A Practical Implementation of the Overlap Dirac Operator

Herbert Neuberger*
Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855-0849
(Received 24 June 1998)

A practical implementation of the overlap Dirac operator $[1 + \gamma_5 \epsilon(H)]/2$ is presented. The implementation exploits the sparseness of H and does not require full storage. A simple application to parity invariant three dimensional SU(2) gauge theory is carried out to establish that zero modes related to topology are exactly reproduced on the lattice. [S0031-9007(98)07537-1]

PACS numbers: 11.15.Ha, 11.30.Rd

The known elementary particles and their interactions are described by a chiral gauge theory. This theory has a large ultraviolet cutoff of unknown nature. It is nevertheless useful because predictions based on Feynman diagrams are independent of this cutoff—the theory is renormalizable. In general, if a renormalizable field theory is not asymptotically free, all its quantitative predictions have a built-in inaccuracy that is negligible for weak couplings but becomes large at high energies where the couplings are strong. On the other hand, an asymptotically free theory like QCD makes infinitely accurate quantitative predictions at all energies. The coupling now becomes strong at low energies, and there Feynman diagrams are useless: One needs nonperturbative ways to calculate. Asymptotically free chiral gauge theories are special in that their Feynman diagrams make useful predictions at high energies only if delicate anomaly cancellations occur. The anomaly cancellation in gauge theories of QCD type is less delicate, and nonperturbative methods exist for low energy calculations—the Euclidean lattice provides a generic nonperturbative tool. But, it is still controversial whether it is at all possible to do meaningful calculations in an asymptotically free chiral gauge theory at low energies. There are no-go theorems on the lattice indicating that chiral gauge theories are fundamentally different from vectorlike gauge theories in this respect. Nevertheless, steady progress has been made during the last five years. Now, after all, it appears that asymptotically free chiral gauge theories do have the same predictive power as their vectorlike relatives. This is relevant even for chiral theories containing some interactions that are not asymptotically free, as long as the related couplings are small enough.

A simpler related long-standing problem has been the preservation of global chiral symmetries on the lattice for vectorlike gauge theories, like QCD. A solution (the overlap) has been found but in the initial stages of the finding practical applicability seemed limited to two dimensions (2D). The overlap was developed starting from [1], which in turn was motivated by [2] and [3]. The formulation of vectorlike theories, which is the main focus of this Letter, can be found in Sec. 9 of Ref. [4]. Work on 2D examples was successful [5] and added confidence in the basic idea. This induced a subsequent search for simplifications. One year ago a new, equivalent but more compact, formulation of exactly massless fermions on the lattice [6] was found. It has a fermion action induced by integrating out all but one of the infinite number of fermions that produce the overlap. The fermions that are integrated out are heavy and the "miracle" is the relative simplicity of the action for the remaining massless fermion. This action conforms to a criterion introduced many years ago by Ginsparg and Wilson (GW) [7] to assure continuumlike Ward identities on the lattice. The simple action produced by the overlap is the first known acceptable closed form solution to the GW requirement. To date, there are no other explicit solutions. Related work can be found in [8,9].

In spite of its compact form, the overlap-Dirac operator is not easy to work with numerically, and its practical usefulness in dimensions higher than two has been unclear until now. One purpose of this Letter is to present a new procedure to simulate the overlap Dirac operator on the computer. This procedure holds the promise to be practicable in three and four dimensional gauge theories.

One of the central issues in gauge theories is the spontaneous breakdown of global chiral symmetries. This issue could not be addressed in a clean way on the lattice until now, but the new development makes this possible.

In three dimensions there are interesting models that are parity invariant and have analogs of four dimensional global chiral symmetries [10]. Intensive investigations have indicated that spontaneous breakdown of these symmetries takes place if the (even) number of flavors is small enough.

In the continuum spontaneous chiral symmetry breakdown is correlated with an enhancement in the spectral density of the Dirac operator around zero. Clean checks of this were impossible on the lattice before the new actions.

In four dimensions it has been suspected that configurations consisting of superposed instanton/anti-instantons are responsible for this accumulation of almost zero eigenvalues. In two dimensions it is quite likely that this is actually true [11]. In three dimensions there are no instantons, but there is another source of low lying levels. Level crossings at zero must occur in backgrounds connecting

 0031-9007/98/81(19)/4060(3)$15.00

topologically distinct pure gauge configurations. These crossings also happen in four dimensions. Until now, the level crossings could not be realized on the lattice since the needed symmetries were broken by the regularization.

Using the new action and the new procedure I shall exhibit below a level crossing in three dimensional SU(2) gauge theory with two flavors. At the moment I see no major difficulties left to overcome on the way to a full dynamical simulation of this exactly "chirally symmetric" model. Generalizations to four dimensions also no longer appear prohibitive. It is possible that this new procedure will revolutionize the way fermions are treated on the lattice.

The compact overlap Dirac operator introduced in [6] is

$$D = \frac{1 + \gamma_5 \epsilon(H)}{2}, \tag{1}$$

where $H = \gamma_5 D_W$. There is some freedom in choosing D_W. The simplest choice is to take D_W as the Wilson-Dirac operator with hopping parameter κ set at $\kappa = 1/(2d - 2)$. This is the choice adopted henceforth.

