A New Perspective on Chiral Gauge Theories

David B. Kaplan∗†
Institute for Nuclear Theory, University of Washington, Seattle, WA 98195-1550, USA
E-mail: dbkaplan@uw.edu

Dorota M. Grabowska
Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720
Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
E-mail: dgrabowska@berkeley.edu

Many different approaches have been pursued in the attempt to define chiral gauge theories on the lattice, all of which either break gauge symmetry explicitly, or are procedures for eliminating unwanted mirror fermions from the theory. I discuss a recently published proposal that combines domain wall fermions with gradient flow as a new idea for how to decouple the mirror fermions, which leads us to a chiral overlap operator.

34th annual International Symposium on Lattice Field Theory
24-30 July 2016
University of Southampton, UK

∗Speaker.
†This work was supported in part by DOE Grant No. DE-FG02-00ER41132 and an NSF Graduate Research Fellowship, Grant No. DGE-1256082
Chiral gauge theories are a fascinating class of quantum field theories thought to exhibit a number of interesting phenomena, including asymptotic freedom, dynamical symmetry breaking and massless composite fermions. Devising a nonperturbative regulator for such theories has proven difficult, however, and neither is there a proof for any perturbative regularization scheme that it is valid to all orders. This should be embarrassing given that the Standard Model is itself a chiral gauge theory, and even our perturbative treatment of it is not on a firm theoretical footing. One might therefore wonder whether devising a nonperturbative regulator for chiral gauge theories in general might not imply new phenomenology for the Standard Model, or resolution of outstanding puzzles such as the strong CP problem. For this reason, even if strongly coupled chiral gauge theories play no role in Nature, we still think it is of importance to settle the problem of nonperturbative regularization.

The path integral for a gauge theory takes the form

\[ Z = \int [dA] e^{-S[A]} \Delta[A] \] (1)

where \( A \) denotes the gauge field – or link fields with a lattice regularization – while \( S[A] \) is the gauge field action and \( \Delta[A] \) is the fermion contribution to the integration measure, the result of performing the path integral over all fermions coupled to the gauge field. For a vector gauge theory containing of \( N_f \) Dirac flavors, \( \Delta \) is just given by the Dirac determinant. In continuum Euclidian spacetime this is given by

\[ \Delta_{\nu}[A] = \det (\mathcal{D} + M) , \] (2)

a positive real quantity, and this definition holds on the lattice where there are suitable definitions of the Dirac operator. The determinant is given as the product of eigenvalues, where the well-defined eigenvalue equation in a chiral basis takes the form

\[ \begin{pmatrix} 0 & D_\mu \sigma_\mu \\ D_\mu \bar{\sigma}_\mu & 0 \end{pmatrix} \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = i \lambda \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} . \] (3)

In contrast, for the chiral gauge theory consisting of left-handed Weyl fermions in a complex representation of the gauge group, a comparable eigenvalue equation

\[ \begin{pmatrix} 0 & D_\mu \sigma_\mu \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = i \lambda \begin{pmatrix} \psi_R \\ 0 \end{pmatrix} . \] (4)

does not lead to an unambiguous definition of \( \lambda \) since \( \psi_R \) can always be redefined by a gauge field dependent phase. Thus we know that

\[ \Delta_{\chi}[A] = e^{i\delta[A]} \sqrt{\det (\mathcal{D})} , \] (5)

where the phase \( \delta[A] \), which encodes both anomalies and dynamics, remains to be determined.

It is evident that the phase \( \delta[A] \) must encode gauge anomalies, since \( \sqrt{\det (\mathcal{D})} \) is gauge invariant. It is also easy to make a heuristic argument that \( \delta[A] \) also affects dynamics in an anomaly-free theory by considering two different \( SO(10) \) gauge theories, one with \( N_f \) Weyl fermions transforming as \( 16 \oplus 16 \), the other with \( N_f \) Weyl fermions transforming as \( 16 \oplus \overline{16} \). Both are free of
gauge anomalies, however the former is a vector theory expected to have a gauge invariant bilinear fermion condensate breaking the $SU(N_f) \times SU(N_f)$ chiral symmetry down to the diagonal $SU(N_f)$, while the other has a global $SU(2N_f)$ symmetry and no possibility for a gauge invariant bilinear condensate. The two theories give rise to the same functional $\Delta[A]$ up to the phase $\delta[A]$. If $\delta[A]$ did not play a role in the dynamics of the second theory, one would predict the two theories to have identical gluon physics, such as identical glueball spectra, which seems very unlikely given that the two theories should be in entirely different universality classes.

