Ginsparg-Wilson Relation and Ultralocality

Ivan Horváth*
Department of Physics, University of Virginia, Charlottesville, Virginia 22903
(Received 12 August 1998)

It is shown that it is impossible to construct a free theory of fermions on infinite hypercubic Euclidean lattice in four dimensions that (a) is ultralocal, (b) respects symmetries of hypercubic lattice, (c) has a corresponding kernel that satisfies $\mathbf{D}\gamma_5 + \gamma_5\mathbf{D} = \mathbf{D}\gamma_5\mathbf{D}$ (Ginsparg-Wilson relation), and (d) describes a single species of massless Dirac fermions in the continuum limit. [S0031-9007(98)07564-4]

PACS numbers: 11.15.Ha, 11.30.Rd

The quest for incorporating chiral symmetry in lattice regularized gauge theory has a long history. Many approaches with various degrees of depth and aesthetic appeal have been tried. Even though remarkable progress has been achieved during the last few years, it is hard to match a striking elegance and clarity of the picture that emerged during the last few months. So far, these developments are mostly relevant for the vectorlike case, but applications to chiral gauge theories appear to be around the corner.

At the heart of the above progress was the realization [1–3] that there exist potentially viable actions, satisfying the Ginsparg-Wilson (GW) relation [4]. If $\mathbf{D}$ is a lattice Dirac kernel, then the simplest form of GW relation is

$$\gamma_5\mathbf{D} + \mathbf{D}\gamma_5 = \mathbf{D}\gamma_5\mathbf{D}. \qquad (1)$$

Lüscher recognized that (1) can be viewed as a symmetry condition [5]. Unlike chiral symmetry, this new symmetry [which we call Ginsparg-Wilson-Lüscher (GWL) symmetry] involves a transformation that couples variables on different lattice sites and becomes a standard chiral symmetry only in the continuum limit. GWL symmetry has virtually the same consequences for the dynamics of the lattice theory as the chiral symmetry does on the dynamics in continuum. In particular, it guarantees the correct anomaly structure and the current algebra predictions for low energy QCD directly on the lattice [2,4–6]. There is no need for tuning to recover aspects of chiral symmetry, there are no complicated renormalizations, and there is no mixing between operators in different chiral representations [2]. Particularly striking are also the completely new avenues for studying topology on the lattice [1,6,7]. All of this can be discussed in the standard local field theory framework as a consequence of GWL symmetry. In view of the Nielsen-Ninomiya theorem [8], it is hard to imagine having things any better than this with respect to chiral issues in lattice QCD.

While all of this definitely holds a promise of extraordinary progress in the near future, the troubling history of chiral symmetry on the lattice indicates that it might come for a price. Indeed, one drawback of the known solutions of the GW relation is that they are not ultralocal, i.e., that the interaction between fermionic variables is nonzero for sites arbitrarily far away from each other [9]. This complicates the perturbation theory with such actions considerably and, also, one looses the obvious numerical advantages stemming from sparcity of the conventional operators such as the Wilson-Dirac operator. Moreover, while locality (exponential decay of interaction at large distances) can be ensured easily for free actions, it is usually not obvious in the presence of the gauge fields if the action is not ultralocal. Consequently, it would be much more preferable to work with ultralocal actions and the question arises whether GWL symmetry and ultralocality can at least in principle coexist.

In this Letter, it is argued that such hopes may not materialize. In particular, we prove that the GW relation (1) cannot be satisfied by a free ultralocal kernel defining a theory with appropriate continuum limit, and respecting the symmetries of the hypercubic lattice. Such a theorem can be extended to a more general GW relation $\gamma_5\mathbf{D} + \mathbf{D}\gamma_5 = 2\mathbf{D}\gamma_5\mathbf{R}\mathbf{D}$, with $\mathbf{R}$ being an ultralocal matrix, trivial in spinor space. Also, analogous statements hold in two dimensions. Discussion of these results as well as a more detailed account of the proof presented here will be given elsewhere [10].

Consider a system of 4-component fermionic degrees of freedom living on the sites of an infinite 4-dimensional hypercubic Euclidean lattice. Free theory of these fermions is described by some kernel $\mathbf{D}$ which can be uniquely expanded in the form

$$\mathbf{D}_{m,n} = \sum_{a=1}^{16} \mathbf{G}^a_{m,n}\Gamma^a. \qquad (2)$$

In the above equation m, n label the space-time lattice points and Γ^a's are the elements of the Clifford basis $\Gamma \equiv \{\mathbb{I}, \gamma_\mu, \gamma_5, \gamma_5\gamma_\mu, \sigma_{\mu\nu,(\mu<\nu)}\}$. Gamma matrices satisfy anticommutation relations $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu,\nu}\mathbb{I}$, and we define $\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$, $\sigma_{\mu\nu} \equiv \frac{i}{2}[\gamma_\mu, \gamma_\nu]$. Because of the completeness of Clifford basis on the space of 4×4 complex matrices, Eq. (2) describes arbitrary kernel and thus arbitrary quadratic action $\bar{\psi}\mathbf{D}\psi$.

