ANNALES DE L'I. H. P., SECTION A

GREGORY B. WHITE

Splitting of the Dirac operator in the nonrelativistic limit

Annales de l'I. H. P., section A, tome 53, nº 1 (1990), p. 109-121
<http://www.numdam.org/item?id=AIHPA_1990__53_1_109_0>

Ann. Inst. Henri Poincaré,

Vol. 53, n° 1, 1990, p. 109-121. *Physique théorique*

Splitting of the Dirac operator in the nonrelativistic limit

by

Gregory B. WHITE

Department of Mathematics,
Occidental College, Los Angeles, CA 90041, U.S.A.

ABSTRACT. — Under the assumption of relative boundedness conditions on the potentials, we show that for c large enough, the Dirac operator $H(c)$ may be expressed as $H_1(c) \oplus H_2(c)$ (ala Foldy and Wouthuysen) where $H_1(c)$ and $H_2(c)(\pm mc^2)$ converge in a rigorous sense (pseudoresolvent convergence) to the corresponding Pauli-Schrödinger operators, H^+ and H^-. We also show that the Dirac operator has a spectral gap of the form $(-mc^2+k, mc^2-l)$ for c large enough, where k and l are any constants greater than the lower bounds of H^- and H^+, respectively. From this proof we find a new formula for estimating the lower bounds of the Pauli-Schrödinger operators and we find a sufficient condition for complete separation of the electron and positron energy levels in the Dirac spectrum.

RÉSUMÉ. — Sous l'hypothèse que les potentiels sont relativement bornés, nous montrons que pour c assez grand, l'opérateur de Dirac $H(c)$ peut s'exprimer comme $H_1(c) \oplus H_2(c)$ (à la Foldy et Wouthuysen) où $H_1(c) \oplus H_2(c)(\pm mc^2)$ convergent (au sens de la convergence des pseudo résolvantes) vers les opérateurs de Pauli-Schrödinger correspondants, H^+ et H^-. Nous montrons aussi que l'opérateur de Dirac a un gap spectral de la forme $(-mc^2+k, mc^2-l)$ pour c assez grand, où k et l sont n'importe quelles constantes plus grandes que la borne inférieure de H^- et H^+ respectivement. Nous obtenons aussi une nouvelle estimation des bornes inférieures des opérateurs de Pauli-Schrödinger et nous donnons une condition suffisante pour la séparation complète des niveaux d'énergie des électrons et des positrons dans le spectre de Dirac.

This paper reports results achieved in [17], the author's doctoral dissertation at U.C.L.A.

The author wishes to thank his dissertation advisor, Donald Babbitt, for suggesting the topic of this research and for his many helpful comments along the way. The author is also indebted to a number of other people who heard early presentations of this work and raised important questions and/or made valuable suggestions which shaped the final form of this paper. These include V. S. Varadarajan, R, Arens, M. Takesaki, S. Busenberg, and the referee.

1. INTRODUCTION

In their 1950 paper, Foldy and Wouthuysen [4] devised a transformation which splits the standard Dirac operator $H(c)$ into a direct sum $H_1(c) \oplus H_2(c)$, where $H_1(c)$ and $H_2(c)$ are each expressed as a formal power series in $1/c$. While $H_1(c)$ and $H_2(c)$ may not be well defined in the sense that the power series do not converge in a rigorous operator sense, they enjoy the property that the formal nonrelativistic limit $(c \to \infty)$ of $H_1(c) - mc^2$ yields the Pauli-Schrödinger operator for the electron (H^+) and the limit of $H_2(c) + mc^2$ yields the Pauli-Schrödinger operator for the positron $(-H^-)$. Further, in their paper they suggest that this method should be expected to be of use in the case of "weak" potentials, i.e. those for which there is a separation of the electron and positron energy levels in the Dirac spectrum. Unfortunately, although the method of Foldy and Wouthuysen is both beautiful and successful (e.g. it yields reasonable relativistic corrections to the Pauli-Schrödinger operators), there is no apparent way to make it rigourous.

The first goal of this paper (Section 4) is to show that by restricting the Dirac operator to the subspaces associated with the positive and negative spectrum, it can indeed be split into the direct sum of two operators which now converge in a rigorous sense (pseudoresolvent convergence—*see* Section 3) to the appropriate limits. Convergence of the Dirac resolvent has been studied in [5], [8] and [16], but restricting it to these subspaces appears to be new.

While the question of how to approach the nonrelativistic limit in a rigorous way seems to have been settled, the question of when the electron and positron energy levels are separated has not. The rest of the paper is devoted to determining where spectra may occur in the interval $(-mc^2, mc^2)$ and to the implications of such results for the limiting operators and for the question of separation of the spectrum.