Alvarez-Gaumé et al. proposed a definition of the phase $[1, 2]$ by equating $\Delta$ to the determinant for a theory with gauged left-handed fermions and non-interacting right-handed fermions:

$$\Delta[A] = \det \begin{pmatrix} 0 & D_\mu \sigma_\mu \\ \partial_\mu \sigma_\mu & 0 \end{pmatrix}.$$  \hfill (6)

This operator now has a well-defined eigenvalue equation, but unlike the eigenvalues of the Dirac theory, its eigenvalues are neither pure imaginary nor gauge invariant, giving rise to the phase $\delta[A]$. The authors showed that $\delta[A]$ was given by the $\eta$-invariant of a five-dimensional operator, whose variation reproduced the expected anomaly, while the right-handed fermions decoupled.

The key problem with constructing chiral symmetries theories on the lattice has to do with anomalies. In the continuum, anomalies are easily understood in terms of charges flowing in and out of the Dirac sea. However, this picture makes use of the “Hilbert hotel” nature of the Dirac sea, a semi-infinite spectrum of negative energy states that can always give up or absorb a fermion while remaining full. On the lattice the Hilbert space is finite and it is not possible to violate charge in this manner; instead the continuum anomaly requires explicit symmetry breaking on the lattice. A well-known example is how Wilson fermions, with explicit chiral symmetry violation in the form of a mass term and a Wilson term, can be fine tuned to yield massless quarks in the continuum limit, with a remnant of chiral symmetry breaking surviving to the continuum limit as the $U(1)_A$ anomaly $[3]$.

For the purposes of this talk, it is worthwhile revisiting how domain wall fermions and overlap fermions reproduce the $U(1)_A$ anomaly in lattice QCD. Domain wall fermions $[4, 5, 6]$ can be formulated as five-dimensional fermions on a compact extra dimension with mass equal to $+\Lambda$ on one side of the circle and $-\Lambda$ on the other, with a gauge field that is constant in the extra dimension (Fig. 1, right). One finds massless chiral modes localized at the mass defects with wavefunction overlap that vanishes exponentially fast with the size of the extra dimension. Thus the effective theory appears to have a $U(1)_A$ symmetry, even though there is no such symmetry in the five-dimensional action. However the fermions still “communicate” via a Chern-Simons operator that is generated on integrating out the bulk modes. This gives rise to bulk currents with a coefficient of this operator was shown to be topologically quantized $[7]$, which can jump at special coupling constants where the gap disappears in the bulk and the number of zeromodes at the defects change $[8]$. It is this quantized current which correctly reproduces the $U(1)_A$ anomaly $[9]$. The system is an example of what condensed matter physicists call a topological insulator, and the Chern-Simons current is the analogue of the Hall current in the integer quantum Hall effect.

Overlap fermions provide the effective four-dimensional theory for domain wall fermions in the limit of infinite extra dimension $[10, 11, 12, 13]$. What makes this approach so powerful is that it reproduces the benefits of taking the limit of infinite extra dimension in the 5D theory, without
Chiral Gauge Theories

Figure 1: Left: the conventional domain wall fermion construction, with negative and positive chiral modes bound to opposite mass defects in the extra dimension, with the gauge field $A$ is constant over the extra dimension. Center: the “waveguide” construction with the gauge symmetry spontaneously broken away from the near defect; this gives rise to a conventional vector-like gauge theory. Right: the gradient flow construction where gauge symmetry is unbroken but only fermions near one defect interact appreciably with physical gauge fields.

