Eigenvalue Inequalities for Fermions in Gauge Theories

Cumrun Vafa[1][*] and Edward Witten[1,2][*][†]

1 Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, USA
2 Institute for Advanced Study, Princeton, NJ 08540, USA

Abstract. We show that QCD with a sufficient number of fermions of zero bare mass has physical massless particles. That result also follows from triangle anomalies, so only our method is novel. Our method involves proving special cases of recently conjectured paramagnetic inequalities for fermions. The proofs are simple applications of the Atiyah-Patodi-Singer theorem on spectral flow.

I. Introduction

Despite intensive study over the years, many fundamental properties of quantum chromodynamics – and strongly interacting gauge theories in general – are imperfectly understood. These include questions of chiral symmetry breaking, dynamical mass generation, and confinement.

There has recently been progress in understanding the properties of strongly interacting gauge theories through the development of rigorous inequalities [1–6]. In particular [2,3], a surprising amount of information about symmetry realization in parity-conserving, vector-like gauge theories (like QCD) follows from relatively simple facts about fermion determinants and propagators in Euclidean space.

In these theories, after integrating out the fermion fields, we are left with a measure for the gauge field integration

$$d\mu(A) = \frac{1}{Z} e^{-[(1/4g^2)\int d^4x \operatorname{Tr} F_{\mu\nu}F_{\mu\nu}]} \det(\not{D} + M)\pi d A_\mu^a(x), \tag{1}$$

which is an ordinary, real, positive measure. (The positivity of the measure depends on the fact that the fermion determinant is positive in four dimensions for fermions in a real representation of the gauge group; see, for instance [3].) Any inequality

that holds pointwise continues to hold after integrating with respect to a positive measure. So, in a vector-like theory, an inequality among fermion matrix elements that holds in an arbitrary background gauge field A also holds, after integrating over A, in the exact quantum field theory. Such inequalities have been used to argue that the axial symmetries are spontaneously broken [2] and the vector symmetries are unbroken [3] in theories like QCD.

In this paper, we will use the same general strategy to prove another significant fact about vector-like theories. We will prove that if there are a sufficient number of quark flavors of zero bare mass (we need at least four flavors), then there is no mass gap in a channel with the quantum numbers of $q\bar{q}$. This absence of mass gap may reflect the existence of physical massless quarks (if there is no confinement) or massless mesons or baryons (if there is confinement). Our methods do not elucidate which option the theory chooses; of course, the real world is believed to prefer the second option (the pion being massless if the quark bare mass is zero). We will establish our result about the absence of mass gap by proving some special cases of the paramagnetic inequalities that have been proposed for Dirac fermions [7].

That there is no mass gap if the quark bare masses are zero has been shown previously, by 't Hooft [8], through consideration of the triangle anomaly. In fact, his method gives more information than we obtain (his argument works for two or more massless flavors; and the anomaly matching gives considerable insight about the quantum numbers of massless particles). Nonetheless, we believe our results are interesting for several reasons. The absence of mass gap in the theory with massless quarks is such a significant result that it is worthwhile to understand it from various viewpoints. We suspect our approach may ultimately have generalizations that may not be possible for the anomaly matching argument. Our results also have some bearing on chiral symmetry breaking, in a way that will be explained in Sect. II; in a certain sense, we will show that the opposite of chiral symmetry breaking (a depletion of the Dirac spectrum near $\lambda = 0$) cannot occur. The paramagnetic inequalities relevant in our work also have interest in their own right. Finally, there is at least one situation in which our method gives results that would not follow from consideration of anomalies; this arises in $2+1$ dimensions, where there is apparently no direct analogue of 't Hooft's anomaly-based argument.

The organization of this paper is as follows. In Sect. II, we describe the inequalities we will prove and their relevance to QCD. In Sect. III, we prove the inequalities. In Sect. IV, we briefly discuss the application of our results to gauge theories in $2+1$ dimensions. The paper is written in such a way that readers only interested in the applications can read only Sects. II and IV, and readers only interested in the mathematics can read only Sect. III.

II. The Inequalities

Consider a theory with n flavors of massless fermions $\psi_1, \psi_2, \ldots, \psi_n$ transforming in the same representation R of the gauge group G. For k even and equal to or less than n, let

$$S(x_1, \ldots, x_k) = -i^k \langle \Omega | \bar{\psi}_1 \psi_2(x_1) \bar{\psi}_2 \psi_3(x_2) \ldots \bar{\psi}_k \psi_1(x_k) | \Omega \rangle . \tag{2}$$

(The phase is for later convenience.) We will be interested in the quantity

$$S(k) = \int d^4x_1 \ldots d^4x_{k-1} S(x_1, x_2, \ldots, x_{k-1}, 0) \tag{3}$$

integrated over $k-1$ variables, with one held fixed.

If the theory has a mass gap, Green's functions show exponential clustering and the integral in (3) converges. We will show that the theory with $n \geq 4$ has no mass gap by showing that this integral diverges for $k \geq 4$. We have chosen the fermion flavors in (2) to be distinct so that S has no disconnected contribution. A disconnected contribution would invalidate the argument by causing the integral in (3) to diverge whether or not there is a mass gap. It is because of this that our method yields interesting results only for $n \geq 4$.

In free massless field theory, $S(k)$ can be evaluated explicitly; it is

$$S_0(k) = d(R) \int \frac{d^4p}{(2\pi)^4} \operatorname{Tr}\left(\frac{1}{\not{p}}\right)^k = 4d(R) \int \frac{d^4p}{(2\pi)^4}\left(\frac{1}{p^2}\right)^{k/2}, \tag{4}$$

where $d(R)$ is the dimension of the representation R. Notice that for $k \geq 4$, $S_0(k)$ is infrared divergent. This divergence is made possible by the absence of mass gap in the free field theory. $S_0(k)$ is also ultraviolet divergent for $k \leq 4$. This ultraviolet divergence is of no interest for our purposes. It can be eliminated, for example, by Pauli-Villars regularization of the fermion kinetic energy, $\not{D} \to \not{D}\left(1 + \frac{1}{\Lambda^2}(i\not{D})^2\right)$. This inserts a factor $(1 + p^2/\Lambda^2)^{-k/2}$ in (4), leaving only the infrared divergence of interest.

When necessary, we will introduce such a regularization later. (For rigor, we should also introduce a suitable cutoff in the gauge field integration, such as the gauge invariant Pauli-Villars procedure that has been rigorously formulated by Asorey and Mitter [9].)

We will prove the absence of mass gap in the interacting theory by proving that

$$S(k) \geq C_k S_0(k), \tag{5}$$

where C_k is a definite, finite, computable, positive constant. Inequality (5) proves the absence of a mass gap in the theory of $n \geq 4$, because it shows that for $k \geq 4$ the integral $S(k)$ in the interacting theory has the same infrared divergence as $S_0(k)$.

To prove (5), we will prove some inequalities for fermions in an arbitrary background gauge field A. Let $S^A(x, y) = \langle x|\not{D}_A^{-1}|y\rangle$ be the fermion propagator in the background field. We will work in an arbitrary finite volume V. Let $S_A(k)$ be the analogue of $S(k)$ in the background field A,

$$S_A(k) = i^{(k+2)}\frac{1}{V}\int d^4x_i \ldots d^4x \langle \bar\psi_1\psi_2(x_1) \ldots \bar\psi_k\psi_1(x_k)\rangle^A$$

$$= i^k\frac{1}{V}\int d^4x_1 \ldots d^4x_k \operatorname{Tr} S^A(x_1, x_2) \ldots S^A(x_k, x_1) = \frac{1}{V}\operatorname{Tr}\left(\frac{1}{i\not{D}_A}\right)^k. \tag{6}$$

We will prove that

$$S_A(k) \geq C_k S_0(k), \tag{7}$$

with a constant C_k that depends on k but not on A. Since $S(k) = \int d\mu(A) S_A(k)$, where $d\mu(A)$ is the normalized, positive measure of Eqs. (1), (7) implies (5).