The requirements of symmetry, ultralocality, and continuum limit constitute the set of restrictions on the above space of fermionic lattice theories. Before we proceed to implement them, let us note that hypercubic lattice structure is invariant under translations by arbitrary lattice vector and under the subset of $O(4)$ transformations—hypercubic rotations and reflections. We refer to the former as *translation invariance* and to the latter as *hypercubic invariance*.

Translation invariance and ultralocality.—Translation invariance restricts the form of the action substantially since it implies

$$\mathbf{G}^a_{m,n} = \mathbf{G}^a_{0,n-m} \equiv g^a_{n-m} \qquad a = 1,2,\ldots,16\,.$$

By ultralocality we mean that the fermionic variables do not interact beyond some finite lattice distance. Let us denote by C_N the set of all lattice sites contained in the hypercube of side $2N$, centered at $n = 0$, i.e., $C_N \equiv \{n{:}|n_\mu| \leq N, \mu = 1,\ldots,4\}$. One convenient way of defining ultralocality for translationally invariant actions is to require the existence of a positive integer N, so that $g^a_n = 0, \forall\, n \notin C_N, \forall\, a$. Translation invariance and ultralocality together imply the existence of the diagonal Fourier image of the space-time part of $\mathbf{D}$. In particular,

$$D(p) = \sum_{a=1}^{16} G^a(p)\Gamma^a, \tag{3}$$

where

$$G^a(p) \equiv \sum_{n\in C_N} g^a_n e^{ip\cdot n}. \tag{4}$$

$G^a(p)$ are thus the complex-valued periodic functions of lattice momenta $p \equiv (p_1,\ldots,p_4)$, whose Fourier series has a *finite* number of terms. It should be emphasized that there is a true mathematical equivalence between the set of all kernels satisfying translation invariance and ultralocality, and the set of kernels defined by Eqs. (3) and (4).

Hypercubic symmetry.—We will discuss hypercubic symmetry directly in the Fourier space which is convenient for our purposes. Let $\mathcal{H}$ be an element of the hypercubic group in defining representation and H the corresponding element of the representation induced on the hypercubic group by spinorial representation of $O(4)$. We require that the action $\bar\psi D\psi$ does not change under $\psi(p) \to H\psi(\mathcal{H}^{-1}p)$, $\bar\psi(p) \to \bar\psi(\mathcal{H}^{-1}p)H^{-1}$. This is equivalent to the requirement

$$D(p) = \sum_{a=1}^{16} G^a(p)\Gamma^a = \sum_{a=1}^{16} G^a(\mathcal{H}p)H^{-1}\Gamma^a H\,. \tag{5}$$

Since any hypercubic transformation $\mathcal{H}$ can be decomposed into products of reflections of single axis ($\mathcal{R}_\mu$) and exchanges of two different axes ($\mathcal{X}_{\mu\nu}$), it is sufficient to require invariance under these operations. Transformation properties of all the elements of the Clifford basis are determined by the fact that γ_μ transforms as p_μ (vector). In particular

$$R_\nu^{-1}\gamma_\mu R_\nu = \begin{cases} -\gamma_\mu, & \text{if } \mu = \nu\,; \\ \gamma_\mu, & \text{if } \mu \neq \nu\,, \end{cases}$$

and

$$X_{\rho\sigma}^{-1}\gamma_\mu X_{\rho\sigma} = \begin{cases} \gamma_\sigma, & \text{if } \mu = \rho\,; \\ \gamma_\rho, & \text{if } \mu = \sigma\,; \\ \gamma_\mu, & \text{otherwise.} \end{cases}$$

The elements of the Clifford basis naturally split into groups with definite transformation properties and the hypercubic symmetry thus translates into definite algebraic requirements on functions $G^a(p)$ some of which we will exploit.

Continuum limit.—Next, it is required that low energy physics happens only for $p \sim 0$, where it corresponds to a single massless relativistic Dirac fermion in the continuum. This implies the following local properties:

$$G^a(p) = \begin{cases} ip_\mu + O(p^2), & \text{if } \Gamma^a = \gamma_\mu\,; \\ O(p^2), & \text{if } \Gamma^a \neq \gamma_\mu, \forall\, \mu\,, \end{cases} \tag{6}$$

and the restriction that $D(p)$ has to be invertible away from the origin of the Brillouin zone (no doublers).

