Stationary direct perturbation theory of relativistic corrections

Werner Kutzelnigg
Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany
(Received 14 November 1994; revised manuscript received 25 September 1995)

The stationary variant of direct perturbation theory of relativistic effects is presented. In this variant neither the unperturbed (nonrelativistic) equation nor the equations for the relativistic corrections are solved exactly, but each of them is replaced by the condition that a certain functional becomes stationary. Let $\psi_0=(\varphi_0,\chi_0)$ be the four-component spinor with modified metric in the nonrelativistic limit and $\psi_2=(\varphi_2,\chi_2)$ the leading relativistic correction of $O(c^{-2})$, then one can define functionals $F_0(\varphi_0,\chi_0)$ and $F_4(\varphi_2,\chi_2)$ called respectively the Lévy-Leblond and the Rutkowski-Hylleraas functional, such that stationarity of F_0 with respect to variation of φ_0 and χ_0 determines φ_0 and χ_0, and stationarity of F_4 with respect to variation of φ_2 and χ_2 determines φ_2 and χ_2. The unperturbed (i.e., nonrelativistic) energy E_0 as well as the leading relativistic correction $c^{-2}E_2$ are expressible through φ_0 and χ_0 while for the next higher corrections $c^{-4}E_4$ and $c^{-6}E_6$, φ_2 and χ_2 are also needed. Either of the two functionals F_0 and F_4 can be decomposed into two contributions, the error of one of which is $\geqslant 0$ while that of the other is $\leqslant 0$. An upper-bound property is obtained if the error of the second part vanishes. A strict variation perturbation theory requires that the approximate $\tilde{\varphi}_2$ and $\tilde{\chi}_2$ reproduce the behavior of the exact φ_2 and χ_2 near a nucleus, which implies terms in $\ln r$. If one regularizes $\tilde{\varphi}_2$ one must also regularize $\tilde{\chi}_2$; otherwise E_6 diverges. If one regularizes both φ_2 and χ_2 in the sense of a kinetic balance, one gets regular results for E_4 and E_6, but one loses the strict upper-bound property. The Breit-Pauli expression for E_2 is shown to be correct only if the nonrelativistic wave equation has been solved exactly. Otherwise there is an extra term. Finally the question as to which extent some of the singularities in the perturbation theory of relativistic effects might be artifacts due to the unphysical assumption of a point nucleus is discussed. It is shown, however, that these singularities are not removed if one uses realistic extended nuclei. For all atoms, the critical radius r_c inside of which the nuclear attraction energy is larger than the rest energy of the electron is larger than the extension of the nucleus. [S1050-2947(96)03907-8]

PACS number(s): 31.30.Jv, 03.65.Ca

I. INTRODUCTION

A few years ago [1,2] the perturbation theory of relativistic effects was analyzed. A main message was that a *direct perturbation theory* (DPT) can be formulated, in which the spurious singularities (in the presence of a Coulomb potential) that arise in the more traditional perturbation treatments are avoided, e.g., in the one based on the Foldy-Wouthuysen transformation [3].

DPT has essentially three roots:

(a) Sewell [4] proposed the change in the metric that makes a c^{-1} expansion of the Dirac equation possible. However, this work was forgotten for decades.

(b) Gesztesy *et al.* [5] have shown that the resolvent of the Dirac operator in the presence of a Coulomb potential is holomorphic in c^{-1}, which implies that the bound-state eigenvalues are analytic in c^{-1} and the eigenfunctions are also, not necessarily pointwise in space, but in the sense of a Hilbert space norm. Inspection of the eigenfunctions of H-like ions, which are known exactly, reveals that these are analytic in c^{-1} everywhere except at the position of the nucleus, where they have a well-known weak singularity as functions of r [1,6] (see Appendix B).

(c) Rutkowski [6,7] has, in an unconventional but ingeneous way, found a scheme which has led to the same working equations as those of Sewell [4] or Gesztesy *et al.* [5], and he was the first to perform nontrivial calculations.

An analysis of these approaches and their close relationships can be found in Refs. [1] and [2]. Another interesting method proposed by Moore [8] that looks similar at first glance must be regarded as a side track. Recently the DPT was combined by Sadlej *et al.* [10] with the regularization method of Chang *et al.* [9].

In the formulation of the DPT as preferred by the present author [1] the essential step is a change of the metric in four-component spinor space [4]. One replaces the bispinor $\psi=(\varphi,\chi)$ by a modified bispinor

$$\overline{\psi}=\begin{pmatrix}\overline{\varphi}\\ \overline{\chi}\end{pmatrix}=\begin{pmatrix}\varphi\\ c\chi\end{pmatrix}, \tag{1.1}$$

where c is the velocity of light in atomic units, in which it is equal to the inverse of the fine-structure constant α. In terms of ψ the Dirac equation $(D-mc^2)\psi=E\psi$ becomes

$$\begin{aligned}\overline{D\psi}&=\left\{\begin{pmatrix}V & \boldsymbol{\sigma}\cdot\mathbf{p}\\ \boldsymbol{\sigma}\cdot\mathbf{p} & -2m\end{pmatrix}+\frac{1}{c^2}\begin{pmatrix}0 & 0\\ 0 & V\end{pmatrix}\right\}\begin{pmatrix}\overline{\varphi}\\ \overline{\chi}\end{pmatrix}\\ &=E\left\{\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}+\frac{1}{c^2}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right\}\begin{pmatrix}\overline{\varphi}\\ \overline{\chi}\end{pmatrix}=ES\overline{\psi}\end{aligned} \tag{1.2}$$

(where V is the potential, $\mathbf{p}$ the momentum, and $\boldsymbol{\sigma}$ the vector consisting of the three Pauli matrices $\sigma_x,\sigma_y,\sigma_x$). It has the natural nonrelativistic limit

$$D_0\psi_0=\begin{pmatrix}V & \boldsymbol{\sigma}\cdot\mathbf{p}\\ \boldsymbol{\sigma}\cdot\mathbf{p} & -2m\end{pmatrix}\begin{pmatrix}\varphi_0\\ \chi_0\end{pmatrix}=E_0\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\varphi_0\\ \chi_0\end{pmatrix}=E_0S_0\psi_0, \tag{1.3}$$

which is nothing but the Lévy-Leblond equation [11], which had originally been proposed as the Galilei-invariant field equation for a spinor field. The natural perturbation parameter in (1.2) is c^{-2}; i.e., a direct expansion in powers of c^{-2} of (1.2) is possible (whence the name ''direct perturbation theory'' [1]).

In previous papers [1,2,12] we have used the notations $\widetilde{\psi}$ and $\widetilde{\chi}$ for what is now symbolized as ψ and χ. We make this change of notation, because we want to reserve here the tilde for *approximate* wave functions and energies.

The expansions of E, ψ, φ, and χ in powers of c^{-1} are

$$E=\sum_{k=0}^{\infty} c^{-2k}E_{2k}, \tag{1.4a}$$

$$\psi=\sum_{k=0}^{\infty} c^{-2k}\psi_{2k}, \tag{1.4b}$$

$$\varphi=\sum_{k=0}^{\infty} c^{-2k}\varphi_{2k}, \quad \chi=\sum_{k=0}^{\infty} c^{-2k}\chi_{2k}. \tag{1.4c}$$

Previously [1,2] the *exact* formulation of perturbation theory was considered, i.e., it was assumed that both the unperturbed Schrödinger equation and the inhomogeneous equations of the various orders of perturbation theory were solved *exactly*. This is hardly realized in practice, where one rather solves both types of equations only approximately. The most common approximations are those based on replacing a Schrödinger (or related) equation by the equivalent *stationarity principle*, but to achieve stationarity only for a limited set of possible variations of the wave function. This stationary approach to DPT is the topic of the present paper.

A compact formulation of stationary perturbation theory in the *nonrelativistic* theory has recently been given [13] in a Lie-algebraic language. One of the essential messages of stationary perturbation theory is that under a certain condition the essential theorems of ''exact'' perturbation theory (e.g., that the first-order energy correction is equal to the expectation value of the perturbation with the unperturbed wave function) remain valid. The condition is that all perturbation corrections are formulated in terms of a ''variational group,'' with respect to which the unperturbed energy expectation value is stationary. Unfortunately, this Lie-algebraic formulation is not trivially applicable to the perturbation theory of relativistic corrections, because it is based on a unitary transformation between the unperturbed and the perturbed equation. Such a unitary transformation does exist in the Foldy-Wouthuysen transformation, but this introduces spurious singularities and is therefore not recommended [2].

We must hence formulate stationary perturbation theory of relativistic effects in a more pedestrian way, in the spirit of the direct-perturbation theory, thus being led to ''stationary direct perturbation theory'' (SDPT). This will be done in Sec. II.

The key quantities of SDPT are two functionals $F_0(\widetilde{\varphi}_0,\widetilde{\chi}_0)$ and $F_4(\widetilde{\varphi}_2,\widetilde{\chi}_2)$, which are made stationary with respect to variations of $\widetilde{\varphi}_0,\widetilde{\chi}_0,\widetilde{\varphi}_2,\widetilde{\chi}_2$, where $\widetilde{\varphi}_0,\widetilde{\chi}_0,\widetilde{\varphi}_2,\widetilde{\chi}_2$ are trial functions that approximate their counterparts $\varphi_0,\chi_0,\varphi_2,\chi_2$ as defined in (1.4). In defining auxiliary two-component spinors,

$$\widetilde{\omega}_k=\widetilde{\chi}_k-\frac{1}{2m}\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_k, \quad k=0,2 \tag{1.5}$$

one can reformulate these functionals in terms of $\widetilde{\varphi}_0$ and $\widetilde{\omega}_0$ or $\widetilde{\varphi}_2$ and $\widetilde{\omega}_2$, respectively, as is outlined in Sec. III. These new functionals decompose into independent functionals of $\widetilde{\varphi}_{2k}$ and $\widetilde{\omega}_{2k}$,

$$\widetilde{F}_0(\widetilde{\varphi}_0,\widetilde{\omega}_0)=F_{01}(\widetilde{\varphi}_0)+F_{02}(\widetilde{\omega}_0), \tag{1.6a}$$

$$\widetilde{F}_4(\widetilde{\varphi}_2,\widetilde{\omega}_2)=F_{41}(\widetilde{\varphi}_2)+F_{42}(\widetilde{\omega}_2)+F_{43}, \tag{1.6b}$$

such that $F_{01}(\widetilde{\varphi}_0)$ or $F_{41}(\widetilde{\varphi}_2)$ has a minimum when $\widetilde{\varphi}_0=\varphi_0$ or $\widetilde{\varphi}_2=\varphi_2$, while $F_{02}(\widetilde{\omega}_0)$ or $F_{42}(\widetilde{\omega}_2)$ has a maximum for $\widetilde{\omega}_0=\omega_0$ or $\widetilde{\omega}_2=\omega_2$ (F_{43} is a constant). The maximum of F_{02} is reached for $\omega_0=0$ and it is trivial to construct a functional of $\widetilde{\varphi}_0$ that is an upper bound for the exact E_0, such that *in the nonrelativistic limit* the theory becomes variational, provided one chooses $\widetilde{\omega}_0=\omega_0=0$.

If a functional to be made stationary is an upper bound to the corresponding exact value, one refers to a *variational* approach. A variational perturbation theory of relativistic effects, not only in the nonrelativistic limit, would be desirable. This could, e.g., be achieved if one could find conditions under which $\widetilde{F}_4(\widetilde{\varphi}_2,\widetilde{\omega}_2)$ is an upper bound to the exact E_4. Such a sufficient condition has been found by Rutkowski [6], namely

$$\widetilde{\chi}_2=\frac{1}{2m}\{\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_2+(V-E_0)\widetilde{\chi}_0\}. \tag{1.7}$$

Rutkowski was also able to show that this very condition (1.7) leads to *divergence* of $\widetilde{E}_6$ for the H atom ground state, provided that a $\widetilde{\varphi}_2$ regular at the position of the nucleus has been used.

The clarification of this dilemma is one of the topics of the present study and is treated in Sec. IV. We shall, in fact, show that the problem lies in the *combination* of the Rutkowski condition (1.7) with the choice of a *regular* trial function $\widetilde{\varphi}_2$, in the presence of a Coulomb singularity. If the trial function $\widetilde{\varphi}_2$ has the correct behavior at $r\to 0$, condition (1.7) leads to well-behaved $\widetilde{E}_4$ and $\widetilde{E}_6$, with $\widetilde{E}_4$ an upper bound to E_4. If, on the other hand, $\widetilde{\varphi}_2$ is regular at $r=0$, (1.7) must not be applied, because it makes $\widetilde{\chi}_2$ *too singular*; one must rather regularize $\widetilde{\chi}_2$ as well. Then one loses the strict upper-bound property for $\widetilde{E}_4$, but the error of $\widetilde{E}_4$ is smaller than if one had applied the Rutkowski condition, and $\widetilde{E}_6$ is well behaved. Details are found in Sec. IV.

The well-known kinetic balance [14–17] is one way of such a regularization of $\widetilde{\chi}$ for a regular $\widetilde{\varphi}$. Contrary to a common belief, it does not guarantee that the expectation value of the Dirac operator is an upper bound to its exact counterpart. This only holds up to $O(c^{-4})$, as originally pointed out by Stanton and Havrilak [14].

Although upper-bound properties in a *non-perturbative* context are not the topic of this paper, its results have some relevance for this topic, with respect to which, in the existing literature, claims (see, e.g., Ref. [18]) are more common than rigorous proofs. One of the few examples of a

strict variational approach is that of Hill and Krauthauser [19], which will, however, hardly be applicable beyond one-electron systems.