Actually, consider the eigenvalue problem $i\not{D}\psi = \lambda\psi$. Let $\lambda_1, \lambda_2, \lambda_3 \ldots$ be the eigenvalues of $i\not{D}$ in order of ascending absolute value. Then (6) can be expressed in terms of λ_i,

$$S_A(k) = \frac{1}{V} \sum_{i=1}^{\infty} \lambda_i^{-k}. \tag{8}$$

In Sect. III, we will prove that in a box of size L (so $V = L^4$),

$$\lambda_1 \leq \frac{C}{L} = CV^{-1/4}, \tag{9}$$

where C is a universal constant, independent of the gauge field A. Since $S_A(k) \geq \frac{1}{V} \lambda_1^{-k}$, this is adequate to prove that $S_A(k) \geq C^{-k} V^{(k-4)/4}$; since $S(k)$ is the average of $S_A(k)$ with a positive measure, it obeys the same lower bound,

$$S(k) \geq C^{-k} V^{(k-4)/4}. \tag{10}$$

For $k > 4$, this shows that $S(k)$ is infrared divergent, blowing up as $V \to \infty$. In fact, since the free Green's function $S_0(k)$ has the same asymptotic behavior as the right-hand side of (10),

$$S_0(k) \sim V^{(k-4)/4}, \qquad k > 4 \tag{11}$$

inequality (10) is equivalent to our previous statement (5). Inequality (9), which will be our first goal in Sect. III, therefore implies the absence of a mass gap if there are at least six massless flavors, so that we may take $k = 6$. At first sight, it is surprising that so much follows from a bound on λ_1 alone.

For $k = 4$, we need a little more care to separate the infrared divergence from the ultraviolet divergence. If one adopts Pauli-Villars regularization, $\not{D} \to \not{D}(1 + (i\not{D})^2/\Lambda^2)$, then (8) is replaced by

$$S_A(k) = \frac{1}{V} \sum_{i=1}^{\infty} \frac{1}{\lambda_i^k} \frac{1}{\left(1 + \dfrac{\lambda_i^2}{\Lambda^2}\right)^k}. \tag{12}$$

In Sect. III, we will prove, with a little extra work, an inequality analogous to (9) for excited states,

$$\lambda_n \leq C n^{1/4} V^{-1/4}. \tag{13}$$

It then follows that

$$S_A(4) \geq \frac{1}{V} \sum_{n=1}^{\infty} \frac{V}{C^4 n} \frac{1}{\left(1 + \dfrac{C^2 n^{1/2}}{V^{1/2} \Lambda^2}\right)^4} \geq \frac{1}{C^4} \ln\left(\frac{V\Lambda^4}{C^4}\right), \tag{14}$$

and $S(4)$ automatically obeys the same lower bound, so

$$S(4) \geq \frac{1}{C^4} \ln\left(\frac{V\Lambda^4}{C^4}\right). \tag{15}$$

Thus $S(4)$ diverges as $V \to \infty$ for fixed Λ; this divergence is at least logarithmic. Inequality (15) is again equivalent to our previous statement (5) since the free Green's function shows exactly the same logarithmic behavior,

$$S_0(4) \overset{\Lambda \to \infty}{\sim} \ln(V \Lambda^4). \tag{16}$$

Inequality (15) shows that there is no mass gap if there are four or more flavors of zero bare mass.

Now let us discuss the connection with chiral symmetry breaking. Here we may be less rigorous, since we can prove no theorems. Turning on a small bare mass m, the induced vacuum expectation value of $\bar\psi\psi$ is

$$\langle \bar\psi\psi \rangle = \frac{1}{V} \mathrm{Tr} \frac{1}{m - \slashed{D}}. \tag{17}$$

If the density of eigenvalues of $i\slashed{D}$ is $V\varrho(\lambda)$, this is

$$\langle \bar\psi\psi \rangle = \int_{-\infty}^{\infty} d\lambda \varrho(\lambda) \frac{1}{m + i\lambda}. \tag{18}$$

We want the limit of $\langle \bar\psi\psi \rangle$ as $m \to 0$. Since $\lim\limits_{m \to 0} \dfrac{1}{m + i\lambda} = \pi\delta(\lambda) - iP\dfrac{1}{\lambda}$ (P being the principal value symbol), and since $\varrho(\lambda) = \varrho(-\lambda)^1$, we arrive at the formula [10],

$$\lim_{m \to 0} \langle \bar\psi\psi \rangle = \pi\varrho(0). \tag{19}$$

The spectral density $\varrho_0(\lambda)$ of free field theory obeys $\varrho_0(\lambda) \sim \lambda^3$ and $\varrho_0(0) = 0$. To establish chiral symmetry breaking one must show an accumulation of eigenvalues at $\lambda = 0$, giving $\varrho_A(0) \neq 0$ for the spectrum of the Dirac operator for the relevant gauge fields A.

Although we cannot prove this, we can show that the opposite does not occur; there is no *depletion* of Dirac eigenvalues near $\lambda = 0$. The behavior of the free Dirac eigenvalues is $\lambda_n \sim (n/V)^{1/4}$, so (13) says that the interacting eigenvalues are no bigger than an overall constant times the free ones. This means

$$\int_0^\lambda d\lambda' \varrho_A(\lambda') \geq \tilde{C} \int_0^\lambda d\lambda' \varrho_0(\lambda')$$

for any λ and some suitable $\tilde{C}$. Up to an overall constant, the spectral density near 0 for the Dirac operator with *any* gauge field is at least as big as the free one.

So far, we have merely stated that the theory with at least four massless fermion flavors has no mass gap. To lowest order in $1/N$ (but not generally), we can be more specific. We can be more specific. If $S(k)$ is cut in any channel, the intermediate state has the flavor quantum numbers of a $\bar{q}q$ pair [in a flavor non-singlet combination since we started with distinct fermion species in Eq. (2)]. Our results really imply the absence of mass gap for states in the adjoint representation of the flavor group.

It would perhaps be worthwhile to conclude this section by clearing up a point that may be confusing. In massless QCD, the most singular infrared divergence in

---

1   This follows from simple Dirac algebra: if λ is an eigenvalue, so is $-\lambda$, since if $i\slashed{D}\psi = \lambda\psi$, then $i\slashed{D}(\gamma_5\psi) = -\lambda(\gamma_5\psi)$

$S(k)$ presumably arises from one pion intermediate states in the various channels. We are accustomed to thinking that $\langle\Omega|\bar{\psi}\psi|\pi\rangle=0$. This is true in the conventional chiral vacuum with $\langle\Omega|\bar{\psi}\psi|\Omega\rangle\neq0$, $\langle\Omega|\bar{\psi}i\gamma_5\psi|\Omega\rangle=0$. In a chirally rotated vacuum with $\langle\Omega|\bar{\psi}i\gamma_5\psi|\Omega\rangle\neq0$, one can have $\langle\Omega|\bar{\psi}\psi|\pi\rangle\neq0$. In key formulas as of the preceding discussion such as Eq. (8), we did not introduce explicit chiral symmetry breaking [which would correspond, for instance, to taking $\lambda_i\rightarrow\lambda_i+im$ in (8) and later taking $m\rightarrow0$], so if chiral symmetry is spontaneously broken, our formulas correspond to averaging over all possible chiral vacuum states, including states with $\langle\Omega|\bar{\psi}\psi|\pi\rangle\neq0$. It is in fact in such states that the most singular possible contribution to S is expected to arise, and the most singular contribution arises from the maximum possible number of pion poles in the various channels.

III. Proof of the Inequalities

In this section, we will prove the inequalities whose implications for QCD were discussed in the previous section.

Consider an arbitrary compact Riemanian manifold M (without boundary) of dimension d, with metric tensor g_{ij}. Let A be an arbitrary gauge field on this manifold (with any compact gauge group G). Let $i\not{D}_A$ be the Dirac operator for fermions ψ in some representation T of G, and let λ_i be the eigenvalues of $i\not{D}_A$, in order of ascending absolute value. We will prove

$$\lambda_1\leqq C,\tag{20}$$

with a constant C that depends on M and on the metric g_{ij} but not on the gauge field A. More generally, we will prove

$$\lambda_n\leqq C'n^{1/d},\tag{21}$$

with some constant C' that depends on M but not on A or on n.

If one rescales the metric of M, $g_{ij}\rightarrow t^2g_{ij}$, the volume V scales as $V\rightarrow t^dV$, and the Dirac eigenvalues scale as $\lambda_i\rightarrow\lambda_i/t=\lambda_iV^{-1/d}$. Inequalities (20) and (21) therefore mean that for manifolds M of fixed shape and variable volume V, $\lambda_1<\bar{C}V^{-1/d}$, $\lambda_n<\bar{C}(n/V)^{+1/d}$. The inequalities were used in this form (with $d=4$) in Sect. II.

For an arbitrary closed manifold M with arbitrary metric, we will show that (20) and (21) hold with *some* constants C and C'. For the special case of a torus with flat metric we will determine the best constant in (20) [but not in (21)].