Our second goal is then to show that the Dirac operator has a spectral gap of the form $(-mc^2+k, mc^2-l)$, for c large enough, where k and l are constants greater than the lower bounds of the corresponding Pauli-Schrödinger operators (Section 5). It has been noted by Cirincione and Chernoff [1] that under the assumption of relatively bounded potentials, the Dirac operator has a spectral gap at zero for c large enough. It is not difficult to see that their method implies that the spectral gap includes an interval of the form $(-a(c), a(c))$, where $a(c) \to \infty$ as $c \to \infty$. Also, convergence of the Dirac resolvent $(\pm mc^2)$ to the Pauli-Schrödinger resolvent together with the lower semiboundedness of the Pauli-Schrödinger operators implies that the Dirac operator has no spectrum in the intervals $(-mc^2+k, -mc^2+b(c))$ and $(mc^2-b(c), mc^2-l)$. Notice that this leaves the two intervals $(-mc^2+b(c), -a(c))$ and $(a(c), mc^2-b(c))$ unaccounted for. Our method deals with the entire interval $(-mc^2+k, mc^2-l)$ uniformly. We also develop from this a new formula for estimating the lower bounds of Pauli-Schrödinger operators.

Finally, in Section 6 we give a sufficient condition for complete separation of the electron and positron energy levels.

2. THE DIRAC OPERATOR IN AN ABSTRACT SETTING

In discussing the Dirac operator, we shall use essentially the same set-up as Cirincione and Chernoff [1] and Gesztesy, Grosse, and Thaller [5]. Much of this section duplicates the analogous section in [1], although some material has been added and some omitted. As noted in [1], the abstract setting for the Dirac equation which we consider is general enough to include the case of curved space as well as the usual Dirac equation over $\mathbb{R}^3$ or $\mathbb{R}^n$ (*see* [1], section 3 for a discussion of how these operators are defined). The reader should also note that A and A* are switched from the usage of [1] (this is in line with the convention of [5]).

Let $\mathscr{H}$ be a Hilbert space and let $\mathbf{A}$ and β be two self-adjoint operators with the following properties:

$$\beta^2 = I \qquad \text{and} \qquad \mathbf{A}\beta + \beta\mathbf{A} = 0. \tag{2.1}$$

The first relation implies that β is bounded and in fact unitary. The operator $\mathbf{A}$ may be unbounded and we assume it is defined and self-adjoint on some dense domain $\mathrm{D}(\mathbf{A})$.

Since $\beta^2 = I$, the operator β has two spectral projections P^+ and P^- corresponding to the eigenvalues $+1$ and -1. We may write $\mathcal{H}$ as the orthogonal direct sum $\mathcal{H}_+ \oplus \mathcal{H}_-$, where $\mathcal{H}_\pm$ is the range of $P^\pm$. With respect to this decomposition, we can represent $\mathbf{A}$ and β by operator matrices. We have

$$\beta = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{2.2}$$

The self-adjointness of $\mathbf{A}$ implies that A_{11} and A_{22} are self-adjoint, while A_{12} and A_{21} are closed and densely defined on $\mathcal{H}_-$ and $\mathcal{H}_+$, respectively, with $A_{12} = A_{21}^*$. The condition $\mathbf{A}\beta = -\beta\mathbf{A}$ tells us that $A_{11} = A_{22} = 0$, so $\mathbf{A}$ is of the form

$$\mathbf{A} = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}, \tag{2.3}$$

where $A : \mathcal{H}_+ \to \mathcal{H}_-$ is a closed densely defined operator with adjoint $A^* : \mathcal{H}_- \to \mathcal{H}_+$. If, for example, $\mathcal{H}_+ = \mathcal{H}_- = \mathcal{H}_0$, a given Hilbert space, so that $\mathcal{H} \cong \mathbb{C}^2 \otimes \mathcal{H}_0$, and if $A = A^*$, then we have $\mathbf{A} = \alpha \otimes A$, where α is the two-by-two matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This is the case which Hunziker [8] discusses.

To define our abstract Dirac operator $H(c)$, we introduce an operator $\mathbf{V}$ representing a potential and set

$$H(c) = c\,\mathbf{A} + mc^2\beta + \mathbf{V}. \tag{2.4}$$

Here m and c are positive constants which we call the "rest mass" and the "velocity of light". We require that $\mathbf{V}$ be self-adjoint, bounded relative to $\mathbf{A}$, and $\mathbf{V}\beta = \beta\mathbf{V}$. The latter condition means that $\mathbf{V}$ is of the form $\begin{bmatrix} V_+ & 0 \\ 0 & V_- \end{bmatrix}$, where V_+ and V_- are self-adjoint operators on $\mathcal{H}_+$ and $\mathcal{H}_-$, respectively. (Usually $\mathcal{H}_+ = \mathcal{H}_-$ and $V_+ = V_-$.) The boundedness condition on V is equivalent to saying that V_+ is bounded relative to A and V_- is bounded relative to A^*. Since $\mathbf{V}$ is bounded relative to $\mathbf{A}$, $\mathbf{V}$ will be bounded relative to $c\,\mathbf{A} + mc^2\beta$ with relative boundless than 1, for c sufficiently large. Then $H(c)$ will be self-adjoint on $D(\mathbf{A})$ by the Kato-Rellich theorem. Henceforth, we will assume that c is large enough so that this is the case.