The problems studied in this paper are related to the singularity of the potential created by a point nucleus. Nevertheless, little changes if one replaces the point nucleus by a realistic finite nucleus. This is the subject of Sec. V.

II. STATIONARY DIRECT PERTURBATION THEORY

We want to make the energy expectation value of the Dirac operator (1.2) with modified metric

$$E=\langle\overline{\psi|D|\psi}\rangle=\langle\varphi|V|\varphi\rangle+2\,\mathrm{Re}\langle\varphi|\boldsymbol{\sigma}\cdot\mathbf{p}|\overline{\chi}\rangle-2m\langle\overline{\chi|\chi}\rangle+c^{-2}\langle\overline{\chi}|V|\overline{\chi}\rangle \tag{2.1}$$

stationary, subject to the *unitary normalization condition* (for an alternative derivation, in which unitary normalization is not assumed, see Appendix A),

$$\langle\overline{\psi|\psi}\rangle=\langle\varphi|\varphi\rangle+c^{-2}\langle\overline{\chi|\chi}\rangle=1 \tag{2.2}$$

for all c^{-2} in the neighborhood of $c^{-2}=0$. To this end, we expand E, φ, and $\overline{\chi}$ in powers of c^{-2} as in (1.4). Ordering powers of c^{-2} leads to

$$E_0=\langle\psi_0|D_0|\psi_0\rangle=\langle\varphi_0|V|\varphi_0\rangle+2\,\mathrm{Re}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_0\rangle-2m\langle\chi_0|\chi_0\rangle, \tag{2.3a}$$

$$\begin{aligned}E_2&=2\,\mathrm{Re}\langle\psi_0|D_0|\psi_2\rangle+\langle\psi_0|D_2|\psi_0\rangle\\&=2\,\mathrm{Re}\langle\varphi_0|V|\varphi_2\rangle+2\,\mathrm{Re}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_2\rangle\\&\quad+2\,\mathrm{Re}\langle\varphi_2|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_0\rangle-4m\,\mathrm{Re}\langle\chi_0|\chi_2\rangle\\&\quad+\langle\chi_0|V|\chi_0\rangle,\end{aligned} \tag{2.3b}$$

$$\begin{aligned}E_4&=2\,\mathrm{Re}\langle\psi_0|D_0|\psi_4\rangle+\langle\psi_2|D_0|\psi_2\rangle+2\,\mathrm{Re}\langle\psi_0|D_2|\psi_2\rangle\\&=2\,\mathrm{Re}\langle\varphi_0|V|\varphi_4\rangle+\langle\varphi_2|V|\varphi_2\rangle+2\,\mathrm{Re}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_4\rangle\\&\quad+2\,\mathrm{Re}\langle\varphi_2|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_2\rangle+2\,\mathrm{Re}\langle\varphi_4|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_0\rangle\\&\quad-4m\,\mathrm{Re}\langle\chi_0|\chi_4\rangle-2m\langle\chi_2|\chi_2\rangle-2\,\mathrm{Re}\langle\chi_0|V|\chi_2\rangle,\end{aligned} \tag{2.3c}$$

$$\begin{aligned}E_6&=2\,\mathrm{Re}\langle\psi_0|D_0|\psi_6\rangle+2\,\mathrm{Re}\langle\psi_2|D_0|\psi_4\rangle+\langle\psi_2|D_2|\psi_2\rangle\\&=2\,\mathrm{Re}\langle\varphi_0|V|\varphi_6\rangle+2\,\mathrm{Re}\langle\varphi_2|V|\varphi_4\rangle+2\,\mathrm{Re}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_6\rangle\\&\quad+2\,\mathrm{Re}\langle\varphi_2|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_4\rangle+2\,\mathrm{Re}\langle\varphi_4|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_2\rangle\\&\quad+2\,\mathrm{Re}\langle\varphi_6|\boldsymbol{\sigma}\cdot\mathbf{p}|\chi_0\rangle-4m\,\mathrm{Re}\langle\chi_0|\chi_6\rangle-4m\,\mathrm{Re}\langle\chi_2|\chi_4\rangle\\&\quad+2\,\mathrm{Re}\langle\chi_0|V|\chi_4\rangle+\langle\chi_2|V|\chi_2\rangle,\end{aligned} \tag{2.3d}$$

etc. The φ_k and χ_k have to satisfy the unitary normalization conditions

$$\langle\varphi_0|\varphi_0\rangle=1, \tag{2.4a}$$

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_2\rangle+\langle\chi_0|\chi_0\rangle=0, \tag{2.4b}$$

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_4\rangle+\langle\varphi_2|\varphi_2\rangle+2\,\mathrm{Re}\langle\chi_0|\chi_2\rangle=0, \tag{2.4c}$$

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_6\rangle+2\,\mathrm{Re}\langle\varphi_2|\varphi_4\rangle+\langle\chi_2|\chi_2\rangle+2\,\mathrm{Re}\langle\chi_0|\chi_4\rangle=0. \tag{2.4d}$$

Note that we take c^{-1} (which in atomic units is equal to the fine-structure constant α) as the formal perturbation parameter; one might as well have taken c^{-2} as perturbation parameter, then what we call E_2,E_4,E_6, respectively, would become E_1,E_2,E_3 Rutkowski [6,7] makes, e.g., the latter choice. One advantage of our counting is that terms of odd order in c^{-1}, which arise, e.g., for QED corrections, are easily symbolized.

All functions and energies in Eqs. (2.1)–(2.4) should carry a tilde to indicate that we consider approximate trial rather than exact functions. In order not to overcharge the notation we have omitted the tilde. Of course Eqs. (2.1)–(2.4) are also valid for the exact wave functions and energies. From now on we will put the tilde whenever we refer to trial functions (which include the exact ones as a special case).

Condition for stationarity of (2.3a) subject to the normalization condition (2.4a) is

$$\langle\delta\widetilde{\varphi}_0|(V-\lambda)\widetilde{\varphi}_0+\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\chi}_0\rangle=0, \tag{2.5a}$$

$$\langle\delta\widetilde{\chi}_0|\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_0-2m\widetilde{\chi}_0\rangle=0. \tag{2.5b}$$

In (2.5a) λ enters as a Lagrange multiplier, which is easily identified with the energy expectation value $\widetilde{E}_0$ as defined by (2.3a).

The functional made stationary by the conditions (2.5) is

$$F_0(\widetilde{\varphi}_0,\widetilde{\chi}_0)=\langle\widetilde{\varphi}_0|V-\lambda|\widetilde{\varphi}_0\rangle+2\,\mathrm{Re}\langle\widetilde{\varphi}_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\widetilde{\chi}_0\rangle-2m\langle\widetilde{\chi}_0|\widetilde{\chi}_0\rangle. \tag{2.6}$$

We shall call $F_0(\widetilde{\varphi}_0,\widetilde{\chi}_0)$ the ''Lévy-Leblond functional'' since it is stationary for all possible variations of $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$ if $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$ satisfy the Lévy-Leblond equation (1.3).

It is easily seen that F_0 is not bounded from below (unlike the functional $\langle\widetilde{\varphi}_0|H_0-\lambda|\widetilde{\varphi}_0\rangle$ corresponding to the nonrelativistic Schrödinger equation). Consider, e.g., a one-electron atom and expand both trial functions $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$ in an $s_{1/2}$ basis. Then the matrix elements of $\boldsymbol{\sigma}\cdot\mathbf{p}$ vanish and we get

$$F_0(\widetilde{\varphi}_0,\widetilde{\chi}_0)=\langle\widetilde{\varphi}_0|V-\lambda|\widetilde{\varphi}_0\rangle-2m\langle\widetilde{\chi}_0|\widetilde{\chi}_0\rangle. \tag{2.7}$$

This functional can become arbitrarily negative. We see that the Lévy-Leblond functional (2.6) suffers from the possibility of a variational collapse [17,20] to the same extent as does the expectation value of the Dirac operator. That the *variational collapse* of the Dirac equation persists in the nonrelativistic limit has probably first been observed by Schwarz and Wechsel-Trakowski [21] and by Mark and Schwarz [22].

We shall study the extremal properties of the functional (2.6) in more detail in Sec. III. At the moment we use a simple argument to avoid the variational collapse without making additional assumptions. We argue that it is trivial to satisfy (2.5b) for *all* possible variations $\delta\widetilde{\chi}_0$ (not just those expressible in the given basis), namely by choosing

$$\widetilde{\chi}_0=\frac{1}{2m}\,\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_0. \tag{2.8}$$

Inserting this into (2.5a) changes this to

$$\langle\delta\widetilde{\varphi}_0|H_0-\lambda|\widetilde{\varphi}_0\rangle=0,\quad H_0=T+V. \tag{2.9a}$$

This must at least hold for $\delta\widetilde{\varphi}_0$ proportional to $\widetilde{\varphi}_0$; hence

$$\lambda = \widetilde{E}_0 = \langle \widetilde{\varphi}_0 | H_0 | \widetilde{\varphi}_0 \rangle. \tag{2.9b}$$

Now $\widetilde{E}_0$ is the expectation value of the nonrelativistic Hamiltonian and is therefore bounded from below and (2.8) is the corresponding stationarity condition.

By eliminating the variational collapse of the Lévy-Leblond functional in this way we have avoided the variational collapse of the Dirac equation, since variational collapse of a relativistic calculation means by definition that one misses the correct nonrelativistic limit [17,20]. This procedure is often referred to as kinetic balance [14–17]. Deviations from an upper bound can now only arise to higher orders in c^{-2} [14].

It looks somewhat unbalanced to satisfy (2.5b) for all possible $\delta\widetilde{\chi}_0$, but (2.5a) only for a limited set of variations $\delta\widetilde{\varphi}_0$. Fortunately this unbalance does not affect the behavior of $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$ at the position of a nucleus, where both φ_0 and χ_0 are regular. This will no longer be the case for the next-higher-order functions φ_2 and χ_2 (see Sec. IV).

We now make the assumption, which is an essential ingredient of stationary perturbation theory [13], namely that the higher-order wave functions $\widetilde{\varphi}_2$, $\widetilde{\varphi}_4$, etc., are expandable in the basis with respect to which $\widetilde{E}_0$ is stationary [6,13], which means, e.g., that (2.9a) holds with $\delta\widetilde{\varphi}_0$ replaced by $\widetilde{\varphi}_2$ or $\widetilde{\varphi}_4$, etc, e.g.,

$$\langle \widetilde{\varphi}_2 | H_0 - \widetilde{E}_0 | \widetilde{\varphi}_0 \rangle = 0. \tag{2.10}$$

Then (2.3b) to (2.3d) will be replaced by

$$\begin{aligned}\widetilde{E}_2 &= 2\,\mathrm{Re}\langle \widetilde{\varphi}_0 | H_0 | \widetilde{\varphi}_2 \rangle + \langle \widetilde{\chi}_0 | V | \widetilde{\chi}_0 \rangle \\ &= 2\widetilde{E}_0\,\mathrm{Re}\langle \widetilde{\varphi}_0 | \widetilde{\varphi}_2 \rangle + \langle \widetilde{\chi}_0 | V | \widetilde{\chi}_0 \rangle = \langle \widetilde{\chi}_0 | V - \widetilde{E}_0 | \widetilde{\chi}_0 \rangle.\end{aligned} \tag{2.11a}$$

$$\begin{aligned}\widetilde{E}_4 &= 2\,\mathrm{Re}\langle \widetilde{\varphi}_0 | H_0 | \widetilde{\varphi}_4 \rangle + \langle \widetilde{\varphi}_2 | V | \widetilde{\varphi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_2 \rangle \\ &\quad - 2m\langle \widetilde{\chi}_2 | \widetilde{\chi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\chi}_0 | V | \widetilde{\chi}_2 \rangle \\ &= \langle \widetilde{\varphi}_2 | V - \widetilde{E}_0 | \widetilde{\varphi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_2 \rangle - 2m\langle \widetilde{\chi}_2 | \widetilde{\chi}_2 \rangle \\ &\quad + 2\,\mathrm{Re}\langle \widetilde{\chi}_0 | V - \widetilde{E}_0 | \widetilde{\chi}_2 \rangle.\end{aligned} \tag{2.11b}$$

$$\begin{aligned}\widetilde{E}_6 &= 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | V - \widetilde{E}_0 | \widetilde{\varphi}_4 \rangle + 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_4 \rangle \\ &\quad + 2\,\mathrm{Re}\langle \widetilde{\varphi}_4 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_2 \rangle - 4m\,\mathrm{Re}\langle \widetilde{\chi}_2 | \widetilde{\chi}_2 \rangle \\ &\quad + 2\,\mathrm{Re}\langle \widetilde{\chi}_0 | V - \widetilde{E}_0 | \widetilde{\chi}_4 \rangle + \langle \widetilde{\chi}_2 | V - \widetilde{E}_0 | \widetilde{\chi}_2 \rangle.\end{aligned} \tag{2.11c}$$

Obviously $\widetilde{E}_2$ is expressible in terms of $\widetilde{\chi}_0$ only, as in the exact theory. For (2.11a) to be valid it is not necessary to assume that $\widetilde{\varphi}_0$ is an exact solution of the Schrödinger equation.

