Index Theorems and Anomalies: a Common Playground for Mathematicians and Physicists†

S.N.M. Ruijsenaars
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

An elementary introduction to the area of index theory and anomalies is presented.

CONTENTS

† Work supported by the Netherlands Organisation for the Advancement of Research (NWO)

1. INTRODUCTION

The following is an attempt to provide a general idea of the closely related areas of index theory and anomalies. Our account is directed towards mathematicians with some background in functional analysis and differential geometry, but we presuppose no previous acquaintance with Atiyah-Singer type theorems and anomalies (or with theoretical physics, for that matter).

The mathematical side of the subject matter is described in Section 2. After some preliminaries concerning Fredholm operators we illustrate index theorems by detailing the simplest case (Toeplitz operators on the circle). Our survey of the far-reaching generalizations of this result which now constitute index theory is however very sketchy. We recommend [3] for further browsing and a host of annotated references, [9] for a recent and elegant introduction, and [6] for a lucid overview of index theorems and the geometric structures associated with this area. We have no occasion to say anything about noncommutative index theory; for information on this setting, see [2].

In Section 3 the physical side of the coin is pictured, mainly by means of a rather detailed discussion of an explicit example. This example has been selected for several reasons. First, it pertains to a situation where various physically important objects can be dealt with in a mathematically rigorous way. Second, it clearly illustrates how index theorems can help in answering questions that naturally arise in quantum field theory. Third, the example can be compared to the so-called chiral anomaly in two space-time dimensions. Last but not least, the context in which the example is set is the only 'anomaly context' the author understands to his satisfaction and hence feels competent to explain. (The last reason is not unrelated to the first; we do present our limited understanding of other contexts towards the end of the paper.)

Since we are not assuming any background in quantum field theory, Section 3 includes a sketch of some basic constructs in this area. This is one reason why much remains to be said about anomaly theory proper. In particular, its applications to real world physics are left out altogether. (There are quite a few.) For more information and a plethora of references one might consult the reprint collection [5] and the Conference Proceedings [1]; to read up on string theories and their (absence of) anomalies we recommend [8].

2. WHAT IS INDEX THEORY ALL ABOUT?

2.1. Fredholm operators and their indices

Let $\mathcal{H}$ be a (separable, complex) Hilbert space and let $A: \mathcal{H} \rightarrow \mathcal{H}$ be a Fredholm operator. That is, A is a bounded (linear) operator with finite-dimensional kernel K and cokernel C. Then the Fredholm index of A is defined by

$$\operatorname{index} A = \dim K - \dim C. \tag{1}$$

In particular, any bounded operator having a bounded inverse is Fredholm with index 0.

If $\mathcal{H}$ has finite dimension, then it follows from elementary linear algebra that any operator on $\mathcal{H}$ is Fredholm and has vanishing index. It is an equally simple fact that this is false when dim $\mathcal{H}=\infty$. In that case there are lots of operators that are not Fredholm, and Fredholm operators need not have vanishing index. The first assertion is evident, whereas the second one is best explained with the following example. Consider the sequence space $l^2(\mathbb{N})$ and let L and R be the left and right shifts, i.e.,

$$L(a_0,a_1,a_2,...) \equiv (a_1,a_2,a_3,...), R(a_0,a_1,a_2,...) \equiv (0,a_0,a_1,...). \quad (2)$$

Then it is clear that L^n and R^n are Fredholm with index n and $-n$, respectively. Note that R is the Hilbert space adjoint of L, $R = L^*$.

It is not hard to see that the set $F(\mathcal{H})$ of all Fredholm operators on $\mathcal{H}$ is closed under taking adjoints, and that

$$\text{index } A^* = -\text{index } A, \quad \forall A \in F(\mathcal{H}). \quad (3)$$

It is less immediate, but true, that $F(\mathcal{H})$ is also closed under operator composition and that

$$\text{index } A_1 A_2 = \text{index } A_1 + \text{index } A_2, \quad \forall A_1, A_2 \in F(\mathcal{H}). \quad (4)$$

Other salient facts include (we assume dim $\mathcal{H}=\infty$ from now on):

- Equipped with the topology derived from the operator norm $F(\mathcal{H})$ is an open subset of $B(\mathcal{H})$ (the bounded operators) with connected components labeled by the index;
- Addition of a compact operator changes neither the Fredholm property nor the index:

$$\text{index } (A + C) = \text{index } A, \quad \forall A \in F(\mathcal{H}), \quad \forall C \in K(\mathcal{H}). \quad (5)$$

- For any $A \in F(\mathcal{H})$ there exist $Q \in F(\mathcal{H})$ and $C_R, C_L \in K(\mathcal{H})$ such that

$$A\,Q = \mathbf{1} + C_R, \quad QA = \mathbf{1} + C_L. \quad (6)$$

As a consequence, any two Fredholm operators with the same index can be connected by a norm continuous path in $F(\mathcal{H})$.

The concepts and results mentioned so far belong to the area of abstract functional analysis, and little more can be said in this general context. However, it has turned out that they can be used to great advantage in a variety of other areas in mathematics and physics, including algebraic topology, differential and algebraic geometry, quantum field theory and string theory. The link with these areas is forged via theorems of Atiyah-Singer type. Roughly speaking, such theorems express the index of Fredholm operators which naturally occur in concrete geometrical or physical contexts in terms of topological invariants. We proceed by discussing this in more detail for Toeplitz and Dirac operators.