Note also that $\frac{1}{2m}\mathbf{A}^2 + \mathbf{V}$ is a self-adjoint operator on $D(\mathbf{A}^2)$, for our hypotheses on $\mathbf{V}$ guarantee that $\mathbf{V}$ is infinitesimally bounded with respect to $\mathbf{A}^2$. Hence the operators

$$H^{\pm}=\frac{1}{2m}(\mathbf{A}^2)_{\pm}\pm V_{\pm} \tag{2.5}$$

are self-adjoint on the spaces $\mathcal{H}_{\pm}$, where $(\mathbf{A}^2)_+=A^*A$ and $(\mathbf{A}^2)_-=AA^*$, are the restrictions of $\mathbf{A}^2$ to the invariant subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$, respectively. We also introduce the notation

$$|A|_+=\sqrt{A^*A} \quad \text{and} \quad |A|_-=\sqrt{AA^*}. \tag{2.6}$$

Note that in this setting we can deal with "magnetic fields" as well as "electrostatic fields" with no additional effort. The magnetic (or "vector") potential is simply a perturbation $\mathbf{B}$ which anticommutes with β, so we may regard $\mathbf{B}$ as a perturbation of $\mathbf{A}$. To be precise, assume that $\mathbf{B}$ is a self-adjoint operator on $\mathcal{H}$ which anticommutes with β and which is bounded relative to $\mathbf{A}$ with relative bound less than 1. Then $\mathbf{V}$ is bounded relative to the self-adjoint operator $\mathbf{A}+\mathbf{B}$, so that

$$H(c)=c(\mathbf{A}+\mathbf{B})+mc^2\beta+\mathbf{V} \tag{2.7}$$

is self-adjoint for sufficiently large c, and all results discussed here will apply verbatim with $\mathbf{A}+\mathbf{B}$ replacing $\mathbf{A}$. Further, the Schrödinger Hamiltonian $H^{\pm}$ introduced above is replaced by the "Pauli" Hamiltonian

$$H^{\pm}=\frac{1}{2m}((\mathbf{A}+\mathbf{B})^2)\pm V_{\pm}. \tag{2.8}$$

We call H^+ and H^- [as defined by (2.5) or (2.8)] the Pauli-Schrödinger operators associated with the Dirac operator [as defined by (2.4)].

3. PSEUDORESOLVENT CONVERGENCE

To quickly put into context the notion of convergence which will be discussed in this section, we remark that pseudoresolvent convergence is to analytic families of pseudoresolvents as resolvent convergence is to analytic families of resolvents. That is, pseudoresolvent convergence is of use when analytic results are either not expected or not needed. As we shall see, this the case for the Dirac operators applications discussed in this paper.

We present here a few theorems which we which we will need later. A full development of theorems (with proofs) which generalize the standard theorems for resolvent convergence (*see* e.g. Reed and Simon [12]) may be found in White [17].

DEFINITION. — Let A_n, $n=1, 2, \ldots$ and A be self-adjoint operators on a Hilbert space $\mathcal{H}$. Let P_n, $n=1, 2, \ldots$ and P be self-adjoint projections on $\mathscr{H}$ such that P_n commutes with A_n for $n=1, 2, \ldots$ and P commutes with A. Then $A_n P_n$ is said to converge to AP in the *norm* (*strong*) *pseudoresolvent sense* if $R_\lambda(A_n) P_n \to R_\lambda(A) P$ uniformly (strongly) for all λ with $\operatorname{Im} \lambda \neq 0$.

NOTATION. — In analogy with the convention for resolvent convergence, we will write $A_n P_n \overset{nps}{\to} AP$ and $A_n P_n \overset{sps}{\to} AP$ to indicate norm and strong pseudoresolvent convergence respectively.

Remarks. — (1) Families of operators of the form $R_\lambda(A) P$ satisfy the first resolvent equation and so by definition are pseudoresolvents. Further, as noted by Veselič [16], any symmetric pseudoresolvent on a Hilbert space is of the form $R_\lambda(A) P$.

(2) There is no explicit requirement on the convergence of the projections P_n. In general, it is not necessary for the P_n to converge to P in any sense in order to have norm pseudoresolvent convergence. For certain theorems on strong pseudoresolvent convergence, strong convergence of the P_n to P is required.

