Some remarks on the fermion determinants in gauge theories

Erhard Seiler
Max-Planck-Institut für Physik und Astrophysik, Föhringer Ring 6, München, Germany
(Received 25 February 1980)

We describe a precise definition of the determinants that arise when fermions interact with a given gauge field. This allows one to clarify the relation between zeros of determinants and eigenvalues of the corresponding Dirac operators which has led to some controversy; sometimes this relation does not conform to naive expectations.

I. INTRODUCTION

In the study of the Euclidean version of gauge theories coupled to fermions, one is naturally led to consider determinants such as "det $i\not{D}$" or "det $(i\not{\partial}+m)$" where $i\not{D}$ is the covariant Dirac operator corresponding to some given gauge field. In the past few years such determinants have been discussed by many authors (see Refs. 1–3 and references given there).

Here I want to make a few remarks of a mathematical nature concerning the definition and the properties of these determinants; these things do not seem to be generally known but are of some relevance in the applications. The main point is that many people have been too cavalier in assuming that the simple relation between determinants and eigenvalues known from finite-dimensional linear algebra carries over to the infinite-dimensional problems considered.

This point is best illustrated by the explicitly known determinant of massless two-dimensional QED (QED_2) (Sec. III). Recently Patrascioiu[3] noticed a discrepancy between the definition based on perturbation theory and a definition using the eigenvalues; Rothe and Schroer[2(b)] blamed this on his use of the "wrong" boundary conditions. Here I want to show that such a discrepancy is to be expected and is in a sense unavoidable; I use an approach that is, I think, very natural and does not depend on any boundary conditions.

This approach is also suitable for two-dimensional quantum chromodynamics (QCD_2), QED_4, and in particular QCD_4 where the situation is qualitatively similar to QED_2.

II. GENERAL REMARKS ON RENORMALIZED DETERMINANTS

The use of renormalized determinants in physics goes back at least to Schwinger's classic papers[4] on QED_4; in mathematics it is even older and can be traced back to Hilbert.[5] In constructive quantum field theory they proved especially useful since they allow one to construct the two-dimensional Yukawa model and verify Wightman's axioms for it.[6,7]

The idea is the following: First we transform our problem by formal manipulations in such a way that we have to compute $\det(1-\lambda K)$ where K is a compact operator. It often happens that $\mathrm{Tr}K^n$ exists for $n \geq n_0$ but diverges for $n < n_0$, so as a first step we discard the divergent terms in the "loop expansion" for $\ln\det(1-\lambda K) = \mathrm{tr}\ln(1-\lambda K)$, defining

$$\ln\det_{n_0}(1-\lambda K) \equiv \mathrm{Tr}\left(\ln(1-\lambda K) + \sum_{k=1}^{n_0-1}\frac{1}{k}(\lambda K)^k\right).$$

The terms that were thrown out are then reinserted after carefully renormalizing them individually (in applications they usually correspond to one-loop Feynman graphs); in this way a renormalized determinant $\det_{\mathrm{ren}}(1-\lambda K)$ is defined.

The mathematics of $\det_n$ is well understood (see Refs. 8, 9, or 10 for a very clear recent exposition); $\det_n(1-\lambda K)$ is usually defined for operators K with $\|K\|_n \equiv [\mathrm{Tr}(K^*K)^{n/2}]^{1/n} < \infty$ (this space of operators is denoted $\mathscr{I}_n$). One crucial property is the following: $\det_n(1-\lambda K)$ is an entire function of λ of order at most n; it has its zeros exactly at the inverse eigenvalues $1/\lambda_i$ of K and

$$\det_n(1-\lambda K) = \prod_{i=1}^{\infty}\left[(1-\lambda\lambda_i)\exp\left(\sum_{k=1}^{n-1}\frac{1}{k}(\lambda\lambda_i)^k\right)\right].$$

$\det_{\mathrm{ren}}$ has the same zeros as $\det_n$; this fact and the close relation to renormalization theory are main advantages of its definition. Both properties are much less obvious (and sometimes false, as shown in this note) for the currently popular ζ-function definition of determinants[11,2]; this definition has of course other virtues that make it interesting in pure mathematics.

