

Chiral anomalies and Wilson fermions

Michael Creutz^{*}

Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA



(Received 4 December 2025; accepted 2 March 2026; published 21 April 2026)

The Wilson formulation of fermions in lattice gauge theory provides a unified description of the chiral anomalies in the standard model. The discrete Dirac operator diagonalizes into a series of 2×2 blocks. In each block the possible eigenvalues either form a complex pair or separate into two real eigenvalues that have specific chirality. The collision of these pairs of eigenvalues occurs outside the perturbative region and provides a path between topological sectors.

DOI: [10.1103/pbpl-5wdb](https://doi.org/10.1103/pbpl-5wdb)

I. INTRODUCTION

The importance of chiral anomalies to the standard model has a long history, going back to the calculation of the decay of the neutral pion to two photons by Steinberger [1]. Naive arguments suggest that this should be suppressed when quark masses are small. This issue was elucidated in the classic papers of Adler *et al.* [2–4].

In the context of the strong interactions, anomalies associated with the axial current emerged as crucial to understanding why the η' meson does not behave as a pseudo-Goldstone boson, unlike the pions and kaons. Also, a portion of the proton mass arises via a similar mechanism.

Perhaps the most unexpected consequence of chiral anomalies is the prediction by 't Hooft [5] that proton decay should occur in the standard model of the weak interactions. This rate is extremely small, perhaps unmeasurable, but its existence is a necessary consequence of the theory.

All these phenomena are tied to an anomaly with a simple change of fermionic variables in path integral

$$\psi \longrightarrow e^{i\phi\gamma_5}\psi, \quad \bar{\psi} \longrightarrow \bar{\psi}e^{i\phi\gamma_5}. \quad (1)$$

This is a symmetry of the classical standard model Lagrangian when the fermion masses vanish. However all renormalization schemes must violate this symmetry. In particular left-handed fields, $\psi_L = (1 - \gamma_5)\psi/2$, ultimately must mix with right-handed ones, $\psi_R = (1 + \gamma_5)\psi/2$, although the mechanism for this depends on the details of the regulator.

^{*}Contact author: mike@latticeguy.net

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

The Wilson lattice approach provides a specific regularization making the Feynman path integral mathematically well defined. As such the approach must allow for the effects of anomalies. A summary of the long history of this problem appears in Ref. [6]. The Smit-Swift model [7,8] (see also [9–11]) is a particular attempt to formulate chiral gauge theories with Wilson fermions. When extended to include an entire generation of fermions, The approach provides a specific lattice formulation for the standard model [12], based on three commuting gauge groups $SU(3) \otimes SU(2) \otimes U(1)$. Here we expand on that presentation with more details on the connection between anomalies and colliding eigenvalues of the Dirac matrix. Among various interesting issues is where does the baryon decay predicted by 't Hooft [5] arise? We will see that it comes from multiple fermion combinations simultaneously coupling in a gauge invariant way with complex configurations of Higgs and gauge fields.

A typical lattice gauge simulation generates configurations for the gauge and Higgs fields and then explores Dirac operator on these configurations. In something like the hybrid Monte Carlo method [13] for a dynamical simulation, feedback is included from the fermion loops. Here we will concentrate on the Dirac operator itself and not be concerned with the details of generating the relevant gauge and Higgs configurations.

II. DOUBLERS AND THE WILSON TRICK

In lattice gauge theory, the free fermion action starts with the derivative term replaced by nearest neighbor differences

$$i\partial_\mu\gamma_\mu\psi = \longrightarrow \sum_\mu \frac{i\gamma_\mu}{a}(\psi_{i+e_\mu} - \psi_i). \quad (2)$$

