

# Chiral anomalies and rooted staggered fermions

Michael Creutz

*Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*

Received 19 January 2007; received in revised form 23 March 2007; accepted 23 March 2007

Available online 18 April 2007

Editor: B. Grinstein

## Abstract

A popular approximation in lattice gauge theory is an extrapolation in the number of fermion species away from the four fold degeneracy natural with the staggered fermion formulation. I show that the procedure mutilates the expected continuum holomorphic behavior in the quark masses. This is due to a chiral symmetry group that is of a higher rank than desired. The conventional resolution proposes canceling the unphysical singularities with a plethora of extra states appearing at finite lattice spacing. This unproven conjecture requires an explicit loss of unitarity and locality. Even if correct, the approach implies large cutoff effects in the low-energy flavor-neutral sector.

© 2007 Elsevier B.V. All rights reserved.

PACS: 11.15.Ha; 11.30.Rd; 12.38.Gc; 12.39.Fe

Lattice gauge theory provides a powerful tool for the investigation of non-perturbative phenomena in strongly coupled field theories, such as the quark confining dynamics of the strong interactions. However numerical calculations are quite computer intensive, strongly motivating approximations that reduce this need. One such, the valence or quenched approximation [1,2], introduces rather uncontrolled uncertainties, but with the growth in computer power, its use is currently being eliminated.

Another popular approximation [3,4] arises from the simplicity of the staggered fermion formulation [5–7]. With only one Dirac component on each site, the large matrix inversions required are substantially faster than with other fermion formulations. However the approach and its generalizations are based on a discretization method that inherently requires a multiple of four fundamental fermions. The reasons for this are related to the cancellation of chiral anomalies. To apply the technique to the physical situation of two light and one intermediate mass quark requires an extrapolation down in the number of fermions. As usually implemented, the approach involves taking a root of the fermion determinant inside standard hybrid Monte Carlo simulation algorithms. This step has not been justified theoretically. The purpose of this Letter is to show that at fi-

nite lattice spacing this reduction inherently mutilates the quark mass dependence expected in the continuum theory. A preliminary discussion of these points appears in Ref. [8].

The method has its roots in the “naive” discretization of the derivatives in the Dirac equation

$$\bar{\psi} \gamma_\mu \partial_\mu \psi \rightarrow \frac{1}{2a} \bar{\psi}_x \gamma_\mu (\psi_{x+ae_\mu} - \psi_{x-ae_\mu}) \quad (1)$$

with  $a$  denoting the lattice spacing. Fourier transforming to momentum space, the momentum becomes a trigonometric function

$$p_\mu \rightarrow \frac{1}{2ia} (e^{iap_\mu} - e^{-iap_\mu}) = \frac{1}{a} \sin(ap_\mu). \quad (2)$$

The natural range of momentum is  $-\pi/a < p_\mu \leq \pi/a$ . The doubling issue is that the propagator has poles not just at small momentum, but also when any component is near  $\pi$  in magnitude. These all contribute as intermediate states in Feynman diagrams; so, the theory effectively has  $2^4 = 16$  fermions. I refer to these multiple states as “doublers” or “flavors” in the following discussion.

Note that the slope of the sine function at  $\pi$  is opposite to that at 0. This can be absorbed by changing the sign of the corresponding gamma matrix. This changes the sign of  $\gamma_5$  as well; so, the doublers divide into different chirality subsets. The

*E-mail address:* [creutz@bnl.gov](mailto:creutz@bnl.gov).

determinant of the Dirac operator is not simply the sixteenth power of a single determinant.

Without a mass, the naive action has an exact chiral symmetry of the kinetic term under

$$\begin{aligned}\psi &\rightarrow e^{i\theta\gamma_5}\psi, \\ \bar{\psi} &\rightarrow \bar{\psi}e^{i\theta\gamma_5}.\end{aligned}\quad (3)$$

The conventional mass term is not invariant under this rotation

$$m\bar{\psi}\psi \rightarrow m\bar{\psi}\psi \cos(2\theta) + im\bar{\psi}\gamma_5\psi \sin(2\theta).\quad (4)$$

Thus any mass term of the form on the right-hand side of this relation can have theta rotated away. This is consistent with known anomalies since this is in reality a flavor non-singlet chiral rotation. The different species use different signs for  $\gamma_5$ . As special cases, in this theory  $m$ ,  $-m$ , and  $\pm i\gamma_5 m$  are all physically equivalent.