having to deal with the infinite extra degrees of freedom of the bulk fermions. Neuberger and Narayanan were able to derive overlap fermions by treating the fifth dimension as time, and using a Hamiltonian approach. The overlap operator $\mathcal{D}_V$ that they derive was shown \cite{13} to obey the Ginsparg-Wilson equation \cite{14}

$$\{D, \gamma_5\} = a\gamma_5 D,$$

where here I have restored the lattice spacing $a$. When $D$ is invertible, this equation is simpler to understand when written as $\{D^{-1}, \gamma_5\} = a\gamma_5$, which in a chiral basis has the general solution

$$D^{-1} = \begin{pmatrix} 0 & S_1 \\ -S_2^\dagger & 0 \end{pmatrix} + \frac{a}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The beauty of this equation is that the second term above encodes explicit chiral symmetry breaking in a way that correctly reproduces the axial anomaly and index theorem \cite{15} while still allowing for an exact symmetry of the action that ensures multiplicative mass renormalization \cite{16} and the absence of fine tuning. For the overlap operator $\mathcal{D}_V$ one has $S_1 = S_2$ and the desired continuum limit

$$\mathcal{D}_V \longrightarrow \mathcal{D} = \begin{pmatrix} 0 & D \cdot \sigma \\ D \cdot \sigma & 0 \end{pmatrix}.$$

The origin of the chiral symmetry breaking term for $\mathcal{D}_V$ may be thought of as the bulk fermion mass in the 5D domain wall formulation.

Now let us revisit the fermion operator in eq. (6) which is proposed to describe chiral fermions in the continuum. We immediately see an obstacle to transcribing it to the lattice: in order to correctly reproduce the $U(1)_A$ anomaly we would like to represent as a Ginsparg-Wilson operator of the more general form eq. (8), with the continuum limit eq. (6); however in this case the chiral symmetry breaking part of eq. (8) will break gauge symmetry, coupling the gauged left-handed fermions to the neutral right-handed ones. We thus face a choice: either (i) gauge the right-handed
“mirror” fermions and hope to decouple them, or else (ii) break gauge symmetry explicitly on the lattice. Both strategies have been followed in the literature. The gauge symmetry breaking scenario has been explored in [17], where there is evidence that the strategy works in perturbation theory. Some models invoking exotic dynamics to decouple mirror fermions have been shown not to work, while for others there is strong evidence for whether it can work or not. The theory we will consider here represents a new direction in the second class: we will gauge the right-handed fermions but have them behave as if they have vanishingly soft form factors, allowing them to (hopefully) decouple from the left-handed fermions which have conventional gauge interactions. Due to the simple geometrical treatment of chirality in the domain wall formulation we return there to take the first step toward constructing a chiral gauge theory constructing a chiral gauge theory.

Something special happens if the five dimensional fermion representation and mass terms are chosen so that at each defect one finds fermion zero-modes which are in an anomaly-free representation of the gauge group: the coefficient of the Chern-Simons term vanishes. This is a valuable piece of information, since any purported lattice regulator must have a built-in path to failure for anomalous fermion representations, and yet it is not easy to find a dynamical effect on the lattice that knows about anomaly cancellation. The vanishing of the Chern-Simons current which communicates between the fermions bound to the two mass defects is an example of such an effect. From the early 1990s such theories were studied as possible starting points for a regularization of chiral gauge theories. The simplest example of such a theory assumes two flavors of five-dimensional fermions with opposite sign masses, so that the 5D theory is $P$- and $T$-invariant and the zero-modes each look like massless Dirac fermions. This (or the 3D analogue) is precisely the Kane-Mele model for the Quantum Spin Hall Effect, proposed for condensed matter systems over a decade later [18], where chiral charges (spin) can be transported between the defects but not gauged charges (electric charge). A slightly less trivial anomaly cancellation model of greater interest to particle theorists is the 3-4-5 model [19].

While it is easy to construct models where the bulk gauge current vanishes algebraically, it is necessary to decouple the zero-modes at one wall from the gauge fields in order to obtain a chiral gauge theory instead of a vector-like one. This was attempted by the “waveguide” model, where the gauge symmetry was spontaneously broken away from one of the defects, but this was subsequently shown to give rise to new zero-modes and a vector-like theory (Fig. 1, center; see [20] for discussion). A new approach for localizing gauge fields was proposed in [21], the central subject of this talk and of a parallel talk given by D. Grabowska at this conference. Instead of introducing dynamical gauge fields in the extra dimension, the idea is to have a purely four-dimensional gauge field extended in the extra dimension via a gradient flow equation which is gauge covariant [22, 23]. An example of such a flow equation in the continuum is

$$\partial_s \mathcal{A}_\mu = \text{sgn}(s) D_\nu \mathcal{F}_{\mu\nu} , \quad \mathcal{A}_\mu(x,0) = A_\mu(x)$$