2.2. Index theorems for Toeplitz operators

Denote the continuous functions on the unit circle by $C(S^1)$. Each $f \in C(S^1)$ defines a bounded multiplication operator M_f on $L^2(S^1)$. Writing $F \in L^2(S^1)$ as

$$F(z) = \sum_{n\in\mathbb{Z}} a_n z^n, \quad z\in S^1, \tag{7}$$

we set

$$(P_+F)(z) = \sum_{n\geqslant 0} a_n z^n, \quad (P_-F)(z) = \sum_{n<0} a_n z^n. \tag{8}$$

Thus we have a direct sum decomposition

$$L^2(S^1) \simeq \mathcal{H}_+ \oplus \mathcal{H}_-, \quad \mathcal{H}_\delta \equiv P_\delta L^2(S^1), \quad \delta=+,- \tag{9}$$

and functions in $\mathcal{H}_+$ ($\mathcal{H}_-$) are L^2-boundary values of holomorphic functions in $|z|<1$ ($|z|>1$). Operators of the form

$$T_f \equiv P_+ M_f P_+, \quad f\in C(S^1) \tag{10}$$

are called Toeplitz operators, and it is not hard to see that they are Fredholm (viewed as operators on $\mathcal{H}_+$), provided f does not vanish on S^1. Assuming this from now on, the curve $f(S^1)\subset\mathbb{C}^*$ has a well-defined winding number $w(f)$ with respect to the origin. The equality

$$\text{index } T_f = -w(f) \tag{11}$$

between objects from the area of analysis on the one hand and from the areas of topology and geometry on the other is the simplest example of an Atiyah-Singer type theorem. Note that the special case $f=z^l$, $l\in\mathbb{Z}$, has already occurred above: The Fourier transforms of T_{z^l} with $l=1,-1$ are just the one-sided shifts R and L, respectively. Therefore, (11) is an easy consequence of the above-mentioned properties of Fredholm operators.

In this example the Fredholm operators arise from continuous multipliers on the L^2-space over S^1; the compression (10) of M_f to $\mathcal{H}_+$ is needed to get a non-zero index (note M_f has a bounded inverse $M_{1/f}$ when $0\notin f(S^1)$). The splitting (9) may be viewed as a decomposition of $L^2(S^1)$ into the positive- and negative-energy spaces of the Dirac operator D on S^1, which reads

$$D = z\frac{d}{dz} \tag{12}$$

in the variables used above.

This picture can be generalized to $(2N+1)$-dimensional oriented compact Riemannian manifolds $\mathfrak{M}$ that admit a spin structure. (We assume $N>0$ from now on.) This is a lifting of the transition functions of the tangent bundle $T\mathfrak{M}$ (which may be assumed to take values in $SO(2N+1)$) to the simply-connected two-fold covering group Spin $(2N+1)$ of $SO(2N+1)$. The spin group has a faithful irreducible representation on $\mathbb{C}^{2^N}$ and correspondingly one obtains a $\mathbb{C}^{2^N}$-bundle over $\mathfrak{M}$, the spinor bundle. The connection on $T\mathfrak{M}$ derived from the metric can now be lifted to a connection on the spinor bundle, and from the covariant derivative corresponding to the spin connection and the generators of a Clifford algebra one can then construct a first order elliptic differential operator that acts on sections of the spinor bundle.

The Dirac operator thus obtained depends on the choice of spin structure, but its most important property is independent of this choice: It is self-adjoint as an operator on the L^2-space $\mathcal{H}$ associated with the spinor bundle and it has infinite-dimensional positive and negative spectral subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$. Just as in the case of S^1, a continuous map from $\mathfrak{M}$ to $\mathbb{C}^*$ gives rise to a Fredholm operator on $\mathcal{H}_+$, and more generally a continuous map from $\mathfrak{M}$ to $GL(k,\mathbb{C})$ yields a Fredholm operator on $\mathcal{H}_+ \otimes \mathbb{C}^k$.

For a smooth map the index of the generalized Toeplitz operator just described can be written in terms of an integral over $\mathfrak{M}$ involving certain closed differential forms. The value of this integral does not change when exact forms are added, since $\mathfrak{M}$ has no boundary. Thus, one is dealing with de Rham cohomology classes. In the above context the class involved ('characteristic class') is determined by the Riemann curvature tensor of $\mathfrak{M}$ and the topological ('winding') characteristics of the map. This state of affairs is the generalization of the fact that for smooth non-vanishing $f \in C(S^1)$ the index theorem (11) can also be written

$$\text{index } T_f = -\int_0^{2\pi} df/2\pi i f. \tag{13}$$

2.3. Index theorems for Dirac operators

On $2N$-dimensional compact oriented Riemannian manifolds admitting a spin structure a spinor bundle can be defined, too. In this case the fibre $\mathbb{C}^{2^N}$ splits into a direct sum of even and odd spinors corresponding to two different irreducible representations of $SO(2N)$ on $\mathbb{C}^{2^{N-1}}$. With respect to this decomposition the Dirac operator can now be written $\begin{bmatrix} 0 & D^* \\ D & 0 \end{bmatrix}$, where D and D^* are again first order elliptic differential operators expressed in terms of Clifford algebra generators and the spin connection. Tensoring the spinor bundle with a (complex) vector bundle equipped with a connection A, one can define a Dirac operator on the tensor product which involves A and which takes the form $\begin{bmatrix} 0 & D_A^* \\ D_A & 0 \end{bmatrix}$ with respect to the even/odd spinor decomposition. Again the index of D_A (which can be viewed as a Fredholm operator between two different Hilbert spaces) can be expressed as an integral over $\mathfrak{M}$ involving characteristic classes that depend on the curvatures of the two connections.