THEOREM 3.1. — *Let* A_n, A, P_n, P, *be as in the definition. Let* λ_0 *be a point in* $\mathbb{C}$. *If* $\operatorname{Im} \lambda_0 \neq 0$ *and* $\| R_{\lambda_0}(A_n) P_n - R_{\lambda_0}(A) P \| \to 0$, *then* $A_n P_n \overset{nps}{\to} AP$.

THEOREM 3.2 (Trotter). — *If* $A_n P_n \overset{sps}{\to} AP$ *and* $P_n \overset{s}{\to} P$, *then* $e^{it A_n} P_n \overset{s}{\to} e^{it A} P$ *for each* t, *and* $e^{it A_n} P_n \varphi \to e^{it A} P \varphi$ *for each* φ *uniformly in* t *in any bounded interval.*

THEOREM 3.3. — *Suppose that* $A_n P_n \overset{nps}{\to} AP$. *Let* $a, b \in \mathbb{R}$, $a<b$ *and suppose that* $a, b \in \rho(A|_{\operatorname{Ran} P})$. *Then*

$$\| P_{(a, b)}(A_n) P_n - P_{(a, b)}(A) P \| \to 0.$$

This last result is of use when we do not know whether the projections P_n converge strongly to P. In fact, as we will see (*cf.* Thm. 4.3), it can be useful in showing that the projections P_n converge.

THEOREM 3.4. — *Suppose that* $A_n P_n \overset{sps}{\to} AP$. *If* $a \in \mathbb{R}$ *and* $a \notin \sigma_{pp}(A|_{\operatorname{Ran} P})$, *then* $P_{(a, \infty)}(A_n) P_n P \varphi \to P_{(a, \infty)}(A) P \varphi$ *and* $P_{(-\infty, a)}(A_n) P_n P \varphi \to P_{(-\infty, a)}(A) P \varphi$ *for all* $\varphi \in \mathscr{H}$.

4. SPLITTING OF THE OPERATOR

As shown by Gesztesy, Grosse, and Thaller [5], the Dirac resolvent $(\mathrm{H}(c)-mc^2-z)^{-1}$ is holomorphic about the Pauli-Schrödinger pseudoresolvent $(\mathrm{H}^+-z)^{-1}\mathrm{P}^+$ at $c=\infty$. Thus, we may consider these together as an analytic family of pseudoresolvents in a neighborhood of $c=\infty$. Similarly, $(\mathrm{H}(c)+mc^2-z)^{-1}$ is holomorphic about $(-\mathrm{H}^--z)^{-1}\mathrm{P}^-$ at $c=\infty$. This holomorphy implies the following weaker statement:

THEOREM 4.1. – *As c goes to infinity*, $\mathrm{H}(c)\pm mc^2 \overset{nps}{\to} \pm\mathrm{H}^\pm\mathrm{P}^\pm$.

Note. – Relating the statement of the theorem to the definition of pseudoresolvent convergence, we have $\mathrm{A}_n=\mathrm{H}(c)\pm mc^2$, $\mathrm{P}_n=\mathrm{I}$, $\mathrm{A}=\pm\mathrm{H}^\pm$ and $\mathrm{P}=\mathrm{P}^\pm$.

As shown in [1] (or see Section 5), the Dirac operator has a spectral gap at zero for c large enough. Thus, for some $\varepsilon>0$ if we define

$$\mathrm{Q}^+(c)=\mathrm{P}_{[\varepsilon,\,\infty]}(\mathrm{H}(c)) \quad \text{and} \quad \mathrm{Q}^-(c)=\mathrm{P}_{[-\infty,\,-\varepsilon]}(\mathrm{H}(c)), \tag{4.1}$$

then we are assured that $\mathrm{Q}^+(c)+\mathrm{Q}^-(c)=\mathrm{I}$ for c large.

Since it is the positive half of the spectrum of $\mathrm{H}(c)$ which converges to the spectrum of H^+, we expect that $(\mathrm{H}(c)-mc^2)\mathrm{Q}^+(c)$ will converge to $\mathrm{H}^+\mathrm{P}^+$. Similarly, we expect convergence of $(\mathrm{H}(c)+mc^2)\mathrm{Q}^-(c)$ to $-\mathrm{H}^-\mathrm{P}^-$. This is the content of the next theorem.