We now put forward the following definition.

Definition (set $\mathcal{U}$).—Let $\mu \in \{1,2,3,4\}$, and let $a \in \{1,2,\ldots,16\}$. Let further $G^a(p)$ be the complex valued functions of real variables p_μ, and let $D(p)$ be the corresponding matrix function constructed as in Eq. (3). We say that the 16-tuple $(G^1,\ldots,G^{16})$ belongs to the set $\mathcal{U}$ if and only if the following holds:

(α) $\exists\, C_N$ such that $G^a(p)$ has the form (4), $\forall\, a$.

(β) $D(p)$ satisfies condition (5).

(γ) $D(p)$ satisfies $D\gamma_5 + \gamma_5 D = D\gamma_5 D$.

(δ) $G^a(p)$ satisfy (6), $\forall\, a$.

(ϵ) $D(p)$ is invertible unless $p_\mu = 0 \pmod{2\pi}$, $\forall\, \mu$.

It should be noted that for every action (kernel) (2) satisfying our requirements there is a corresponding element of $\mathcal{U}$ and vice versa. If the requirement of ultralocality is replaced by a weaker condition of locality (at least exponential decay at large distances), then there exist free actions satisfying the rest of the conditions. However, there do not appear to be examples of ultralocal actions enjoying the same level of symmetry. In fact, we now prove the following statement.

Theorem.—Set $\mathcal{U}$ is empty.

Proof: We will proceed by contradiction. Assume that there is at least one element $(G^1,\ldots,G^{16}) \in \mathcal{U}$. To such an element we can assign a 16-tuple of functions of *single variable* $(\overline{G}^1,\ldots,\overline{G}^{16})$ by restricting $G^a(p)$ to the points $p \equiv (q,q,0,0)$, i.e.,

$$G^a(p) \xrightarrow{p=(q,q,0,0)} \overline{G}^a(q)\,.$$

We now investigate the consequences of conditions (α)–(ϵ) on restrictions $\overline{G}^a(q)$.

(α) As a consequence of Eq. (4), functions $\overline{G}^a(q)$ have Fourier series with a finite number of terms, i.e., there exist non-negative integers K, L, such that

$$\overline{G}^a(q) = \sum_{-L\leq m\leq K} \overline{g}^a_m e^{iq\cdot m}, \quad \forall\, a\,. \tag{7}$$

(β) Consider the terms in $D(p)$ of the form $iB_\mu(p)\gamma_\mu$. Invariance under reflections implies

$$B_\mu(\ldots,-p_\nu,\ldots) = \begin{cases} -B_\mu(\ldots,p_\nu,\ldots), & \mu = \nu\,; \\ +B_\mu(\ldots,p_\nu,\ldots), & \mu \neq \nu\,. \end{cases}$$

Applying this to the reflection of p_3 or p_4, we have

$$\overline{B}_4(q) = B_4(q,q,0,0) = -B_4(q,q,0,0) = 0\,,$$

and similarly $\overline{B}_3(q) = 0$. Furthermore, since under $\mathcal{X}_{12}$, γ_1 exchanges with γ_2, we must have

$$B_1(q,q,0,0) = B_2(q,q,0,0) \equiv \overline{B}(q)\,.$$

Next, consider the term $C(p)\gamma_5$. Since $\gamma_5 \to -\gamma_5$ under $\mathcal{R}_\mu$, it is required that

$$C(\ldots,-p_\mu,\ldots) = -C(\ldots,p_\mu,\ldots), \quad \forall\ \mu\,.$$

Reflecting the component p_4, for example, this gives

$$\overline{C}(q) = C(q,q,0,0) = -C(q,q,0,0) = 0\,.$$

For the terms of the form $iE_\mu(p)\gamma_5\gamma_\mu$ invariance under reflections demands

$$E_\mu(\ldots,-p_\nu,\ldots) = \begin{cases} +E_\mu(\ldots,p_\nu,\ldots), & \mu = \nu\,; \\ -E_\mu(\ldots,p_\nu,\ldots), & \mu \neq \nu\,, \end{cases}$$

and using similar arguments as above, we can infer from this that $\overline{E}_\mu(q) = 0$, $\forall\ \mu$. Finally, consider the terms $F_{\mu\nu}\sigma_{\mu\nu}$. Invariance under reflections implies

$$F_{\mu\nu}(\ldots,-p_\rho,\ldots) = \begin{cases} -F_{\mu\nu}(\ldots,p_\rho,\ldots), & \rho = \mu \text{ or } \nu\,; \\ +F_{\mu\nu}(\ldots,p_\rho,\ldots), & \text{otherwise}, \end{cases}$$