Probably the simplest example of this situation is the torus $\mathfrak{M} = S^1 \times S^1$ with its flat metric. Employing the above coordinates on S^1 and the obvious spin structure ('periodic boundary conditions') one can take

$$\mathcal{H} = L^2(S^1) \otimes L^2(S^1) \otimes \mathbb{C}^2, \quad D = z_1 \frac{\partial}{\partial z_1} + i z_2 \frac{\partial}{\partial z_2}. \tag{14}$$

Since the curvature vanishes, the index theorem implies index $D=0$. (In fact, this is always the case when N is odd, as can be read off from the relevant cohomology class. Moreover, in the case at hand this is immediate from (14): both the kernel and the cokernel of D are spanned by the constant sections.)

However, when one tensors the spinor bundle with a line bundle with connection A, the index theorem reads

$$\text{index } D_A = -\frac{1}{2\pi}\int_{\mathcal{M}} F \tag{15}$$

where F is the curvature 2-form corresponding to A.

The framework just exemplified is very general. It encompasses such results as the Gauss-Bonnet-Chern theorem, the Hirzebruch signature theorem and (when $\mathcal{M}$ is a Kähler manifold) Riemann-Roch type theorems. The index theorems can be used to obtain information on various geometric questions, such as the existence of positive scalar curvature metrics or zeros of vector fields on the given manifold $\mathcal{M}$. Other applications include conclusions concerning topological invariants of manifolds obtained from 'simple' manifolds (for example, spheres and tori) by glueing or covering operations; This hinges on the additive properties of the index that are evident from its being given by an integral over the manifold. In the other direction, the fact that Fredholm indices are integral can be used to conclude that certain rational cohomology classes are actually integral on manifolds admitting the structure necessary for the relevant index theorem to apply, or that certain manifolds do not admit such structures, since one knows already that the relevant characteristic class is not integral.

The scope of applications of index theory and its unifying character can be further enhanced by generalizing it to pseudo-differential operators and families thereof, and to a noncommutative setting. However, we shall not go into this and refer instead to the literature cited in the introduction.

3. What are anomalies all about?

3.1. Preamble

The notion of anomaly currently plays an important role in theoretical elementary particle physics. This is the case not only in the well-established area of relativistic quantum field theory, but even more so in string theories, which are sweeping (but so far highly speculative) generalizations of relativistic quantum field theory.

Here, we shall restrict ourselves to the context of quantum field theory. This may be viewed as a generalization of quantum mechanics accomodating the phenomenon of creation and annihilation of elementary particles, which is experimentally observed in cosmic radiation and high energy colliders. For the quantum field models that are most relevant for the description of such real world processes a nonperturbative and mathematically rigorous grip on the dynamics and scattering is not likely to be achieved in this century. However, physicists cannot afford to wait til mathematics catches up and, moreover, they have a good reason not to worry much about existence questions: Perturbative quantum field theory yields numbers that agree very well with experiment (to an accuracy of better than one part in 10^7 in some cases). Furthermore, perturbative field theory is in great shape from a mathematical viewpoint, too.

On the other hand, the advent of nonabelian gauge field theories in the seventies led to the opinion that one must try and go beyond perturbation theory if one is ever to understand why quarks (the constituents of the hadrons, a large class of elementary particles) are not observed in high energy collision processes. This interest in nonperturbative phenomena led to a reappraisal of certain inconsistencies known since the late sixties. These so-called anomalies occurred in the perturbation series for certain models, and formed an obstruction to a consistent renormalization of the infinities such series are fraught with. In the late seventies physicists came to the conclusion that anomalies should really be viewed as being nonperturbative features of a quantum field model, which indicate that certain symmetries (such as gauge invariance) of its classical version do not survive the quantization procedure sketched below. Moreover, the connection between anomaly theory and index theory surfaced at that time, and has been further worked out ever since.

After having provided some context, we now turn to an example that may be viewed as the simplest version of the so-called chiral anomaly. (Chirality = 'handedness'; this term refers to a two-fold degree of freedom for massless Dirac particles, cf. below.) The physical situation we wish to consider is the interaction of Dirac particles on the line $\mathbb{R}$ with a time-dependent external field. This can be described at the levels of classical and quantum field theory (also called the first- and second-quantized levels in physics).

3.2. Scattering at a gauge field: classical version

At the classical level one is concerned with complex-valued functions ψ on Minkowski space-time $\mathbb{R}^2$ that satisfy the (massless, positive chirality) Dirac equation

$$[i\partial_t + i\partial_x - \lambda A(t,x)]\psi(t,x) = 0. \tag{16}$$

Here, $\lambda\in\mathbb{R}$ is the coupling constant and the external field A is a given smooth real-valued function on $\mathbb{R}^2$ with compact support in x for fixed t.

The object that is most relevant to the present discussion is the scattering operator S associated to the hyperbolic partial differential equation (16). In order to introduce S we first rewrite (16) as a Hilbert space evolution equation

$$i\frac{d}{dt}\psi = (H_0+\lambda A(t,\cdot))\psi, \quad H_0 \equiv -i\frac{d}{dx} \tag{17}$$

in $L^2(\mathbb{R},dx)$. Then a uniquely determined unitary evolution operator $U(T_2,T_1)$ exists satisfying

$$U(T_3,T_2)U(T_2,T_1) = U(T_3,T_1), \quad U(T,T) = \mathbf{1} \tag{18}$$

and solving (17) in the sense that

$$i\partial_t U(t,s)\psi = (H_0+\lambda A(t,\cdot))U(t,s)\psi \tag{19}$$

This propagator can be obtained by iteration from the integral equation associated with (17) and is norm continuous in λ. The S-operator is then defined by

$$S = \lim_{\substack{t\to\infty \\ s\to-\infty}} e^{itH_0} U(t,s) e^{-isH_0}, \tag{20}$$

whenever the (strong) limit exists. The operator S codes the change in the wave function ψ of the Dirac particle due to its scattering at the external field λA. Note, in particular, the special case

$$\lambda = 0 \Rightarrow U(t,s) = e^{-i(t-s)H_0}, \quad S = \mathbf{1}. \tag{21}$$