$\widetilde{E}_4$ is expressed in terms of $\widetilde{\chi}_0$ (which we know already) and $\widetilde{\chi}_2$ as well as $\widetilde{\varphi}_2$, with respect to variations of which we have to make $\widetilde{E}_4$ stationary, subject to the normalization condition (2.4b); i.e., the functional to be made stationary is

$$\begin{aligned}F_4(\widetilde{\varphi}_2, \widetilde{\chi}_2) &= \langle \widetilde{\varphi}_2 | V - \widetilde{E}_0 | \widetilde{\varphi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_2 \rangle \\ &\quad - 2m\langle \widetilde{\chi}_2 | \widetilde{\chi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\chi}_0 | V - \widetilde{E}_0 | \widetilde{\chi}_2 \rangle \\ &\quad - \widetilde{E}_2\{2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \widetilde{\varphi}_0 \rangle + \langle \widetilde{\chi}_0 | \widetilde{\chi}_0 \rangle\} \\ &= \langle \widetilde{\psi}_2 | D_0 - \widetilde{E}_0 S_0 | \widetilde{\psi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\psi}_0 | D_2 - \widetilde{E}_0 S_2 \\ &\quad - \widetilde{E}_2 S_0 | \widetilde{\psi}_2 \rangle - \widetilde{E}_2 \langle \widetilde{\psi}_0 | S_2 | \widetilde{\psi}_0 \rangle.\end{aligned} \tag{2.12}$$

This may be regarded as the Hylleraas functional of a stationary direct relativistic perturbation theory. We propose to call it the Rutkowski-Hylleraas functional since it has first been studied by Rutkowski [6], though in a somewhat different notation (a precursor to that of direct perturbation theory).

Conditions for the stationarity of (2.12) are

$$\langle \delta\widetilde{\varphi}_2 | (V - \widetilde{E}_0)\widetilde{\varphi}_2 + \boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\chi}_2 - \widetilde{E}_2\widetilde{\varphi}_0 \rangle = 0, \tag{2.13a}$$

$$\langle \delta\widetilde{\chi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_2 - 2m\widetilde{\chi}_2 + (V - \widetilde{E}_0)\widetilde{\chi}_0 \rangle = 0. \tag{2.13b}$$

We shall always require that variation of $\widetilde{\varphi}_2$ (or $\widetilde{\chi}_2$) includes variation of a factor by which $\widetilde{\varphi}_2$ (or $\widetilde{\chi}_2$) is multiplied, which is the most trivial case of a linear variation parameter. Stationarity with respect to such a parameter implies that (2.13) holds with $\delta\widetilde{\varphi}_2$ (or $\delta\widetilde{\chi}_2$) replaced by $\widetilde{\varphi}_2$ (or $\widetilde{\chi}_2$).

This allows us to rewrite the stationary $F_4(\widetilde{\varphi}_2)$ alternatively as

$$\widetilde{E}_4 = \mathrm{Re}\langle \widetilde{\chi}_0 | V - \widetilde{E}_0 | \widetilde{\chi}_2 \rangle - \tfrac{1}{2}\widetilde{E}_2 \langle \widetilde{\chi}_0 | \widetilde{\chi}_0 \rangle \tag{2.14}$$

or as

$$\begin{aligned}\widetilde{E}_4 &= -\langle \widetilde{\varphi}_2 | V - \widetilde{E}_0 | \widetilde{\varphi}_2 \rangle - 2\,\mathrm{Re}\langle \widetilde{\varphi}_2 | \boldsymbol{\sigma}\cdot\mathbf{p} | \widetilde{\chi}_2 \rangle \\ &\quad + 2m\langle \widetilde{\chi}_2 | \widetilde{\chi}_2 \rangle + \widetilde{E}_2 \langle \widetilde{\chi}_0 | \widetilde{\chi}_0 \rangle \\ &= -\langle \widetilde{\psi}_2 | D_0 - \widetilde{E}_0 S_0 | \widetilde{\psi}_2 \rangle + \widetilde{E}_2 \langle \widetilde{\psi}_0 | S_2 | \widetilde{\psi}_0 \rangle.\end{aligned} \tag{2.15}$$

The result (2.14) is formally identical to that which one gets in the exact perturbation theory, for the same (unitary) normalization [1,2,6], just with the exact χ_0 and χ_2 replaced by the approximative ones.

We can now use the stationarity conditions (2.13) for F_4 to simplify $\widetilde{E}_6$ as given by (2.11c). Requiring that $\widetilde{\varphi}_4$ and $\widetilde{\chi}_4$ be expandable in the bases used for the expansion of $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ respectively, we get

$$\widetilde{E}_6 = \langle \widetilde{\chi}_2 | V - \widetilde{E}_0 | \widetilde{\chi}_2 \rangle - \widetilde{E}_2\{\langle \widetilde{\varphi}_2 | \widetilde{\varphi}_2 \rangle + 2\,\mathrm{Re}\langle \widetilde{\chi}_0 | \widetilde{\chi}_2 \rangle\}. \tag{2.16}$$

Again, we reproduce the result for the exact perturbation theory [1,2,6] in unitary normalization.

The critical step prior to the construction of $\widetilde{E}_4$ and $\widetilde{E}_6$ is the determination of $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ from (2.13). In analogy to the solution of (2.5) one can think of two possibilities.

(a) One solves (2.13b) exactly, i.e., chooses the condition (1.7).

(b) One expands $\widetilde{\chi}_2$ in a basis (usually of functions regular at $r=0$) and solves (2.13a) and (2.13b) via their matrix representations.

Rutkowski [6] has realized that (1.7) guarantees that the approximate $\widetilde{E}_4$ is an upper bound to the exact E_4 and has

therefore favored choice (a). However, he has also found that for this very choice (1.7) $\widetilde{E}_6$ diverges, provided that $\widetilde{\varphi}_2$ was chosen regular at the origin. This observation disqualifies choice (a) and lets choice (b) appear to be preferable, in particular, since then $\widetilde{E}_6$ does not diverge and $\widetilde{E}_4$ comes out closer to the exact E_4 (although it is no longer a rigorous upper bound to E_4). We have done our numerical calculations [12,23] entirely in terms of this variant. We shall analyze this problem in more detail in Secs. III and IV.

To conclude this section let us wonder whether the second-order correction $\widetilde{E}_2$ given by (2.11a) can be transformed to the well-known Pauli form. If we insert (2.8) and apply the turnover rule for $\boldsymbol{\sigma}\cdot\mathbf{p}$—which must be done in the distribution sense [24]—we can rewrite (2.11a) as

$$\begin{aligned}\widetilde{E}_2&=\frac{1}{4m^2}\langle\widetilde{\varphi}_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\widetilde{\varphi}_0\rangle\\&=\frac{1}{8m^2}\langle\widetilde{\varphi}_0|\boldsymbol{\sigma}\cdot\mathbf{p}[V,\boldsymbol{\sigma}\cdot\mathbf{p}]|\widetilde{\varphi}_0\rangle\\&\quad+\frac{1}{8m^2}\langle\widetilde{\varphi}_0|[\boldsymbol{\sigma}\cdot\mathbf{p},V]\boldsymbol{\sigma}\cdot p|\widetilde{\varphi}_0\rangle\\&\quad+\frac{1}{4m}\langle\widetilde{\varphi}_0|T(V-E_0)+(V-E_0)T|\widetilde{\varphi}_0\rangle\\&=-\frac{1}{8m^2}\langle\widetilde{\varphi}_0|[\boldsymbol{\sigma}\cdot\mathbf{p},(\boldsymbol{\sigma}\cdot\mathbf{p},V)]|\widetilde{\varphi}_0\rangle\\&\quad+\frac{1}{4m}\langle\widetilde{\varphi}_0|[T,V-E_0]+|\widetilde{\varphi}_0\rangle.\end{aligned}\tag{2.17}$$

The first term on the last right-hand side (rhs) of (2.17) can easily be reformulated to the sum of Darwin and spin-orbit terms (for V spherically symmetric)

$$\frac{\hbar^2}{8m^2}\langle\widetilde{\varphi}_0|\nabla^2V|\widetilde{\varphi}_0\rangle-\hbar\left\langle\widetilde{\varphi}_0\middle|\boldsymbol{\sigma}\cdot\mathbf{l}\,\frac{1}{r}\,\frac{\partial V}{\partial r}\middle|\widetilde{\varphi}_0\right\rangle.\tag{2.18}$$

The second term on the last rhs of (2.17) is equal to the velocity-mass term plus a correction

$$-\frac{1}{2m}\langle\widetilde{\varphi}_0|T^2|\widetilde{\varphi}_0\rangle+\frac{1}{4m}\langle\widetilde{\varphi}_0|[T,H_0-E_0]_+|\widetilde{\varphi}_0\rangle.\tag{2.19}$$

The correction vanishes if $(H_0-E_0)\varphi_0=0$, i.e., if $\widetilde{\varphi}_0=\varphi_0$, but is not negligible otherwise. In other words, if $\widetilde{\varphi}_0$ is not the exact nonrelativistic wave function, the expression for $\widetilde{E}_2$ in terms of the Pauli Hamiltonian is a poorer approximation than (2.11a). This has been confirmed numerically [12].

One may argue that the correction term even vanishes if $(H_0-E_0)\widetilde{\varphi}_0$ is not exactly zero, but that the stationarity condition (2.9a) should be sufficient. However, this is usually not the case. Condition (2.9a) would be sufficient to guarantee that the second term in (2.19) vanishes, if the variations $\delta\varphi_0$ considered in (2.9a) would include those proportional to $T\varphi_0$. In other words, wave functions of the type $T\varphi_0$ ought to be included in the basis into which one expands φ_0. This is usually not done, and not even recommended in view of the singularity of these functions.

III. EXTREMAL PROPERTIES OF THE LEVY-LEBLOND FUNCTIONAL AND OF THE HYLLERAAS-RUTKOWSKI FUNCTIONAL

The Lévy-Leblond equation (1.3) is the condition for stationarity of the functional (2.6) with respect to arbitrary variations of φ_0 and χ_0. Taking the variations of (2.6) with respect to φ_0 and χ_0 and equating them to zero lead immediately to (2.5). The λ in (2.5a) plays the role of a Lagrange parameter, taking care of the normalization of φ_0.

In order to study the extremal properties of this functional it is recommended to replace $\widetilde{\chi}_0$ by another two-component spinor

$$\widetilde{\omega}_0=\widetilde{\chi}_0-\frac{1}{2m}\,\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_0\tag{3.1}$$

and to formulate F_0 given by (2.6) in terms of $\widetilde{\varphi}_0$ and $\widetilde{\omega}_0$,

$$\begin{aligned}F_0(\widetilde{\varphi}_0,\widetilde{\chi}_0)=\widetilde{F}_0(\widetilde{\varphi}_0,\widetilde{\omega}_0)&=\langle\widetilde{\varphi}_0|H_0-\lambda|\widetilde{\varphi}_0\rangle-2m\langle\widetilde{\omega}_0|\widetilde{\omega}_0\rangle\\&=\widetilde{F}_{01}(\widetilde{\varphi}_0)+\widetilde{F}_{02}(\widetilde{\omega}_0).\end{aligned}\tag{3.2}$$

The tilde on $\widetilde{F}_0$ (and analogously later on $\widetilde{F}_4$) shall indicate that it is taken as a functional of $\widetilde{\varphi}_0$ and $\widetilde{\omega}_0$ rather than of $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$. It has hence a different meaning than a tilde on $\widetilde{\varphi}$, $\widetilde{\chi}$, or $\widetilde{E}$.

We see that $\widetilde{F}_0$ decomposes into a sum of two independent functionals, one of $\widetilde{\varphi}_0$, the other of $\widetilde{\omega}_0$. One realizes immediately that

$$\widetilde{F}_{01}(\widetilde{\varphi}_0)=\langle\widetilde{\varphi}_0|H_0-\lambda|\widetilde{\varphi}_0\rangle\geqslant 0,\tag{3.3a}$$

$$\widetilde{F}_{02}(\widetilde{\omega}_0)=-2m\langle\widetilde{\omega}_0|\widetilde{\omega}_0\rangle\leqslant 0.\tag{3.3b}$$

The stationarity conditions for $\widetilde{F}_{01}$ and $\widetilde{F}_{02}$ are respectively

$$\langle\delta\widetilde{\varphi}_0|H_0-\lambda|\widetilde{\varphi}_0\rangle=0,\tag{3.3c}$$

$$\langle\delta\widetilde{\omega}_0|\widetilde{\omega}_0\rangle=0.\tag{3.3d}$$

Stationarity with respect to arbitrary variations of $\widetilde{\varphi}_0$ and $\widetilde{\omega}_0$ is achieved if

$$(H_0-E_0)\varphi_0=0,\quad \omega_0=0.\tag{3.4}$$

For $\widetilde{\varphi}_0=\varphi_0$ and $\widetilde{\chi}_0=\chi_0$, $\widetilde{F}_{01}$ has its minimum, $\widetilde{F}_{02}$ its maximum. Assume now that we approximate both φ_0 by $\widetilde{\varphi}_0$ and χ_0 by $\widetilde{\chi}_0$, with $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$ completely independent, which implies that the exact $\omega_0\equiv 0$ is approximated by $\widetilde{\omega}_0=\widetilde{\chi}_0-\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_0/2$, which is generally not identically zero, then the error of $\widetilde{E}_0$ consists of two parts (for $\widetilde{\varphi}_0$ normalized to unity)

$$\Delta E_0=\widetilde{E}_0-E_0=\Delta E_{01}+\Delta E_{02},\tag{3.5a}$$

$$\Delta E_{01}=\langle\widetilde{\varphi}_0|H_0|\widetilde{\varphi}_0\rangle-E_0\geqslant 0,\tag{3.5b}$$

$$\Delta E_{02}=-2m\langle\widetilde{\omega}_0|\widetilde{\omega}_0\rangle\leqslant 0.\tag{3.5c}$$

We see that although ΔE_0 is not positive semidefinite, it is easy to enforce that it becomes so, simply by choosing $\widetilde{\omega}_0=0$ i.e., by enforcing the kinetic balance condition (2.8) between $\widetilde{\varphi}_0$ and $\widetilde{\chi}_0$.