The lattice sites are labeled i , space-time directions μ , and lattice spacing a . Before including bosonic fields,

this difference takes a particularly simple form in momentum space,

$$p_\mu \longrightarrow \frac{1}{a} \sin(ap_\mu) = p_\mu + O(a). \quad (3)$$

The natural range of each component of lattice momentum is $-\frac{\pi}{a} < p_\mu \leq \frac{\pi}{a}$. Unfortunately within this range there is an extra pole in the propagator at $p_\mu \sim \frac{\pi}{a}$. The Wilson [14] solution to this issue removes this pole by giving the doubler a large mass

$$i\gamma_\mu p_\mu \longrightarrow \frac{1}{a} (i\gamma_\mu \sin(ap_\mu) + 1 - \cos(ap_\mu)). \quad (4)$$

Effectively this adds a $O(ap^2)$ correction to the naive result, a formally irrelevant operator $\sim \bar{\psi} \partial^2 \psi$.

The added heavy states explicitly break chiral symmetry and are ultimately the source of anomalies. One might think of them as an analogy of the heavy states in a Pauli-Villars [15] approach.

In the usual continuum theory, $i\not{D} + m$ is an anti-Hermitian kinetic term plus a real constant mass term. As such, all eigenvalues lie on a line at fixed real part m . On the lattice the Wilson term smears the mass term at high momentum. For the free lattice theory the sin and cosine factors in the above equations give rise to a set of nested circles as shown in Fig. 1.

When interactions turn on, the eigenvalues move around and momentum is no longer a useful parameter. Indeed, the Dirac operator is a non-normal operator with $[D, D^\dagger] \neq 0$. This comes about because of the complex gauge factors as the fermions hop around. Indeed, different eigenvectors of D need not be orthogonal. Also the right and left eigenvectors are not generally the same.

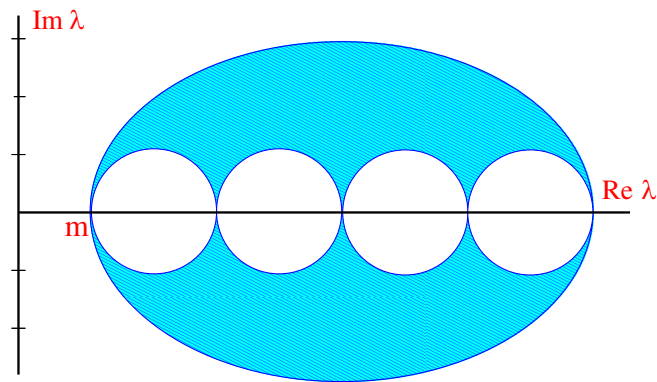


FIG. 1. The spectrum of the free four-dimensional Dirac operator. The crossings of the real axis correspond to how many doubler momentum components are at π/a . The multiplicities at these points follow the pattern 1, 4, 6, 4, 1.

III. γ_5 HERMITICITY

Although the lattice Dirac operator is neither Hermitian or normal, it does satisfy what is called γ_5 Hermiticity

$$\gamma_5 D = D^\dagger \gamma_5. \quad (5)$$

The kinetic or “hopping” terms involving gauge fields are anti-Hermitian and contain a γ matrix factor while the mass and Wilson term are Hermitian and commute with γ_5 .

This “symmetry” has important consequences for the eigenvalue spectrum of the Dirac operator. Consider some particular complex right eigenvector $D\psi = \lambda\psi_1$. Playing with the characteristic equation tells us

$$0 = |D - \lambda| = |D^\dagger - \lambda^*| = |D - \lambda^*| = |D^\dagger - \lambda|. \quad (6)$$

Thus the complex conjugate of λ is also an eigenvalue, and both D and $D^\dagger$ have the same eigenvalues. Of course since our matrix is not normal, the various eigenvectors are not in general the same or even orthogonal.