Next, we impose the further restriction

$$A(t,x) = \begin{cases} \alpha(x) \in C_0^\infty(\mathbb{R}) & t > T > 0 \\ 0 & t < -T \end{cases} \tag{22}$$

Hence, A is a smooth interpolation of two time-independent fields. The assumption (22) suffices for S to exist, as will now be shown. Note first that one has

$$H \equiv H_0 + \lambda\alpha(\cdot) = M^* H_0 M \tag{23}$$

where M denotes the unitary operator of multiplication by

$$M(x) \equiv \exp(-i\lambda \int_x^\infty \alpha(y)dy). \tag{24}$$

Since $U(t,T) = e^{-i(t-T)H}$ for $t > T$, it now follows from (18) that

$$U(t,s) = M^* e^{-i(t-T)H_0} M U(T,-T) e^{i(T+s)H_0}, \; t \geqslant T, \;\; s \leqslant -T. \tag{25}$$

Next, note e^{itH_0} is just translation over a distance t, so that

$$s\text{-}\lim_{t\to\infty} e^{itH_0} M^* e^{-itH_0} = \mathbf{1}. \tag{26}$$

Therefore, S exists and is given by

$$S = e^{iTH_0} M U(T,-T) e^{iTH_0}. \tag{27}$$

At this point there is still no hint that the upshot (27) of our reasoning is related to index theory or that it has anything anomalous about it. However, this can only be made clear after we have discussed the second-quantized, many-particle framework.

3.3. Dirac's second quantization

To this end we begin by pinpointing the reason why the first-quantized description is inadequate (physically speaking): The one-particle Dirac Hamiltonian H_0 has a spectrum that is not bounded below, but particles with negative energies have never been observed. The quantum field description arises naturally in curing this physical disease of the one-particle Dirac theory. The picture that leads up to a many-particle framework is referred to as the 'Dirac sea': since no negative energy states are observed, Dirac postulated that the negative energy states are all filled by a sea of unobservable particles. Annihilating such a negative energy particle with a given charge should then amount to creating a 'hole in the sea', which can be observed as a new type of positive energy particle with an opposite charge. This highly intuitive picture led Dirac to predict that charged particles described by his equation (such as electrons) have oppositely charged partners (antiparticles), a prediction that has been abundantly confirmed by experiment.

The mathematical arena in which all this can be formalized is the so-called fermion Fock space $\mathcal{F}_a(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space that may be viewed as a one-particle space. This Fock space is defined by

$$\mathcal{F}_a(\mathcal{H}) = (\mathbb{C} \oplus \mathcal{H} \oplus \wedge^2 \mathcal{H} \oplus \wedge^3 \mathcal{H} \oplus ...)^{-}, \tag{28}$$

i.e., it equals the completion of the antisymmetric tensor algebra over $\mathcal{H}$ in the obvious inner product. The tensor (1,0,0,...) is viewed as the vacuum (the 'filled Dirac sea') and denoted by Ω. To get around in Fock space one employs the creation and annihilation operators $c^*(F)$ and $c(F)$, $F \in \mathcal{H}$. The former operator acts by wedging with F, entailing

$$c^*(F_1)...c^*(F_n)\Omega = F_1 \wedge ... \wedge F_n, \tag{29}$$

and $c(F)$ is its adjoint (contraction with F).

In the Dirac theory $\mathcal{H}$ is an L^2-space on which the Dirac operator is a multiplication operator, and one usually writes

$$c(F) = a(P_+ F) + b(P_- F), \quad F \in \mathcal{H} \tag{30}$$

where $\mathcal{H}_\delta \equiv P_\delta \mathcal{H}$, $\delta = +, -$, are the positive and negative spectral subspaces of $\mathcal{H}$. Thus, a and b are to be viewed as particle and antiparticle annihilators, respectively. The subspace of $F_a(\mathcal{H})$ spanned by tensors of the form

$$\prod_{i=1}^{k} a^*(f_i) \prod_{j=1}^{l} b^*(g_j)\Omega, \quad k - l = q, \; f_i \in \mathcal{H}_+, \; g_j \in \mathcal{H}_-, \tag{31}$$

is called the charge-q sector and the operator Q with eigenvalue q on this sector is called the charge operator, in keeping with the physical picture sketched above.

Let us now return to the special Dirac equation (16). In the free case $\lambda = 0$ any of its L^2-solutions can be written

$$\psi(t,x) = (2\pi)^{-1/2} \int_0^\infty dp\,[e^{ip(x-t)} a(p) + e^{-ip(x-t)} b(p)] \tag{32}$$

where $a, b \in L^2([0, \infty), dp)$. Correspondingly one can take

$$\mathcal{H}_\delta = L^2([0, \infty), dp), \quad \delta = +, -. \tag{33}$$

The transform of H_0 to $\mathcal{H}$ (again denoted H_0) then acts as multiplication by p on $\mathcal{H}_+$ and by $-p$ on $\mathcal{H}_-$. Dirac's second quantization of the first-quantized field (32) is now given by

$$\psi(t,x) = (2\pi)^{-1/2} \int_0^\infty dp\,[e^{ip(x-t)} a(p) + e^{-ip(x-t)} b^*(p)]. \tag{34}$$

Here, $a(p)$ and $b^*(p)$ are quadratic forms that may be viewed as the densities of the operators $a(f)$ and $b^*(g)$ on $F_a(\mathcal{H})$, defined above: One has, e.g.,

$$b^*(g) = \int_0^\infty dp b^*(p) g(p), \quad g \in \mathcal{H}_-. \tag{35}$$

(The integrals in (34) and (35) have a rigorous meaning as quadratic form integrals on a dense subspace of $F_a(\mathcal{H})$, but only in the second case the resulting form is the form of an operator.)