Let us now try to decompose the Rutkowski-Hylleraas functional (2.12) in a similar way. We define in analogy to (3.1),

$$\widetilde{\omega}_2=\widetilde{\chi}_2-\frac{1}{2m}\,\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_2, \tag{3.6}$$

such that (2.12) can be reformulated to

$$\widetilde{F}_4(\widetilde{\varphi}_2,\widetilde{\omega}_2)=F_4(\widetilde{\varphi}_2,\widetilde{\chi}_2)=\widetilde{F}_{41}(\widetilde{\varphi}_2)+\widetilde{F}_{42}(\widetilde{\omega}_2)+\widetilde{F}_{43}, \tag{3.7}$$

$$\widetilde{F}_{41}(\widetilde{\varphi}_2)=\langle\widetilde{\varphi}_2|H_0-\widetilde{E}_0|\widetilde{\varphi}_2\rangle + \frac{1}{2m^2}\,\mathrm{Re}\langle\widetilde{\varphi}_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-\widetilde{E}_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\widetilde{\varphi}_2\rangle - 2\widetilde{E}_2\,\mathrm{Re}\langle\widetilde{\varphi}_2|\widetilde{\varphi}_0\rangle, \tag{3.8a}$$

$$\widetilde{F}_{42}(\widetilde{\omega}_2)=-2m\langle\widetilde{\omega}_2|\widetilde{\omega}_2\rangle+2\,\mathrm{Re}\langle\widetilde{\chi}_0|V-\widetilde{E}_0|\widetilde{\omega}_2\rangle, \tag{3.8b}$$

$$\widetilde{F}_{43}=-\widetilde{E}_2\langle\widetilde{\chi}_0|\widetilde{\chi}_0\rangle. \tag{3.8c}$$

Stationarity conditions for (3.8) are

$$\left\langle \delta\widetilde{\varphi}_2\middle|(H_0-\widetilde{E}_0)\widetilde{\varphi}_2+\frac{1}{4m^2}\,\boldsymbol{\sigma}\cdot\mathbf{p}(V-\widetilde{E}_0)\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}_0-\widetilde{E}_2\middle|\widetilde{\varphi}_0\right\rangle = 0, \tag{3.9a}$$

$$\langle\delta\widetilde{\omega}_2|-2m\widetilde{\omega}_2+(V-\widetilde{E}_0)|\widetilde{\chi}_0\rangle=0. \tag{3.9b}$$

Choosing $\delta\widetilde{\varphi}_2$ in (3.9a) proportional to $\widetilde{\varphi}_2$ implies that

$$\langle\widetilde{\varphi}_2|H_0-\widetilde{E}_0|\widetilde{\varphi}_2\rangle=\widetilde{E}_2\,\mathrm{Re}\langle\widetilde{\varphi}_2|\widetilde{\varphi}_0\rangle - \frac{1}{4m^2}\,\mathrm{Re}\langle\widetilde{\varphi}_2|\boldsymbol{\sigma}\cdot\mathbf{p}(V-\widetilde{E})\boldsymbol{\sigma}\cdot\mathbf{p}|\widetilde{\varphi}_0\rangle. \tag{3.10a}$$

and hence

$$\widetilde{F}_{41}(\widetilde{\varphi}_2)=-\langle\widetilde{\varphi}_2|H_0-\widetilde{E}_0|\widetilde{\varphi}_2\rangle\leqslant 0. \tag{3.10b}$$

Similarly we get for $\delta\widetilde{\omega}_2$ proportional to $\widetilde{\omega}_2$

$$2m\langle\widetilde{\omega}_2|\omega_2\rangle=\langle\widetilde{\omega}_2|V-\widetilde{E}_0|\widetilde{\chi}_0\rangle, \tag{3.11a}$$

$$\widetilde{F}_{42}(\widetilde{\omega}_2)=2m\langle\widetilde{\omega}_2|\widetilde{\omega}_2\rangle\geqslant 0. \tag{3.11b}$$

The stationary $\widetilde{F}_{41}$ is negative, the stationary $\widetilde{F}_{42}$ is positive, while $\widetilde{F}_{43}$ is positive as well, since $\widetilde{E}_2$ is negative.

Let us now assume that φ_2 and ω_2 are exact, i.e., make $\widetilde{F}_4$ stationary for all variations, and let $\widetilde{\varphi}_2$ and $\widetilde{\omega}_2$ be some approximations. Then

$$\widetilde{F}_{41}(\widetilde{\varphi}_2)-\widetilde{F}_{41}(\varphi_2)=\langle\widetilde{\varphi}_2-\varphi_2|H_0-\widetilde{E}_0|\widetilde{\varphi}_2-\varphi_2\rangle\geqslant 0, \tag{3.12a}$$

$$\widetilde{F}_{42}(\widetilde{\omega}_2)-\widetilde{F}_{42}(\omega_2)=-2m\langle\widetilde{\omega}_2-\omega_2|\widetilde{\omega}_2-\omega_2\rangle\leqslant 0, \tag{3.12b}$$

i.e., $\widetilde{F}_{41}(\widetilde{\varphi}_2)$ is an upper bound to its exact counterpart, while $\widetilde{F}_{42}$ is a lower bound.

We have thus gotten a simple expression for the error in $\widetilde{E}_4$, namely,

$$\Delta E_4=\widetilde{E}_4-E_4=\langle\widetilde{\varphi}_2-\varphi_2|H_0-\widetilde{E}_0|\widetilde{\varphi}_2-\varphi_2\rangle - 2m\langle\widetilde{\omega}_2-\omega_2|\widetilde{\omega}_2-\omega_2\rangle. \tag{3.13}$$

We can, of course, achieve an overall upper bound, i.e., $\widetilde{E}_4-E_4>0$ if $\widetilde{\omega}_2$ is chosen such that $\widetilde{F}_{42}$ attains its maximum, which is the case for $\widetilde{\omega}_2=\omega_2$. We see from (3.9b) that this means

$$\widetilde{\omega}_2=\frac{1}{2m}\,(V-\widetilde{E}_0)\widetilde{\chi}_0, \tag{3.14}$$

which is equivalent to relating $\widetilde{\chi}_2$ and $\widetilde{\varphi}_2$ via (1.7). We have thus given an alternative proof to that of Rutkowski [6] that (1.7) guarantees that $\widetilde{E}_4$ is an upper bound to E_4. We remember, of course, that for this very choice $\widetilde{E}_6$ diverges, provided that $\widetilde{\varphi}_2$ is chosen regular at the position of a point nucleus. We come back to this problem in Sec. IV.

The decomposition of $\widetilde{\chi}_2$ according to (3.6) into two parts has allowed us to understand the stationarity properties of the Hylleraas functional F_4 better. However, this decomposition requires some care, since it is, in some sense, contradictory to our philosophy in previous papers [1,2]. There we have insisted that in order to avoid singularities, the two contributions to the χ_{2n} should never be separated. We see, in fact, that the separation of E_6 along the same lines becomes problematic. Formally we can write, of course

$$E_6=E_{61}(\varphi_2)+E_{62}(\omega_2)+E_{63}(\varphi_2,\omega_2)+E_{64}, \tag{3.15a}$$

$$E_{61}(\varphi_2)=\frac{1}{4m^2}\,\langle\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_2|V-E_0|\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_2\rangle-E_2\langle\varphi_2|\varphi_2\rangle - \frac{1}{m}\,E_2\mathrm{Re}\langle\chi_0|\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2\rangle-2E_4\mathrm{Re}\langle\varphi_2|\varphi_0\rangle, \tag{3.15b}$$

$$E_{62}(\omega_2)=\langle\omega_2|V-E_0|\omega_2\rangle-2E_2\mathrm{Re}\langle\chi_0|\omega_2\rangle, \tag{3.15c}$$

$$E_{63}(\varphi_2,\omega_2)=\frac{1}{2m}\mathrm{Re}\langle\omega_2|V-E_0|\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_2\rangle, \tag{3.15d}$$

$$E_{64}=-E_4\langle\chi_0|\chi_0\rangle. \tag{3.15e}$$

Unlike for the analogous decomposition of E_4 there is also a mixed term, both depending on φ_2 and ω_2. The main difference from the decomposition of E_4 is, however, that now all contributions diverge in the presence of a point nucleus, even for the exact φ_2. The decomposition (3.15) can hence only be legitimate if all integrals are first taken from r_c to ∞ and if the limit $r_c\to 0$ in the sum (3.15a) is taken after all contributions that diverge in this limit have been cancelled. A generalization of the decoupling of F_0 and of F_4 and the

related minimax property is, unfortunately, not possible for the higher-order functionals F_6, etc.

IV. ROLE OF THE REGION NEAR A POINT NUCLEUS

The choice of the trial function in a stationary approach is not fully arbitrary. This function must at least be square integrable to all orders in the expansion parameter c^{-1}, and the expectation value of the Dirac operator must also exist to all orders. It is usually not necessary that the trial function satisfy the same boundary conditions as the exact wave function. The classical example is a Gaussian trial function for the nonrelativistic ground state of the H atom, which neither satisfies the correct boundary condition at $r\to 0$ nor at $r\to\infty$ and is nevertheless acceptable. In the context of the perturbation theory of relativistic effects, the behavior near the positon of a nucleus matters, because it affects the square integrability and the existence of expectation values of the Dirac operator. A careful study of this behavior is hence necessary.

Since deviations from square integrability or divergences of matrix elements only arise in a small neighborhood around a point nucleus, it is sufficient to have a look at this region. For the sake of simplicity we now consider a one-electron atom, but the generalization to arbitrary one-electron molecules or even to many-electron systems in the Hartree-Fock approximation is straightforward, because the behavior of the exact wave function near any nucleus is of the same type, simply as a consequence of the Coulomb singularity [25]. Without explicitly stating this, we consider in this section always point nuclei.

In Appendix B the behavior of the exact wave function in a c^{-1} expansion near the nucleus is outlined. We repeat the result here for the case $\kappa=-1$, $l=0$. (The cases $\kappa<0$ and $l>0$ are less critical, since in view of factors r^l in the components of the wave functions no singularities of the contributions to the energy arise to the orders considered here. The case $\kappa=1$, $l=1$ is somewhat analogous to the case $\kappa=-1$, $l=0$ discussed here.) The large and small components can be expanded according to (B1) and (B2) and we get, as special cases of (B13) and (B14) for $\kappa=-1$,

$$g_0=a_0(1-Zr)+O(r^2), \tag{4.1a}$$

$$g_2=-a_0\,\frac{Z^2}{2}\,\ln r+O(r), \tag{4.1b}$$

$$g_4=-a_0\,\frac{Z^4}{8}\,(\ln r-\ln^2 r)+O(r), \tag{4.1c}$$

$$f_0=a_0\,\frac{Z}{2}+O(r), \tag{4.2a}$$

$$f_2=a_0\,\frac{Z^3}{8}\,(1-2\,\ln r)+O(r), \tag{4.2b}$$

$$f_4=a_0\,\frac{Z^5}{16}\,(1-2\,\ln r+\ln^2 r)+O(r) \tag{4.2c}$$

with a_0 a common numerical factor (not to be confused with the atomic unit of length).

If we expand ω in analogy to (B1) and (B2) as

$$\omega=i\;\;h(r)\,\eta_\kappa^m(\vartheta,\varphi,s), \tag{4.3a}$$

$$h(r)=\frac{1}{c}\,r^\nu[c_0+c_1 r+O(r^2)]=\sum_k\;c^{-2k-1}h_{2k}(r), \tag{4.3b}$$

the exact h_{2k} become

$$h_0=0, \tag{4.4a}$$

$$h_2=\frac{a_0Z^2}{4}\,r^{-1}+O(r^0), \tag{4.4b}$$

$$h_4=\frac{a_0Z^4}{16}\,\{r^{-1}-2r^{-1}\ln r\}+O(r^0). \tag{4.4c}$$

We also note that, for a hydrogenlike ion, as a consequence of the normalization of g_0,

$$a_0=2Z^{3/2}. \tag{4.5}$$

Let us now study the contributions of a region $0<r<r_c$ around a point nucleus at $\mathbf{r}=\mathbf{0}$ to E_0,E_2,E_4,E_6 for the exact functions (so far we do not specify r_c).

$$E_0:\quad \int_0^{r_c} g_0H_0g_0r^2dr=E_0a_0^2\int_0^{r_c}\{r^2-2Zr^3+O(r^4)\}dr=E_0a_0^2\left\{\frac{1}{3}\,r_c^3-\frac{Z}{2}\,r_c^4+O(r_c^5)\right\}, \tag{4.6a}$$

$$E_2:\quad \int_0^{r_c} f_0\left(-\frac{Z}{r}-E_0\right)f_0r^2dr=-a_0^2\,\frac{Z^3}{4}\int_0^{r_c} r[1+O(r)]dr=-a_0^2\,\frac{Z^3}{8}\,r_c^2+O(r_c^3), \tag{4.6b}$$

$$E_4:\quad \int_0^{r_c} f_0\left(-\frac{Z}{r}-E_0\right)f_2r^2dr-\tfrac{1}{2}E_2\int\,(f_0)^2r^2dr=-a_0\,\frac{Z^5}{16}\int_0^{r_c}\{r-2r\,\ln r\}dr+O(r_c^3)=-a_0\,\frac{Z^5}{16}\,r_c^2(1-\ln r_c)+O(r_c^3), \tag{4.6c}$$

$$E_6:\quad \int_0^{r_c} f_2\left(-\frac{Z}{r}\right)f_2r^2dr+O(r_c^3)=a_0^2\,\frac{Z^6}{64}\int_0^{r_c}(r-4r\,\ln r+4r\,\ln^2 r)dr+O(r_c^3)=a_0^2\,\frac{Z^6}{64}\,r_c^2(\tfrac{3}{2}-2\,\ln r_c)+O(r_c^3). \tag{4.6d}$$

All these contributions are regular, in spite of the weak singularities in g_2 and f_2. The contribution to E_2 of the region $r<r_c$ is proportional to the volume of this region, the contributions to E_4 and E_6 are relatively larger than the volume of the region [noting that r_c is regarded as small and hence $O(r_c^2)$ is larger than $O(r_c^3)$].