The first attempt to use D directly, rather than the original overlap formula [4], was made in two dimensions for a U(1) gauge theory by Chiu [12]. He used a Newton iteration to find $\sqrt{H^2}$. His method required storage of the full matrix H in memory. It is easy to bypass the computation of $\sqrt{H^2}$ and establish an iteration for $\epsilon(H)$ directly:

$$\frac{1}{X_{k+1}} = \frac{1}{2}\left(X_k + \frac{1}{X_k}\right); \qquad k = 0, 1, 2, \ldots; \qquad X_0 = H. \tag{2}$$

Observe now that the iteration can be "solved" by the replacement $W_k = (1 - X_k)/(1 + X_k)$. I assume that X_0 has no eigenvalue equal to 0, and if this is true it will also be true of all X_k. Equation (2) immediately leads to $W_k = (W_0)^{2^k}$. Setting $n = 2^{k-1}$, we write the solution to (2) as $X_k = f_n(H)$. The function $f_n(z)$ is given by

$$f_n(z) = \frac{(1 + z)^{2n} - (1 - z)^{2n}}{(1 + z)^{2n} + (1 - z)^{2n}}. \tag{3}$$

$f_n(H)$ is a rational approximant to $\epsilon(H)$ of the Padé type. The following identity is easily derived by calculating all the poles of $f_n(z)$ and their residues:

$$f_n(z) = \frac{z}{n} \sum_{s=1}^{n} \frac{1}{z^2 \cos^2 \frac{\pi}{2n}(s - \frac{1}{2}) + \sin^2 \frac{\pi}{2n}(s - \frac{1}{2})}. \tag{4}$$

$f_n(H)$ is relatively easy to calculate using, for example, a CG (conjugate gradient) algorithm for the inversions. Equation (4) has been previously derived by applied mathematicians [13] in a different context.

The computational cost is not much higher than a single inversion since the inversions for all s are related by shifts [14]. The convergence of the CG iteration is controlled by the $s = 1$ term. There is no reason to require n to be a power of 2 any more, and $f_n(H)$ is a truncation of $\epsilon(H)$ for any $n \geq 1$. In practice, below, I used $n = 48$ to get ten digits precision when acting on random vectors. Essentially all the precision required of the CG was maintained. Since f_n is odd all that one needs to check is how close $\|f_n(H)b\|^2/\|b\|^2$ is to unity for randomly chosen vectors b. Since $\epsilon(\lambda H) = \epsilon(H)$ for positive λ we also have scaled versions of f_n, with slightly different convergence properties. Since physics is concentrated in the spectrum of H close to zero it is not recommended to choose a λ to ensure optimal overall convergence. For simplicity, I have kept $\lambda = 1$ in what follows.

The formula for H appropriate for three dimensions was obtained by dimensional reduction along direction 4 from four dimensions [15]. The boundary conditions obeyed by the fermions are picked periodic in direction 4 and antiperiodic in directions 1, 2, 3. I used the real form of H introduced in [16]. Starting from a single Dirac fermion in four dimensions we end up with two Dirac fermions in three dimensions. Let us denote $\gamma_5 \epsilon(H)$ by V. V is orthogonal and obeys $\gamma_5 V \gamma_5 = \gamma_4 V \gamma_4 = V^T$, and consequently $[\gamma_4 \gamma_5, V] = 0$. The first two identities are of GW type. The three properties together constitute the lattice realization of the global "chiral" SU(2) symmetry known in the continuum. The reality of V implies that parity is conserved at the action level. In the continuum, it is believed that parity does not break down spontaneously, but the global SU(2) symmetry does. The breaking is accompanied by two massless "pions."

The crossing we wish to see occurs when we interpolate smoothly unit link variables to a lattice pure gauge configuration of special type. The gauge transformation defining the latter is a discretized form of a known nontrivial continuum map from the three torus in the continuum to the SU(2) group manifold [17]. The interpolation parameter is denoted by t and goes from 0 to 1. There is a built-in symmetry in the configuration under $t \to 1 - t$. Thus, the crossing is expected exactly at $t = 0.5$.

To see the crossing I compute the lowest few eigenvalues of $1 + \frac{1}{2}(V + V^T)$ by a CG based variational method [18]. The lowest state is found to be doubly degenerate. It corresponds to two conjugate eigenvalues of V, $e^{\pm i\theta}$. For t close to 0.5, θ is close to π, and the crossing can easily be seen from a plot of the square root of the eigenvalue $1 + \cos\theta(t)$ as a function of t. The flow is shown in Fig. 1.

The crossing takes place on lattices as small as 6^3 and has been confirmed by the indirect method of [17]. That method is based on the Herzberg–Longuet-Higgins effect [19]. Without any effort I increased the volume to 16^3. At this point H is a $32\,768 \times 32\,768$ real matrix. QCD on a 6^4 lattice has a complex fermionic matrix of size $15\,552 \times 15\,552$. This is comparable to our example.