The unsmeared free Dirac field (34) may be viewed as an operator-valued tempered distribution: Smearing with a function $f(t,x)$ in the Schwartz space $S(\mathbb{R}^2)$ yields a bounded operator that can be written

$$a(P_+F) + b^*(\overline{P_-F}) \equiv \Phi(F), \quad F \in \mathcal{H} \tag{36}$$

(The bar denotes complex conjugation; the antilinearity of Φ in F is conventional.) The 'abstract' Dirac field $\Phi(F)$ and its adjoint generate a C^*-algebra, the so-called CAR algebra (CAR = canonical anticommutation relations). Unitary operators U on $\mathcal{H}$ give rise to automorphisms $\Phi(F) \mapsto \Phi(UF)$ of the CAR algebra, and whenever a unitary operator $\mathcal{U}$ on $\mathcal{F}_a(\mathcal{H})$ exists satisfying

$$\mathcal{U}\Phi(F) = \Phi(UF)\mathcal{U}, \quad \forall F \in \mathcal{H} \tag{37}$$

it is regarded as the many-particle version of U. In particular, the 'unphysical' one-particle Dirac dynamics e^{itH_0} is implemented by a unitary one-parameter group on $\mathcal{F}_a(\mathcal{H})$, whose generator (the second-quantized Dirac Hamiltonian) has positive spectrum.

The (heuristics of the) formalism just sketched has been standard fare in physics for half a century. Dirac's 'hole theory' substitution $b(p) \mapsto b^*(p)$ turning $c(F)$ into $\Phi(F)$ is a key step: It is not only indispensable for the great success of interacting relativistic quantum field theory in describing high energy physics phenomena, but is also responsible for the highly singular character of quantum field theory and for its intimate relations to seemingly disconnected parts of mathematics, such as index theory.

3.4. Scattering at a gauge field: quantum version

We are now prepared to return to our example. Recall we have a unitary operator S on $L^2(\mathbb{R},dx)$ that codes the single particle scattering. Transforming to $\mathcal{H}$ and denoting the resulting operator again by S, the first question to answer is whether a unitary Fock space operator $\mathbb{S}$ exists implementing the automorphism $\Phi(F) \mapsto \Phi(SF)$, cf. (37).

The necessary and sufficient condition for this is that the off-diagonal parts of

$$S \simeq \begin{pmatrix} S_{++} & S_{+-} \\ S_{-+} & S_{--} \end{pmatrix} \tag{38}$$

be Hilbert-Schmidt (HS). (Here, S_{++} stands for P_+SP_+, e.g. Recall that an operator A is HS when A^*A has finite trace.) Whenever this holds, S_{++} is

Fredholm (as an operator on $\mathcal{H}_+$). Indeed, by unitarity one has

$$\begin{aligned} S_{++}S_{++}^* &= P_+ - S_{+-}S_{-+}^* \\ S_{++}^*S_{++} &= P_+ - S_{+-}^*S_{-+}, \end{aligned} \tag{39}$$

so that S_{++} is invertible modulo compacts. A key point is now that the structure of the unitary implementer $\mathbb{S}$ strongly depends on the Fredholm index i of S_{++}: It maps the charge-q sector in $\mathcal{F}_a(\mathcal{H})$ onto the charge-$(q-i)$ sector.

To investigate these issues for the case at hand, let us recall that S is explicitly given by (27). Under our assumptions on A it can be shown that the operators $U(T,-T)_{\delta,-\delta}$ are HS. Moreover, the index of $U(T,-T)_{++}$ vanishes, since the propagator is norm continuous in λ and equals 1 for $\lambda=0$. Since the off-diagonal parts of the free Dirac evolution vanish, the operators $S_{\delta,-\delta}$ and $S_{\delta\delta}$ are HS and Fredholm, respectively, if and only if $M_{\delta,-\delta}$ and $M_{\delta\delta}$ are. Furthermore, one has

$$\text{index } S_{++} = \text{index } M_{++} \tag{40}$$

whenever both operators are Fredholm, in view of the properties of the Fredholm index mentioned in Subsection 2.1.

To establish whether the diagonal and off-diagonal parts of M are Fredholm and HS, respectively, it is expedient to unitarily transform from $L^2(\mathbb{R})$ to $L^2(S^1)$ by using the Cayley transform $S^1 \to \mathbb{R}$, $z \mapsto (z-i)/(z+i)$. Then the splitting of $L^2(\mathbb{R})$ associated with H_0 transforms into the splitting of $L^2(S^1)$ associated with D, cf. (9), (12). (Recall that functions holomorphic in the unit disc turn into functions holomorphic in the upper half-plane under the Cayley transform.) Moreover, M turns into multiplication by a smooth function on $L^2(S^1)$ if and only if

$$n \equiv \frac{\lambda}{2\pi}\int_{-\infty}^{\infty} \alpha(y)dy \in \mathbb{Z} \tag{41}$$

(recall (24) to see this). It is also readily verified that M has HS off-diagonal parts when (41) holds. (If $n \notin \mathbb{Z}$, one can show $M_{\delta,-\delta}$ is not even compact.) Invoking now the index theorem (11) and recalling (40) one concludes

$$n \in \mathbb{Z} \Rightarrow S_{++} \text{ Fredholm, index } S_{++} = -n. \tag{42}$$

The above chain of arguments exemplifies how index theory naturally emerges from quantum field theoretic considerations, but the relation to a breakdown of gauge invariance at the quantum level is not yet apparent. Before coming to that, we would like to mention that the representation theory of Virasoro and Kac-Moody algebras is also closely tied up with the one-dimensional, massless, single chirality Dirac theory considered thus far. The crux is, that the central extensions that characterize these algebras naturally arise in the transition from the classical to the quantum level, again as a consequence of the 'hole theory' substitution $b(p) \mapsto b^*(p)$. Physicists view such central extensions as anomalies in quantum commutators ('Schwinger terms'), but since this type of anomaly is less directly related to index theorems, we shall not work this out any further.