We will first prove the desired inequalities for odd d. The proof will depend on simple considerations about the spectral flow of a family of Dirac operators [11]; the relevant facts have been discussed in the physics literature [12]. For any odd $d>1$, $\pi_d(\mathrm{SU}(N))=Z$ for large enough N. Pick such a value of N. The fermion field $\psi_{\alpha a}$ already has a spin index α and a G index a; we now take N copies of the fermi field, giving it a new index $i=1\ldots N$. Embedding G in $G\times\mathrm{SU}(N)$ in the trivial way, we regard A as a $G\times\mathrm{SU}(N)$ gauge field, and we let $i\not{D}_A$ act on the fermion field $\psi_{\alpha ai}$ that lives in an enlarged Hilbert space. The effect of this on the Dirac spectrum is trivial; the eigenvalues are the same as before, but the multiplicity is bigger by a factor of N.

Since $\pi_d(\mathrm{SU}(N))=Z$, we can choose a smooth mapping $U:M\rightarrow\mathrm{SU}(N)$ of non-zero winding number; for definiteness, choose the winding number to be one.
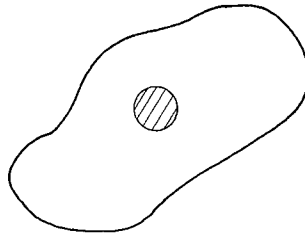
Fig. 1. A d dimensional ball B^d can be embedded in an arbitrary d manifold M. The topological classification of mappings $U : M \to \mathrm{SU}(N)$ that are constrained to be constant outside B^d is given by $\pi_d(\mathrm{SU}(N))$

[Perhaps it is worthwhile to specify that maps of non-zero winding number exist on any closed manifold, not necessarily a sphere. For instance, one can choose $U(x)$ to equal 1 outside a ball, as in Fig. 1.] We will think of U as an operator acting on the fermion field $\psi_{\alpha a i}$. Now, we define the operator

$$X = U^{-1} i\not{D}_A U - i\not{D}_A = i \sum_{j=1}^{d} \gamma^j U^{-1} \frac{\partial U}{\partial \phi^j}, \tag{22}$$

where ϕ^j are a local coordinate system on M. Since M is compact and U is a smooth function, X is a bounded operator. Let

$$C = \mathrm{norm}\, X . \tag{23}$$

Notice that X is independent of the gauge field A, as the last formula in (22) makes clear. Therefore, C is independent of A.

We will study the one parameter family of operators

$$Q(t) = i\not{D}_A + tX . \tag{24}$$

Notice that $Q(0) = i\not{D}_A$, while $Q(1) = U^{-1} i\not{D}_A U$, which is unitarily equivalent to $Q(0)$. Hence the spectrum of $Q(1)$ coincides with the spectrum of $Q(0)$. However, according to the Atiyah-Patodi-Singer theorem, as t is varied smoothly from 0 to 1, the individual eigenvalues of $Q(t)$ do not return to their starting points but are shifted by one or more units (Fig. 2). The eigenvalues are shifted up or down depending on whether U has winding number one or minus one. The magnitude of the shift is the magnitude of the winding number of U times the dimension q of the representation T^2.

The proof of this statement can be found in the original reference [11] and also in the physics literature [12]. Here is a brief sketch. We look at a Dirac operator on the $d+1$ dimensional manifold $M \times R$, R being the real line. Parametrize R by a variable τ. Let $t(\tau)$ be a slowly varying, monotonic function such that $t(-\infty) = 0$ and $t(+\infty) = 1$. On $M \times R$ we take the gauge field $A_i(\phi, \tau) = A_i(\phi) - it(\tau)U^{-1}\partial_i U$, $A_\tau(\phi, \tau) = 0$, and we look at the Dirac operator

$$D = i \sum_{j=1}^{d} \gamma^j(\partial_j + iA_j + tU^{-1}\partial_j U) + i\gamma^d \frac{\partial}{\partial \tau} = Q(t) + i\gamma^d \frac{\partial}{\partial \tau} . \tag{25}$$

---

2   This factor appears because as far as $\mathrm{SU}(N)$ is concerned, the fermion field consists of q multiplets in the fundamental representation of $\mathrm{SU}(N)$
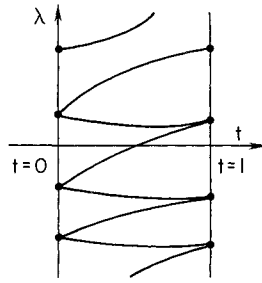
Fig. 2. The spectrum of $Q(t)$ for $0 \leqq t \leqq 1$ in a case where $N = 2$, the original fermion representation has dimension $q = 1$ (A is an abelian gauge field) and U has winding number $k = 1$. Several points should be noted. $Q(1)$ and $Q(0)$ are unitarily equivalent and have the same spectrum. At $t = 0$ and $t = 1$, but not for other t, the spectrum of $Q(t)$ is doubled since (with $N = 2$) we took two copies of the original Dirac field in beginning the construction described in the text. Since $qk = 1$, the Atiyah-Patodi-Singer theorem predicts an upward flow of the eigenvalues by one unit between $t = 0$ and $t = 1$

Now, the first observation is that by the index theorem [13] the index of D equals the winding number of U times the dimension of T, so if U has non-zero winding number, D has at least one zero eigenvalue. Second, if $t(\tau)$ is slowly varying, the equation $D\psi = \lambda\psi$ can be solved in an adiabatic approximation. Let the eigenvalues of $Q(t)$ be $\lambda_i(t)$ with eigenfunctions $\psi_i^t(\phi)$. Then, since $\{\gamma_d, Q(t)\} = 0$, the hermitian operator $-i\gamma^d Q(t)$ has eigenvalues $\pm\lambda_i(t)$; call the eigenfunctions $\bar\psi_i^t(\phi)$. To solve $D\psi = 0$, we try $\psi(\phi, t) = F(t)\bar\psi_i^t(\phi)$ and find $\dfrac{dF}{d\tau} = \pm\lambda_i(t)F(\tau)$; the solution is

$$F(\tau) = \exp \pm \int_0^\tau d\tau' \lambda_i(t(\tau')) . \tag{26}$$

This is normalizable only if λ_i is positive for $t = 0$ and negative for $t = 1$, or vice-versa. But (26) *must* be normalizable for some i, because D is known to have a zero eigenvalue. So for some i, λ_i changes sign between $t = 0$ and $t = 1$, and we get the picture of Fig. (2).

The desired inequality is an immediate consequence. Arrange the eigenvalues $\lambda_a(t)$ of $Q(t)$ in ascending order,

$$-\infty < \dots \lambda_{a-1}(t) < \lambda_a(t) < \lambda_{a+1}(t) \leqq \dots < +\infty \quad \text{with} \quad -\infty < a < +\infty;$$

normalize a so that at $t = 0$, $\lambda_1(0)$ is the lowest non-negative eigenvalue. Since $Q(0) = i\slashed{D}_A$, the $\lambda_a(0)$ are just the eigenvalues λ_a of $i\slashed{D}_A$. From Fig. 2, $\lambda_a(1) = \lambda_{a+1}(0)$, so

$$|\lambda_{a+1}(0) - \lambda_a(0)| \leqq \int_0^1 dt \left|\frac{d\lambda_a}{dt}\right| . \tag{27}$$

But $\left|\dfrac{d\lambda_a}{dt}\right| = |\langle\psi_a^t|X|\psi_a^t\rangle| \leqq C$, since norm $X = C$, so $|\lambda_{a+1} - \lambda_a| < C$. In particular, for $a = 0$, $\lambda_1 \geqq 0$, $\lambda_0 < 0$, so this says $\lambda_1 \leqq C$, $|\lambda_0| \leqq C$. This is the desired inequality, since by construction C depends on the manifold M (and its metric) but not on the gauge field A. Actually, the more general result $|\lambda_{a+1} - \lambda_a| < C$ is also of note; it shows that, regardless of A, there can be no large "gaps" in the Dirac spectrum.
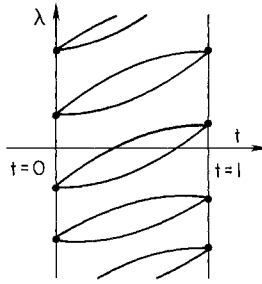
Fig. 3. The spectrum of $Q(t)$ for $N=2$, $q=1$ (an abelian gauge field), and $k=2$ (U has winding number 2). The only significant change from Fig. 2 is that now the spectrum flows upward by two units

This completes the proof of our first inequality for odd dimensional manifolds. What about the even dimensional case? The following device reduces the even dimensional problem to the odd dimensional case. Consider two manifolds N_1 and N_2 with gauge fields A_1 and A_2 and with Dirac eigenvalues λ_i and $\tilde{\lambda}_j$, respectively. If we regard A_1 and A_2 as gauge fields in commuting groups G_1 and G_2, they induce in a trivial way a $G_1 \times G_2$ gauge field on the product $N_1 \times N_2$. The eigenvalues of the corresponding Dirac equation on $N_1 \times N_2$ are $\lambda_{i,j}^{\pm} = \pm\sqrt{\lambda_i^2 + \tilde{\lambda}_j^2}$.

