

The $|0\pm\rangle_U$ are ground states of $\hat{H}_{\pm}(U)$ and the phases $\Phi_{\pm}(U)$ are defined by a convention borrowed from Wigner-Brillouin perturbation theory:

$$e^{i\Phi_{\pm}(U)} = \frac{_U\langle 0 \pm |0 \pm\rangle_1}{|_U\langle 0 \pm |0 \pm\rangle_1|} . \qquad (5)$$

For all gauge fields $_U\langle 0 - |0-\rangle_1$ is nonzero. But $_U\langle 0 + |0+\rangle_1$ can be zero for some gauge fields because the numbers of positive eigenvalues of $\mathbf{H}_+(U)$ and $\mathbf{H}_+(1)$ can be different. In such an instance, $_U\langle 0 - |0+\rangle_U$ also

vanishes for the same reason and a definition of the phase, $\Phi_+(U)$, for such gauge fields is not needed.

Equation (4) completely defines the real part of $S_i(U)$. Equation (5) is one choice to fix the phase ambiguity in (4). It is only in (5) that gauge invariance is broken; any other definition of $\Phi_{\pm}(U)$ which differs from the above by a local functional of $U - 1$ (i.e., when expanded in $U - 1$ the coefficients are analytic at the origin of lattice momentum space) will be acceptable. The real part of S_i is naturally gauge invariant; another way to enforce the gauge invariance of $\text{Re}(S_i)$ is proposed in [9].

A continuum variant of the overlap formula was shown recently to give the correct anomaly in two Euclidean dimensions with a U(1) gauge group [8]. The calculation is tedious because there are no straightforward Feynman-diagram techniques and one has to use some form of Hamiltonian perturbation theory. To do the same calculation on the lattice would be even more tedious due to the extra nonlinearity in the expressions for $e^{\mathbf{H}_{\pm}}$. In addition, such a calculation would not tell us whether nonperturbatively S_i behaves correctly at the semiclassical level. Therefore we proceed numerically.

We set $d = 2$ and work on a square lattice of size $L \times L$. We use antiperiodic boundary conditions for the fermions (relative to the gauge backgrounds introduced below) to avoid divergences associated with constant spinors. For a given gauge configuration the matrices $e^{\mathbf{H}_{-\pm}}$ are constructed in momentum basis and diagonalized by the Jacobi method. The corresponding one particle states are filled to construct the vacua entering the overlaps. The overlaps in (4) and (5) are evaluated by computing the appropriate determinants.

For sufficiently small $|U - 1|$ the gauge fields are perturbative and each of the $\mathbf{H}_{\pm}$ has L^2 positive eigenvalues. The determinants are nonvanishing and $S_i(U)$ in (4) is obtained. To check the anomaly we choose

$$U_{n,\mu}^{(k)} = \exp\left[i\frac{A_{\mu}(\phi)}{L}\cos\left(\frac{2\pi n \cdot k}{L} + \frac{\pi k_{\mu}}{L}\right)\right]$$

as the gauge configuration on the lattice. $k_{\mu} \in Z$ is kept fixed as $L \to \infty$. $\{U_{n,\mu}^{(k)}\}$ represents a standing plane wave with a fixed number of nodes on the torus in the continuum limit. $A_{\mu}(\phi) = A_{\mu} + 2\phi\sin(\pi k_{\mu}/L)$ represents a family of gauge equivalent connections. When ϕ is varied, the imaginary part of $S_i(U)$ should change and the change in the continuum limit is

$$\lim_{L \to \infty} -\frac{i}{\pi}\frac{\partial S(A_{\mu}(\phi))}{\partial\phi}\bigg|_{\phi=0} = \frac{1}{4\pi}[A_1 k_2 - A_2 k_1] . \qquad (6)$$

We fix $k = (1,0)$, $A_1 = A_2 = 0.32$ and, in preparation for $L \to \infty$ limit, compute the left-hand side in (6) as a function of L. The computations are performed at two different values of m. The data points are shown in Fig. 1 as a function of $1/L^2$. As expected, the continuum limit is independent of m. Extrapolation to $L = \infty$ is smooth
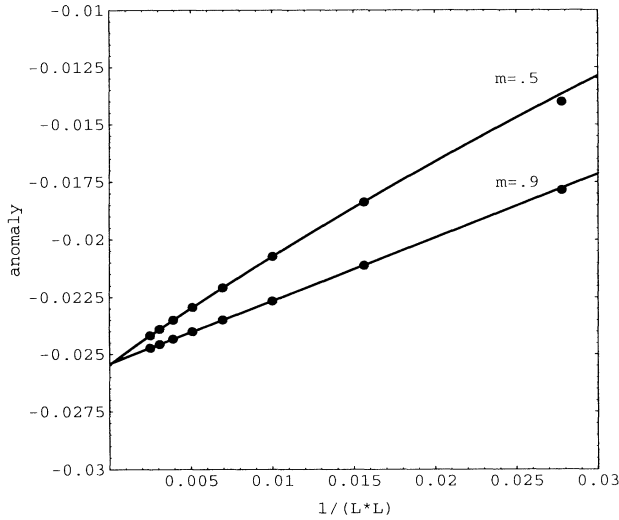
FIG. 1. The anomaly [left-hand side of (6)] as a function of lattice size. The points are for $L=6$ to 20 in steps of 2 at $m=0.5, 0.9$. The lines are fits to the data for $L \geq 10$.

and yields $-0.02545(5)$ for the anomaly which agrees with the continuum value of -0.02546.