(A recent account in the spirit of this paper can be found in [4].)

3.5. The chiral anomaly in two dimensions

Let us now come to the two-dimensional chiral anomaly and its relation to the example just discussed. To this end we need to consider massless Dirac particles of positive and negative chirality (even and odd spinors) in interaction with external vector and axial vector fields A^v_μ and A^a_μ, $\mu=0,1$. These terms refer to transformation properties under the Lorentz group that are tied up with Clifford algebra lore, but for brevity we shall not spell this out. Rather, we choose a representation of the two-dimensional Clifford algebra such that the corresponding classical Dirac equation reads

$$\begin{bmatrix} i\partial_0+i\partial_1-A^v_0-A^v_1-A^a_0-A^a_1 & 0 \\ 0 & i\partial_0-i\partial_1-A^v_0+A^v_1+A^a_0-A^a_1 \end{bmatrix}\psi = 0. \tag{43}$$

(Here and from now on we use $x^0=t$, $x^1=x$ as space-time coordinates.) We make assumptions on A^v_μ and A^a_μ that parallel those on the above external field A. This guarantees the existence of the classical evolution and scattering operators as unitary operators on $L^2(\mathbb{R},dx)\otimes\mathbb{C}^2$.

Clearly, (43) is invariant under separate gauge transformations of the potentials (connections) A^v_μ and A^a_μ. That is, if ψ solves (43), then

$$\tilde{\psi} \equiv \begin{bmatrix} \exp(i\Lambda^v + i\Lambda^a) & 0 \\ 0 & \exp(i\Lambda^v - i\Lambda^a) \end{bmatrix}\psi \tag{44}$$

solves (43) with

$$A^i_\mu \mapsto A^i_\mu - \partial_\mu\Lambda^i, \quad i=v,a. \tag{45}$$

Note that these transformations do not change the field strengths (curvatures)

$$F^i_{\mu\nu} \equiv \partial_\mu A^i_\nu - \partial_\nu A^i_\mu, \quad i=v,a. \tag{46}$$

These invariance properties can also be formulated in a Lagrangean framework and imply via a standard variational argument ('Noether's theorem') that the vector and axial vector currents j^v_μ and j^a_μ given by

$$\begin{aligned} j^v_0 &= -j^a_1 = \bar{\psi}_+\psi_+ + \bar{\psi}_-\psi_- \\ j^a_0 &= -j^v_1 = \bar{\psi}_+\psi_+ - \bar{\psi}_-\psi_- \end{aligned} \tag{47}$$

are conserved (i.e., $\partial_0 j^i_0 - \partial_1 j^i_1 = 0$). This in turn leads via Gauss' theorem to the corresponding charges

$$q^i = \int dx\, j^i_0(t,x), \quad i=v,a \tag{48}$$

being time-independent. (Of course, both properties can be directly verified from (43), as well.)

Next, let us pass to the quantum level. The relevant Fock space is now the tensor product of two copies of the previous one, corresponding to the two-fold chirality degree of freedom. It is expedient to introduce two chiral charge operators Q_+ and Q_- referring to the two species of Dirac particles, and to set

$$Q^v = Q_+ + Q_-, \quad Q^a = Q_+ - Q_-. \tag{49}$$

As before, an appropriate choice of the time-independent fields to which A^v_μ and A^a_μ reduce for $t>T>0$ ensures existence of the Fock space $\mathbb{S}$-operator. Moreover, the winding numbers of the corresponding multipliers determine the charge changes ΔQ^v and ΔQ^a effected by $\mathbb{S}$.

We are finally ready to connect all this to the notion of 'breakdown of gauge invariance at the quantum level'. First, let us point out that the rigorous and nonperturbative approach followed thus far does not fix the phase of the unitary implementer defined by (37). In particular, the phase of $\mathbb{S}$ is left undetermined. This constitutes a crucial difference with the formal approach, which uses perturbation theory. Here, one obtains an expression for the logarithm of the expectation value E of $\mathbb{S}$ with respect to the vacuum vector Ω,

$$E = (\Omega, \mathbb{S}\Omega). \tag{50}$$

This logarithm is given by a series of integrals associated with closed loop Feynman diagrams with n vertices, which code an n-fold scattering at the external fields. Thus, the imaginary part of the series formally fixes the phase of $\mathbb{S}$.