To arrive at the staggered formulation, note that whenever a fermion hops between neighboring sites in direction  $\mu$ , it picks up a factor of  $\gamma_\mu$ . An arbitrary closed fermion loop on a hypercubic lattice gives a product of many gamma factors, but any particular component always appears an even number of times. Bringing them through each other using anti-commutation, the net factor for any loop is proportional to unity. Gauge fields do not change this fact since they just involve  $SU(3)$  phases on the links. So if a fermion starts in one spinor component, it returns to the same component after the loop. The 4 Dirac components give 4 independent theories. There is an exact  $SU(4)$  symmetry. Without a mass term, this is actually an exact  $SU(4) \otimes SU(4)$  chiral symmetry [9].

Staggered fermions single out one component on each site. Which component depends on the gamma factors to get to the site in question from one chosen starting site. Ignoring the other components reduces the degeneracy from 16 to 4. The process brings in various oscillating phases from the gamma matrix components. One explicit projection that accomplishes this is (using integer coordinates and the convention  $\gamma_5 = -\gamma_1\gamma_2\gamma_3\gamma_4$  with Euclidean gamma matrices)

$$\begin{aligned}P &= P^2 \\ &= \frac{1}{4}(1 + i\gamma_1\gamma_2(-1)^{x_1+x_2} + i\gamma_3\gamma_4(-1)^{x_3+x_4} \\ &\quad + \gamma_5(-1)^{x_1+x_2+x_3+x_4}).\end{aligned}\quad (5)$$

Note that some degeneracy must remain. No chiral breaking appears in the action, and all infinities are removed. The conventional axial anomaly is canceled between the remaining species. Furthermore, the naive replacement  $\psi \rightarrow \gamma_5\psi$  exactly relates the theory with mass  $m$  and mass  $-m$ . With 4 flavors this symmetry is allowed since it still represents a flavored chiral rotation. The doublers appear in chiral pairs.

To proceed I sketch how a typical simulation with fermions proceeds. For a generic fermion matrix  $D$ , the goal of the simulation is to generate configurations of gauge fields  $A$  with a probability

$$P(A) \propto \exp(-S_g(A) + N_f \text{Tr} \log(D(A))).\quad (6)$$

Here  $S_g$  is the pure gauge part of the action and  $N_f$  is the number of fermion species. With some algorithms additional commuting “pseudo-fermion” fields are introduced [10,11], but these details are not important to the following discussion. With staggered or naive fermions the eigenvalues of  $D$  all appear in complex conjugate pairs; thus, the determinant is non-negative as necessary for a probability density.

In hybrid Monte Carlo schemes [12] auxiliary “momentum” variables  $P$  are introduced, one for each degree of freedom in  $A$ . The above distribution is generalized into

$$P(A, P) \propto \exp(-S_g(A) + N_f \text{Tr} \log(D(A)) - \sum P_i^2/2).\quad (7)$$

As the momenta are Gaussian random variables, it is easy to generate a new set at any time. For the gauge fields one sets up a “trajectory” in a fictitious “Monte Carlo” time variable  $\tau$  and uses the exponent in (7) as a classical Hamiltonian

$$H = \sum P_i^2/2 + V(A)\quad (8)$$

with the “potential”

$$V(A) = S_g(A) - N_f \text{Tr} \log(D(A)).\quad (9)$$

The Hamiltonian dynamics

$$\begin{aligned}\frac{dA_i}{d\tau} &= P_i, \\ \frac{dP_i}{d\tau} &= F_i(A) = -\frac{\partial V(A)}{\partial A}\end{aligned}\quad (10)$$

conserves energy and phase space. Under such evolution the equilibrium ensemble stays in equilibrium, a sufficient condition for a valid Monte Carlo algorithm. After evolution along a trajectory of some length  $\tau$ , discretized time steps  $\delta\tau$  can introduce finite step errors and give a small change in the “energy”. The hybrid Monte Carlo algorithm corrects for this with a Metropolis accept/reject step on the entire the trajectory. The trajectory length and step size are parameters to be adjusted for reasonable acceptance. After the trajectory one can refresh the momenta by generating a new set of gaussianly distributed random numbers. The procedure requires the “force” term

$$F_i(A) = -\frac{\partial V(A)}{\partial A} = -\frac{\partial S_g(A)}{\partial A} + N_f \text{Tr} \left( D^{-1} \frac{\partial D(A)}{\partial A} \right).\quad (11)$$

To calculate the second term requires an inversion of the sparse matrix  $D$  applied to a fixed vector. Standard linear algebra techniques such as a conjugate gradient algorithm can accomplish this. In practice this step is the most time consuming part of the algorithm.