Let us have a look at the two contributions $\widetilde{F}_{41}$ and $\widetilde{F}_{42}$ to E_4, which are given by (3.10b) and (3.11b). It is convenient to replace $\int\varphi T\varphi d\tau$ by $\frac{1}{2}\int(\nabla\varphi)^2 d\tau$ to avoid spurious boundary terms at $r=r_c$.

$$\widetilde{F}_{41}: \quad -\int_0^{r_c}\{\tfrac{1}{2}(\nabla g_2)^2-E_0(g_2)^2\}r^2dr$$

$$=-a_0^2\frac{Z^4}{8}\int_0^{r_c}\{r^{-2}+O(r^0)\}r^2dr$$

$$=-a_0^2\frac{Z^4}{8}r_c+O(r_c^2), \tag{4.7a}$$

$$\widetilde{F}_{42}: \quad 2\int_0^{r_c}(h_2)^2r^2dr=a_6^2\frac{Z^4}{8}r_c+O(r_c^2). \tag{4.7b}$$

Either contribution is now much larger in absolute value [being of $O(r_c)$] than their sum E_4 [of $O(r_c^2)$], in which the $O(r_c)$ terms cancel.

We now consider approximate wave functions. As long as $\widetilde{g}_2$ behaves as (4.1b) and as $\widetilde{f}_2$ is constructed from $\widetilde{g}_2$ via (1.7), which means

$$\widetilde{f}_2=-\frac{1}{2}\left\{\frac{\partial}{\partial r}\widetilde{g}_2+\left(\frac{Z}{r}+E_0\right)f_0\right\}$$

$$=-a_0\frac{Z^2}{2}r^{-1}+a_0\frac{Z^2}{2}r^{-1}+O(r^0)$$

$$=O(r^0), \tag{4.8a}$$

$$\widetilde{h}_2=\left(-\frac{Z}{r}-E_0\right)f_0=\frac{a_0Z^4}{4}r^{-1}+O(r^0). \tag{4.8b}$$

(4.6c) and (4.6d) as well as (4.7a) and (4.7b) remain essentially unchanged. Both $\widetilde{E}_4$ and $\widetilde{E}_6$ are regular.

Let us next choose $\widetilde{g}_2$ regular at $r=0$, i.e., of the form

$$\widetilde{g}_2=A_0+A_1r+O(r^2) \tag{4.9a}$$

and $\widetilde{f}_2$ related to $\widetilde{g}_2$ by the Rutkowski condition (1.7)

$$\widetilde{f}_2=\frac{1}{2}\left\{\frac{\partial}{\partial r}\widetilde{g}_2+\left(\frac{Z}{r}+E_0\right)f_0\right\}=a_0\frac{Z^2}{4}r^{-1}+O(r^0), \tag{4.9b}$$

$$\widetilde{h}_2=\frac{a_0Z^4}{4}r^{-1}+O(r^0). \tag{4.9c}$$

Instead of (4.7), (4.6c), and (4.6d) we get

$$\widetilde{F}_{41}: \quad O(r_c^2), \tag{4.10a}$$

$$\widetilde{F}_{42}: \quad -a_0^2\frac{Z^4}{8}r_c+O(r_c^2), \tag{4.10b}$$

$$\widetilde{E}_4: \quad -a_0^2\frac{Z^4}{8}r_c+O(r_c^2), \tag{4.10c}$$

$$\widetilde{E}_6: \quad a_0^2\frac{Z^5}{16}\int_0^{r_c}r^{-1}dr=\infty. \tag{4.10d}$$

The leading contributions of the terms $O(r_c)$ to $\widetilde{E}_4$ no longer cancel, and the total contribution to $\widetilde{E}_4$ is of $O(r_c)$ rather than of $O(r_c^2)$ as for the correct wave function, while the contribution to $\widetilde{E}_6$ now diverges. The origin of both deteriorations is that $\widetilde{f}_2$ as given by (4.0) is now too singular (it goes for small r as $\sim r^{-1}$ instead of $\sim\ln r$). Choosing $\widetilde{g}_2$ regular and imposing (1.7) makes $\widetilde{f}_2$ too singular.

To avoid this spurious singularity there are two possible remedies. The best is to insist on the correct behavior (4.1b) of $\widetilde{g}_2$. Then all contributions of the region $r<r_c$ are regular, $\widetilde{E}_4$ is an upper bound to the exact $\widetilde{E}_4$, and $\widetilde{E}_6$ exists. (It has previously been pointed out in the nonpertubative context [26,27] that satisfaction of the correct boundary conditions by the trial function is essential for getting variational results.) The second best remedy is to regularize $\widetilde{f}_2$ as well, if $\widetilde{g}_2$ is regular. In fact, in not insisting on (1.7) in the region of small r and choosing

$$\widetilde{f}_2=B_0+B_1r+O(r^2), \tag{4.11a}$$

$$\widetilde{h}_2=B_0-\tfrac{1}{2}A_1+O(r), \tag{4.11b}$$

we get

$$\widetilde{F}_{41}: \quad O(r_c^2), \tag{4.12a}$$

$$\widetilde{F}_{42}: \quad O(r_c^3), \tag{4.12b}$$

$$\widetilde{E}_4: \quad -B_0\frac{a_0}{4}Z^2r_c^2+O(r_c^3), \tag{4.12c}$$

$$\widetilde{E}_6: \quad -B_0^2Zr_c^2+O(r_c^3) \tag{4.12d}$$

and we have both avoided spurious contributions $O(r_c)$ to $\widetilde{E}_4$ and the divergence of $\widetilde{E}_6$. Of course, $\widetilde{E}_4$ is no longer a rigorous upper bound to E_4.

How large deviations from an upper-bound property does one then expect? If one regularizes only in a region $r<r_c$, the error of $\widetilde{E}_4$ is of $O(r_c^2)$. If one chooses r_c as the critical radius (see Sec. V) where the Coulomb attraction becomes equal to the rest mass, we conclude that the error is of order c^{-4}, while without regularization the error is of order c^{-2}.

Let us consider a simple example of a brute-force regularization for the ground state of a H-like ion. It consists in cutting off (the exact) φ_2 at some cut-off radius r_c. Of course φ_0 and χ_0, which are regular anyway, are not cut off, such that E_0 and E_2 are not affected. The cutoff of φ_2 must be done smoothly enough, such that boundary terms in $\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_2$ (which is part of χ_2) are negligible. In expressing χ_2 exactly in terms of φ_2 in the sense of (1.7), there will be no cutoff in $\omega_2=1/2m(V-E_0)\chi_0$. The error of E_4 is then entirely given by the first term in (3.13) with

$$\widetilde{\varphi}_2=\begin{cases}\varphi_2 & \text{for } r>r_c\\ 0 & \text{for } r<r_c,\end{cases} \tag{4.13}$$

$$\widetilde{E}_4-E_4=\langle\widetilde{\varphi}_2-\varphi_2|H_0-E_0|\widetilde{\varphi}_2-\varphi_2\rangle$$
$$=\int_0^{r_c}r^2dr\int d\Omega\,ds\,\varphi_2^*(H_0-E_0)\varphi_2$$
$$=\frac{Z^7}{2}\int_0^{r_c}e^{-2Zr}dr=\frac{Z^7}{2}r_c-\frac{Z^8}{2}r_c^2+\tfrac{1}{3}Z^9r_c^3+O(r_c^4), \tag{4.14}$$

where $d\Omega$ is the solid angle element. This is in agreement with (4.7a) or (4.10c) and (4.5). For the evaluation of (4.14) the following useful auxiliary relation [28] has been used:

$$\langle\Phi|f(r)(H_0-E_0)f(r)|\Phi\rangle=\tfrac{1}{2}\left\langle\Phi\left|\left(\frac{\partial f}{\partial r}\right)^2\right|\Phi\right\rangle, \tag{4.15}$$

which holds when $(H_0-E_0)\Phi=0$. The error (4.14) is positive (i.e., we get an upper bound for E_4) and it is linear in r_c for small r_c. A reasonable value for r_c is the critical radius given by (5.1b). For this choice we get from (4.14)

$$\widetilde{E}_4-E_4=\frac{Z^8}{2mc^2}-\frac{Z^{10}}{2m^2c^4}+O(c^{-6}). \tag{4.16}$$

The error of $c^{-4}\widetilde{E}_4$ is then of the order of $c^{-6}E_6$. Even if $\widetilde{E}_6$ did not diverge (which it actually does for this choice), it would not make sense to go beyond $\widetilde{E}_4$ since its error is of the order of E_6.

Let us now regularize *both* φ_2 and χ_2 in the same brute-force way. This means we also cut off ω_2 at r_c, i.e., in addition to (4.13) we choose

$$\widetilde{\omega}_2=\begin{cases}\omega_2 & \text{for } r>r_c\\ 0 & \text{for } r<r_c.\end{cases} \tag{4.17}$$

Now the error of $\widetilde{E}_4$ consists of both terms in (3.13), i.e., in addition to (4.14) we have

$$2\langle\widetilde{\omega}_2-\omega_2|\widetilde{\omega}_2-\omega_2\rangle=\int_0^{r_c}r^2dr\int d\Omega\;ds|\omega_2|^2$$
$$=\frac{Z^7}{2}\int_0^{r_c}e^{-2Zr}\left(1-Zr+\frac{Z^2}{4}r^2\right)$$
$$=\frac{Z^7}{2}r_c-\tfrac{3}{4}Z^8r_c^2+\tfrac{17}{24}Z^9r_c^3+O(r_c^4). \tag{4.18}$$

From (4.14) and (4.18) we get

$$\widetilde{E}_4-E_4=\tfrac{1}{4}Z^8r_c^2-\tfrac{3}{8}Z^9r_c^3+O(r_c^4). \tag{4.19}$$

The terms linear in r_c have cancelled, such that the leading term is now $O(r_c^2)$ or for the choice (5.1b) of r_c,

$$\widetilde{E}_4-E_4=\frac{1}{4}\frac{Z^{10}}{m^2c^4}-\frac{3}{8}\frac{Z^{12}}{m^3c^6}+O(c^{-8}). \tag{4.20}$$

The error in $c^{-4}\widetilde{E}_4$ is now of the order of $c^{-8}E_8$, i.e., it does make sense to consider also E_6 (but not terms beyond E_6). Moreover, by plotting the error (4.19) as a function of r_c one finds that it increases monotonically to the value $3Z^2/32$ for $r_c\to\infty$. For this regularization procedure (for both φ_2 and χ_2) one still gets an upper bound to the exact E_4.

In practice one will not regularize by means of a cut-off radius, but rather by expanding both $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ (or $\widetilde{\omega}_2$) in a basis of functions that are regular at $r=0$. The contributions of the region $r<r_c$ will then essentially be the same as for the brute-force regularization, since they are determined by the weak singularities of the exact wave function, which one does not duplicate. However, there will be, in addition, a contribution of the region $r>r_c$ to the error, since in this region neither $\widetilde{\varphi}_2$ nor $\widetilde{\chi}_2$ will be exact. Since there are no singularities left, these error contributions will diminish with increasing basis size in the usual way. There is, however, no reason why the error in F_{41} should always be larger in absolute value than that of F_{42}. Hence there is no guarantee that one has an upper bound to the exact E_4; only convergence to the exact E_4 can be expected if the basis is chosen appropriately.

The strategy to follow in this regularized SDPT is obvious. One must not try to satisfy the Rutkowski condition (1.7) pointwise, in particular not at $r=0$, in order not to make χ_2 too singular. On the other hand, one must try to make the second term of the error ΔE_4 given by (3.13) as small as possible in order to be as close as possible to an upper bound for E_4. This means that the basis into which $\widetilde{\chi}_2$, and hence $\widetilde{\omega}_2$, is expanded must be such that (3.14) is well satisfied *in the mean*. If the basis for $\widetilde{\varphi}_2$ is well suited for approximating $\ln r\varphi_0$ in the mean, and is kinetically balanced, then the basis for $\widetilde{\chi}_2$ and $\widetilde{\omega}_2$ should well represent $r^{-1}\chi_0$, as required by (3.14).

One may refer to such a regularized stationary perturbation theory for E_4 as *quasivariational*. Although an upper-bound property is not guaranteed, deviations from an upper bound for E_4 are not serious. In order to avoid deviations from an upper bound, the basis for the expansion of χ_2 should not be poorer than that for the expansion of φ_2 because then the second term in (3.13) would become dominating.

Our numerical experience with the regularized SDPT has been the following.

(a) For standard Gaussian basis sets we have for the one-electron system H-like ions, H_2^+ and HeH^{2+}, never found any deviations from an upper bound for E_4 [12].

(b) For Stater-type orbital (STO) basis sets for H-like ions, upper bounds were obtinaed or not obtained, depending on the choice of the basis and on the type of state (e.g., $\kappa>0$ or $\kappa<0$), but the results are consistent with the conjecture that approaching basis completeness there is convergence to the exact E_4. Details are planned to be published in the future.

(c) In relativistic Hartree-Fock calculations both with Gaussian-type orbitals (GTOs) or STOs there was no indication of deviations from an upper bound, although, admittedly, no exact reference values are available.

The quasivariational regularized SDPT has one disadvantage with respect to a rigorous variation perturbation theory. The vanishing of the error of $\widetilde{E}_4$ does not imply that $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ are exact.