However, the integral corresponding to the $n=2$ loop turns out to be divergent, and a renormalization must be made to render it finite. It is at this point that the anomaly arises: There is no choice of renormalization guaranteeing that the imaginary part of the integral is invariant under gauge transformations of both A^v and A^a. Physicists interpret the corresponding gauge variations of $i\ln E$ as the divergences of the second-quantized currents, and it is customary to insist on vector current conservation

$$\partial^\mu J^v_\mu = 0. \tag{51}$$

The price one then pays is that the divergence of the axial vector current is non-zero (anomalous). For instance, when $A^a=0$ one finds

$$\partial^\mu J^a_\mu = \frac{1}{\pi} F^v_{01}. \tag{52}$$

Hence, under scattering at the external field A^v_μ one expects (using Gauss' theorem) a charge change

$$\Delta Q^a = \frac{1}{\pi} \int dt \; dx F^v_{01}(t,x). \tag{53}$$

This is precisely what happens when one has $A^a_\mu=0$, $A^v_0=0$ and

$$A^v_1(t,x) = \lambda\alpha(x), \quad n \equiv \frac{\lambda}{2\pi}\int dy \; \alpha(y) \in \mathbf{Z}, \quad t>T \tag{54}$$

in addition to the previous conditions. Indeed, as we have seen above (cf. (41), (42)) $\mathbb{S}$ then maps the (q_+, q_-)-sector onto the $(q_+ + n, \; q_- - n)$-sector, so that

$\Delta Q^v = 0$, $\Delta Q^a = 2n$. Since $F^v_{01} = \partial_0 A^v_1$ in the case at hand, the right-hand side of (53) can be evaluated and yields $2n$, too.

Readers who got their feathers ruffled by the formalities leading from (50) to (53) are hopefully somewhat reassured by this agreement. On the other hand, the heuristic arguments just referred to (which are not seriously questioned in particle physics) also lead to the conclusion that the charge changes effected by $\mathbb{S}$ need not necessarily be integers, and depend on how a divergent integral is rendered finite. This is at variance with the rigorous result: Whenever $\mathbb{S}$ exists, it can only change both Q^v and Q^a in integer units, and these changes are independent of the (arbitrary) phase choice. In fact, whenever charge changes are involved, perturbation theory cannot be rigorized.

3.6. Generalizations and outlook

The above gauge fields can be viewed as two-dimensional Maxwell (electromagnetic) fields, the gauge group being $U(1)$. Generalization to a nonabelian gauge group amounts to a consideration of external Yang-Mills gauge fields. For instance, when the gauge group equals $U(k)$ for $k > 1$, then the gauge potentials $A_\mu(t,x)$ are $k \times k$ matrix-valued multipliers on $L^2(\mathbb{R}) \otimes \mathbb{C}^2 \otimes \mathbb{C}^k$. The above considerations can be generalized to this situation, and one again concludes that at the quantum level insistence on vector current conservation entails nonconservation of the axial vector current.

At the level of Feynman perturbation theory the same phenomenon occurs in any even space-time dimension D. The axial vector anomaly can be calculated, and one finds again a gauge-invariant density expressed in terms of the $k \times k$ matrix-valued field strength

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]. \tag{55}$$

If one makes assumptions on the gauge potentials A_μ that parallel those made above for $D = 2$, then $F_{\mu\nu}$ has compact support and the integral of the anomaly density is finite.

This formal result can again be compared to a nonperturbative analysis of the scattering along the lines sketched above for $D = 2$. The classical S-operator exists when the Yang-Mills potential A_μ is pure gauge and time-independent for $t > T$, and just as for $D = 2$ one finds that the chiral parts of S_{++} have Fredholm indices $q_+ = -q_-$, which are determined by a generalized winding number of the relevant unitary matrix multiplier. Moreover, the relation of the winding number to the integrated anomaly I is known, and in this way one obtains $q_+ - q_- = I$, just as for $D = 2$.

However, for $D > 2$ there are two important changes compared to the $D = 2$ situation. First, the off-diagonal parts of S are never Hilbert-Schmidt when homotopically nontrivial multipliers are involved. Hence, a unitary Fock space $\mathbb{S}$-operator does not exist in that case. (However, a quadratic form $\mathbb{S}$ implementing the CAR automorphism in quadratic form sense does exist and has the same charge structure as for $D = 2$.) Second, for $D > 2$ the above assertions concerning the chiral Fredholm indices of S_{++} cannot be reduced to the index theorems discussed in Section 2, since these all hinge on compactness of

$\mathfrak{M}$. Put differently, no $D>2$ analog is known of the Cayley transform, which made it possible to obtain the index of the Wiener-Hopf operator M_{++} from that of the unitarily equivalent Toeplitz operator T_M. In point of fact, the index theorem needed to substantiate our $D>2$ assertions has only recently been proven [10].

So far, our account of anomalies and their relation to index theory has emphasized Minkowski space scattering theory. As regards the chiral anomaly, this setting has been predominant in physics til the late seventies. However, at that time the connection was made with the flamboyant world of index theorems for Dirac operators on compact Riemannian manifolds, and ever since the austere Minkowski space setting appears to play a minor role in the particle theorist's views on anomaly theory.

The connection just mentioned involves two steps, and we proceed to sketch these. First, one performs a Wick rotation $x^0 \mapsto ix^0 \equiv x^2$, turning Minkowski space into Euclidean space. Doing this e.g. in the Dirac equation (43) with $A^v = A^a = 0$ and multiplying by $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, one obtains an equation that can be written

$$\mathfrak{D}\psi = 0, \quad \mathfrak{D} = \begin{pmatrix} 0 & D^* \\ D & 0 \end{pmatrix}, \quad D = -i\partial_1 + \partial_2 \cdot \tag{56}$$

Thus the hyperbolic evolution equation turns into a zero-mode equation for the elliptic Dirac operator $\mathfrak{D}$ on $\mathbb{R}^2$. More generally, in the Euclidean world it is customary to consider the zero-mode equation

$$\mathfrak{D}_A\psi = 0, \quad \mathfrak{D}_A = \begin{pmatrix} 0 & D_A^* \\ D_A & 0 \end{pmatrix} \tag{57}$$

with gauge potentials that are coupled in the same way as the connections dealt with in Subsection 2.3; moreover, one views the determinant of $\mathfrak{D}_A$ as the analytic continuation of the vacuum expectation value (50).