It should be stressed that for our definition of $\det_{\mathrm{ren}}$ K does not have to be self-adjoint—in fact in applications it almost never is—in contrast to the situation for the ζ-function definition.[2]

III. QED$_2$

Here the goal is to define $\Delta \equiv \det \not{D}/\det \not{\partial}$ where $\not{D} = \not{\partial} - ie\not{A}$ and for definiteness we may use the γ matrices

$$\gamma_1 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It might seem natural to try to interpret Δ as $\det[1 - (e/\not{p})\not{A}]$ (with $\not{p} = i\not{\partial}$); in fact a Hilbert space on which $(1/\not{p})\not{A}$ is a compact operator can be found. It is easier (and equivalent), however, to make a formal similarity transformation and interpret Δ as $\det_{\text{ren}}(1 - eK)$ with $K(A) \equiv (\not{p}/|p|^{3/2}) \times \not{A}(1/|p|^{1/2})$ considered as an operator on two-component square-integrable functions on $\mathbb{R}^2$ [i.e., $L^2(\mathbb{R}^2) \oplus L^2(\mathbb{R}^2)$].

We first consider the "nonwinding" case; for simplicity we assume $\int |A_\mu|^q d^2x < \infty$ for all $q \geq \frac{1}{2}$ (i.e., $|A_\mu|^{1/2} \in \bigcap_{q \geq 1} L^q$). Then the following facts are true:

(1) $K \in \mathcal{I}_q$ for all $q > 2$.

(2) The spectrum of K consists only of the origin (i.e., K is quasinilpotent).

(1) follows from a theorem proved for instance in Ref. 12 which asserts that an operator of the form $f(x)g(p)$, i.e., multiplication by g in p space followed by multiplication by f in x space is in $\mathcal{I}_q$ $(2 \leq q < \infty)$ provided f and g are in L^q. To apply this here one has to write K as a product of two factors of the form $f(x)g(p)$ and $g(p)f(x)$, respectively:

$$K = \left(\frac{\not{p}}{|p|}\frac{1}{|p|^{1/2}}|A|^{-1/2}\not{A}\right)\left(|A|^{1/2}\frac{1}{|p|^{1/2}}\right)$$

and then split $1/|p|^{1/2}$ into $1/(1+|p|)^{1/2}$ and $1/|p|^{1/2} - 1/(1+|p|)^{1/2}$; $K \in \mathcal{I}_q$ $(q > 2)$ follows then from some elementary properties of $\mathcal{I}_q$ spaces.

(2) expresses the triviality of the Schwinger model. To prove it one only has to prove $\det_4(1 - eK) = 1$ according to the remark made in Sec. II. This follows from the following facts:

(a) $\det_4(1 - eK)$ is gauge invariant,
(b) $\det_4(1 - eK)$ is even in e,
(c) $\det_4(1 - eK) = \det_4(1 - e\gamma_5 K)$.

(a) should be clear; (b) follows from charge-conjugation invariance (Furry's theorem); (c) follows from (b) by noting $\det_4(1 - eK) = \det_2(1 - e^2K^2)$ and $K^2 = (\gamma_5 K)^2$.

Now by (a) we may assume $\partial_\mu A_\mu = 0$; by (c) and the fact that $\gamma_5 \not{A} = \not{B}$ with $B_\mu = i\epsilon_{\mu\nu}A_\nu$ we see that $\det_4[1 - eK(A)] = \det_4[1 - eK(B)]$. But B_μ is a pure gauge and so $\det_4[1 - eK(B)] = \det_4[1 - eK(0)] = \det_4(1) = 1$.

To define $\det_{\text{ren}}(1 - eK)$ we only have to interpret the graph (Fig. 1) corresponding formally to $\text{Tr}K^2$ in the standard gauge-invariant way (the most convincing way to achieve this might be to start with the lattice approximation, cf. Ref. 19); the well-known answer is

$$\frac{e^2}{2\pi}\int (A_\mu^{\text{trans}})^2 d^2x,$$

where A_μ^{trans} denotes the transverse part of A_μ, uniquely defined by $\partial_\mu A_\mu^{\text{trans}} = 0$, $\epsilon_{\mu\nu}\partial_\nu(A_\mu - A_\mu^{\text{trans}}) = 0$ and $\int (A_\mu^{\text{trans}})^2 < \infty$. Accordingly $\ln \det_{\text{ren}}(1 - eK) = -(e^2/2\pi)\int (A^{\text{trans}})^2 d^2x$ which is the well-known result of Schwinger.[13]

We want to stress here that, obviously,

$$\det_{\text{ren}}(1 - eK) \neq \prod_i (1 - e\lambda_i)$$

in spite of the fact that no counterterm was needed; the reason is that K is not in $\mathcal{I}_2$ even though the graph in Fig. 1 is conditionally convergent. It is *not* equal to $\text{Tr}K^2 = \sum e^2\lambda_i^2 = 0$.