Returning to staggered fermions, one would like to eliminate the unwanted degeneracy by a factor of four. One attempt to do this reduction involves an extrapolation in the number of flavors. In the molecular dynamics trajectories for the simulation of the gauge field, the coefficient of the fermionic force term in Eq. (11) is arbitrarily reduced from  $N_f$  to  $N_f/4$ , where  $N_f$  is the desired number of physical flavors. Although not proven, this seems reasonable when  $N_f$  is itself a multiple of four. The controversy arises for other values of  $N_f$ .

Here I argue that the procedure is an approximation that inherently mutilates the analytic structure expected in the quark masses. To see this consider the case of two flavor QCD with quark masses  $m_u$  and  $m_d$ . Complexifying the mass terms in the usual way

$$\sum_{a=u,d} \text{Re } m_a \bar{\psi}^a \psi^a + i \text{Im } m_a \bar{\psi}^a \gamma_5 \psi^a \quad (12)$$

the physical theory is invariant under the flavored chiral rotation

$$\begin{aligned} m_u &\rightarrow e^{i\theta} m_u, \\ m_d &\rightarrow e^{-i\theta} m_d. \end{aligned} \quad (13)$$

Due to the chiral anomaly, it must not be invariant under the singlet chiral rotation

$$\begin{aligned} m_u &\rightarrow e^{i\theta} m_u, \\ m_d &\rightarrow e^{i\theta} m_d. \end{aligned} \quad (14)$$

The symmetry in mass parameter space requires that the rotations of the up and down quark masses be in opposite directions. Indeed, this limitation is correlated with there only being one neutral Goldstone boson for the two flavor theory.

Now formulate this theory with two independent staggered fermions, one for the up and one for the down quark, each reduced using the rooting procedure. From Eq. (5), the corresponding complexification of the staggered mass term takes the form

$$\sum_{a=u,d} (\text{Re } m_a + i S(j) \text{Im } m_a) \psi^\dagger(j) \psi(j) \quad (15)$$

with  $S(j)$  being  $\pm 1$  depending on the parity of the site  $j$ . The issue arises from the fact that the staggered fermion determinant, and therefore the path integral, are exactly invariant under  $m \rightarrow e^{i\theta} m$  for either the up or the down quark. This is too much symmetry in parameter space. The physical  $SU(2)$  chiral symmetry group is of rank one, while the chiral symmetry of the two flavored staggered fermion formulation has rank two. It requires two neutral Goldstone bosons in the massless limit, rather than the one of the physical theory.

The issues become particularly severe in the chiral limit when  $N_f$  is odd. For the staggered theory, the fermion determinant is a function of  $m^2$ . The surviving chiral symmetry gives equivalent physics for either  $m$  or  $-m$ . However, it is well known that with an odd number of flavors, physics has no symmetry under changing the sign of the mass [13–15]. The most dramatic demonstration of this appears in the one flavor theory [16]. In this case anomalies break all chiral symmetries and no Goldstone bosons are expected. The theory behaves smoothly as the mass parameter passes through zero. The lightest meson, call it the  $\eta'$ , acquires a mass through anomaly effects, and the lowest order quark mass corrections are linear

$$m_{\eta'}^2(m) = m_{\eta'}^2(0) + cm. \quad (16)$$

Such a linear dependence in a physical observable is immediately inconsistent with  $m \leftrightarrow -m$  symmetry.

The one flavor case is perhaps a bit special, but there are similar problems with the three flavor situation [15]. Identify

the quark bi-linear with an effective chiral field  $\bar{\psi}_a \psi_b \sim \Sigma_{ab}$ . Here  $a$  and  $b$  are flavor indices. The  $SU(3) \otimes SU(3)$  chiral symmetry of the massless theory is embodied in the transformation

$$\Sigma \rightarrow g_L^\dagger \Sigma g_R \quad (17)$$

with  $g_L, g_R \in SU(3)$ . For positive mass,  $\Sigma$  should have an expectation value proportional to the  $SU(3)$  identity  $I$ . This is not equivalent to the negative mass theory because  $-I$  is not in  $SU(3)$ . Indeed, for negative mass it is expected that the infinite volume theory spontaneously breaks CP symmetry, with  $\langle \Sigma \rangle \propto e^{\pm 2\pi i/3}$  [15,17].

These qualitative effective Lagrangian arguments are quite powerful and general. Another way to see the one flavor behavior is to start with a larger number of flavors, say 3 or 4, and make the masses non-degenerate. As only one of the masses passes through zero, the behavior for the lightest meson mimics that in Eq. (16). Extrapolated staggered quarks with their symmetry under taking any quark mass to its negative will miss the linear term.