This first step may appear quite baffling: There is no reason why the external fields should have any analyticity properties in x^0, and the appearance of $\det\mathfrak{D}_A$ seems a deus ex machina. Therefore, we must slightly digress to render the above more plausible. First of all, it should be pointed out that to a particle theorist the external field picture is only a preliminary: He really wants to understand the theory in which the second-quantized Dirac field is coupled to a Yang-Mills field that is second-quantized, too. This fully interacting quantum field theory is now reduced to a consideration of the theory where the Yang-Mills field is viewed as a classical field (a function on space-time) by using Feynman path integrals. The path integral is a widely used tool in theoretical physics, which rests on the idea that it is possible to view quantum mechanical transitions between an initial and a final state as being obtainable by summing over all classical trajectories connecting the given states. In this way physicists can exploit various features of the classical world to gain insight into the quantum world. In the path integral formalism the above Wick rotation can be formally performed and the determinant then arises when one fixes the classical Yang-Mills field and integrates out the classical Dirac fields.

Before coming to the second step, we would like to mention that the formalities just sketched have been given a rigorous mathematical interpretation for the simplest interacting quantum field models. An introduction to the extraordinarily deep and difficult results obtained by exploiting such Euclidean path integrals (among many other tools) can be found in [7]. One lesson that can be drawn from this branch of mathematics (called constructive field theory) is that the measure in the path integral assigns zero weight to the set consisting of continuous classical configurations. For constructive field theory this forms a so far unsurmountable barrier in using the rich and well-studied geometry of classical gauge field theories (which hinges on continuity, if not smoothness) to obtain rigorous conclusions about their quantum versions. However, at this point in time constructive field theory is by and large ignored. It is a long and venerable tradition in physics not to be daunted by analysis when it gets in the way, and the second step (to which we now turn) is another case in point.

Euclidean space-time $\mathbb{R}^D$ is not compact and topologically uninteresting. Therefore, it would appear that after the first step one is still very far removed from the compact manifold context of the index theorems described in Section 2. Moreover, on $\mathbb{R}^D$ the Dirac operator has continuous spectrum, so that the expression $\det \mathcal{D}_A$ is a bit problematic even to a particle theorist. The second step which cures both problems is a great stride indeed: One simply compactifies Euclidean space-time! Then $\mathcal{D}_A$ has discrete spectrum, but since the eigenvalues are not bounded, $\det \mathcal{D}_A$ is still ill defined. However, via a zeta function renormalization one can now tie in $\det \mathcal{D}_A$ with solid mathematics: The result is, that the gauge variation of $\ln(\det \mathcal{D}_A)$ can on the one hand be interpreted as twice the difference of the positive and negative chirality zero modes of $\mathcal{D}_A$ or, equivalently, as twice the Fredholm index of D_A; On the other hand, the variation can also be calculated explicitly via path integral techniques, and in this way one can 'prove' the index theorems described in Subsection 2.3. (The curvature of the D-dimensional compact manifold can be taken into account, too; its contribution to the index density is referred to as the gravitational anomaly.) For instance, if $\mathbb{R}^2$ is compactified to a torus, then one can obtain (15) in this way.

All this may strike an analyst as having the advantage of theft over honest toil. Just the same, using quantum field theory in a systematic way, physicists are able to rederive the integrands that occur in the index theorems (after a lot of toil, to be sure), and this is certainly a spectacular and intriguing achievement. As a consequence of these developments and the related development of 'instanton physics' particle theorists and geometers have become increasingly aware of each other's existence since the mid-seventies. This interaction has been beneficial not only to the physicists but also to the mathematicians. A prime example of this is the dramatic progress in differential topology due to the use of instantons. Staying with the subject of this paper one can mention e.g. that the simplest rigorous proof of the Atiyah-Singer theorem for Dirac operators now available (by Getzler) uses some ideas that are borrowed from quantum field theory, and that superstring theory is currently yielding new insights into the budding area of elliptic cohomology (a new branch of index theory). It may be expected that this confrontation between the two cultures will continue to lead to new and unexpected insights for some time to come.

REFERENCES

1. W.A. BARDEEN, A.R. WHITE (eds.) (1985). *Symposium on Anomalies, Geometry, Topology,* World Scientific, Singapore.
2. B. BLACKADAR (1986). *K-theory for Operator Algebras,* Springer, New York.
3. B. BOOSS, D.D. BLEECKER (1985). *Topology and Analysis. The Atiyah-Singer Index Formula and Gauge-Theoretic Physics,* Springer, New York.
4. A.L. CAREY, S.N.M. RUIJSENAARS (1987). On fermion gauge groups, current algebras and Kac-Moody algebras. *Acta Appl. Math. 10,* 1-86.
5. *Current Algebra and Anomalies* (1985), World Scientific, Singapore.
6. T. EGUCHI, P.B. GILKEY, A.J. HANSON (1980). Gravitation, gauge theories and differential geometry. *Phys. Reports 66,* 213-393.
7. J. GLIMM, A. JAFFE (1981). *Quantum Physics. A Functional Integral Point of View,* Springer, New York.
8. M.B. GREEN, J.H. SCHWARZ, E. WITTEN (1987). *Superstring Theory,* Vols. 1,2, Cambridge University Press, Cambridge.
9. J. ROE (1988). *Elliptic Operators, Topology and Asymptotic Methods,* Longman Group, Harlow.
10. S.N.M. RUIJSENAARS (1989). Index Formulas for Generalized Wiener-Hopf Operators and Boson-Fermion Correspondence in $2N$ Dimensions, *Commun. Math. Phys. 124,* 553-593.