The situation changes drastically for a "winding" field A_μ. A_μ can only fall off as $1/|x|$ and our method fails. In fact $K(A)$ cannot be compact because there is a whole (open) disk of eigenvalues. [Note, however, that K is still bounded: $\|K\| = \| \, |p|^{-1/2}(|x|+1)^{-1/2}\|^2 = \Gamma(\frac{1}{4})^4(2\pi)^{-2}$ according to a result of Herbst.[14]] The disk of eigenvalues can be obtained from the well-known zero solution of $\not{D}\psi = 0$: To be specific, assume $A_\mu = (1/2e)\epsilon_{\mu\nu}x_\nu/(x^2 + \rho^2)$. $K\varphi = \lambda\varphi$ $(\varphi \in L^2 \oplus L^2)$ is equivalent to

$$\left(\not{p} - \frac{1}{\lambda}\not{A}\right)\psi = 0 \quad (|p|^{-1/2}\psi \in L^2).$$

For $\lambda = 1$ this has the well-known solution

$$\psi^{(1)} = \begin{pmatrix} \psi_L^{(1)} \\ 0 \end{pmatrix}, \quad \psi_L^{(1)} = \frac{1}{\rho^2 + x^2}.$$

For general λ we have solutions

$$\psi^{(\lambda)} = \begin{pmatrix} \psi_L^{(1)} \\ 0 \end{pmatrix}$$

with $\psi_L^{(\lambda)} = (\rho^2 + x^2)^{-1/\lambda}$. The condition $|p|^{-1/2}\varphi \in L^2$ requires $\text{Re}(1/\lambda) > \frac{3}{4}$ or $|\lambda - \frac{2}{3}| < \frac{2}{3}$ which defines the above-mentioned disk.

So for winding A_μ there is no possibility to define $\det_{\text{ren}}[1 - eK(A)]$ in such a way that its zeros reflect the eigenvalues of $K(A)$. The ζ-function

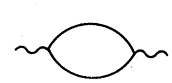

FIG. 1. The graph corresponding formally to $\text{Tr}K^2$.

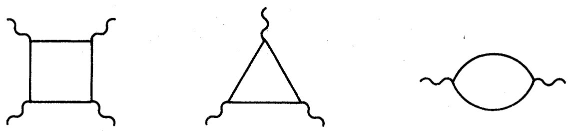

FIG. 2. The graphs corresponding formally to $\mathrm{Tr}K^4$, $\mathrm{Tr}K^3$, $\mathrm{Tr}K^2$.

definition given in Ref. 2(a) exists of course but its field-theoretic meaning is not clear.

IV. QCD_2, QED_4, QCD_4

The general formalism carries over unchanged, but of course explicit computation is no longer possible.

QCD_2 does not need any further discussion here, so we turn to four dimensions. For gauge fields that fall off sufficiently fast, e.g., $|A_\mu|^{1/2} \in \bigcap_{q\geq 1} L^q$, we have now $K(A) \in \mathcal{J}_q$ for $q>4$ and the definition of $\det_{\mathrm{ren}}$ is straightforward:

$$\ln \det_{\mathrm{ren}}(1-eK) \equiv \ln \det_5(1-eK) - \tfrac{1}{4}e^4(\mathrm{tr}K^4)_{\mathrm{ren}} - \tfrac{1}{3}e^3(\mathrm{tr}K^3)_{\mathrm{ren}} - \tfrac{1}{2}e^2(\mathrm{tr}K^2)_{\mathrm{ren}},$$

where $(\mathrm{tr}K^n)_{\mathrm{ren}}$ $(n=2,3,4)$ stands for the standard gauge invariantly renormalized (and interpreted, respectively) graphs in Fig. 2; the graph in the middle is of course zero in QED_4.

For winding fields in QCD_4 one has again the disastrous situation encountered before: K fails to be compact while still being bounded; it has spectrum everywhere in a disk. This can be seen in a way that is exactly analogous to the "winding" QED_2 discussion; again one uses the well-known "zero solution" of $\not{D}\psi=0$.[1,15] So in this case again there is no fully satisfactory definition of the determinant.

It should also be noted that the problem cannot be solved by factoring off a finite number of zero modes because changing a noncompact operator by a finite-rank operator cannot make it compact. Introducing a mass in the fermion propagator, on the other hand, makes K compact, and a satisfactory definition of the determinant is possible using a suitable infrared regularization for the (finite number of) divergent loops.

V. CONCLUSIONS

The connection between fermion determinants and eigenvalues of the corresponding Dirac operators is subtle and requires careful examination in each case. The renormalized determinants described here allow us to answer such questions unambiguously, and they show that in some instances this relation is more complicated than what might be expected naively. In the situation that received the widest attention, namely, where massless fermions are coupled to topologically nontrivial gauge fields, these renormalized determinants cease to exist and no natural generalization seems to be available. Thus the meaning of the often heard statement that the existence of zero modes for the Dirac operator implies vanishing of the fermion determinant[1,2,16] is not clear.