Now, on a circle S^1 of circumference 2π, the Dirac eigenvalues are integers m. Let N be an even dimensional manifold with gauge field A and Dirac eigenvalues λ_i. Then $N \times S^1$ is an odd dimensional manifold with Dirac eigenvalues $\pm\sqrt{\lambda_i^2 + m^2}$. By our previous construction, there is some constant C, independent of A, such that the smallest positive eigenvalue of $i\not{D}_A$ on $N \times S^1$ is less than C. With Dirac eigenvalues $\pm\sqrt{\lambda_i^2 + m^2}$ on $N \times S^1$, the modulus of the smallest eigenvalue on $N \times S^1$ is the same as the smallest modulus of an eigenvalue on N, and this shows that also in the even dimensional case, the smallest eigenvalue obeys an upper bound independent of A.

Notice that this argument does *not* prove that in the even dimensional case $|\lambda_a - \lambda_{a+1}| \leq C$, and that in fact is not true in general for even dimensional manifolds. (For a constant magnetic field in two dimensions, the spectrum has gaps of width $2|B|$, which can be arbitrarily big.)

This reduction of the even dimensional problem to the odd dimensional problem may seem opaque. A direct proof in the even dimensional case is given in Appendix I.

Now we turn to the more general problem of proving that $\lambda_n \leq C' n^{1/d}$ on a manifold of dimension d. The previous discussion is the special case $n = 1$. We again assume first that d is odd.

The proof of the stronger inequality is a relatively simple modification of what we have already done. We choose U in the previous construction to have not winding number 1 but winding number k. We again define X as in Eq. (22), and let $C = \text{norm} X$. More precisely, let C_k be the lowest possible norm of X as U ranges over mappings of winding number k.

We again study the family of operators $i\not{D}_A + tX$, $0 \leq t \leq 1$. Now as t varies from 0 to 1, there are not q but qk eigenvalues of $i\not{D}_A$ that flow through zero (Fig. 3);

recall that q is the dimension of the fermion representation T. However, as one switches on the perturbation X, of norm C_k, no eigenvalue can change by more than C_k. The fact that qk eigenvalues of $i\not{D}_A$ change sign when t goes from 0 to 1 means that $i\not{D}_A$ has at least qk eigenvalues that are negative but more than $-C_k$ and at least qk that are positive but less than $+C_k$.

Before drawing conclusions about the spectrum of the original Dirac operator $i\not{D}_A$, we must remember that in passing from group G to gauge group $G \times \mathrm{SU}(N)$, the multiplicity of the spectrum was increased by a factor of N. What we have actually proved about the spectrum of the original Dirac operator with gauge group G is therefore that there are at least qk/N eigenvalues between 0 and C_k. More generally, the same argument shows that there are at least qk/N eigenvalues of the original Dirac operator between f and $f + C_k$, for any real number f.

To proceed, we must learn how C_k depends on k. We will now prove that for a manifold M of d dimensions, $C_k \leqq \hat{C} k^{1/d}$, with a constant $\hat{C}$ that depends on M but not on k. The inequality $C_k \leqq \hat{C} k^{1/d}$ will complete our proof of the general upper bound (21) on Dirac eigenvalues, because it means that there are at least qk/N non-negative eigenvalues less than $\hat{C} k^{1/d}$, or in other words that $\lambda_n \leqq C' n^{1/d}$ with

$$C' = \hat{C} \left(\frac{N}{q} \right)^{1/d} .$$

It is enough to prove $C_k \leqq \hat{C} k^{1/d}$ for the special case $k = r^d$ (integer r); the case of general k follows from this, with a slightly different value of $\hat{C}$. To show that $C_{r^d} \leqq \hat{C} r$, suppose first that M is a torus with flat metric. Let ϕ_i be standard angular coordinates spanning M, so that $\phi_i \simeq \phi_i + 2\pi$, $i = 1 \ldots d$. Let $U : M \to \mathrm{SU}(N)$ be a degree one mapping chosen so that $U(\phi_1 \ldots \phi_d) = 1$ if any of the ϕ_i is 0 or 2π. Since the ϕ_i are angles, we have

$$U(\phi_1, \ldots, \phi_d) = U(\phi_1 + 2\pi n_1, \ldots, \phi_d + 2\pi n_d)$$

for any integers n_i. Define X as before, $X = iU^{-1}\gamma^i \partial_i U$, and let $\hat{C} = \mathrm{norm}\, X$.

Define a mapping $\alpha_r : M \to M$ by $\alpha_r(\phi_1, \ldots, \phi_d) = \alpha(r\phi_1, \ldots, r\phi_d)$. Thus, α_r is a mapping from M to M of degree r^d; it maps M around itself r^d times.

Now, let $U_r = U \circ \alpha_r$; i.e. $U_r(\phi_1, \ldots, \phi_d) = U(r\phi_1, \ldots, r\phi_d)$. Then U_r is a mapping of degree r^d. Moreover, if $X_r = iU_r^{-1}\gamma^i \partial_i U_r$, then the norm of X_r is precisely r times the norm of X [since the derivatives of $U(r\phi)$ are exactly r times the derivatives of $U(\phi)$ in absolute value]. So $\mathrm{norm}\, X_r = r\hat{C}$. Since $C_{r^d} \leqq \mathrm{norm}\, X_r$, we have $C_{r^d} \leqq r\hat{C}$, as desired.

We must proceed now from the case where M is a torus to the general case. In any d dimensional manifold M, one can embed a cube (Fig. 4). Identify the cube with the d-fold product $I^d = I \times I \ldots \times I$, where I is the interval $[0, 2\pi]$ on the real line. Thus, we have coordinates ϕ_i, $i = 1 \ldots d$ on I^d in M, but we cannot identify $\phi_i = 0$ with $\phi_i = 2\pi$.

Now for any integer we define a mapping $U_r : M \to \mathrm{SU}(N)$ as follows. In I^d we take U_r to be the same function of the ϕ_i that we used previously. Outside of I^d we take $U_r = 1$. The two definitions agree on the boundary of I^d, since we previously took $U = 1$ (and so $U_r = 1$) for $\phi_i = 0$ or 2π. The winding number of U_r is still r^d, and the norm of the corresponding operator X is still $\hat{C} r$, so we have again $C_{r^d} \leqq \hat{C} r$.

This completes our proof that for odd d, $\lambda_n \leqq C' n^{1/d}$, and more generally that there are always at least n eigenvalues between f and $f + C' n^{1/d}$, for any real f. We
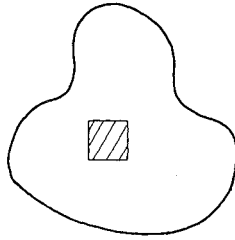
Fig. 4. A cube I^d mapped into a general d manifold. The induced metric on the cube may not be the standard one, but the cube is embedded in the topological sense

still must prove the former statement (but not the latter one, which is not true in general) for even d. A direct approach to even d can be found in Appendix I; here, we will show how to infer the even d case form the odd d case.