One should note however, that in restricting $\mathrm{H}(c)$ to the subspaces corresponding to the positive and negative halves of the spectrum, holomorphy is lost. This can be seen by following the behavior of the essential spectrum of $\mathrm{H}(c)$ as c rotates through a circle in the complex plane which encompasses the origin. It becomes clear that $\mathrm{Q}^+(c)$ and $\mathrm{Q}^-(c)$ cannot be continued analytically as c passes through the imaginary axis. Thus, we should not expect an analytic result. Rather, the best sense of convergence we can hope for is:

THEOREM 4.2. – $(\mathrm{H}(c)\pm mc^2)\mathrm{Q}^\pm(c)\overset{nps}{\to}\pm\mathrm{H}^\pm\mathrm{P}^\pm$ *as* $c\to\infty$.

Proof. – To see that $(\mathrm{H}(c)-mc^2)\mathrm{Q}^+(c)\overset{nps}{\to}\mathrm{H}^+\mathrm{P}^+$ we note that

$$\begin{aligned}&\|(\mathrm{H}(c)-mc^2-i)^{-1}\mathrm{Q}^+(c)-(\mathrm{H}^+-i)^{-1}\mathrm{P}^+\|\\&\qquad\leqq\|(\mathrm{H}(c)-mc^2-i)^{-1}-(\mathrm{H}^+-i)^{-1}\mathrm{P}^+\|+\|(\mathrm{H}(c)-mc^2-i)^{-1}\mathrm{Q}^-(c)\|\end{aligned}$$

where we have used the fact $\mathrm{Q}^+(c)+\mathrm{Q}^-(c)=\mathrm{I}$. The norm of the first term on the right goes to zero since by Theorem 4.1, $\mathrm{H}(c)-mc^2\overset{nps}{\to}\mathrm{H}^+\mathrm{P}^+$. The norm of the second term on the right goes to zero since for c large enough, the distance from the spectrum of $(\mathrm{H}(c)-mc^2)\mathrm{Q}^-(c)$ to i is $>mc^2$ (in fact, $>2mc^2-l$ as we shall see in

Section 5). Hence, the norm of the second term goes to zero as $c \to \infty$, and by Theorem 3.1 we now have $(\mathrm{H}(c) - mc^2)\mathrm{Q}^+(c) \xrightarrow{nps} \mathrm{H}^+ \mathrm{P}^+$.

Similarly, $(-\mathrm{H}(c) - mc^2)\mathrm{Q}^-(c) \xrightarrow{nps} \mathrm{H}^- \mathrm{P}^-$. □.

Remarks. – (1) Note that if we set $\mathrm{H}_1(c) = \mathrm{H}(c)\mathrm{Q}^+(c)$ and $\mathrm{H}_2(c) = \mathrm{H}(c)\mathrm{Q}^-(c)$, then for c large, $\mathrm{H}(c) = \mathrm{H}_1(c) \oplus \mathrm{H}_2(c)$ where $\mathrm{H}_1(c) - mc^2$ and $\mathrm{H}_2(c) + mc^2$ have the appropriate nonrelativistic limits, ala Foldy and Wouthuysen.

(2) We may also view this as splitting into an "electron" term, a "positron" term and a "rest mass" term by writing

$$\mathrm{H}(c) = (\mathrm{H}(c) - mc^2)\mathrm{Q}^+(c) + (\mathrm{H}(c) + mc^2)\mathrm{Q}^-(c) + mc^2(\mathrm{Q}^+(c) - \mathrm{Q}^-(c)).$$

Convergence of the first two terms has been discussed. The next theorem implies that $\mathrm{Q}^+(c) - \mathrm{Q}^-(c) \xrightarrow{s} \mathrm{P}^+ - \mathrm{P}^- = \beta$, so the third term is much like rest mass term in the Dirac equation.

THEOREM 4.3. – $\mathrm{Q}^\pm(c) \xrightarrow{s} \mathrm{P}^\pm$ as $c \to \infty$.

Proof. – Rearranging the signs in Theorem 4.1, we have $\pm\mathrm{H}(c) - mc^2 \xrightarrow{nps} \mathrm{H}^\pm \mathrm{P}^\pm$. Let l be less than the lower bounds of H^+ and H^-. With $\mathrm{P}_n = \mathrm{I}$ in theorem 3.4, we have

$$\mathrm{P}_{(-l,\,\infty)}(\pm\mathrm{H}(c) - mc^2)\mathrm{P}^\pm \xrightarrow{s} \mathrm{P}_{(-l,\,\infty)}(\mathrm{H}^\pm)\mathrm{P}^\pm.$$

But, $\mathrm{P}_{(-l,\,\infty)}(\mathrm{H}^\pm) = \mathrm{I}$ and $\mathrm{P}_{(-l,\,\infty)}(\pm\mathrm{H}(c) - mc^2) = \mathrm{Q}^\pm(c)$ for c large enough (*see* Theorem 5.3), so $\mathrm{Q}^\pm(c)\mathrm{P}^\pm \xrightarrow{s} \mathrm{P}^\pm$. Similarly, considering the interval $(-\infty, -l)$ we find $\mathrm{Q}^\mp(c)\mathrm{P}^\pm \xrightarrow{s} 0$. The result follows by adding the previous two equations appropriately and using $\mathrm{P}^+ + \mathrm{P}^- = \mathrm{I}$. □

Note. – The above result appears in [1] with a rather different proof.