The error $\widetilde{E}_4-E_4$ is not a direct criterion of the quality of $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$. A better criterion is given by the two error

contributions in (3.13) independently or by the sum of their absolute values. To minimize this sum might be an interesting starting point for an alternative to SDPT.

With the regularized SDPT it may happen that the two error contributions in (3.13) cancel each other more or less by chance, such that $|\widetilde{E}_4 - E_4|$ becomes very small or vanishes, although $\widetilde{\varphi}_2$ and $\widetilde{\omega}_2$ are far from their exact counterparts. It can even occur that for a particular ansatz with a small number of parameters, $\widetilde{E}_4 - E_4$ is non-negative in the entire parameter space and has its minimum equal to 0 for one point in this space. One is then in the frame of a *pseudovariational* method.

A striking example of such a pseudovariational situation will be given in a separate paper [29]. Of course in such a situation Eq. (3.14) will be very poorly satisfied in the mean, as is seen in a large (in absolute value) second part of $\Delta\widetilde{E}_4$ as given by (3.13).

Let us repeat at this point that a strict variation perturbation theory for $\widetilde{E}_4$ is possible if one chooses $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ with the correct $\ln r$ singularities near $r=0$. If one prefers a regular $\widetilde{\varphi}_2$, one must also choose $\widetilde{\chi}_2$ regular at $r=0$. In trying to impose a strict upper-bound condition for $\widetilde{\chi}_2$, one would make $\widetilde{\chi}_2$ too singular and lose more than one would gain.

V. EXTENDED NUCLEI

Since most problems with the perturbation theory of relativistic effects are related to the singularity of the Coulomb potential, one may wonder whether on worrying about this singularity one is not studying a purely academic problem, since real nuclei are extended rather than point charges. This objection needs some comments.

Let us first observe that for a relativistic one-electron atom in its ground state there are four characteristic lengths.

(1) The Bohr radius

$$r_B = Z^{-1} a_0 \tag{5.1a}$$

with a_0 the atomic unit of length (1 bohr), not to be confused with the parameter a_0 used in Sec. IV and Appendix B.

(2) The critical radius, for which the Coulomb attraction becomes equal to the rest mass, which is Z times the *classical radius* of the electron, and is sometimes called the Thomson radius

$$r_c = Z\alpha^2 a_0 . \tag{5.1b}$$

(3) The nuclear radius r_N. According to a popular model [30] this is related to the nuclear mass M as

$$r_N = 2.27\times 10^{-5} M^{1/3} a_0 . \tag{5.1c}$$

If we choose roughly $M \approx 2.5\ Z$ (for $Z \gg 1$), this becomes

$$r_N \approx 3.1\times 10^{-5} Z^{1/3} a_0 \tag{5.1d}$$

and we get the following ratios:

$$r_c / r_B = Z^2\alpha^2 \approx 5\times 10^{-5} Z^2, \tag{5.2a}$$

$$r_N / r_B \approx 3\times 10^{-5} Z^{4/3}, \tag{5.2b}$$

$$r_c / r_N \approx 1.6\times Z^{2/3}. \tag{5.2c}$$

For all Z the critical radius, inside which the Coulomb attraction is larger in absolute value than the rest energy of the electron, is larger than the nuclear radius, in particular so for heavy nuclei (see, e.g., Fig. 1 of Ref. [20]).

(4) The Compton wavelength of the electron,

$$\lambda_c = \frac{\hbar}{mc} = \alpha\ \ a_0 , \tag{5.3}$$

which happens to be the geometrical mean of r_B and r_c. Note that λ_c is linear in α and is usually quite larger than r_c, which is quadratic in α.

We have to consider three independent issues. (a) the accuracy of the point nucleus as an approximation to the physical extended nucleus; (b) the rate of convergence of the perturbation expansion of relativistic corrections; (c) the question whether for finite nuclei the singularity problems typical for point nuclei disappear. For issue (a) the ratio r_N/r_B should be small, which is related to the probability of the electron being inside the nucleus. This ratio is very small for light nuclei and not so small for heavy ones, such that for these it is more important to use a finite nucleus (independent of the fact that a finite nucleus extends the existence of solutions of the Dirac equation beyond $Z=137$). The rate of convergence of perturbation theory (b) depends on the magnitude of $Z\alpha$, i.e., indirectly on the ratio r_c/r_B. For light nuclei this ratio is small and perturbation theory converges fast [1,2].

Whether a finite nucleus removes the problems with the Coulomb singularity depends on the ratio r_c/r_N. As long as $r_c > r_N$—and this is the case for all Z—there is a region near the nucleus with $r < r_c$, where the Coulomb attraction of the electron by the nucleus is larger than its rest energy and where the expansion of certain operators in powers of c^{-1} diverges. The ratio r_c/r_N even increases with Z, and hence the problems related to the almost-point-nuclear behavior of the nuclear potential become more serious with increasing Z (although in a different context the finite-size effects of the nucleus also become more important with increasing Z). We conclude that the problems related to the strong Coulomb force at small distance between nucleus and electron are not removed if one accounts for the finite size of the nucleus.

VI. CONCLUSIONS

In exact perturbation theory one assumes that the unperturbed equation is solved exactly and that the same holds for the inhomogeneous differential equations for the perturbative corrections to the wave function. In stationary perturbation theory the exact equations are replaced by stationarity conditions for some functionals. Even if stationarity is achieved only for a limited class of variations, some important results of the exact theory can be taken over, e.g., that the first-order correction to the energy is simply an expectation value evaluated with the unperturbed wave function.

Although the Dirac operator is not bounded from below, stationary direct perturbation theory combined with the kinetic balance condition lead to an upper bound of the energy expectation value in the nonrelativistic limit, i.e., the variational collapse is avoided. In order to get un upper bound for the functional $F_4(\widetilde{\varphi}_2, \widetilde{\chi}_2)$ that determines the lowest-order

relativistic corrections to the wave function, a condition has to be satisfied that leads to regular results only if $\widetilde{\varphi}_2$ and $\widetilde{\chi}_2$ have the correct logarithmic behavior at the position of a point nucleus. If one decides to choose a regular trial function $\widetilde{\varphi}_2$, one must also regularize $\widetilde{\chi}_2$, in order to avoid divergence of $\widetilde{E}_6$, i.e., the energy contribution to $O(c^{-6})$. The price to pay is that $\widetilde{E}_4$ is then no longer a rigorous upper bound to the exact E_4.

It is interesting that the energy as functional of the wave function can up to $O(c^{-4})$ be divided into two independent parts, one depending on the large component φ, the other on $\widetilde{\omega}=\widetilde{\chi}-\boldsymbol{\sigma}\cdot\mathbf{p}\widetilde{\varphi}/(2\mathrm{m})$, with $\widetilde{\chi}$ the small component, such that the first part gives an upper bound to its exact counterpart, the second one a lower bound. Such a minimax property of the relativistic energy functional, as, e.g., discussed by Talman [31] does not hold to higher orders.

We have here explicitly only considered the lower orders of SDPT, i.e., to $O(c^{-2})$ for the wave function and to $O(c^{-6})$ for the energy, but the essential results persist to higher orders, especially as far as the validity of exact expressions and the regularization at the position of a point nucleus are concerned. We have finally seen that although most problems that arise with the perturbation theory of relativistic effects are related to the Coulomb singularity, the use of realistic extended nuclei does not make a substantial change since the critical radius, inside of which the Coulomb attraction is larger than the rest energy of the electron, is always larger than the nuclear radius.

Although we were concerned with stationary and variational approaches, in the framework of direct perturbation theory, the present results are also relevant for nonperturbative approximative solutions of the Dirac equation. In particular, one understands better under which conditions a quasivariational theory of the Dirac equation is possible, and the meaning of the use of *kinetically balanced* basis sets becomes more transparent. Stationary direct perturbation theory can be directly applied to relativistic Hartree-Fock theory [12,21,25].

ACKNOWLEDGMENTS

The author thanks W. H. E. Schwarz, Ch. van Wüllen, E. Ottschofski, R. Franke, and W. Cencek for useful comments. This work was supported by Deutsche Forschungsgemeinschaft and Fond der Chemischen Industrie.

APPENDIX A: ALTERNATIVE DERIVATION OF STATIONARY DIRECT PERTURBATION THEORY FOR ARBITRARY NORMALIZATION

We omit the tilde here, although we refer to trial wave functions. The bar on ψ indicates the change of the metric [see Eqs. (1.1) and (1.2)].

We start from the energy expectation value

$$E=\langle\overline{\psi}|D|\overline{\psi}\rangle/\langle\overline{\psi}|S|\overline{\psi}\rangle. \tag{A1}$$

The condition for its stationarity with respect to variations of ψ is

$$\langle\delta\overline{\psi}|D-ES|\overline{\psi}\rangle=0. \tag{A2}$$

We can expand this condition in powers of c^{-2} using the expansion (1.4b) for ψ:

$$\langle\delta\psi_0|D_0-E_0S_0|\psi_0\rangle=0, \tag{A3a}$$

$$\langle\delta\psi_0|D_0-E_0s_0|\psi_2\rangle+\langle\delta\psi_0|D_2-E_2S_0-E_0S_2|\psi_0\rangle+\langle\delta\psi_2|D_0-E_0S_0|\psi_0\rangle=0. \tag{A3b}$$

Since we expand ψ_2 and ψ_0 in the same basis, the last term in (A3b) vanishes if (A3a) is satisfied. Noting that (A3a) also holds with $\delta\psi_0$ replaced by ψ_2, we can put $\delta\psi_0=\psi_0$ in (A3b) and get

$$E_2\langle\psi_0|S_0|\psi_0\rangle=E_2=\langle\psi_0|D_2-E_0S_2|\psi_0\rangle=\langle\chi_0|V-E_0|\chi_0\rangle. \tag{A4}$$

Although we keep the normalization arbitrary, it is really no loss of generality to choose φ_0 normalized to unity i.e., to require

$$\langle\psi_0|S_0|\psi_0\rangle=\langle\varphi_0|\varphi_0\rangle=1. \tag{A5}$$

The next equation in the series of (A3) is, omitting already those terms that vanish if (A3a) and (A3b) are satisfied,

$$\langle\delta\psi_0|(D_0-E_0S_0)\psi_4+(D_2-E_0S_2-E_2S_0)\psi_2-(E_4S_0+E_2S_2)\psi_0\rangle=0. \tag{A6}$$

By an analogous argument as before we get

$$E_4\langle\psi_0|S_0|\psi_0\rangle=E_4=\langle\psi_0|D_2-E_0S_2-E_2S_0|\psi_2\rangle-E_2\langle\psi_0|S_2|\psi_0\rangle=\langle\chi_0|V-E_0|\chi_2\rangle-E_2[\langle\varphi_0|\varphi_2\rangle+\langle\chi_0|\chi_0\rangle]. \tag{A7}$$

We can consider the two special cases of intermediate normalization for the large component

$$\langle\varphi_0|\varphi_2\rangle=0\Rightarrow E_4=\langle\chi_0|V-E_0|\chi_2\rangle-E_2\langle\chi_0|\chi_0\rangle \tag{A8a}$$

and unitary normalization

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_2\rangle+\langle\chi_0|\chi_0\rangle=0\Rightarrow E_4=\langle\chi_0|V-E_0|\chi_2\rangle-\tfrac{1}{2}E_2\langle\chi_0|\chi_0\rangle. \tag{A8b}$$

Continuing in the hierarchy of (A3) we get

$$\langle\delta\psi_0|(D_0-E_0S_0)\psi_6+(D_0-E_0S_2-E_2S_0)\psi_4-(E_4S_0+E_2S_2)\psi_2+(E_0S_0+E_4S_2)\psi_0\rangle=0,$$

$$E_6=\langle\psi_0|D_2-E_0S_2-E_2S_0|\psi_4\rangle-\langle\psi_0|E_4S_0+E_2S_2|\psi_2\rangle-\langle\psi_0|E_4S_2|\psi_0\rangle. \tag{A9}$$

Using (A3b) (omitting, of course, the last term) for $\delta\psi_0=\psi_4$ and then (A6), or rather its complex conjugate with $\delta\psi_0=\psi_2$, the following reformulation of E_6 is possible:

$$E_6=\langle\chi_2|V-E_0|\chi_2\rangle-E_2\langle\varphi_2|\varphi_2\rangle-2E_4\ \mathrm{Re}\langle\varphi_2|\varphi_0\rangle - 2E_2\ \mathrm{Re}\langle\chi_2|\chi_0\rangle-E_4\langle\chi_0|\chi_0\rangle. \tag{A10}$$

For the two normalizations analogous to (A8a) and (A8b) we get

$$E_6=\langle\chi_2|V-E_0|\chi_2\rangle-E_2\{\langle\varphi_2|\varphi_2\rangle+2\ \mathrm{Re}\langle\chi_2|\chi_0\rangle\} - E_4\langle\chi_0|\chi_0\rangle, \tag{A11a}$$

$$E_6=\langle\chi_2|V-E_0|\chi_2\rangle-E_2\{\langle\varphi_2|\varphi_2\rangle+2\ \mathrm{Re}\langle\chi_2|\chi_0\rangle\}. \tag{A11b}$$

The Hylleraas functional corresponding to the stationarity condition (A6) is

$$F_4(\psi_2)=\langle\psi_2|D_0-S_0E_0|\psi_2\rangle+2\ \mathrm{Re}\langle\psi_2|D_2-E_0S_2 - E_2S_0|\psi_0\rangle-E_2\langle\psi_0|S_2|\psi_0\rangle. \tag{A12}$$

The last term in (A12) is independent of ψ_2, but in order to identify $F_4(\psi_2)$ with E_4 for the ψ_2 that makes $F_4(\psi_2)$ stationary, this term has to be kept.