To proceed it is convenient to consider a two-dimensional space spanned by our eigenvector ψ_1 and $\psi_2 \equiv \gamma_5 \psi_1$. Since $\gamma_5^2 = 1$, these each have the same norm, which we can take to be unity. By definition these satisfy

$$D\psi_1 = \lambda\psi_1, \quad D^\dagger\psi_2 = \lambda\psi_2. \quad (7)$$

We have a pair of right eigenvectors, one for D and one for $D^\dagger$. Conjugation gives

$$\psi_1^\dagger D^\dagger = \lambda^* \psi_1^\dagger, \quad \psi_2^\dagger D = \lambda^* \psi_2^\dagger. \quad (8)$$

Thus we also have a pair of left eigenvectors, one for $D^\dagger$ and one for D . We see that D maps pairs of right eigenvectors for D and $D^\dagger$ into a similar pair of left eigenvectors. Furthermore, when λ is complex, we have an orthogonality condition

$$\begin{aligned} \psi_2^\dagger D\psi_1 &= \lambda \psi_2^\dagger \psi_1 = \lambda^* \psi_2^\dagger \psi_1 = 0, \\ \psi_1^\dagger D^\dagger \psi_2 &= \lambda^* \psi_1^\dagger \psi_2 = \lambda \psi_1^\dagger \psi_2 = 0. \end{aligned} \quad (9)$$

The one combination we do not know is

$$\psi_1^\dagger D\psi_2 = (\psi_2^\dagger D^\dagger \psi_1)^* = \psi_1^\dagger D\gamma_5 \psi_1 \equiv d \quad (10)$$

where d is an *a priori* unknown complex parameter.

The Dirac matrix thus breaks down into 2×2 blocks coupling pairs of right eigenvectors with pairs of left eigenvectors

$$\begin{aligned} \psi_i^\dagger D\psi_j &= \psi_i^\dagger \begin{pmatrix} \lambda & d \\ 0 & \lambda^* \end{pmatrix}_{ij} \psi_j, \\ \psi_i^\dagger D^\dagger \psi_j &= \psi_i^\dagger \begin{pmatrix} \lambda^* & 0 \\ d^* & \lambda \end{pmatrix}_{ij} \psi_j. \end{aligned} \quad (11)$$

These depend on two complex parameters, λ and d .

This matrix has determinant $\lambda^*\lambda$ independent of d . This is a factor of the full determinant of D . Repeating this process on further eigenvalues, we find that the determinant of D can be expressed as a product of determinants of 2×2 complex matrices.

A complication occurs when the basis becomes singular with $\psi_1 \propto \psi_2$. This can happen when the eigenvalues become real. To treat this situation it is useful to mix the modes into a “chiral” basis

$$\begin{aligned}\psi_+ &= (\psi_1 + \psi_2)/2 = \frac{1 + \gamma_5}{2} \psi_1, \\ \psi_- &= (\psi_1 - \psi_2)/2 = \frac{1 - \gamma_5}{2} \psi_1.\end{aligned}\quad (12)$$

In this basis γ_5 reduces to the Pauli matrix σ_3 .

We now write our submatrix of D as the most general 2×2 matrix satisfying $D^\dagger = \sigma_3 D \sigma_3$,

$$D \rightarrow a_0 + ia_1\sigma_1 + ia_2\sigma_2 + a_3\sigma_3 \quad (13)$$

where the coefficients a_i are real. This is easily diagonalized with eigenvalues

$$\lambda = a_0 \pm i\sqrt{a_1^2 + a_2^2 - a_3^2}. \quad (14)$$

The contribution to fermion determinant is the expected

$$|D| = a_0^2 + a_1^2 + a_2^2 - a_3^2 = \lambda\lambda^*. \quad (15)$$

The γ_5 Hermiticity also implies

$$H = \gamma_5 D \quad (16)$$

is a Hermitian quantity and always has real eigenvalues. This is also contained in our 2×2 blocks, with eigenvalues

$$\lambda_H = a_3 \pm \sqrt{a_0^2 + a_1^2 + a_2^2}. \quad (17)$$

The real parts of the eigenvalues of D are constrained to lie between these.