As before, let d be even and let N be a d dimensional manifold with Dirac eigenvalues λ_i.[3] Then $N \times S^1$ is a $d+1$ dimensional manifold on which (if S^1 has circumference 2π) the Dirac eigenvalues are $\pm\sqrt{\lambda_i^2 + m^2}$, for arbitrary integers m. (Each m appears with unit multiplicity, the eigenfunction on S^1 being $e^{im\phi}$.) Let $W(S)$ be the number of Dirac eigenvalues on N that are between 0 and S (because of facts noted in the footnote, this equals the number of Dirac eigenvalues on N that are between $-S$ and 0). Let $Y(S)$ be the number of Dirac eigenvalues on $N \times S^1$ that are between 0 and S. For $\pm\sqrt{\lambda_i^2 + m^2}$ to be between 0 and S, we need $|\lambda_i| \leq S$ and $|m| \leq S$, so $Y(S) \leq 4SW(S)$. But for the $d+1$ dimensional manifold $N \times S^1$ we have proved that the nth Dirac eigenvalue is less than $C'n^{1/(d+1)}$, so that $Y(S) \geq S^{d+1}/(C')^{d+1}$, where C' is independent of A. With $Y(S) \leq 4SW(S)$, this gives $W(S) \geq S^d/(4C'^{d+1})$. But that is the desired result, because it means that, with a new constant C'', $\lambda_n < C''n^{1/d}$. This result, which we have now established for both even and odd dimensional manifolds, is a relatively good bound, in the sense that it is actually true for any A that asymptotically, for large n, $\lambda_n \sim n^{1/d}$. (This is related to the classical limit of quantum statistical mechanics.) However, the proof that we have sketched certainly does not give the best possible value of the constant C. The best value of C depends on the topology and geometry of the manifold M. To determine the best constant in general would be extremely difficult. However, there is a special case in which methods similar to those we have described can be used to prove a sharp inequality, with the best possible constant.

Consider a torus P with periodic coordinates ϕ^i, $i=1 \ldots d$ $(0 \leq \phi^i \leq 2\pi)$, and with a flat metric; thus, the line element is $ds^2 = g_{ij}d\phi^id\phi^j$, where g is a constant matrix, not necessarily diagonal. We will consider fermion fields that obey periodic boundary conditions (the discussion can easily be modified). On the torus, there is a d parameter family of $U(1)$ gauge fields which are constant in the sense that each component of the gauge field A_i is constant, $A_i = (a_1, a_2, \ldots, a_d)$ with some real numbers a_i. For such constant gauge fields, $F_{ij} = \partial_iA_j - \partial_jA_i = 0$; nonetheless, A_i cannot be gauged away. The gauge transformation $\psi \to e^{i\sum n_i\phi^i}\psi$ (n_i being integers)

---

3   The λ_i are paired. For every non-zero eigenvalue λ there is an eigenvalue $-\lambda$, the eigenfunctions being related by $\psi \to \Gamma\psi$, where $\Gamma = \gamma_1\gamma_2 \ldots \gamma_d$ is the operator that distinguishes the two fermion chiralities

shift the a_i by integers, $a_i \rightarrow a_i + n_i$; but this is the best one can do. The a_i can thus be regarded as periodic variables, a_i being equivalent to $a_i + n_i$ for $n_i \in Z$. We have, in effect, a family of $U(1)$ gauge fields labeled by the torus (d-fold product) $Q = S^1 \times S^1 \times \ldots \times S^1$, with the a_i being coordinates on this torus. The key facts of this family of operators have been described by Gromov and Lawson [14], who used it to prove that the torus does not admit a metric of strictly positive scalar curvature. The general idea for the argument we will give is similar to theirs.

For $a_i = 0$, the lowest Dirac eigenvalue is $\lambda_1 = 0$, corresponding to an eigenfunction which is a constant spinor. In general, the eigenfunctions are $\psi(\phi^i) = (\exp i\Sigma m_i \phi^i)\eta$, where η is a constant spinor and m_i are integers. On such a state the Dirac operator reduces to $\gamma^j(m_j - a_j)$. The square of that operator is $g^{ij}(m_i - a_i)(m_j - a_j)$ (since $\gamma^i \gamma^j + \gamma^j \gamma^i = 2g^{ij}$), so the eigenvalues are $\pm \sqrt{g^{ij}(m_i - a_i)(m_j - a_j)}$.

If $a_i \neq 0$, the lowest eigenvalue λ_1 is not zero. For given a_i, λ_1 is the minimum of $\sqrt{g^{ij}(m_i - a_i)(m_j - a_j)}$ for integers m_i. Keeping the metric g^{ij} fixed, but varying a^i, let λ_M be the maximum value of λ_1 that occurs for any a^i. For example if one takes the standard rectangular metric on the torus, $g^{ij} = \delta^{ij}$, the maximum of λ_1 occurs for $a_1 = a_2 = \ldots = a_d = 1/2$ and is $\lambda_M = 1/2\sqrt{d}$.

For future reference, λ_M may be characterized as follows. For given a^i, choose integers m^i to minimize the norm of $\gamma^j(m_j - a_j)$. Let $a'_j = a_j - m_j$. As the $\{a_j\}$ vary in R^d, the $\{a'_j\}$ lie in a bounded region in R^d which we may call the "first Brillouin zone." The maximum of the norm of $\gamma^j a_j$ in the first Brillouin zone is λ_M.

Now, λ_M is the largest possible value of λ_1 for any constant $U(1)$ gauge field $A_i = (a_1, a_2, \ldots, a_d)$. What we will prove is that for *any* gauge field with *any* group G, λ_1 is at most equal to λ_M. To prove this we will, as before, study not a single Dirac operator but a whole family of Dirac operators.

If G does not contain a $U(1)$ factor, we extend it to $G \times U(1)$, complexifying the fermions if they are not already complex. We add to the given gauge field A_i a constant piece a_i, $A_i \rightarrow (A_1 + a_1, A_2 + a_2, \ldots, A_d + a_d)$, the a_i being real numbers. Instead of a single Dirac operator $i\not{D}_A$, we study the d parameter family of Dirac operators

$$D^{(a_1 \ldots a_n)} = i \sum_{j=1}^{d} \gamma^j \left(\frac{\partial}{\partial \phi^j} + iA_j + ia_j \right) = i\not{D}_A - \sum_{j=1}^{d} \gamma^j a_j. \tag{28}$$

The spectrum of $D^{(a_1 \ldots a_n)}$ is a periodic function of the a_i, with period one, since the a_i can be shifted by integers by a gauge transformation. Now, Gromov and Lawson have shown, from the Atiyah-Patodi-Singer theorem, that regardless of the choice of A, there is always some value of the a_i such that $D^{(a_1, \ldots, a_n)}$ has a zero eigenvalue. We will sketch a proof of this shortly, but first discuss the consequences.

If $D^{(a_1 \ldots a_n)}$ has a zero eigenvalue, then $i\not{D}_A$ has an eigenvalue that differs from zero by at most the norm of $X^{(A_i)} = \sum_{j=1}^{d} \gamma^j a_j$. The norm of $X^{(a_i)}$ is $\sqrt{\sum_{i,j} g^{ij} a_i a_j}$, as we discussed before. Given that $D^{(a_1 \ldots a_n)}$ has a zero eigenvalue for some a_i, it has a zero eigenvalue somewhere in the first Brillouin zone, since the spectrum of $D^{(a_1, \ldots, a_n)}$ is periodic in the a_i, and a_i can be shifted into the first Brillouin zone by subtracting integers m_i. The maximum norm of $X^{(a_i)}$ in the first Brillouin zone is what we

previously called λ_M, so the magnitude of the lowest eigenvalue of $i\not{D}_A$ is at most λ_M, as we wished to prove.

Let us finally sketch a simple proof of the statement that $D^{(a_1, \ldots, a_n)}$ has a zero eigenvalue for some a_i. To do this, we will study a Dirac operator on a torus $P \times Q$ of $2d$ dimensions, spanned by the original d angular variables ϕ^i and d additional variables, the a^i. On this product space we choose the metric $ds^2 = \varepsilon g_{ij} d\phi^i d\phi^j + \Sigma da_i^2$, for suitable small positive ε. For the gauge field, we take the components in the ϕ^i directions to be $\tilde{A}_i(\phi^j, a^k) = A_i(\phi^j) + a_i$ – where $A_i(\phi^j)$ is the original gauge field on P that we are really interested in studying. The gauge field components in the a_i directions we take to be zero. Thus, the $2d$ dimensional gauge field is $(A_1 + a_1, A_2 + a_2, \ldots, A_d + a_d, 0, \ldots, 0)$. The Dirac operator is

$$\hat{D} = \frac{1}{\varepsilon} \sum_{j=1}^{d} i\gamma^j \left(\frac{\partial}{\partial \phi^j} + iA_j + ia_j \right) + \sum_{k=1}^{d} i\gamma^{d+k} \frac{\partial}{\partial a^k} = \frac{1}{\varepsilon} D^{(a_1, \ldots, a_d)} + i \sum_{k=1}^{d} \gamma^{d+k} \frac{\partial}{\partial a^k}. \quad (29)$$

Now, the first step is to show that $\hat{D}$ has a zero eigenvalue. This follows from the fact that the index of $\hat{D}$ is non-zero. In fact, since the index is a topological invariant, we can compute it by setting $A_j = 0$. (This tacitly assumes A_j was a connection on a trivial bundle on P, in the same topological class as $A_j = 0$. However, by further study of the index formula we are about to discuss, it can be shown that the index of $\hat{D}$ does not depend on which bundle over P we started with.) With $A_j = 0$, the index can be computed as a simple integral:

$$\text{index}\,\hat{D} = \frac{\dim T(-1)^d}{(4\pi)^d (d!)^2} \int_0^{2\pi} d\phi_i \int_0^1 da_i \varepsilon^{i_1 \ldots i_{2d}} F_{i_1 i_2} \ldots F_{i_{2d-1} i_{2d}}. \quad (30)$$

[The factor of $\dim T$ comes from taking the trace of the product of U(1) field strengths in the fermion representation. An index i on F_{ij} refers to ϕ^i if $i \leq d$, and to a_{i-d} if $i > d$.] With $A_i = 0$, the only non-zero components of F_{ij} are $F_{i,d+j} = -F_{d+j,i} = -\delta_{ij}$, and one readily finds $\text{index}\,\hat{D} = \dim T \neq 0$.