We close this section by stating a related result (also in [1]) which now follows easily from Theorems 4.2, 4.3 and 3.2.

THEOREM 4.4. – *For each* $t \in \mathbb{R}$, $e^{it(\pm\mathrm{H}(c) - mc^2)}\mathrm{Q}^\pm(c) \xrightarrow{s} e^{it\mathrm{H}^\pm}\mathrm{P}^\pm$ *and for each* $\varphi \in \mathscr{H}$, $e^{it(\pm\mathrm{H}(c) - mc^2)}\mathrm{Q}^\pm(c)\varphi \to e^{it\mathrm{H}^\pm}\mathrm{P}^\pm\varphi$ *uniformly on bounded t-intervals as* $c \to \infty$.

5. THE SPECTRAL GAP

We will prove in this section that given a relative boundness condition on the potential **V**, the Dirac operator has a spectral gap of the form

$[-mc^2+k,\ mc^2-l]$ for c large enough, where k and l are related to the lower bounds of the operators H^- and H^+, respectively.

To make precise what we mean by "spectral gap", we make the following

DEFINITION. — Let I be an interval in $\mathbb{R}$ and let $H(c)$ be a family of self-adjoint operators depending on a real parameter c. We say I is a *spectral gap* for $H(c_0)$ if $I \subseteq \rho(H(c))$ for all $c \geqq c_0$.

Remarks. — The condition "for all $c \geqq c_0$" is to insure that the spectrum remains separated as we go the nonrelativistic limit.

NOTATION. — In order to write much of what follows in a more concise form, we introduce the following operators:

$$H_1^0(c)=c^2 A^* A+m^2 c^4-\lambda^2$$

and

$$H_2^0(c)=c^2 AA^*+m^2 c^4-\lambda^2. \tag{5.1}$$

We note that if $l>0$ and $\lambda \in [mc^2+l,\ mc^2-l]$, then $H_1^0(c)$ and $H_2^0(c)$ are invertible.

In the following theorem we will make use of two commutation relations proved by Deift [3]:

$$\begin{gathered}(A^* A-z)^{-1} A^* \subseteq A^*(AA^*-z)^{-1}\\ \text{and}\\ (AA^*-z)^{-1} A \subseteq A(A^* A-z)^{-1}\\ \text{for } z \in \rho(A^* A)\setminus\{0\}=\rho(AA^*)\setminus\{0\}.\end{gathered} \tag{5.2}$$

THEOREM 5.1. — *Let $H(c)=cA+mc^2\beta+V$ be a Dirac operator and assume that there exist constants $a,\ b>0$ such that $\|V\varphi\| \leqq a\|A\varphi\|+b\|\varphi\|$ for all $\varphi \in D(A)$. Let $l_0=a^2 m+b+\sqrt{(a^2 m+b)^2-b^2}$, let $\varepsilon>0$, let $l=l_0+\varepsilon$ and $I_l(c)=[-mc^2+l,\ mc^2-l]$. Then for c large enough, $I_l(c) \subset \rho(H(c))$. Furthermore, for $\lambda \in I_l(c)$ and for c large enough, we have*

$$\begin{aligned}(H(c)-\lambda I)^{-1}=&\left(I+\begin{pmatrix}(\lambda+mc^2)(H_1^0(c))^{-1}V_+ & cA^*(H_2^0(c))^{-1}V_-\\ cA(H_1^0(c))^{-1}V_+ & (\lambda-mc^2)(H_2^0(c))^{-1}V_-\end{pmatrix}\right)^{-1}\\ &\times\begin{pmatrix}(\lambda+mc^2)(H_1^0(c))^{-1} & cA^*(H_2^0(c))^{-1}\\ cA(H_1^0(c))^{-1} & (\lambda-mc^2)(H_2^0(c))^{-1}\end{pmatrix}.\end{aligned} \tag{5.3}$$

Proof. — In matrix form we may write

$$H(c)-\lambda I=\begin{pmatrix}-\lambda+mc^2 & cA^*\\ cA & -\lambda-mc^2\end{pmatrix}+\begin{pmatrix}V_+ & 0\\ 0 & V_-\end{pmatrix}.$$