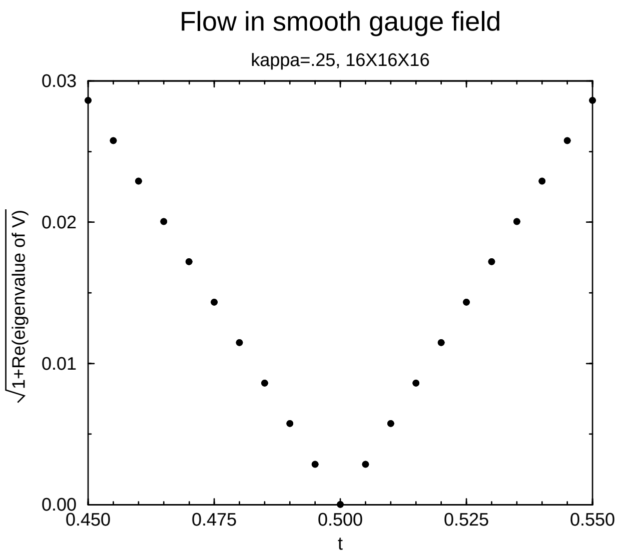

FIG. 1. The level crossing.

The preceding computation of low eigenvalues is similar in effort to an inversion of $1 + V$.

Many refinements of the procedure outlined here are possible. Lurking behind the scenes are Chebyshev polynomials, and they have appeared in this context before, in the analysis of another truncation of the overlap, studied in [20,21]. The light fermions in the latter truncation are usually referred to as domain wall fermions. There also are heavy fermions which would have to be explicitly removed in simulations of dynamic fermions. Although I suspect that the method presented here is superior to domain wall fermions, numerical efforts employing the latter should also be pursued because they will produce useful results. Space does not permit a more thorough discussion at this time.

In this Letter a new and direct approach to treat the overlap action $[1 + \gamma_5 \epsilon(H)]/2$ on potentially realistic lattices in three and four dimensions has been defined. The feasibility of a serious numerical study in $d = 3$ of a certain class of interesting models has been established. These results open up many directions for future research.

This work was supported in part by the DOE under Grant No. DE-FG05-96ER40559. I am grateful to Claudio Rebbi for providing me with the qcdf90 package [22] which I adapted to my needs in this work. I am grateful to Artan Boriçi for sending me a copy of Ref. [13] and for related comments. I became aware of Ref. [13] following the initial submission of this manuscript.

*Email address: neuberg@physics.rutgers.edu

[1] R. Narayanan and H. Neuberger, Phys. Lett. B **302**, 62 (1993).
[2] D. B. Kaplan, Phys. Lett. B **288**, 342 (1992).
[3] S. A. Frolov and A. A. Slavnov, Phys. Lett. B **309**, 344 (1993).
[4] R. Narayanan and H. Neuberger, Nucl. Phys. **B443**, 305 (1995).
[5] R. Narayanan, H. Neuberger, and P. Vranas, Phys. Lett. B **353**, 507 (1995).
[6] H. Neuberger, Phys. Lett. B **417**, 141 (1998).
[7] P. Ginsparg and K. Wilson, Phys. Rev. D **25**, 2649 (1982).
[8] P. Hasenfratz, V. Laliena, and F. Niedermayer, Phys. Lett. B **427**, 125 (1988).
[9] M. Lüscher, Phys. Lett. B **428**, 342 (1988).
[10] T. Appelquist and D. Nash, Phys. Rev. Lett. **64**, 721 (1990).
[11] A. Casher and H. Neuberger, Phys. Lett. **139B**, 67 (1984).
[12] Ting-Wai Chiu, Phys. Rev. D **58**, 074511 (1998).
[13] See equation (6.5) in the review by Nicholas J. Higham, in *Linear Algebra and Applications, Proceedings of the ILAS Conference "Pure and Applied Linear Algebra: The New Generation," Pensacola, 1993* (Elsevier, New York, 1994), and references therein.
[14] A. Frommer, S. Güsken, T. Lippert, B. Nöckel, and K. Schilling, Int. J. Mod. Phys. C **6**, 627 (1995).
[15] Y. Kikukawa and H. Neuberger, Nucl. Phys. **B513**, 735 (1998).
[16] H. Neuberger, hep-lat/9803011.
[17] H. Neuberger, hep-lat/9805027.
[18] B. Bunk, K. Jansen, M. Lüscher, and H. Simma, DESY report, 1994; T. Kalkreuter and H. Simma, Comput. Phys. Commun. **93**, 33 (1996).
[19] G. Herzberg and H. C. Longuet-Higgins, Discuss. Faraday Soc. **35**, 77 (1963).
[20] H. Neuberger, Phys. Rev. D **57**, 5417 (1998).
[21] Y. Kikukawa, H. Neuberger, and A. Yamada, Nucl. Phys. **B526**, 572 (1998).
[22] I. Dasgupta, A. R. Levi, V. Lubicz, and C. Rebbi, Comput. Phys. Commun. **98**, 365 (1996).