The parameter space of the four a 's separates nicely into two domains. When $a_1^2 + a_2^2 > a_3^2$ we have two complex eigenvalues as above. However when $a_1^2 + a_2^2 < a_3^2$ the eigenvalues become real [16]

$$\lambda = a_0 \pm \sqrt{a_3^2 - a_1^2 - a_2^2}. \quad (18)$$

As the gauge and Higgs fields evolve in a simulation, we can transit between these situations. In the process the two eigenvalues collide and move off in opposite directions on the real axis, as sketched in Fig. 2.

Large values of the parameter a_3 give rise to real modes of D . If we now cool our configuration down to classical

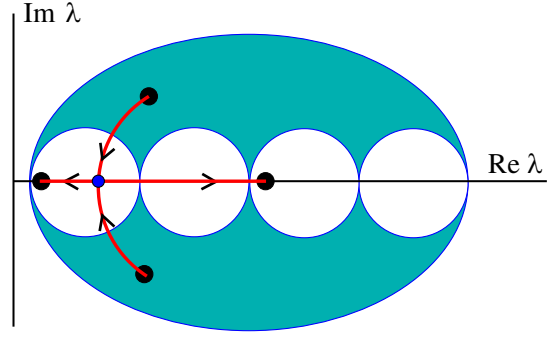


FIG. 2. As gauge fields evolve, two complex eigenvalues can collide and move off on the real axis.

fields, these become the topological objects commonly called “instantons” or “sphalerons,” the famous zero modes of the index theorem. With real eigenvalues, we lose the orthogonality condition of Eq. (9), and the combination $\psi^* \gamma_5 \psi$ need not vanish. Indeed this must happen in the continuum limit for a single instanton. In general real modes can mix right- and left-handed fermions. This is at the heart of the anomalies that break chiral symmetry.

While real simulations do involve finite jumps in configurations, the space of gauge and Higgs fields is simply connected and there do exist smooth paths between successive configurations. Going between different topologies must involve violating Luscher’s “admissibility” condition [17], as discussed in [18]. No assumptions are being made about the smoothness of the underlying fields.

In our two-dimensional subspace, the role of γ_5 is played by σ_3 . The free theory contains no terms directly involving γ_5 , so this term can only be generated from fermion loops in the determinant of D . Indeed a loop that involves hoppings in all four dimensions can generate such a term. Reference [18] used the loops circumnavigating hypercubes to define a local topological susceptibility. See also Ref. [19]. These loops necessarily mix kinetic hoppings with hoppings involving the Wilson term.

Note that the eigenvalue collision does not occur in perturbation theory. The effect does not occur until chiral effects are sufficiently large. Isolated modes with nonzero chirality appear immediately after the real modes separate. It is not required to go to the zero mode limit. Classical instantons and the index theorem do prove chiral modes must exist for smooth fields. Different lattice gauge configurations however are in general simply connected and these colliding eigenvalues provide a mechanism for flowing between different topologies.

If one were to construct Neuberger’s overlap operator [20] by projecting all eigenvalues onto a circle, singularities will appear as either of the modes pass through the projection point [21]. In general these singularities do not occur when the eigenvalues collide, but when either mode crosses the circle center.

Small eigenvalues are formally suppressed in the path integral

$$\begin{aligned} Z &= \int (dA)(d\bar{\psi}d\psi) e^{-S_g + \bar{\psi}D\psi} \\ &= \int (dA) e^{-S_g(A)} \prod \lambda_i. \end{aligned} \quad (19)$$

Naively that would suggest that zero modes might be irrelevant. However 't Hooft [5] showed how their effects could be enhanced in physical observables. This can be intuitively seen by adding source terms to the path integral

$$\begin{aligned} Z(\eta, \bar{\eta}) &= \int (dA)(d\bar{\psi}d\psi) e^{-S_g + \bar{\psi}D\psi + \bar{\eta}\psi + \bar{\psi}\eta} \\ &= \int (dA) e^{-S_g + \bar{\eta}D^{-1}\eta/4} \prod \lambda_i. \end{aligned} \quad (20)$$

The presence of D^{-1} is how the zero modes remain relevant to anomalies even as the quark masses go to zero.