Since $H_1^0(c)$ and $H_2^0(c)$ are invertible, we have

$$\begin{pmatrix} -\lambda+mc^2 & cA^* \\ cA & -\lambda-mc^2 \end{pmatrix}^{-1}=\begin{pmatrix} (\lambda+mc^2)(H_1^0(c))^{-1} & cA^*(H_2^0(c))^{-1} \\ cA(H_1^0(c))^{-1} & (\lambda-mc^2)(H_2^0(c))^{-1} \end{pmatrix},$$

where the commutation relations (5.2) are required to show that this is a left-inverse. Then

$$\begin{aligned} H(c)-\lambda I=&\begin{pmatrix} -\lambda+mc^2 & cA^* \\ cA & -\lambda-mc^2 \end{pmatrix} \\ &\times\left(I+\begin{pmatrix} (\lambda+mc^2)(H_1^0(c))^{-1}V_+ & cA^*(H_2^0(c))^{-1}V_- \\ cA(H_1^0(c))^{-1}V_+ & (\lambda-mc^2)(H_2^0(c))^{-1}V_- \end{pmatrix}\right). \end{aligned} \tag{5.4}$$

Now $H(c)-\lambda I$ will be invertible if

$$\left\|\begin{pmatrix} (\lambda+mc^2)(H_1^0(c))^{-1}V_+ & cA^*(H_2^0(c))^{-1}V_- \\ cA(H_1^0(c))^{-1}V_+ & (\lambda-mc^2)(H_2^0(c))^{-1}V_- \end{pmatrix}\right\|<1. \tag{5.5}$$

In estimating the norms of the matrix elements above, we note that in fact we are estimating the norm of the closures of the operators.

For $\lambda\in I_l(c)$ we may decompose $H_1^0(c)$ as

$$H_1^0(c)=(c|A|_+-i\sqrt{m^2c^4-\lambda^2})(c|A|_++i\sqrt{m^2c^4-\lambda^2}).$$

Then,

$$\begin{aligned} &\|(\lambda+mc^2)(H_1^0(c))^{-1}V_+\| \\ &\quad=(\lambda+mc^2)\left\|\left(c|A|_++i\sqrt{m^2c^4-\lambda^2}\right)^{-1}\right\| \\ &\quad\quad\times\left\|V_+(c|A|_++i\sqrt{m^2c^4-\lambda^2})^{-1}\right\| \\ &\quad\leqq(\lambda+mc^2)\left(\frac{1}{\sqrt{m^2c^4-\lambda^2}}\right)\left(\frac{a}{c}+\frac{b}{\sqrt{m^2c^4-\lambda^2}}\right) \\ &\quad=\frac{a}{c}\sqrt{\frac{mc^2+\lambda}{mc^2-\lambda}}+\frac{b}{mc^2-\lambda}. \end{aligned}$$

Similarly,

$$\|(\lambda-mc^2)(H_2^0(c))^{-1}V_-\|\leqq\frac{a}{c}\sqrt{\frac{mc^2-\lambda}{mc^2+\lambda}}+\frac{b}{mc^2+\lambda}.$$

Note that these norm estimates are maximized at the right and left endpoints of $I_l(c)$ respectively. Each has a maximum value of $\frac{a}{c}\sqrt{\frac{2mc^2-l}{l}}+\frac{b}{l}$.

Similar calculations show that $\|cA(H_1^0(c))^{-1}V_+\|$ and $\|cA^*(H_2^0(c))^{-1}V_-\|$ are both less than $\frac{a}{c}+\frac{b}{\sqrt{2mc^2l-l^2}}$ for $\lambda\in I_l(c)$.

From the above estimates it is easily seen that for $\lambda \in I_l(c)$,

$$\|(5.5)\| \leqq \frac{a}{c}\sqrt{\frac{2mc^2-l}{l}}+\frac{b}{l}+\frac{a}{c}+\frac{b}{\sqrt{2mc^2 l-l^2}} \tag{5.6}$$

where $\|(5.5)\|$ is the norm of the matrix in equation (5.5). Now the third and fourth terms on the right go to zero as $c \to \infty$ and the first term is less than $a\sqrt{2m/l}$ for all c, so we will have $\|(5.5)\|<1$ for c sufficiently large if

$$\frac{a\sqrt{2m}}{\sqrt{l}}+\frac{b}{l}<1.$$

Elementary algebra shows that this inequality holds if $l>l_0$.

Thus, $H(c)-\lambda I$ is invertible for $\lambda \in I_l(c)$ and c large. Finally, taking the inverse of equation (5.4) gives us equation (5.3). □

The pseudoresolvent convergence of $H(c) \pm mc^2$ to $\pm H^{\pm} P^{\pm}$ (Theorem 4.1 and 3.3) together with the above imply the following:

COROLLARY 5.2. — *Let l_0 be as in Theorem* (5.1), *then $-l_0$ is a lower bound for* H^+ *and* H^-.