From the fact that $\hat{D}$ has a zero eigenvalue, we now wish to show that $D^{(a_1, \ldots, a_n)}$ has a zero eigenvalue for some value of $(a_1, \ldots, a_n)$. To do so, note

$$\hat{D}^2 = \frac{1}{\varepsilon^2} (D^{(a_1, \ldots, a_n)})^2 + \frac{1}{2\varepsilon} \sum_{j=1}^{d} i[\gamma^j, \gamma^{j+d}] + \sum_{i=1}^{d} \left(-\frac{\partial^2}{\partial a_i^2} \right). \quad (31)$$

In (31), the last term $\sum \left(-\dfrac{\partial^2}{\partial a_i^2} \right)$ is a non-negative operator. In the second term $\sum_j [\gamma^j, \gamma^{j+d}]$ has negative eigenvalues, but is bounded below by $-2d$. If $D^{(a_1, \ldots, a_n)}$ had no zero eigenvalue for any value of the a_i, it would be bounded away from zero, and the first, positive term in (31) would dominate for sufficiently small ε. If so, $\hat{D}$ would be for small ε a strictly positive operator, contradicting the fact that it has a nonzero index. It must therefore be that $D^{(a_1, \ldots, a_n)}$ has a zero eigenvalue for some value of the a_i, as was to be proved.

The same argument can be pushed farther, to yield the stronger statement of the Atiyah-Patodi-Singer theorem relating the index of $\hat{D}$ to the behavior of $D^{(a_1, \ldots, a_n)}$ near values of the a_i where it has zero eigenvalues. We will not enter into that, however. Similar matters have been discussed elsewhere [15].

It should be noted that the inequalities we have proved in this section go in the *opposite* direction from certain analogous inequalities – the Kato inequalities – that hold in the spin zero case [16]. (For field dependent upper bounds for eigenvalues of the scalar operator see Appendix II.) Qualitative reasons that the inequalities for fermions go in the opposite direction from those for bosons were proposed in connection with the original suggestions of paramagnetic inequalities for fermions [7]. We have, however, proved elsewhere [3] an upper bound on the Dirac propagator that goes in the same direction as the Kato inequality for bosons.

IV. Gauge Theories in Three Dimensions

In this section, let us briefly discuss one of the few areas in which our work has implications that cannot already be deduced from anomaly considerations[4]. This is the case of gauge theories in $2+1$ dimensions, which have recently attracted much interest [17–22].

Let us first discuss the basic kinematics of fermions in $2+1$ dimensions. The basic fermi field is the two component Majorana spinor – a real, two dimensional representation ψ_α of $O(2, 1)$. Such a field can have a mass term $\pm im\varepsilon^{\alpha\beta}\psi_\alpha\psi_\beta$. The two signs of the mass are physically inequivalent. They give massive fermion states with eigenvalue $+\hbar/2$ or $-\hbar/2$ of the one rotation generator of a world that has two spatial dimensions. A non-zero mass for a Majorana fermion violates certain discrete symmetries: time reversal, and the space reflection symmetry $(t, x, y) \rightarrow (t, x, -y)$. For fermions of spin $+\hbar/2$ spin in a right-handed sense; those of spin $-\hbar/2$ spin in a left-handed sense. Of course, if one has *two* Majorana fermions, one of positive mass and one of negative mass, the discrete symmetries are preserved, as long as one defines time reversal or space deflection to exchange the two Majorana fermions.

When one introduces gauge fields, it is possible [18] to give them a parity violating, gauge invariant bare mass with a quantized coefficient. This term is imaginary in Euclidean space, so if it is introduced the Feynman path integrand is not positive definite. In such a case, our method of proving the absence of a mass gap when fermions of zero bare mass are introduced does not work. This is just as well, because when the gauge fields have a parity-violating bare mass, fermions of zero bare mass get a non-zero mass at the one loop level, and there is in fact a mass gap.

Setting the gauge field bare mass to zero, we now proceed to couple the gauge field to fermions. Let the fermions consist of N copies of a real irreducible representation R of some gauge group G. Then there is a global $O(N)$ flavor symmetry. (The symmetry is larger if $R = Q + \bar{Q}$ for some complex or pseudo-real representation Q, but we will assume for the sake of exposition that that is not so.)

---

4   Actually, if one assumes parity is not spontaneously broken, then in $2+1$ dimensions, when fermions are in a real, odd dimensional representation of the gauge group, there is a mod-2 version of the anomaly matching condition. Gauging of the $SU(N_f)$ flavor symmetry gives in such cases an anomalous breakdown of parity [21], and requiring that this could be understood at the composite level implies that the theory must have physical massless fermions or breakdown of flavor symmetry

Suppose the fermion bare mass is zero. Then the Euclidean Dirac operator $i\not{D}$ is Hermitian, so the eigenvalues are real. Therefore the fermion determinant is real. For Majorana fermions, the basic fermion integral is $\sqrt{\det i\not{D}}$, which is real but not necessarily positive[5]. If there are N flavors, the fermion integral is $(\det i\not{D})^{N/2}$. This is positive if N is even, and that is the case for which our methods apply.

For N even and at least four, the arguments of Sects. II and III show that – for any group G and representation R – the theory with no bare mass for fermions or gauge bosons has no mass gap. We do not know of another way to prove this with such generality.

Now let us discuss whether the absence of mass gap can be understood in terms of symmetry breaking. If fermion masses are generated dynamically, these may be positive or negative. [We assume the fermion dynamical masses have the same absolute value, an assumption we will justify shortly by showing there must be an unbroken $O(N/2) \times O(N/2)$ symmetry.] If k species have positive masses and $N - k$ have negative masses, this breaks $O(N)$ to $O(k) \times O(N-k)$. There are Goldstone bosons (and no mass gap) unless $k = 0$ (or $k = N$). Time reversal and reflection symmetry are spontaneously broken unless $k = N/2$.

To determine what value of k is plausible, first suppose the fermions have non-zero bare masses of equal absolute value m. The fermion operator is now $i\not{D} \pm im$ which is *not* hermitian; $\det(i\not{D} + im)$ and $\det(i\not{D} - im)$ are complex – and are complex conjugates of each other. The fermion integral is real and positive only if $N/2$ fermions have positive bare masses and $N/2$ have negative bare masses. In that case the fermion integral is $(\det(i\not{D} + im)\det(i\not{D} - im))^{N/4}$, which is strictly positive.

Such equal and opposite fermion bare masses explicitly break $O(N)$ to $O(N/2) \times O(N/2)$. However, since we have kept the fermion determinant positive, we can invoke the results of reference [3] which show that for $m \neq 0$ the $O(N/2) \times O(N/2)$ symmetry which is not broken *explicitly* by the fermion bare masses is also not broken *spontaneously*. Hence – for any non-zero m and presumably also for $m = 0$ – the theory has an unbroken $O(N/2) \times O(N/2)$ symmetry.

Turning off the bare masses, we previously saw that fermion dynamical masses would break $O(N)$ to $O(k) \times O(N - k)$. We now see that the only plausible values are $k = 0$, $N/2$, or N, since there must be an unbroken $O(N/2) \times O(N/2)$ symmetry.