(10)

where $s$ is the coordinate of the extra dimension, $\mathcal{F}_{\mu\nu}(x,s)$ is the field strength tensor for the gauge field $\mathcal{A}_\mu(x,s)$, and the boundary value $A_\mu(x)$ is the 4D gauge field being integrated over in the path integral. The effect of this flow equation is most readily exhibited for a $U(1)$ gauge theory in two Euclidean dimensions, where we can decompose

$$\mathcal{A}_\mu(x,s) = \partial_\mu \omega(x,s) + \epsilon_{\mu\nu} \partial_\nu \lambda(x,s) ,$$

(11)
where $\lambda$ is the physical gauge field, $\omega$ is the gauge degree of freedom, and $\mu, \nu = 1, 2$. The solution is

$$\omega(x,s) = \omega(x,0), \quad \tilde{\lambda}(p,s) = e^{-p^2 s} \tilde{\lambda}(p,0),$$  \hspace{1cm} (12)

where $\tilde{\lambda}(p,s)$ is the Fourier transform of $\lambda(x,s)$. Evidently the physical gauge field experiences Gaussian damping, while the gauge degree of freedom is unaffected by flow. When the fifth dimension is discretized, the $s$ coordinate looks like a flavor index from the 4D perspective, and when coupled to this flowed gauge field it appears that as $|s|$ increases, each subsequent flavor of fermion has an increasingly soft form factor. The flavors of fermion behave as if they are bigger and less able to couple to high momentum gauge bosons the farther one travels from $s = 0$. The zeromodes at the far defect are therefore called “fluff” and appear to entirely decouple from the physical gauge fields in the limit of infinite extra dimension. The result is indicated on the right side of Fig. 1 where we have succeeded in localizing the gauge fields near one mass defect without breaking the gauge symmetry, or producing new sharp boundaries where a new fermion zeromodes might form.

This construction raises some questions: If the construction is manifestly gauge invariant, independent of the fermion representation, how is this consistent with $\Delta$ having a gauge variant phase $\delta[A]$ for anomalous 4D theories? Aren’t gaussian form factors as in eq. (12) a problem when continuing back to Minkowski spacetime, where $p^2$ need not be positive? The example of gauge flow given above ignores topology...since topology obviously cannot change under continuous flow, won’t the unwanted mirror fermions still couple to the fermions via instantons?

The answer to the first question is addressed in [21] and will only be summarized here: for an anomalous fermion representation there will be a nonzero Chern Simons operator induced by the bulk fermions, and it was shown that this term cannot look like a local 4D operator due to the gradient flow of the gauge fields. Thus there is no way this construction can be used to investigate anomalous 4D chiral gauge theories. For the second question: indeed, the theory looks sick at finite extra dimension, and this motivates taking the strict infinite extra dimension limit as the definition of the Euclidian theory, which in turn begs devising an overlap construction which has been presented in [24] and will be summarized below.

As for the third question: the flow takes the initial gauge field $A$ at $s = 0$ to a gauge field $A_*$ at the far defect; as the extra dimension gets large, $A_*$ will generally correspond to an attractive fixed point of the flow equation. Since the right side of the flow equation eq. (10) is the Euclidian equation of motion, it follows that all attractive fixed points of the flow equation will be stable gauge field solutions to the Euclidian equations of motion. Therefore for each topological sector with winding number $\nu$ there will be at least the fixed points corresponding to the exact $\nu$-instanton solutions (or the exact $\nu$-anti-instanton solutions for negative $\nu$). There are no fixed points expected containing instanton-anti-instanton pairs, however, and so the interaction of the mirror fermions at the far defect with the gauge field $A_*$ will indeed induce topological correlations with the ordinary fermions, but they will not be local or extensive, and it is not clear whether they spoil decoupling in the infinite volume limit.