IV. PHYSICAL CONSEQUENCES

In a simulation D is gauge dependent; so, for physical results we must combine the fields into gauge invariant observables. As usual in lattice gauge theory, we do not consider fixing a gauge. As such the various fields are highly fluctuating. To avoid these fluctuations we need to look at gauge invariant combinations of the fields.

Here we concentrate in observables that are sensitive to the chiral modes under discussion. We briefly discuss effects with each of the three commuting standard model groups $SU(3) \otimes SU(2) \otimes U(1)$.

A. Electrodynamics

First we discuss electrodynamics $U(1)$ where it is well known that one cannot conserve both the vector and axial current. In this case the above modes immediately break the axial symmetry of Eq. (1). In this case it is perhaps simplest to think of the doublers as playing the role of a Pauli-Villars field [15].

B. QCD

Here we consider only one generation of fermions. Thus we have u and d quarks which represent two $SU(3)$ triplets. When they are degenerate they couple to exactly the same gauge and Higgs fields and thus both see the same eigenvalues of D . A chiral mode will flip both spins together. This generates an effective mass term for the η' meson.

The chiral modes can influence the masses of other gauge singlet particles including the proton. The effective operator affects each quark color, generating an effective proton mass even if the quarks are massless.

The magnitude of these effects is controlled by the renormalization scale of the strong interactions Λ_{QCD} ,

$$m_{\eta'}, m_p \propto \Lambda_{\text{QCD}} \propto \frac{1}{a} g^{-\beta_1/\beta_0^2} \exp\left(\frac{-1}{2\beta_0 g^2}\right). \quad (21)$$

This scale is not from classical instantons alone with $\exp(-S_I) = \exp(-\frac{8\pi^2}{g^2})$, but rather is enhanced by fermionic loops and quantum fluctuations. The precise numerical values can in principle be determined via numerical simulations.

C. Weak $SU(2)$

Now we turn to the group of the weak interactions. Here 't Hooft [5] has shown there is a small contribution to baryon decay. Because the basic electroweak coupling is small, the effects are probably too small to observe.

The first generation standard model is built on four $SU(2)$ left-handed doublets

$$\begin{pmatrix} u^r \\ d^r \end{pmatrix}_L, \begin{pmatrix} u^g \\ d^g \end{pmatrix}_L, \begin{pmatrix} u^b \\ d^b \end{pmatrix}_L, \begin{pmatrix} \nu \\ e^- \end{pmatrix}_L \quad (22)$$

where the superscripts r, g, b refer to the strong interaction “colors.” Their antiparticles form four right-handed doublets $\psi_R^c = \tau_2 \gamma_2 \psi_L^*$. Explicitly these are

$$\begin{pmatrix} d^{r,c} \\ r^{r,c} \end{pmatrix}_R, \begin{pmatrix} d^{g,c} \\ r^{g,c} \end{pmatrix}_R, \begin{pmatrix} d^{b,c} \\ r^{b,c} \end{pmatrix}_R, \begin{pmatrix} e^+ \\ \nu^c \end{pmatrix}_R. \quad (23)$$

All of the above fermion variables have conjugate fields which are independent Grassmann variables in the path integral.

When the gauge and Higgs fields of a configuration form the chiral modes as discussed earlier, all of these doublets can be sensitive to them. To form gauge invariant observables, combine the left-handed doublets with the right-handed ones to define a 4×4 matrix of $SU(2)$ singlet combinations

$$T_{ij} = \bar{\psi}_i^R \psi_j^L. \quad (24)$$

To proceed we need to obtain a quantity that is also singlet under the strong and electromagnetic groups. This is accomplished in the determinant of this matrix

$$\det(T_{ij}). \quad (25)$$

This antisymmetrizes over $SU(3)$ colors, making an $SU(3)$ singlet. The quark and lepton charges also cancel making it electrically neutral. So this combination is a fully gauge invariant operator sensitive the chiral modes. This is the determinant that 't Hooft [5] found to summarize the weak anomalies in the standard model.