APPENDIX B: BEHAVIOR OF THE EXACT WAVE FUNCTION FOR ARBITRARY BOUND STATES OF H-LIKE IONS FOR SMALL r

We make the ansatz

$$\varphi=g(r)\,\eta^m_\kappa(\vartheta,\varphi,s), \tag{B1a}$$

$$\chi=if(r)\,\eta^m_{-\kappa}(\vartheta,\varphi,s), \tag{B1b}$$

$$g(r)=r^\nu[a_0+a_1r+a_2r^2+O(r^3)]=\sum_k\ c^{-2k}g_{2k}(r), \tag{B2a}$$

$$f(r)=\frac{1}{c}\,r^\nu[b_0+b_1r+b_2r^2+O(r^3)]=\sum_k\ c^{-2k-1}f_{2k}(r), \tag{B2b}$$

where the η^m_κ are normalized functions of spin- and angular variables, with $-\kappa$ the eigenvalue of the operator K defined by

$$K=\beta(\boldsymbol{\sigma}\cdot\boldsymbol{l}+1) \tag{B3}$$

(note that in this context a_0 has not the meaning of the Bohr radius, but is just a coefficient).

Insertion of (B.1) into the Dirac equation leads to (for $e=1$, $\hbar=1$, $m=1$)

$$\left(\frac{Z}{r}+E\right)g-c\left(\frac{\partial}{\partial r}+\frac{1-\kappa}{r}\right)f=0, \tag{B4a}$$

$$c\left(\frac{\partial}{\partial r}+\frac{1+\kappa}{r}\right)g+\left(\frac{Z}{r}+2c^2+E\right)f=0. \tag{B4b}$$

For g and f given by (B.2) one gets after ordering in powers of r

$$b_0(1-\kappa+\nu)-a_0Z=0, \tag{B5a}$$

$$b_1(2-\kappa+\nu)-a_0E-a_1Z=0, \tag{B5b}$$

$$b_2(3-\kappa+\nu)-a_1E-a_2Z=0, \tag{B5c}$$

$$a_0(1+\kappa+\nu)+b_0Zc^{-2}=0, \tag{B6a}$$

$$a_1(2+\kappa+\nu)+[b_0(E+2c^2)+b_1Z]c^{-2}=0, \tag{B6b}$$

$$a_2(3+\kappa+\nu)+[b_1(E+2c^2)+b_2Z]c^{-2}=0. \tag{B6c}$$

The condition for the existence of a solution of (B5a) and (B6a) is that [32]

$$\nu=\sqrt{\kappa^2-Z^2c^{-2}}-1. \tag{B7}$$

We expand E and ν in powers of c^{-2}

$$E=E_0Z^2+E_2Z^4c^{-2}+E_4Z^6c^{-4}+\cdots, \tag{B8a}$$

$$\nu=|\kappa|-1-Z^2/(2c^2|\kappa|)-Z^4/(8c^4|\kappa|^3)+\cdots. \tag{B8b}$$

The expansion of r^ν is

$$r^\nu=r^{|\kappa|-1}\left\{1-\frac{Z^2\ln r}{2c^2|\kappa|}+\frac{Z^4}{8c^4}\left[\frac{\ln^2 r}{|\kappa|^2}-\frac{\ln r}{|\kappa|^3}\right]+O\left(\frac{Z^6}{c^6}\right)\right\}. \tag{B9}$$

This expansion is analytic in c^{-2} for all $r>0$. Solving (B5) and (B6) one gets b_0/a_0, a_1/a_0, etc., in powers of c^{-1} and inserting this into (B1) and (B2) the asymptotic expansion (in powers of r and $\ln r$) of the coefficients of the c^{-1} expansion of φ_{2k},χ_{2k} for $r\to0$ is obtained.

This asymptotic expansion is defined at least for all r on the positive real axis. An asymptotic expansion of a function $y(x)$ around $x=0$ means that coefficients c_k exist such that

$$y(x)=\sum_{k=n}^{n}\ c_ky_k(x)+O(x^{n+1}), \tag{B10a}$$

$$y_k(x)=O(x^k), \tag{B10b}$$

where (B10b) is a shorthand notation for

$$\lim_{x\to0}x^{1-k}y_k(x)=0. \tag{B10c}$$

Note that

$$x^\nu=O(x^0)\quad\text{if}\quad -1<\nu\leqslant0, \tag{B10d}$$

$$\ln x=O(x^0). \tag{B10e}$$

The normalization of φ is not determined by the condition (B4). We fix the normalization by choosing a_0 independent of c (for $\kappa>0$, a_1 instead of a_0). One can, e.g., add an arbitrary multiple of g_0 to g_2 and the same multiple of f_0 to g_2, which only affects the normalization of φ.

One must distinguish the two cases $\kappa<0$ and $\kappa>0$. We start with

$$\kappa<0,\quad \kappa=-l-1,\quad j=l+1/2 \tag{B11}$$

for which

$$b_0 = a_0[Z/(2|\kappa|) + Z^3/(8c^2|\kappa|^3) + O(Z^5/c^4)], \tag{B12a}$$

$$a_1 = -a_0\left[\frac{Z}{|\kappa|} + \frac{(2E_0|\kappa|^2+1)Z^3(1+4|\kappa|)}{4c^2|\kappa|^3(1+2|\kappa|)} + O\left(\frac{Z^5}{c^4}\right)\right], \tag{B12b}$$

$$b_1 = -a_0\left[\frac{(1-E_0|\kappa|)Z^2}{|\kappa|(1+2|\kappa|)} + O\left(\frac{Z^4}{c^2}\right)\right], \tag{B12c}$$

and finally

$$g_0 = a_0 r^l\left\{1 - \frac{Z}{l+1}\, r + O(r^2)\right\}, \tag{B13a}$$

$$g_2 = -a_0 Z^2 r^l\left\{\frac{1}{2(l+1)}\ln r + \frac{Z[(4l+5)[2E_0(l+1)^2+1]}{4(2l+3)(l+1)^3}\, r - \frac{Z}{2(l+1)^2}\, r\ln r + O(r^2)\right\}, \tag{B13b}$$

$$g_4 = -a_0 Z^4\left\{\frac{1}{8(l+1)^3}\ln r - \frac{1}{8(l+1)^2}\ln^2 r + O(r)\right\}, \tag{B13c}$$

$$f_0 = a_0 r^l Z\left\{\frac{1}{2(l+1)} + \frac{[E_0(l+1)-1]Z}{(l+1)(2l+3)}\, r + O(r^2)\right\}, \tag{B14a}$$

$$f_2 = a_0 r^l Z^3\left\{\frac{1}{8(l+1)^3} - \frac{1}{4(l+1)^2}\ln r + O(r)\right\}, \tag{B14b}$$

$$f_4 = a_0 r^l Z^5\left\{\frac{1}{16(l+1)^5} - \frac{1}{8(l+1)^4}\ln r + \frac{1}{16(l+1)^8}\ln^2 r + O(r)\right\}. \tag{B14c}$$

We now come to the case

$$\kappa > 0, \quad \kappa = l, \quad j = l - \tfrac{1}{2}. \tag{B15}$$

The counterpart of (B12a) is

$$b_0 = a_0\left[-\frac{2c^2\kappa}{Z} + \frac{Z}{2\kappa} + O\left(\frac{Z^3}{c^3}\right)\right]. \tag{B16}$$

This means that a_0 is smaller by $O(c^2)$ than b_0. If we want to express all coefficients in terms of a_0, we get a c^{-1} expansion with a leading term $O(c^2)$. We can avoid this by expressing everything in terms of b_0, or of a_1, which are of the same order. Then

$$a_0 = a_1\left[\frac{2l+1}{4l}\,\frac{Z}{c^2} + \frac{4l^2+1-E_0(2l^2+4l^3)}{16l^3}\,\frac{Z^3}{c^4} + O\left(\frac{Z^5}{c^6}\right)\right], \tag{B.17a}$$

$$b_0 = -a_1\left\{\frac{2l+1}{2} - \frac{Z^2}{4c^2 l}[1-2l+E_0(l+2l^2)] + O\left(\frac{Z^4}{c^4}\right)\right\}, \tag{B.17b}$$

$$b_1 = a_1\left[Z + \frac{Z^3}{c^2}\,\frac{2+E_0(2l+1)}{4l} + O\left(\frac{Z^5}{c^4}\right)\right], \tag{B.17c}$$

$$a_2 = -a_1\left\{\frac{Z}{l+1} + \frac{(l+3)+E_0(1+6l+5l^2)}{4l(l+1)^2}\,\frac{Z^3}{c^2} + O\left(\frac{Z^5}{c^4}\right)\right\}, \tag{B.17d}$$

$$b_2 = -a_1\left\{\frac{1-E_0(l+1)}{2(l+1)}\, Z^2 + O\left(\frac{Z^4}{c^2}\right)\right\}. \tag{B.17e}$$

For the same order in r as before we now also need a_2 and b_2. The final result is

$$g_0 = a_1 r^l\left[1 - \frac{Z}{l+1}\, r + O(r^2)\right], \tag{B.18a}$$

$$g_2 = a_1 r^l\left[\frac{2l+1}{4l}\, Z r^{-1} - \frac{Z^2}{2l}\ln r + O(r)\right], \tag{B.18b}$$

$$g_4 = a_1 r^l Z^3\left\{\frac{4l^2+1+2E_0 l^2(2l-1)}{16l^2}\, r^{-1} - \frac{(2l+1)}{8l^2}\, r^{-1}\ln r - \frac{Z}{8l^3}\ln r + \frac{Z}{8l^2}\ln^2 r + O(r)\right\}, \tag{B.18c}$$

$$f_0 = -a_1 r^l\left[\frac{2l+1}{2}\, r^{-1} - Z + O(r)\right], \tag{B.19a}$$

$$f_2 = a_1 r^l\left[\frac{1-2l+E_0(2l^2+1)}{4l}\, Z^2 r^{-1} + \frac{2l+1}{4l}\, Z^2 r^{-1}\ln r + O(r^0)\right], \tag{B.19b}$$

$$f_4 = a_1 r^l Z^4 \times\left\{\frac{-4l^2+1-E_0 l^2 - 2E_0^2 l^3(2l+1) + 4E_2 l^3(2l+1)}{16l^3} \times r^{-1} + \frac{4l^2+1-2E_0 l^2(2l+1)}{16l^3}\, r^{-1}\ln r - \frac{2l+1}{16l^2}\, r^{-1}\ln^2 r + O(r^0)\right\}. \tag{B.19c}$$

All those expressions in (B13) or (B18) that involve E_0 or E_2 are specific to a particular state, while those where E_0 or E_2 do not enter are universal and hold for all states (with either $\kappa<0$ or $\kappa>0$) and are even valid—*mutandis mutatis*—for molecules.

APPENDIX C: KINETIC BALANCE AS CHANGE OF THE METRIC

As first proposed by the present author [17] and worked out in detail by Dyall [33] the kinetic balance condition can be formulated by defining an auxilary large-component-type wave function φ', related to the small component χ of $\psi=(\varphi,\chi)$,

$$\chi=\frac{1}{2mc}\,\boldsymbol{\sigma}\cdot\mathbf{p}\varphi'. \tag{C1}$$

In terms of φ and φ' the Dirac equation can be written as

$$\begin{pmatrix} V & T \\ T & -T+\dfrac{1}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p} \end{pmatrix}\begin{pmatrix}\varphi\\ \varphi'\end{pmatrix}=E\begin{pmatrix}1 & 0\\ 0 & T/(2mc^2)\end{pmatrix}. \tag{C2}$$

This is another Dirac equation with modified metric, and an alternative to (1.2). We define

$$\overline{D_0}=\begin{pmatrix}V & T\\ T & -T\end{pmatrix},\quad \overline{D_2}=\begin{pmatrix}0 & 0\\ 0 & \dfrac{1}{4m^2}\,\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}\end{pmatrix}, \tag{C3a}$$

$$\overline{S_0}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\quad \overline{S_2}=\begin{pmatrix}0 & 0\\ 0 & T/2m\end{pmatrix}, \tag{C3b}$$

$$\overline{\psi}=\begin{pmatrix}\varphi\\ \varphi'\end{pmatrix}. \tag{C3c}$$

Note the changed meaning of $\overline{\psi}$. Then (C2) can be written as

$$\left(\overline{D_0}+\frac{1}{c^2}\,\overline{D_2}\right)\overline{\psi}=E\left(\overline{S_0}+\frac{1}{c^2}\,\overline{S_2}\right)\overline{\psi} \tag{C4}$$

and we can take over the entire formalism of direct perturbation theory [1], just replacing D_0, S_0, etc., by $\overline{D_0}$, $\overline{S_0}$, etc. In particular, for $\langle\psi_0|S_0|\psi_0\rangle=1$

$$(\overline{D_0}-E_0\overline{S_0})\,\overline{\psi_0}=0, \tag{C5a}$$

$$(\overline{D_2}-E_0\overline{S_2}-E_2\overline{S_0})\,\overline{\psi_0}+(\overline{D_0}-E_0\overline{S_0})\,\overline{\psi_2}=0. \tag{C5b}$$

$$E_0=\langle\overline{\psi_0}|\overline{D_0}|\overline{\psi_0}\rangle=\langle\varphi_0|V|\varphi_0\rangle+2\,\mathrm{Re}\langle\varphi_0|T|\varphi_0'\rangle-\langle\varphi_0'|T|\varphi_0'\rangle, \tag{C6a}$$

$$E_2=\langle\overline{\psi_0}|\overline{D_2}-E_0\overline{S_2}|\overline{\psi_0}\rangle=\frac{1}{4m^2}\,\langle\varphi_0'|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_0'\rangle, \tag{C6b}$$

$$\begin{aligned}E_4&=\langle\overline{\psi_0}|\overline{D_2}-E_0\overline{S_2}-E_2\overline{S_0}|\overline{\psi_2}\rangle-E_2\langle\overline{\psi_0}|\overline{S_0}|\overline{\psi_2}\rangle\\ &=\frac{1}{4m^2}\,\langle\varphi_0'|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}\rangle|\varphi_2'\rangle-E_2\left\{\left\langle\varphi_0'\middle|\frac{T}{2m}\middle|\varphi_0'\right\rangle+\langle\varphi_0|\varphi_2\rangle\right\},\end{aligned} \tag{C6c}$$

$$\begin{aligned}E_6&=\langle\psi_0|\overline{D_2}-E_0\overline{S_2}-E_2\overline{S_0}|\psi_4\rangle-E_2\langle\psi_0|\overline{S_2}|\psi_2\rangle\\ &\quad-E_4\{\langle\psi_0|\overline{S_2}|\psi_0\rangle+\langle\psi_0|\overline{S_0}|\psi_2\rangle\}\\ &=\langle\psi_2|\overline{D_2}-E_0\overline{S_2}|\psi_2\rangle-E_2\{2\,\mathrm{Re}\langle\psi_0|\overline{S_2}|\psi_2\rangle\\ &\quad+\langle\psi_2|\overline{S_0}|\psi_2\rangle\}-E_4\{\langle\psi_0|\overline{S_2}|\psi_0\rangle+2\,\mathrm{Re}\langle\psi_0|\overline{S_0}|\psi_2\rangle\}\\ &=\frac{1}{4m^2}\,\langle\varphi_2'|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2'\rangle-E_2\left\{\frac{1}{m}\,\langle\varphi_0'|T|\varphi_2'\rangle+\langle\varphi_2|\varphi_2'\rangle\right\}-E_4\left\{\frac{1}{2m}\,\langle\varphi_0'|T|\varphi_0'\rangle+2\,\mathrm{Re}\langle\varphi_0|\varphi_2\rangle\right\}.\end{aligned} \tag{C6d}$$