Two comments are in order here. Since we used plane waves, once several k's are checked, the numerical work is as good a check as an analytical calculation would be. The coefficient in front of the curvature in the anomaly equation is $1/4\pi$ and not $1/2\pi$ as obtained by Jansen [10] in the three-dimensional wall setup. This reflects the fact that the current considered by Callan and Harvey is affected by the covariant form of the anomaly, not the consistent one. Since we deal with the effective action directly we have to obtain the "consistent" value [8,11].

The overlap in (4) will vanish when the number of positive eigenvalues of H_+ differs from the free case. This situation is expected if the set of gauge fields represents a U(1) connection on a nontrivial principal bundle over the two-dimensional torus. The first Chern number, represented by the "lattice topological charge,"

$$C_1 = \sum_{\Diamond} \frac{\ln(U_\Diamond)}{2\pi i}, \qquad (7)$$

should be nonzero in such an instance [12]. In (7), $\ln(1)=0$, the cut is along the negative real axis, and the sum runs over all elementary plaquettes. $U_\Diamond$ is the parallel transporter around the plaquette $\Diamond$. A simple choice that produces a nonzero C_1 is $U_\Diamond = \exp(2\pi i q/L^2)$ for all plaquettes with q some fixed integer as $L \to \infty$. $C_1 = q$ and this configuration has a smooth continuum limit. A gauge configuration that produces this uniform $U_\Diamond$ is $U_{(n_1,n_2),1} = \exp[(2\pi i q/L^2)n_2]$, $U_{(n_1,L-1),2} = \exp[(2\pi i q/L)n_1]$ and all other $U_{n,2}$ are equal to unity. For this gauge configuration at $L=6$, we found that H_+ had $36-q$ positive eigenvalues for all $|q| \leq 5$. That only H_+

is affected and not H_- is consistent with the observation by Golterman, Jansen, and Kaplan [13] that the Goldstone-Wilczek currents are present only on one side of the defect. We added random noise to the above gauge configuration, multiplying each link variable U by $e^{i\phi}$ with ϕ randomly chosen in the interval $(-\pi/6, \pi/6)$ and found that the qualitative behavior was robust. Generically, the behavior we obtain is clearly of the kind we would expect if instanton effects are to be reproduced. It would be interesting to insert the appropriate fermion operators to obtain a nonzero result and carry over the 't Hooft computation [4] of fermion number violating processes to the lattice.

The extension of the results in this paper to four dimensions is just a question of computer time. We think it would be feasible within current computer capacities. In two dimensions, for modest volumes, it appears feasible to proceed to the next stage, i.e., estimate the average over gauge fields. This by itself might be quite interesting because even in two dimensions chiral theories are either exactly soluble or, if also strongly coupled, quite intractable. One may also revisit attempts to construct anomalous gauge theories: For example, one may check whether the $a=1$ case of the Jackiw and Rajaraman [14] continuum construction can be reproduced on the lattice.

We should stress that we have invested little time to date in thinking about efficient methods of computation in four dimensions. We were only concerned with what one could do if computer power were unlimited. We believe that the approach outlined in this paper has sufficient potential to warrant more intense scrutiny.

We would like to thank T. Banks, D. B. Kaplan, M. F. Golterman, S. Shenker, and E. Witten for discussions. E. Witten was probably aware of a connection between chiral determinant and some form of overlap since the mid 1980's. This research was supported in part by DOE under Grant No. DE-FG05-90ER40559.

Note added:—Since the original submission of this manuscript, we have carried out similar tests on four-dimensional tori and obtained correct behavior in perturbative and nonperturbative Abelian backgrounds.

[1] J. Smit, Acta Phys. Pol. B 17, 531 (1986); L. Karsten and J. Smit, Nucl. Phys. B183, 103 (1981); H. B. Nielsen and M. Ninomiya, Nucl. Phys. B185, 20 (1981); B193, 173 (1981); F. Wilczek, Phys. Rev. Lett. 59, 2397 (1987); C. J. Griffin and T. D. Kieu, Phys. Rev. Lett. 70, 3844 (1993).

[2] A. Borrelli, L. Maiani, G. C. Rossi, R. Sisto, and M. Testa, Nucl. Phys. B333, 335 (1990); G. C. Rossi, R. Sarno, and R. Sisto, Nucl. Phys. B398, 101 (1993).

[3] L. Alvarez-Gaume, in Proceedings of the International School on Mathematical Physics, Erice, July 1985 (unpublished); and ENS Summer Workshop, August 1985 (unpublished).

[4] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976).

[5] D. B. Kaplan, Phys. Lett. B **288**, 342 (1992).

[6] R. Narayanan and H. Neuberger, Phys. Lett. B **302**, 62 (1993).

[7] C. G. Callan, Jr., and J. A. Harvey, Nucl. Phys. **B250**, 427 (1985).

[8] R. Narayanan and H. Neuberger, Rutgers University Report No. RU-93-25 (to be published).

[9] G. T. Bodwin and Eve V. Kovacs, Nucl. Phys. B (Proc. Suppl.) **30**, 617 (1993).

[10] K. Jansen, Phys. Lett. B **288**, 348 (1992).

[11] S. G. Naculich, Nucl. Phys. **B296**, 837 (1988); W. Bardeen and B. Zumino, Nucl. Phys. **B244**, 421 (1984).

[12] A. Phillips, Ann. Phys. (N.Y.) **161**, 399 (1985).

[13] M. F. L. Golterman, K. Jansen, and D. B. Kaplan, Phys. Lett. B **301**, 219 (1993).

[14] R. Jackiw and R. Rajaraman, Phys. Rev. Lett. **54**, 1219 (1985).