This determinant includes terms that couple two quarks to an antiquark and a lepton [22]. This allows for baryon-lepton mixing while preserving all gauged symmetries. This includes $p \leftrightarrow e^+$ and $n \leftrightarrow \bar{\nu}$.

V. SUMMARY

Given a configuration of gauge and Higgs fields in the path integral, the corresponding Dirac operator can be reduced to a series of 2×2 matrices, each representing two possible eigenvalues contributing to the determinant of D . As the configuration evolves in a simulation, these modes can collide and transition between complex pairs and real chiral modes. This provides a common mechanism for chiral anomalies in all three gauge symmetries of the standard model. These real modes occur outside the perturbative limit. Cooling procedures to smooth fields drives such modes to known topological configurations, although the anomalies do not require this limit.

An interesting question concerns the counting of nonperturbative topological parameters. With Wilson fermions these are related to relative phases between

the fermion masses and the Wilson terms [23,24]. In QCD flavored chiral rotations can move this phase Θ between different quark species. For example it can be placed entirely in the phase of a single quark, i.e. m_u or even m_t . How many such parameters are in the full standard model? Do both $SU(3)$ of QCD and $SU(2)$ of the weak interactions have independent Θ parameters? And how does QED fit into this picture?

ACKNOWLEDGMENTS

This manuscript has been authored under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

DATA AVAILABILITY

No data were created or analyzed in this study.

-
- [1] J. Steinberger, *Phys. Rev.* **76**, 1180 (1949).
 - [2] S. L. Adler, *Phys. Rev.* **177**, 2426 (1969).
 - [3] S. L. Adler and W. A. Bardeen, *Phys. Rev.* **182**, 1517 (1969).
 - [4] J. Bell and R. Jackiw, *Nuovo Cimento A* **60**, 47 (1969).
 - [5] G. 't Hooft, *Phys. Rev. D* **14**, 3432 (1976).
 - [6] J. Smit, *Eur. Phys. J. H* **50**, 5 (2025).
 - [7] J. Smit, *Acta Phys. Pol. B* **17**, 531 (1986).
 - [8] P. V. D. Swift, *Phys. Lett.* **145B**, 256 (1984).
 - [9] J. Smit and J. C. Vink, *Nucl. Phys.* **B286**, 485 (1987).
 - [10] S. Itoh, Y. Iwasaki, and T. Yoshie, *Phys. Lett. B* **184**, 375 (1987).
 - [11] S. Itoh, Y. Iwasaki, and T. Yoshie, *Phys. Rev. D* **36**, 527 (1987).
 - [12] M. Creutz, *Phys. Rev. D* **109**, 034514 (2024).
 - [13] S. Duane, A. Kennedy, B. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216 (1987).
 - [14] K. G. Wilson, *Phys. Rev. D* **10**, 2445 (1974).
 - [15] W. Pauli and F. Villars, *Rev. Mod. Phys.* **21**, 434 (1949).
 - [16] Notice an analogy with the connection between the Lorentz group and the group $SL(2)_C$, where vectors also divide into two classes: time- and spacelike.
 - [17] M. Luscher, *J. High Energy Phys.* **06** (2000) 028.
 - [18] M. Creutz, *Ann. Phys. (Amsterdam)* **326**, 911 (2011).
 - [19] M. Teper, *Phys. Lett.* **162B**, 357 (1985).
 - [20] H. Neuberger, *Phys. Lett. B* **417**, 141 (1998).
 - [21] M. Creutz, *Nucl. Phys. B, Proc. Suppl.* **119**, 837 (2003).
 - [22] This is consistent with the strong gauge group since in $SU(3)$ the product of two triplet representations contains an antitriplet, i.e. $3 \otimes 3 = 6 \oplus \bar{3}$.
 - [23] E. Seiler and I. Stamatescu, *Phys. Rev. D* **25**, 2177 (1982).
 - [24] F. Karsch, E. Seiler, and I. O. Stamatescu, *Nucl. Phys.* **B271**, 349 (1986).