The unperturbed equation (C5a) is in component form,

$$V\varphi_0+T\varphi_0'=E_0\varphi_0, \tag{C7a}$$

$$T\varphi_0-T\varphi_0'=0. \tag{C7b}$$

From (C7b) we conclude that φ_0 and φ_0' can only differ by a function for which

$$T(\varphi_0-\varphi_0')=0, \tag{C8a}$$

i.e., which is essentially the solution of a Laplace equation. No nontrivial solution of (C8a) is square integrable; we hence conclude that

$$\varphi_0'=\varphi_0, \tag{C8b}$$

$$H_0\varphi_0=(V+T)\varphi_0=E_0\varphi_0, \tag{C9}$$

i.e., (C5a) is equivalent to the Schrödinger equation, in the same sense (i.e., with the same reservations) in which the Lévy-Leblond equation is equivalent to the Schrödinger equation [1,2,13]. Equation (C5b) is in component form

$$-E_2\varphi_0+(V-E_0)\varphi_2+T\varphi_2'=0, \tag{C10a}$$

$$\frac{1}{4m^2}\,\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_0'+T\varphi_2-T\varphi_2'=0. \tag{C10b}$$

Substitution of (C10b) into (C10a) leads to (noting that $\varphi_0=\varphi_0'$)

$$(H_0-E_0)\varphi_2=\frac{1}{4m^2}\,\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_0+E_2\varphi_0, \tag{C11}$$

which can be solved for φ_2. One can obtain $\varphi_2'-\varphi_2$ from the Poisson-like equation (C10b). φ_2' is required for the construction of E_4 according to (C6c). Actually only $\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_2'$ is needed and one can replace (C10b) by

$$\frac{1}{2m}\,(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}\varphi_0+\boldsymbol{\sigma}\cdot\mathbf{p}(\varphi_2-\varphi_2')=0, \tag{C12}$$

such that

$$E_4=\frac{1}{4m^2}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2'\rangle - E_2\left\{\left\langle\varphi_0\left|\frac{T}{2m}\right|\varphi_0\right\rangle+\langle\varphi_0|\varphi_2\rangle\right\}$$
$$=\frac{1}{4m^2}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2\rangle+\frac{1}{8m^3}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V - E_0)^2\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_0\rangle+E_2\left\{\left\langle\varphi_0\left|\frac{T}{2m}\right|\varphi_0\right\rangle+\langle\varphi_0|\varphi_2\rangle\right\} \tag{C13}$$

which is, of course, the same result as from classical direct perturbation theory.

Let us now consider the stationary variant of this approach. For ψ normalized to unity, i.e.,

$$\langle\varphi|\varphi\rangle+\frac{1}{2mc^2}\langle\varphi'|T|\varphi'\rangle=1, \tag{C14}$$

the energy expectation value is

$$\langle D\rangle=\langle\varphi|V|\varphi\rangle+2\,\mathrm{Re}\langle\varphi|T|\varphi'\rangle-\langle\varphi'|T|\varphi'\rangle+\frac{1}{4m^2c^2}\langle\varphi'|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi'\rangle. \tag{C15}$$

Expanding this in powers of c^{-2} we get

$$\langle D\rangle_0=\langle\varphi_0|H_0\varphi_0\rangle-\langle\varphi_0|T|\varphi_0\rangle+2\,\mathrm{Re}\langle\varphi_0|T|\varphi_0'\rangle-\langle\varphi_0'|T|\varphi_0'\rangle$$
$$=\langle\varphi_0|H_0|\varphi_0\rangle-\langle\varphi_0-\varphi_0'|T|\varphi_0-\varphi_0'\rangle. \tag{C15a}$$

Evidently this expection value is not bounded from below, since the first term is negative and the second non-negative. For arbitrary φ_0 the maximum of (C15a) with respect to variation of $\varphi_0-\varphi_0$ is reached for $\varphi_0=\varphi_0$. Imposing $\varphi_0=\varphi_0'$ the minimum is reached for φ_0 an eigenfunction of H_0. We further get (using $\varphi_0=\varphi_0'$)

$$\langle D\rangle_2=2\,\mathrm{Re}\{\langle\varphi_0|V|\varphi_2\rangle+\langle\varphi_0|T|\varphi_2'\rangle+\langle\varphi_2|T|\varphi_0'\rangle\}-2\,\mathrm{Re}\langle\varphi_0'|T|\varphi_2'\rangle+\frac{1}{4m^2}\langle\varphi_0'|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_0'\rangle$$
$$=2\,\mathrm{Re}\langle\varphi_0|H_0|\varphi_2\rangle+\frac{1}{4m^2}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_0\rangle, \tag{C15b}$$

$$\langle D\rangle_4=2\,\mathrm{Re}\langle\varphi_0|V|\varphi_4\rangle+2\,\mathrm{Re}\{\langle\varphi_0|T|\varphi_4'\rangle+\langle\varphi_4|T|\varphi_0\rangle\}-2\,\mathrm{Re}\langle\varphi_0'|T|\varphi_4'\rangle+\langle\varphi_2|V|\varphi_2\rangle+2\,\mathrm{Re}\langle\varphi_2|T|\varphi_2'\rangle-\langle\varphi_2'|T|\varphi_2\rangle+\frac{1}{4m^2}2\,\mathrm{Re}\langle\varphi_0'|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2'\rangle$$
$$=2\,\mathrm{Re}\langle\varphi_0|H_0|\varphi_4\rangle+\langle\varphi_2|H_0|\varphi_2\rangle-\langle\varphi_2-\varphi_2'|T|\varphi_2-\varphi_2'\rangle+\frac{1}{2m^2}\mathrm{Re}\langle\varphi_0'|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2'\rangle. \tag{C15c}$$

The stationary condition for (C15a) (taking $\varphi_0=\varphi_0'$) subject to the normalization condition $\langle\varphi_0|\varphi_0\rangle=1$ is

$$\langle\delta\varphi|H_0-E_0|\varphi_0\rangle=0. \tag{C16a}$$

This together with the next normalization conditions

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_2\rangle+\frac{1}{2m}\langle\varphi_0'|T|\varphi_0'\rangle=0, \tag{C17a}$$

$$2\,\mathrm{Re}\langle\varphi_0|\varphi_4\rangle+\langle\varphi_2|\varphi_2\rangle+\frac{1}{m}\mathrm{Re}\langle\varphi_2|T|\varphi_0\rangle=0, \tag{C17b}$$

allows us to rewrite (C15b) and (C15a) as

$$\langle D\rangle_2=\frac{1}{4m^2}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_0\rangle, \tag{C18a}$$

$$\langle D\rangle_4=\langle\varphi_2|H_0-E_0|\varphi_2\rangle-\langle\varphi_2-\varphi_2'|T|\varphi_2-\varphi_2'\rangle+\frac{1}{2m^2}\mathrm{Re}\langle\varphi_0|\boldsymbol{\sigma}\cdot\mathbf{p}(V-E_0)\boldsymbol{\sigma}\cdot\mathbf{p}|\varphi_2'\rangle. \tag{C18b}$$

The stationarity of (C18b) for arbitrary variations subject to the normalization condition leads to (C10).

[1] W. Kutzelnigg, Z. Phys. D **11**, 15 (1989).
[2] W. Kutzelnigg, Z. Phys. D **15**, 27 (1990).
[3] J. D. Morrison and R. E. Moss, Mol. Phys. **41**, 491 (1980).
[4] G. L. Sewell, Proc. Cambridge Philos. Soc. **45**, 631 (1949).
[5] F. Gesztesy, H. Grosse, and B. Thaller, Phys. Lett. **116B**, 155 (1982); Ann. Inst. Henri Poincaré **40**, 159 (1984); Adv. Appl. Math. **6**, 159 (1985); Phys. Rev. D **30**, 2189 (1984); S. Gesztesy, B. Thaller, and H. Grosse, Phys. Rev. Lett. **50**, 625 (1983); see also B. Thaller, *The Dirac Equation* (Springer, Berlin, 1992).
[6] A. Rutkowski, J. Phys. B **19**, 149 (1986); **19**, 3431 (1986).
[7] A. Rutkowski, J. Phys. B **19**, 3443 (1986); K. Jankowski and A. Rutkowski, Phys. Scr. **36**, 464 (1987); A. Rutkowski, K. Jankowski, and B. Mikielewicz, J. Phys. B **21**, L147 (1988).
[8] R. A. Moore, Can. J. Phys. **53**, 1240 (1975); **53**, 1247 (1975); **53**, 1251 (1975); R. A. Moore and T. C. Scott, *ibid*. **64**, 297 (1980).
[9] Ch. Chang, M. Pelissier, and Ph. Durand, Phys. Scr. **34**, 314 (1987).
[10] A. J. Sadlej and J. G. Snijders, Chem. Phys. Lett. **229**, 435 (1994); A. J. Sadlej, J. G. Snijders, E. van Lenthe, and E. J. Baerends, J. Chem. Phys. **102**, 1758 (1995).
[11] J. M. Lévy-Leblond, Commun. Math. Phys. **6**, 286 (1967).
[12] R. Franke and W. Kutzelnigg, Chem. Phys. Lett. **199**, 561 (1992); E. Ottschofski and W. Kutzelnigg, J. Chem. Phys. **102**, 1752 (1995).

[13] W. Kutzelnigg, Theor. Chim. Acta **83**, 263 (1992).
[14] R. E. Stanton and S. J. Havrilak, J. Chem. Phys. **76**, 735 (1982).
[15] Y. S. Lee and A. P. McLean, J. Chem. Phys. **76**, 735 (1982).
[16] P. J. C. Aerts and W. C. Nieuwpoort, Int. J. Quantum Chem. S **19**, 267 (1986).
[17] W. Kutzelnigg, Int. J. Quantum Chem. **25**, 107 (1984).
[18] I. P. Grant, J. Phys. B **1986**, 3187.
[19] R. N. Hill and C. Krauthauser, Phys. Rev. Lett. **72**, 2151 (1994).
[20] W. H. E. Schwarz and H. Wallmeier, Mol. Phys. **46**, 1045 (1982).
[21] W. H. E. Schwarz and E. Wechsel-Trakowski, Chem. Phys. Lett. **85**, 94 (1982).
[22] F. Mark and W. H. E. Schwarz, Phys. Rev. Lett. **48**, 673 (1982).
[23] W. Kutzelnigg, R. Franke, E. Ottschofski, and W. Klopper, in *New Challenges in Computational Quantum Chemistry*, edited by P. C. J. Aert *et al.* (University of Groningen, Groningen, 1994).
[24] W. Kutzelnigg, Theor. Chim. Acta **73**, 173 (1988).
[25] W. Kutzelnigg, in *Aspects of Many-Body Effects in Molecules and Extended Systems*, edited by D. Mukherjee, Lecture Notes in Chemistry Vol. 50 (Springer, Berlin, 1989).
[26] S. P. Goldman, Phys. Rev. **31**, 3541 (1985).
[27] Z. Chen, G. Fonte and S. P. Goldman, Phys. Rev. A **50**, 3838 (1994).
[28] M. Karplus and H. J. Kolker, J. Chem. Phys. **38**, 1263 (1963); H. J. Kolker and M. Karplus, *ibid*. **39**, 2011 (1963).
[29] E. Ottschofski and W. Kutzelnigg (unpublished).
[30] See, e.g., R. Hofstadter, Rev. Mod. Phys. **28**, 214, (1956). D. G. Ravenhall, Rev. Mod. Phys. **30**, 214 (1958); J. P. Desclaux, Comput. Phys. Commun. **21**, 31 (1975).
[31] J. D. Talman, Phys. Rev. Lett. **57**, 1091 (1986).
[32] See standard textbooks, e.g., E. M. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
[33] K. Dyall, J. Chem. Phys. **100**, 2118 (1994).