which in turn ensures that $\overline{F}_{\mu\nu}(q) = 0$, except for $\overline{F}_{12}(q)$. However, under the exchange $\mathcal{X}_{12}$ of p_1 and p_2 we have $\sigma_{12} \to -\sigma_{12}$, while $\overline{F}_{12}(q) \to \overline{F}_{12}(q)$, implying that even this term has to vanish. Summarizing the relevant implications of hypercubic symmetry, restriction $\overline{D}(q)$ of $D(p)$ must have the form

$$\overline{D}(q) = (1 - \overline{A}(q))\mathbb{1} + i\overline{B}(q)(\gamma_1 + \gamma_2)\,. \tag{8}$$

(γ) GW relation for $\overline{D}(q)$ given in Eq. (8) takes a simple form

$$\overline{A}^2 + 2\overline{B}^2 = 1\,. \tag{9}$$

(δ) The local properties (6) imply

$$\overline{A}(q) = 1 + O(q^2) \qquad \overline{B}(q) = q + O(q^2)\,. \tag{10}$$

To proceed, we will rely on the lemma stated below this proof. According to the lemma, the solutions of Eq. (9) that have the form (7) (with some minimal K,L) exist only if $K = L$. If $K = L = 0$ (case of constant functions), then condition (10) cannot be satisfied and to avoid contradiction, we have to assume that $K = L > 0$. If that is the case, then the lemma states that the necessary (but not sufficient) condition for $(\overline{A},\overline{B})$ to be the solution of Eq. (9) is that only the highest frequency modes are present in their Fourier expansion, i.e.,

$$\overline{A}(q) = \overline{a}_{-K}e^{-iq\cdot K} + \overline{a}_K e^{iq\cdot K},$$

$$\overline{B}(q) = \overline{b}_{-K}e^{-iq\cdot K} + \overline{b}_K e^{iq\cdot K}.$$

Conditions (10) then dictate uniquely what the coefficients in the above equations have to be. In particular, $\overline{a}_{-K} = \overline{a}_K = 1/2$ and $\overline{b}_{-K} = -\overline{b}_K = i/2K$, which corresponds to $\overline{A}(q) = \cos(Kq), \overline{B}(q) = \sin(Kq)/K$. For these functions we have

$$\overline{A}^2 + 2\overline{B}^2 = \cos^2(Kq) + \frac{2}{K^2}\sin^2(Kq)\,,$$

and consequently, Eq. (9) can be satisfied only if $2/K^2 = 1$. However, there is no positive integer K so that this condition is satisfied. We have therefore arrived at the contradiction with the existence of $(G^1,\ldots,G^{16}) \in \mathcal{U}$ and the proof is thus complete.

In essence, the above proof relies on two major ingredients: First is the fact that it is sufficient to consider a single periodic direction in the Brillouin zone and that the hypercubic symmetry is powerful enough to render the problem tractable. The second ingredient is a perhaps surprising result that periodic solutions of equations of type (9) either involve a single Fourier component or infinitely many of them. This is summarized by the following lemma, whose complete proof will be given in the detailed account of this work [10].

Lemma.—Let K,L be non-negative integers and d a positive real number. Consider the set $\mathcal{F}^{K,L}$ of all pairs of functions $[A(q),B(q)]$ that can be written in the form

$$A(q) = \sum_{-L\le n\le K} a_n e^{iq\cdot n} \qquad B(q) = \sum_{-L\le n\le K} b_n e^{iq\cdot n},$$

where $q \in \mathbb{R}, n \in \mathbb{Z}$, and $a_n, b_n \in \mathbb{C}$ are such that a_K, b_K do not vanish simultaneously and a_{-L}, b_{-L} do not vanish simultaneously. Further, let $\mathcal{F}_d^{K,L} \subset \mathcal{F}^{K,L}$ denote the set of all solutions on $\mathcal{F}^{K,L}$ of the equation

$$A(q)^2 + dB(q)^2 = 1\,. \tag{11}$$

Then the following holds:

(a) If $K = L = 0$, then

$$\mathcal{F}_d^{0,0} = \{[a_0,b_0]{:}a_0^2 + db_0^2 = 1\}\,.$$

(b) If $K = L > 0$, then $\mathcal{F}_d^{K,K} = \{[A(q),B(q)]\}$, with

$$A(q) = a_{-K}e^{-iq\cdot K} + a_K e^{iq\cdot K},$$

$$B(q) = b_{-K}e^{-iq\cdot K} + b_K e^{iq\cdot K},$$

and

$$a_K = ci\sqrt{d}\,b_K \qquad a_{-K} = \frac{c}{4i\sqrt{d}\,b_K} \qquad b_{-K} = \frac{1}{4db_K}\,,$$

where $b_K \neq 0$, $\sqrt{d} > 0$, and $c = \pm 1$.