To gain more insight, consider probing for a dynamically generated mass by introducing a tiny bare mass m and taking $m \to 0$. In the two cases of (x) all bare masses of same sign or (y) $N/2$ positive and $N/2$ negative bare masses, the fermion determinant becomes

$$\Delta_x = \lim_{m \to 0} (\det(i\not{D} + im))^{N/2}, \qquad \Delta_y = \lim_{m \to 0} \det((i\not{D} + im)(i\not{D} - im))^{N/4}. \qquad (32)$$

---

5   In four dimensions, $\det i\not{D}$ is positive for real representations because of γ_5 symmetry. In three dimensions $\det i\not{D}$ is positive for real representations for an analogous reason. We can choose a 2×2 basis of gamma matrices $(\sigma_x, \sigma_y, \sigma_z)$. Of these, σ_y is imaginary, and σ_x, σ_z are real. If $i\not{D}\psi = \lambda\psi$, then $i\not{D}(\sigma_y\psi^*) = \lambda(\sigma_y\psi^*)$. This pairing of eigenvalues ensures that $\det i\not{D}$ is positive, so $\sqrt{\det i\not{D}}$ is real, but nothing forces $\sqrt{\det i\not{D}}$ to be positive. (In certain cases with a pseudoreal rather than real representation, there is a discrete anomaly [21] as a result of which $\sqrt{\det i\not{D}}$ cannot be real and gauge invariant. In such cases our methods do not apply)

We see that Δ_y is real and non-negative, while Δ_x has the same absolute value as Δ_y but may have any phase. The vacuum energy $E_{x,y}$ computed in the two cases (in a box of volume V) is

$$e^{-E_x V} = \int d\mu(A)\Delta_x, \qquad e^{-E_y V} = \int d\mu(A)\Delta_y. \tag{33}$$

Since $d\mu(A)$ is a positive measure, the phase in Δ_x can only make the integral in E_x smaller, so $E_x \geq E_y$. Thus the vacuum energy is less with fermion dynamical masses that break $O(N)$ to $O(N/2) \times O(N/2)$ and do not break time reversal than with dynamical masses that break time reversal but not $O(N)$. [The possibility $E_x = E_y$ is implausible unless there are no dynamical masses and neither $O(N)$ nor time reversal is broken.]

We conclude that if fermions get dynamical masses, the theory has Goldstone bosons, accounting for the absence of a mass gap.

In a beautiful calculation, Pisarski recently showed [22] that at least for large N, the fermions get dynamical masses in $2+1$ dimensions. Though Pisarski did not check the sign for these masses, we expect further study to show that for even N his mechanism generates equal numbers of positive and negative fermion masses.

Appendix I

In this appendix we will briefly explain how to prove that $\lambda_1 \leq c$ directly for an even dimensional manifold, rather than inferring this from the odd dimensional case as was done in Sect. III. We will also generalize to show $\lambda_n \leq cn^{1/d}$.

Consider a gauge field A with gauge group G on an even dimensional manifold M. As in Sect. III, we first extend the gauge group from G to $G \times U(P)$ (for suitable P, to be chosen) and think of $i\slashed{D}_A$ as acting on P identical copies of the original Hilbert space; this does not change the value of λ_1.

Our idea is to find a bounded operator X, of norm c, such that $i\slashed{D}_A + X$ has a zero eigenvalue. X and c will not depend on A. For $i\slashed{D}_A + X$ to have a zero eigenvalue, $i\slashed{D}_A$ must have an eigenvalue within c of zero, so $\lambda_1 \leq c$ if we can find a suitable X.

For any even d, there is some N and some $U(N)$ gauge field B_μ such that $i\slashed{D}_B$ has a non-zero index. For $d = 2$ one takes $N = 1$ and the magnetic monopole; for $d = 4$ one takes $N = 2$ and the instanton. The corresponding objects for $d > 4$ are related to $\pi_{d-1}(U(N))$ and are not so well known in physics.

For our first try, we take $N = P$ and take $X = i\gamma^\mu B_\mu$, where $i\slashed{D}_B$ has non-zero index. Thus $i\slashed{D}_A + X = i\gamma^\mu(D_\mu^{(A)} + B_\mu)$. So $i\slashed{D}_A + X$ is the Dirac operator associated with the $G \times U(N)$ gauge field $A_\mu + B_\mu$. It is indeed true that $i\slashed{D}_A + X$ has zero eigenvalues, for any A, because of the non-zero index due to the choice of B_μ[6].

However, this choice of X is not suitable. In fact, $i\slashed{D}_A$ and $i\slashed{D}_A + X$ do not even act on the same Hilbert space, since B_μ is a connection on a non-trivial bundle. A more prosaic way to state the difficulty is that X as defined is an unbounded operator. The norm of X would have been the pointwise upper bound of $i\gamma^\mu B_\mu$, but this is infinite, because of the gauge singularities in B_μ on a compact manifold.

---

6   We assume here that the index of $i\slashed{D}_A + X$ comes only from B. If $i\slashed{D}_A$ had non-zero index, then $\lambda_1 = 0$ and there was nothing to prove to begin with

To avoid the difficulty, we pass from $U(N)$ to $U(2N)$. We introduce a $U(N)$ gauge field $\bar{B}_\mu$ of topological quantum number opposite to that of B_μ, and we define the $U(2N)$ gauge field

$$\tilde{B}_\mu = \left(\begin{array}{c|c} B_\mu & 0 \\ \hline 0 & \bar{B}_\mu \end{array} \right). \tag{34}$$

Thus, the total winding number of $\tilde{B}_\mu$ is zero; it is a connection on a topologically trivial bundle.

Now, to refine our previous effort, let $\tilde{X} = i\gamma^\mu \tilde{B}_\mu$. It is again true that $i\not{D}_A + \tilde{X}$ has zero eigenvalues for any A. (Although the index of $i\not{D}_A + \tilde{X}$ is zero, it is presented as a direct sum of two operators of equal and opposite index.) We have improved on the previous situation from a conceptual point of view, because at least $i\not{D}_A$ and $i\not{D}_A + \tilde{X}$ act on the same Hilbert space. However, $\tilde{X}$ is still an unbounded operator, because of the gauge singularities in B_μ and $\bar{B}_\mu$.

To get a bounded operator, note that since $\tilde{B}_\mu$ has zero net winding number, there is a $U(2N)$ gauge transformation that would remove the singularities from $\tilde{B}_\mu$. Thus, for some mapping $V: M \to U(2N)$, $B'_\mu = V^{-1}\tilde{B}_\mu V - V^{-1}\partial_\mu V$ is nonsingular. [Of course, one cannot choose V to lie in $U(N) \times U(N)$.] Now we let $X' = i\gamma^\mu B'_\mu$; X' is, at last, a bounded operator. And $iD_A + X'$ has zero eigenvalues since $i\not{D}_A + X' = V^{-1}(i\not{D}_A + \tilde{X})V$, and we already saw that $i\not{D}_A + \tilde{X}$ has zero eigenvalues. If $C = \mathrm{norm}\, X'$, these facts show $\lambda_1 \leq C$ for any A.

Now we wish to show $\lambda_n \leq C'n^{1/d}$ for any n. The generalization is as it was in Sect. III. If N is a torus with angular coordinates ϕ^i, let $B'_r(\phi^i) = rB'(r\phi^i)$, $X'_r = i\gamma B'_{ir}$. Then the winding number of B'_r is r^d and the norm of X'_r is rC. These facts, applied to $iD_A + X'_r$ show $\lambda_{r^d} \leq Cr$, so $\lambda_n \leq C'n^{1/d}$. If N is not a torus, one picks a cube in N and carries out a similar argument, as in Sect. III.

Appendix II

In this appendix we will obtain upper bounds for eigenvalues of the scalar operator $-D_\mu D^\mu$ on the torus with a flat metric. The Kato inequality [16] implies that on the average the eigenvalues of the interacting scalar operator go up, compared to the non-interacting case, in the sense that

$$\mathrm{Tr}\, e^{-\beta(-D_\mu D^\mu)} \leq \mathrm{Tr}\, e^{-\beta(-\partial_\mu \partial^\mu)}, \tag{35}$$

where $(-\partial_\mu \partial^\mu)$ is the free scalar operator. In particular this inequality implies that the lowest eigenvalue goes up as we turn on the gauge field.