Small real eigenvalues of the Dirac operator are responsible for these effects. The odd terms come from topological structures in the gauge fields [18]. For small mass in the traditional continuum discussion,  $|D| \sim m^\nu$  with  $\nu$  the winding number of the gauge field. The condensate

$$\langle \bar{\psi} \psi \rangle = \frac{1}{Z} \int (dA) |D|^{N_f} e^{-S_g(A)} \text{Tr } D^{-1} \quad (18)$$

receives a contribution going as  $m^{N_f-1}$  from the  $\nu = 1$  sector. For the one flavor case, this is an additive constant. This constant will be missing from the extrapolated staggered theory because of the symmetry in Eq. (3). This phenomenon is also responsible for the fact that a single massless quark is not a well-defined concept [19].

For the general odd flavor case, the odd winding number terms have the opposite symmetry under the sign of the mass than the even terms, although with more flavors this starts at a higher order in the mass. For 3 flavors the condensate at finite volume will display an  $m^2$  correction to the leading linear behavior. The extrapolation down from the staggered 4 flavor theory will not see this term.

It is during the transitions between topological sectors that the unrooted theory behaves quite differently than the target theory. In particular, with a smooth gauge field of unit winding number near the continuum limit, the unrooted theory should have four small eigenvalues representing the zero modes from the index theorem. Considering an evolution of the gauge fields from zero to unit winding number, two of these drop down from positive imaginary part and two move up from below. Any approximate four fold degeneracies between the higher eigenvalues must break down during this evolution. Attempts to define the rooting procedure by selecting one fourth of the eigenvectors will necessarily involve ambiguities.

While I have shown diseases with the chiral behavior of extrapolated staggered fermions at finite cutoff, it has been suggested [21–23] that these problems go away as the cutoff is removed. Indeed, in quantum field theory we are accustomed to the non-commutation of certain limits, such as vanishing mass

and infinite volume when a symmetry is being spontaneously broken. In that case the mass and the volume are both infrared issues. As the lattice is an ultraviolet regulator and the chiral issues raised here involve long distance physics, it seems peculiar for the order of these limits to affect each other. Nevertheless, suppose that taking the cutoff to zero before taking the massless limit does give the correct physics. Then the regulator must introduce singularities that are not present in the continuum theory.

The issue is again clearest for the one flavor theory, where in the continuum the condensate,  $\langle \bar{\psi} \psi \rangle$  appropriately renormalized, does not vanish and is smoothly behaved around  $m = 0$ . Analyticity in the mass is expected with a radius of order the eta-prime mass-squared over the typical scale of the strong interactions,  $\Lambda_{\text{QCD}}$ . Now turn on the extrapolated staggered regulator. At  $m = 0$ ,  $\langle \bar{\psi} \psi \rangle$  must suddenly jump to zero. For every eigenvalue of the staggered fermion matrix at vanishing mass, its negative is also an eigenvalue. Configuration by configuration the trace of  $D^{-1}$ , and thus the condensate, is incorrectly predicted to be identically zero. Furthermore, due to confinement and the chiral anomaly, this unphysical jump occurs both at finite volume and in the absence of any massless physical particles for the continuum theory.

This problem generalizes to the multi-flavor theory with non-degenerate quark masses. The proposed regulator forces the condensate associated with any given species to vanish with the corresponding mass, in direct contradiction with the continuum behavior expected from effective Lagrangian analysis. Physical observables at specific points in parameter space where continuum physics is smooth are forced to develop infinite derivatives with respect to the cutoff as it is removed. Even if this occurs only in the vicinity of isolated points, this seems an absurd behavior for an ultraviolet regulator and is in strong contrast to more sensible schemes such as Wilson fermions [20].

It has recently been argued [22] that this unphysical behavior could be avoided in the continuum limit as long as one stays away from these singularities. Consider the two flavor theory discussed earlier. Due to the doubling, the unrooted theory has 32 neutral pseudo-scalar mesons. The anomaly should give one of these a mass of order the QCD scale, and this becomes the eta prime. At finite lattice spacing the remaining 31 particles divide into two exact Goldstone bosons corresponding to the exact chiral symmetries and 29 approximate Goldstone bosons. If we now give only one of the quarks a small mass, one of the massless pseudo-scalars should acquire a mass and represent the neutral pion. The second, however, must remain massless due to the remaining symmetry. Ref. [22] argues that at finite lattice spacing the 29 extra mesons at finite mass are still there after rooting. They suggest, without proof, that if the second quark is given a small mass and as the lattice spacing is taken zero, it is possible that this plethora of extra states could move down in energy and cancel the unwanted extra Goldstone boson. This scheme requires a loss of unitarity; indeed, the production cross sections for some pairs of the unwanted mesons must be negative so the total production can add to zero. And before this happens the theory is non-local because

of long range forces due to the one unwanted massless particle.