I want to make the (somewhat unorthodox) suggestion that it may not be necessary to find a fully satisfactory definition of fermion determinants for "winding" gauge fields. θ states have been constructed without getting into that problem at least in two dimensions[17,18]; generally they should be obtainable by adding a term $i\theta\int_V q(x)dx$ ($q(x)$ = topological charge density) to the action, first for a finite volume V, and then expanding V to all of space-time. 't Hooft's mechanism for the avoidance of the U(1) problem[1,16] might also be unaffected by that difficulty. In the one case where it is rigorously understood—the (Thirring-) Schwinger model, it can be reinterpreted as the absence of any symmetry, spontaneously broken or not, corresponding to chiral transformations on the physical Hilbert space (cf. Ref. 18); this conclusion is of course not dependent on any construction of the fermion determinant for topologically nontrivial gauge fields.

Note added. We should make more precise in what sense our definition is independent of boundary conditions. The reader has noticed that we imposed a certain falloff on the functions on which our operator K acts by requiring them to be square-integrable and this might be considered by some to be a boundary condition. The crucial point is, however, that both the operator K and the corresponding determinant can be approximated by finite-volume expressions with various boundary conditions (such as periodic, antiperiodic, Dirichlet) and that these approximations converge to a limit independent of those boundary conditions. This can be seen rather easily by adaptation of the methods used in Sec. II to show that $K \in J_q$.

ACKNOWLEDGMENT

I thank A. Patrascioiu for correspondence.

[1]G. 't Hooft, Phys. Rev. Lett. 37, 8 (1976); Phys. Rev. D 14, 3432 (1976); 18, 2199(E) (1978).

[2](a) M. Hortacsu, K. D. Rothe, and B. Schroer, Phys. Rev. D 20, 3203 (1979); (b) Report No. FUB/HEP, 1979 (unpublished); (c) N. K. Nielsen, K. D. Rothe, and B. Schroer, Nucl. Phys. B160, 330 (1979); (d) K. D. Rothe and J. A. Swieca, Ann. Phys. (N.Y.) 117, 382 (1979); (e) B. Schroer, in *Facts and Prospects of Gauge Theories*, proceedings of the XVII Schladming Conference on Nuclear Physics, edited by P. Urban (Springer, Berlin, 1978) [Acta Phys. Austriaca Suppl. 19 (1978)], p. 155; (f) K. D. Rothe and B. Schroer, in *Field Theoretical Methods in Particle Physics*, edited by W. Rühl (Plenum, New York, 1980); (g) R. Schrader and R. Seiler,

[3]A. Patrascioiu, Phys. Rev. D 20, 491 (1979); Phys. Rev. D (to be published).

[4]J. Schwinger, Phys. Rev. 93, 615 (1953).

[5]D. Hilbert, Nachr. Akad. Wiss. Goettingen Math. Phys. KL., 49 (1904).

[6]E. Seiler, Commun. Math. Phys. 42, 163 (1975).

[7]J. Magnen and R. Sénéor, Commun. Math. Phys. 51, 297 (1976); A. Cooper and L. Rosen, Trans. Ann. Math. Soc. 234, 93 (1977).

[8]N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963), Vol. II.

[9]I. C. Goh'berg and M. G. Krein, *Introduction to the Theory of Linear Non-Selfadjoint Operators*, Translations AMS 18, Providence 1969.

[10]B. Simon, Adv. Math. 24, 244 (1977).

[11]R. T. Seeley, in *Proceedings of the Symposium on Pure and Applied Mathematics* (American Mathematical Society, Providence, 1967), Vol. 10, p. 288; D. Ray and I. Singer, Adv. Math. 7, 145 (1971).

[12]E. Seiler and B. Simon, Commun. Math. Phys. 45, 99 (1975).

[13]J. Schwinger, Phys. Rev. 128, 2425 (1962).

[14]I. W. Herbst, Commun. Math. Phys. 53, 285 (1977).

[15]R. Jackiw and C. Rebbi, Phys. Rev. D 14, 517 (1976).

[16]S. Coleman, in *New Phenomena in Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1977).

[17]S. Coleman, R. Jackiw, and L. Susskind, Ann. Phys. (N.Y.) 93, 267 (1975).

[18]J. Fröhlich and E. Seiler, Helv. Phys. Acta 49, 889 (1976).

[19]D. H. Weingarten and J. L. Challifour, Ann. Phys. (N.Y.) 123, 61 (1979).