It is not difficult to see that unlike the case for the Dirac operator there does not exist any field independent upper bound for the eigenvalues of the scalar operator. Consider for example, turning on a constant magnetic field B in 2 dimensions. For a very large box $\left(\text{i.e., } L^2 \gg \dfrac{1}{B} \right)$ one obtains the Landau levels shown in Fig. 5a. The lowest eigenvalue is proportional to the magnitude of the magnetic field. This is basically a consequence of the uncertainty principle. The stronger the field, the more localized the state in space and thus the higher the uncertainty in momentum and energy. So by increasing the field strength the lowest eigenvalue could be made
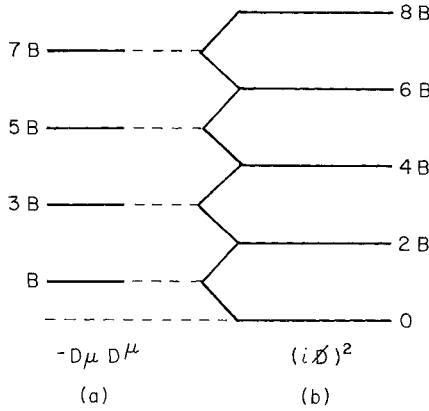
Fig. 5a and b. The spectrum of the scalar operator **a** and the Dirac operator **b** for a constant magnetic field B in 2 dimensions. The magnetic moment interaction splits the levels of the scalar operator to give the spectrum of the Dirac operator

as large as one pleases and therefore there is no field independent upper bound for the eigenvalues of the scalar operator.

However, as we will see, field dependent upper bounds do exist for the eigenvalues of the scalar operator. We consider the following identity:

$$(i\slashed{D})^2 = -D_\mu D^\mu + \tfrac{1}{2}\sigma^{\mu\nu}F_{\mu\nu}, \tag{36}$$

where $\sigma^{\mu\nu} = \dfrac{i}{2}[\gamma^\mu, \gamma^\nu]$. This identity implies that the eigenvalues of $(i\slashed{D})^2$ and $-D_\mu D^\mu$ cannot differ by more than the norm of $\tfrac{1}{2}\sigma^{\mu\nu}F_{\mu\nu}$, which we denote by F $\Big($i.e., $F = \max\limits_x |\tfrac{1}{2}\sigma^{\mu\nu}F_{\mu\nu}(x)|$, where $\|$ denotes the norm of the matrix$\Big)$. If we label the eigenvalues of $(i\slashed{D})^2$ by λ_n and those of $(-D_\mu D^\mu)$ by λ_n^0 in an ascending order, we have

$$\lambda_n^0 \leqq \lambda_n + F. \tag{37}$$

From Sect. III we have $\lambda_n \leqq \left(\dfrac{C}{L}\right)^2 n^{2/d}$, so we get

$$\lambda_n^0 \leqq \left(\dfrac{C}{L}\right)^2 n^{2/d} + F. \tag{38}$$

In particular in the limit that $L \to \infty$ we have $\lambda_1^0 \leqq F$, which is saturated by a constant magnetic field. In the particular case of constant B field in 2 dimensions the eigenvalues of the Dirac operator are simply obtained from the eigenvalues of the scalar operator by noting that the spin term in Eq. (36) splits the levels, as shown in Fig. 5b. In particular we see that the spectrum of the Dirac operator starts from zero, as it should $\Big($from our inequalities $\lambda_1 \leqq \dfrac{C}{L} \xrightarrow{L \to \infty} 0\Big)$. Thus the

fact that magnetic moment interaction cancels the zero point energy is not limited to this example.

For odd dimensional tori we have a further result. In these cases the number of eigenvalues of the scalar operator between $\varrho^2 - F$ and $(\varrho + \delta)^2 + F$ is at least $C'(L\delta)^d$. This follows from the corresponding estimate for the spectrum of the Dirac operator.

Looking at Fig. 5a and b and noting that the first eigenvalue of the scalar operator cannot be raised above zero by more than what it has been for the case of the constant field, makes one speculate whether all the gaps in the spectrum are maximal for a constant B field. We do not know the answer to this question.

References

1. Tomboulis, E.T.: Permanent confinement in four-dimensional non-abelian lattice gauge theory. Phys. Rev. Lett. **50**, 88 (1983)
2. Weingarten, D.: Mass inequalities for quantum chromodynamics. Phys. Rev. Lett. **51**, 1830 (1983)
3. Vafa, C., Witten, E.: Restrictions on symmetry breaking in vector-like gauge theories. Nucl. Phys. B**234**, 173 (1984)
4. Witten, E.: Some inequalities among hadron masses. Phys. Rev. Lett. **51**, 2351 (1983)
5. Nussinov, S.: Baryon-meson mass inequality. Phys. Rev. Lett. **51**, 1081 (1983); Mass inequalities in quantum chromodynamics **52**, 966 (1984)
6. Tomboulis, E.T., Yaffe, L.: Princeton preprint (1983)
7. Hogreve, H., Schrader, R., Seiler, R.: A conjecture on the spinor functional determinant. Nucl. Phys. B**142**, 525 (1978);
   Brydges, D., Fröhlich, J., Seiler, E.: On the construction of quantized gauge fields. I. General results. Ann. Phys. **121**, 227 (1979);
   Avron, J., Simon, B.: A counterexample to the paramagnetic conjecture. Phys. Lett. **75A**, 41 (1979)
8. 't Hooft, G.: In: Recent developments in gauge theories, 't Hooft, G. et al. (eds.). New York: Plenum Press 1980
9. Asorey, M., Mitter, P.K.: Regularized, continuum Yang-Mills process and Feynman-Kac functional integral. Commun. Math. Phys. **80**, 43 (1981)
10. Banks, T., Casher, A.: Chiral symmetry breaking in confining theories. Nucl. Phys. B**169**, 102 (1980)
    Kogut, J., Stone, M., Wyld, H.W., Gibbs, W.R., Shigemitsu, J., Shenker, S.H., Sinclair, D.K.: Deconfinement and chiral symmetry restoration at finite temperature in SU(2) and SU(3) gauge theories. Phys. Rev. Lett. **50**, 393 (1983);
    Scales of chiral symmetry breaking in quantum chromodynamics. Phys. Rev. Lett. **48**, 1140 (1982)
11. Atiyah, M.F., Patodi, V., Singer, I.: Math. Proc. Camb. Philos. Soc. **79**, 71 (1976)
12. Callan, C.G., Dashen, R., Gross, D.J.: Toward a theory of the strong interactions. Phys. Rev. D**17**, 2717 (1978)
    Kiskis, J.: Fermion zero modes and level crossing. Phys. Rev. D**18**, 3690 (1978)
13. Atiyah, M.F., Singer, I.M.: Ann. Math. **87**, 485, 546, (1968); **93**, 1, 119, 139 (1971)
    Atiyah, M.F., Segal, G.B.: Ann. Math. **87**, 531 (1968)
14. Gromov, M., Lawson, H.B., Jr.: Ann. Math. **111**, 209 (1980)
15. Witten, E.: J. Diff. Geom. **17**, 661 (1982); Alvarez-Gaumé, L., Ginsparg, P.: The topological meaning of non-Abelian anomalies. Harvard preprint (1983)

16. Kato, T.: Israel J. Math. **13**, 135 (1972)
17. Feynman, R.P.: The qualitative behavior of Yang-Mills theory in $2+1$ dimensions. Nucl. Phys. B**188**, 479 (1981)
    Singer, I.M.: Phys. Scripta **24**, 817 (1981)
18. Jackiw, R., Templeton, S.: How super-renormalizable interactions cure their infrared divergences. Phys. Rev. D**23**, 2291 (1981)
    Schönfeld, J.: A mass term for three-dimensional gauge fields. Nucl. Phys. B**185**, 157 (1981); Deser, S., Jackiw, R., Templeton, S.: Three-dimensional massive gauge theories. Phys. Rev. Lett. **48**, 975 (1982); Topologically massive gauge theories, Ann. Phys. (N.Y.) **140**, 372 (1982)
19. Appelquist, T., Pisarski, R.D.: High-temperature Yang-Mills theories and three-dimensional quantum chromodynamics. Phys. Rev. D**23**, 2305 (1981)
20. Affleck, I., Harvey, J., Witten, E.: Instantions and (super-)symmetry breaking In $(2+1)$ dimensions. Nucl. Phys. B**206**, 413 (1982)
21. Redlich, N.: Gauge noninvariance and parity nonconservation of three-dimensional fermions, Phys. Rev. Lett. **52**, 18 (1984)
    Alvarez-Gaumé, L., Witten, E.: Gravitational anomalics, Nucl. Phys. B**234**, 269 (1984)
22. Pisarski, R.D.: University of California preprint (1984)