Such a mechanism appears to me as rather contrived, but Ref. [22] suggests that it is merely an ugly feature of the algorithm. Even if the proposed cancellation does occur, at finite lattice spacing we have a factor of 16 more neutral pseudo-scalar mesons than in the physical theory. This suggests that the lattice corrections to physics in the flavor singlet sector are potentially rather large.

To summarize, at finite lattice spacing the holomorphic behavior in the quark masses for rooted staggered quarks is qualitatively incompatible with continuum physics. The chiral symmetry group with rooted fermions is of a higher rank than desired. This gives rise to unphysical singularities when any single quark mass vanishes. For the extra symmetry to disappear as the lattice spacing is taken to zero requires rather subtle cancellations which have not been demonstrated. The approximation may still be reasonable for some observables, most particularly those involving only flavor non-singlet particles. But any predictions for which anomalies are important are particularly suspect. This includes the  $\eta'$  mass, but also more mundane quantities such as the lightest baryon mass, which in the chiral limit is entirely non-perturbative.

## Acknowledgements

This manuscript has been authored under contract number DE-AC02-98CH10886 with the US Department of Energy. Accordingly, the US Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

## References

- [1] D.H. Weingarten, D.N. Petcher, Phys. Lett. B 99 (1981) 333.
- [2] H. Hamber, G. Parisi, Phys. Rev. Lett. 47 (1981) 1792.
- [3] C. Aubin, et al., Fermilab Lattice Collaboration, Phys. Rev. Lett. 94 (2005) 011601, hep-ph/0408306.
- [4] C. Aubin, et al., MILC Collaboration, Phys. Rev. D 70 (2004) 114501, hep-lat/0407028.
- [5] J.B. Kogut, L. Susskind, Phys. Rev. D 11 (1975) 395.
- [6] L. Susskind, Phys. Rev. D 16 (1977) 3031.
- [7] H.S. Sharatchandra, H.J. Thun, P. Weisz, Nucl. Phys. B 192 (1981) 205.
- [8] M. Creutz, hep-lat/0603020, unpublished.
- [9] L.H. Karsten, J. Smit, Nucl. Phys. B 183 (1981) 103.
- [10] F. Fucito, E. Marinari, G. Parisi, C. Rebbi, Nucl. Phys. B 180 (1981) 369.
- [11] D.J. Scalapino, R.L. Sugar, Phys. Rev. Lett. 46 (1981) 519.
- [12] S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, Phys. Lett. B 195 (1987) 216.
- [13] P. Di Vecchia, G. Veneziano, Nucl. Phys. B 171 (1980) 253.
- [14] M. Creutz, Phys. Rev. D 52 (1995) 2951, hep-th/9505112.
- [15] M. Creutz, Phys. Rev. Lett. 92 (2004) 201601, hep-lat/0312018.
- [16] M. Creutz, hep-th/0609187, Ann. Phys., in press.
- [17] I. Montvay, Nucl. Phys. B (Proc. Suppl.) 83 (2000) 188, hep-lat/9909020.
- [18] G. 't Hooft, Phys. Rev. D 14 (1976) 3432;  
G. 't Hooft, Phys. Rev. D 18 (1978) 2199, Erratum.
- [19] M. Creutz, Phys. Rev. Lett. 92 (2004) 162003, hep-ph/0312225.

- [20] K.G. Wilson, in: A. Zichichi (Ed.), CLNS-321 New Phenomena in Subnuclear Physics, Part A, Proceedings of the First Half of the 1975 International School of Subnuclear Physics, Erice, Sicily, July 11–August 1 1975, Plenum Press, New York, 1977, p. 69.
- [21] S. Durr, C. Hoelbling, Phys. Rev. D 71 (2005) 054501, hep-lat/0411022.
- [22] C. Bernard, M. Golterman, Y. Shamir, S. Sharpe, hep-lat/0603027; S.R. Sharpe, PoS LAT2006 (2006) 022, hep-lat/0610094.
- [23] S. Durr, C. Hoelbling, hep-lat/0604005.