Remarks. — (1) Using the Kato-Rellich theorem (*see* e.g. Reed and Simon [13], p. 162) and the relative boundness condition, one can predict that $-l_{KR}$ is a lower bound for H^+ and H^-, where

$$l_{KR}=2a^2 m+b+\sqrt{(2a^2 m+b)^2-b^2}.$$

We see that the bound developed above is better than this by almost a factor of 2.

(2) In the case of the Coulomb potential, $V=-z/r$, $A=\text{grad}$, we may take our relative boundedness condition to be $\|V\varphi\| \leqq 2z\|A\varphi\|$ (*cf.* Kato [9], p. 307). In c.g.s. units this yields a lower bound estimate of $l_0=16(me^4 z^2/2h^2)$, or 16 times the actual ground state energy. We note that any estimate of a in $\|V\varphi\| \leqq a\|A\varphi\|$ must be at least $z/2$, since $a=z/2$ yields the actual ground state energy as a lower bound estimate.

We can now sharpen the result of Theorem 5.1.

THEOREM 5.3. — *Let $-K$ be the greatest lower bound for H^-, let $-L$ be the g.l.b. for H^+, and let $\varepsilon>0$. Then for c large enough,* $[-mc^2+K+\varepsilon, mc^2-L-\varepsilon] \subset \rho(H(c))$.

Proof. — Let l_0 be defined as in Theorem 5.1. Note that $[-l_0-\varepsilon, -L-\varepsilon] \subset \rho(H^+)$, so $P_{[-l_0-\varepsilon,\, -L-\varepsilon]}(H^+)=0$. By Theorem 3.3, we have $\|P_{[-l_0-\varepsilon,\, -L-\varepsilon]}(H(c)-mc^2)\| \to 0$ as $c \to \infty$. But the norm of a spectral projection is either 1 or 0, so, for c large enough, the projection must be zero. That is, $[mc^2-l_0-\varepsilon, mc^2-L-\varepsilon] \subset \rho(H(c))$ for c large enough. Similarly, $[-mc^2+K+\varepsilon, \ -mc^2+l_0+\varepsilon]\subset\rho(H(c))$ for c large enough.

Since, by Theorem 5.1, $[-mc^2+l_0+\varepsilon, mc^2-l_0-\varepsilon]\subset\rho(H(c))$ for c large enough, the theorem is proved. □

6. CLASSIFICATION OF THE SPECTRUM

It is natural to ask under what conditions we will have a spectral gap. The next theorem gives a sufficient condition.

THEOREM 6.1. – *Let* $H(c)=c\mathbf{A}+mc^2\beta+\mathbf{V}$ *be a Dirac operator. Let a and b be positive constants with*

$$\frac{a}{c}+\frac{b}{mc^2}<\frac{1}{2}. \tag{6.1}$$

If $\|\mathbf{V}\varphi\|\leqq a\|\mathbf{A}\varphi\|+b\|\varphi\|$ *for all* $\varphi\in D(\mathbf{A})$, *then* $H(c)$ *has a spectral gap at zero.*

Proof. – First we note that these conditions are sufficient to give self-adjointness of $H(c)$ on $D(\mathbf{A})$ by the Kato-Rellich Theorem, since $a/c<1/2$ and for all $\varphi\in D(\mathbf{A})$

$$\|\mathbf{V}\varphi\|\leqq\frac{a}{c}\|c\mathbf{A}\varphi\|+\mathbf{b}\|\varphi\|.$$

From equation (5.6) of the proof of Theorem 5.1, $H(c)-\lambda I$ will be invertible for all $\lambda\in[-mc^2+l, mc^2-l]$ if

$$\frac{a}{c}\sqrt{\frac{2mc^2-l}{l}}+\frac{b}{l}+\frac{a}{c}+\frac{b}{\sqrt{2mc^2l-l^2}}<1. \tag{6.2}$$

By differentiating the sum of the first and the third term with respect to c and noting that the other two terms are non-increasing, it is easy to see that the left-hand side of (6.2) is decreasing in c. Hence, if (6.2) holds for some $c=c_0$, it holds for all $c>c_0$ and $H(c_0)$ will have a gap a zero.

Thus, we will have a spectral gap if equation (6.2) holds for some l with $l<mc^2$. Since the left-hand side of (6.2) is continuous, in l, we will be able to find such an l if there is strict inequality for $l=mc^2$. Setting $l=mc^2$ in equation (6.2) yields equation (6.1). Hence, there will be a spectral gap if equation (6.1) holds. □

Remarks. – For the purpose of classifying the spectrum, the importance of a spectral gap is that the spectrum cannot cross the gap as c increases [since $H(c)$ is a holomorphic family in c, the spectrum moves analytically]. Hence, existence of a spectral gap insures that the spectrum is clearly