(c) If $K \neq L$, then $\mathcal{F}_d^{K,L} = \varnothing$.

Outline of the proof.—Using the completeness and othogonality of the Fourier basis, Eq. (11) is equivalent to the following set of conditions on Fourier coefficients:

$$\sum_{\substack{-L\le n\le K\\ -L\le k-n\le K}} a_n a_{k-n} + d \sum_{\substack{-L\le n\le K\\ -L\le k-n\le K}} b_n b_{k-n} = \delta_{k,0}\,,$$

where $-2L \le k \le 2K$.

Case (a) is obvious and we start with case (b): The idea is to explicitly solve the above equations by analyzing them in the appropriate sequence. We start with the group $K \le k \le 2K$, which involves only coefficients of nonnegative frequencies. By induction, starting from $k = 2K$ and continuing down, it is possible to show that this group of conditions is equivalent to

$$a_n = ci\sqrt{d}\, b_n \qquad \sqrt{d} > 0,\ c = \pm 1\,, \tag{12}$$

where $n = 0, 1, \ldots, K$. Similarly, analyzing the group involving only coefficients of nonpositive frequencies, i.e., $-2K \le k \le -K$, we arrive at

$$a_{-n} = \bar{c} i\sqrt{d}\, b_{-n} \qquad \sqrt{d} > 0,\ \bar{c} = \pm 1\,, \tag{13}$$

for $n = 0, 1, \ldots, K$. Inserting results (12) and (13) in condition for $k = 0$, implies $\bar{c} = -c$, and consequently,

$$a_0 = b_0 = 0\,.$$

Using these results, we can start induction at $k = K - 1$ to show that conditions for $1 \le k \le K - 1$ lead to

$$b_{-n} = 0 = a_{-n} \qquad n = 1, 2, \ldots, K - 1\,,$$

and, analogously, for $-K + 1 \le k \le -1$ we arrive at

$$b_n = 0 = a_n \qquad n = 1, 2, \ldots, K - 1\,.$$

Finally, the last condition that was not fully exploited is the one for $k = 0$, which now simplifies to

$$b_K b_{-K} = \frac{1}{4d}\,.$$

The above steps establish the result (b).

Case (c).—Technically, this is arrived at in a completely analogous manner to case (b). However, due to the asymmetry between positive and negative frequencies, the equation for $k = 0$ can never be satisfied.

Let us close by noting that in the proof of the Theorem, condition (ϵ) was not used at all. In other words, there are no acceptable ultralocal solutions of (1) with or without doublers. This is not true if the requirement of hypercubic symmetry is relaxed. In that case, there exist ultralocal solutions with doublers and it is still an open question whether doubler-free solutions do exist. Since breaking the hypercubic symmetry carries with itself the necessity of tuning to recover rotation invariance in the continuum limit, it is not obvious whether such a possibility would be practically viable. On the other hand, theoretically it would be quite interesting to know whether hypercubic symmetry can be traded for GWL symmetry.

The author thanks Wolfgang Bietenholz, Mike Creutz, Peter Hasenfratz, Robert Mendris, Martin Mojžiš, Tony Kennedy, and Hank Thacker, whose input was useful at various stages of this work. This work was supported in part by the U.S. Department of Energy under Grant No. DE-AS05-89ER40518.

*Email address: ih3p@virginia.edu

[1] P. Hasenfratz, V. Laliena, and F. Niedermayer, Phys. Lett. B **427**, 125 (1998).
[2] P. Hasenfratz, Nucl. Phys. **B525**, 401 (1998).
[3] H. Neuberger, Phys. Lett. B **427**, 353 (1998).
[4] P. Ginsparg and K. Wilson, Phys. Rev. D **25**, 2649 (1982).
[5] M. Lüscher, Phys. Lett. B **428**, 342 (1998).
[6] S. Chandrasekharan, hep-lat/9805015.
[7] T. W. Chiu and S. V. Zenkin, hep-lat/9806019; T. W. Chiu, Phys. Rev. D **58**, 074 511 (1998).
[8] H. B. Nielsen and M. Ninomiya, Phys. Lett. **105B**, 219 (1981); Nucl. Phys. **B185**, 20 (1981); **B195**, 541(E) (1982).
[9] W. Bietenholz, hep-lat/9803023.
[10] I. Horváth (to be published).