The above discussion suggests that while our new construction may provide a plausible path to the fermion determinant $\Delta[A]$ for a chiral gauge theory, it will certainly require taking the strict limit of infinite extra dimension first. From our experience with vector-like theories, the way to do this is to construct the corresponding overlap operator for the 5D theory.
The overlap operator for a vector-like gauge theory takes a beautifully simple form \[ (12) \]
\[
D_V = 1 + \gamma_5 \epsilon(H)
\]
where \( \epsilon(x) = x / \sqrt{x^2} \) is the sign function, and \( H \) is the Wilson Hamiltonian,

\[
\gamma_5 H = \frac{1}{2} [\gamma_\mu (\nabla_\mu + \nabla^*_\mu) - \nabla_\mu \nabla^*_\mu] - m \equiv D_w - m
\]

where \( \nabla_\mu \) and \( \nabla^*_\mu \) are the forward and backward lattice covariant derivatives respectively. By using the property that \( \epsilon^2 = 1 \) it is easy to show that \( D_V \) obeys the Ginsparg-Wilson equation; from the definition of \( H_w \) it is also easy to show that \( D_V \) obeys the desired continuum limit, eq. (9). The way \( D_V \) can be derived from the conventional domain wall theory with an \( s \)-independent gauge field is to represent it in terms of the transfer matrix \( T \) for Hamiltonian evolution in the fifth dimension; the way \( \epsilon \) arises is as the limit

\[
\epsilon(H) = \lim_{L \to \infty} \frac{1 - T_L}{1 + T_L}
\]

where \( L \) is the number of sites in the fifth dimension.

In the model we are interested in \( A(x) \) is replaced by the flowed gauge field \( \mathcal{A}(x,s) \), which means that \( T \) is now a function of \( s \). If we were to make the incautious but simplifying assumption that the gauge field is roughly constant as \( A \) for half the lattice, and then quickly transitions to the fixed point \( A_* \) for the second half of the lattice, then we find

\[
D_\chi = 1 + \gamma_5 \epsilon_\chi
\]

where

\[
\epsilon_\chi = \lim_{L \to \infty} \frac{1 - T^{L/2}_L T^{L/2}_L}{1 + T^{L/2}_L T^{L/2}_L} = \left[ 1 - (1 - \epsilon_*) \frac{1}{1 + \epsilon \epsilon_*} (1 - \epsilon) \right].
\]

where \( \epsilon = \epsilon(H[A]) \) and \( \epsilon_* = \epsilon(H[A_*]) \). As discussed in [24] there are problems associated with the rapid change from \( A \) to \( A_* \), as well as subtleties in inverting the matrix \( (1 + \epsilon \epsilon_*) \), but we demonstrate that the above expression for \( D_\chi \) possesses two important properties: (i) it obeys the Ginsparg-Wilson equation, and (ii) it has the continuum limit

\[
D_\chi \quad \xrightarrow{a \to 0} \quad \begin{pmatrix} 0 & D \cdot \sigma \\ D_* \cdot \sigma & 0 \end{pmatrix}
\]

which looks similar to our starting point in eq. (6), except that the right-handed fermions, instead of being neutral, couple to the gauge field \( A_* \). They are what we call fluff. Since the flow equation is covariant, the fluff transforms under gauge transformations the same way as do the fermions, which is why gauge invariance is not broken by the diagonal terms in \( D_\chi \) which are mandated by the Ginsparg-Wilson equation. However, because \( A_* \) is the fixed point of the gradient flow equation, the fluff cannot interact with gauge bosons, although possibly with residual topology.

This bring us to the last topic of discussion: the nature of topology and gradient flow on the lattice. Unlike in the continuum, there is no notion of topology for general gauge fields on the
lattice, and if we use the Wilson action to generate our gradient flow, the only fixed points expected are pure gauge configurations; however other actions can be devised to preserve the topology of at least smooth field configurations \[25, 26, 27, 28, 29, 30, 31\]. It is tempting to use the Wilson action and topology destroying flow so that the fluff only sees pure gauge fields. The conclusion of \[24\], however, is that when \(A\) transitions abruptly to \(A\), one obtains an incorrect index theorem is \(\text{Tr} e \neq \text{Tr} e\), where these traces serve as a surrogate for gauge field winding number for overlap fermions. Whether adiabatic Wilson flow makes more sense is an open question.

Whether or not one is restricted to topology preserving flow equations, the overriding question is whether fluff successfully decouples or not. Hopefully progress can be made to better understand analytically the case of gradual gradient flow. An interesting computational question is whether or not the fluff develops topology-induced interactions with the fermions which persists to large volume. It would seem that the simplest theory to start with will be of the “Quantum Spin Hall Effect” type, namely a theory where the “chiral” gauge theory in fact consists of a single Dirac fermion localized at the \(s = 0\) defect, since in this case one has a chance of avoiding sign problems, and one thoroughly understands the expected phenomenology for the target 4D theory.

References


