Eq. (C-3) to get: For the ground state,

$$u_1{}^0(r)=-\frac{1}{6\pi}\psi_1{}^0(0)e^{-\beta_1 r}[-(1/r)+2\beta_1\ln(2\beta_1 r)+\beta_1(2\gamma-5)+2\beta_1{}^2 r], \quad \text{(C-5a)}$$

$$v_1{}^0(r)=-\frac{1}{6\pi}\psi_1{}^0(0)\frac{\beta_1}{2\kappa}e^{-\beta_1 r}[-(3/r)+2\beta_1\ln(2\beta_1 r)+\beta_1(2\gamma-7)+2\beta_1{}^2 r], \quad \text{(C-5b)}$$

and for the excited state

$$u_2(r)=-\frac{1}{6\pi}\psi_2{}^0(0)e^{-\beta_2 r}[-(1/r)+4\beta_2\ln(2\beta_2 r)+\beta_2(4\gamma-3)-4\beta_2{}^2 r\ln 2\beta_2 r+\beta_2{}^2 r(13-4\gamma)-2\beta_2{}^3 r^2], \quad \text{(C-6a)}$$

$$v_2(r)=-\frac{1}{6\pi}\psi_2{}^0(0)\frac{\beta_2}{2\kappa}e^{-\beta_2 r}[-(6/r)+8\beta_2\ln(2\beta_2 r)+\beta_2(8\gamma-12)-4\beta_2{}^2 r\ln 2\beta_2 r+\beta_2{}^2 r(17-4\gamma)-2\beta_2{}^3 r^2], \quad \text{(C-6b)}$$

where $\gamma=2.5777\cdots$ is Euler's constant.

The integrals involving u and v that are required are

$$\Delta\int_0^\infty \psi^0(r)u^0(r)\frac{dr}{|\psi^0(0)|^2}=-\frac{1}{6\pi}(\tfrac{3}{2}-\ln 2), \quad \text{(C-7)}$$

$$\Delta\int_0^\infty [\psi^0(r)v^0(r)+\eta^0(r)u^0(r)]\times\frac{dr}{|\psi^0(0)|^2}=\frac{Z\alpha}{3\pi}(\ln 2-17/16), \quad \text{(C-8)}$$

$$\Delta\int_0^\infty \psi^0(r)u^0(r)\frac{dr}{|\psi^0(0)|^2}=\frac{1}{8\pi\kappa Z\alpha}, \quad \text{(C-9)}$$

$$\Delta\int_0^\infty u^0(r)\left(r\frac{d^2}{dr^2}+\frac{d}{dr}\right)\psi^0(r)\times\frac{dr}{|\psi^0(0)|^2}=-\frac{\kappa Z\alpha}{6\pi}(\tfrac{9}{16}-\ln 2). \quad \text{(C-10)}$$

PHYSICAL REVIEW VOLUME 107, NUMBER 4 AUGUST 15, 1957

Convergent Schrödinger Perturbation Theory

GERALD SPEISMAN
Florida State University, Tallahassee, Florida
(Received February 18, 1957)

A formulation of Schrödinger perturbation theory is developed that gives a unified treatment of nondegenerate and degenerate cases, is unique, and has a nonzero radius of convergence under very general conditions. Two alternative procedures are given for finding perturbed eigenvectors, one of which is simpler for the nondegenerate case or for small finite degeneracy, the other simpler for infinite or large finite degeneracy. The low-order terms in the perturbation expansions of quantities used in applications are given. The perturbation theory formulated in this paper has the following advantages over the conventional Schrödinger and Brillouin-Wigner perturbation theories: (i) When the convergence criterion is satisfied, bounds on the error made in replacing an appropriate infinite perturbation series by its first n terms can be obtained. (ii) For the case of degeneracy, the conventional Schrödinger perturbation theory can break down under conditions to which the convergence of the perturbation theory developed in this paper are insensitive. (iii) There is no implicit dependence on the eigenvalue, such as appears in the Brillouin-Wigner perturbation theory. (iv) For the case of degeneracy, statistical information about the distribution of certain eigenvalues can be obtained without finding the individual eigenvalues. (v) The theory is applicable to a wider class of problems than the conventional Schrödinger and Brillouin-Wigner perturbation theories.

I. INTRODUCTION

THE conventional Schrödinger perturbation theory is concerned with finding the eigenvectors and eigenvalues in a Hilbert space of a Hermitean operator of the form $\mathbf{H}_0+\epsilon\mathbf{V}$ as a power series in the real parameter ϵ.[1] We want to go into this theory in some detail to point out the relation between it and the theory developed in this paper. The advantages of the latter will be pointed out as we go along. To avoid difficulties of a purely mathematical nature, we will assume that the Hermitean operator $\mathbf{H}_0$ possesses a complete orthonormal set of eigenvectors $\xi_0, \xi_1, \cdots, \xi_n, \cdots$ with eigenvalues $E_0, E_1, \cdots, E_n, \cdots$, respectively. We fix our attention on the eigenvalue E_0, and require that, if $E_n\neq E_0$, then in fact $|E_n-E_0|>\delta>0$ for some fixed δ. In other words, E_0 is an isolated point in the spectrum of $\mathbf{H}_0$.

Let $\mathbf{P}$ be the projection operator onto the closed linear manifold M_{E_0} of all solutions ψ_0 of the equation $\mathbf{H}_0\psi_0=E_0\psi_0$. Then $\mathbf{P}\xi_n=\xi_n$ for $E_n=E_0$, and $\mathbf{P}\xi_n=0$ for $E_n\neq E_0$, and thus $\mathbf{H}_0\mathbf{P}\xi_n=E_0\mathbf{P}\xi_n=E_n\mathbf{P}\xi_n=\mathbf{P}\mathbf{H}_0\xi_n$, so

[1] E. Schrödinger, Ann. Physik 80, 437 (1926).

that, since the ξ_n's form a complete set,

$$\mathbf{H}_0\mathbf{P}=\mathbf{P}\mathbf{H}_0=E_0\mathbf{P}.$$

We define the operator $(\mathbf{1}-\mathbf{P})/(\mathbf{H}_0-E_0)$ as follows:

$$\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\xi_n=\begin{cases}(E_n-E_0)^{-1}\xi_n & \text{for } E_n\neq E_0,\\ 0 & \text{for } E_n=E_0.\end{cases}$$

Note that

$$\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}=\frac{1}{\mathbf{H}_0-E_0}(\mathbf{1}-\mathbf{P})=(\mathbf{1}-\mathbf{P})\frac{1}{\mathbf{H}_0-E_0}(\mathbf{1}-\mathbf{P})=\lim_{z\to 0}(\mathbf{1}-\mathbf{P})\frac{1}{\mathbf{H}_0-E_0-z},$$

where z is a complex number, and that

$$\left|\left(\xi_n,\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\xi_n\right)\right|<\delta^{-1}.$$

Although $1/(\mathbf{H}_0-E_0)$ is a singular operator, $(\mathbf{1}-\mathbf{P})/(\mathbf{H}_0-E_0)$ is not.

We proceed as follows: In the eigenvector equation $(\mathbf{H}_0+\epsilon\mathbf{V})\psi=E\psi$ we let $E=E_0+\epsilon\Lambda$ and $\psi=A(\psi_0+\psi_1)$, where $\mathbf{P}\psi_0=\psi_0$, $\mathbf{P}\psi_1=0$, and A is a normalization constant. Note that we are requiring that $E\to E_0$ as $\epsilon\to 0$. The eigenvector equation can then be split into two component equations as follows:

$$\mathbf{PVP}\psi_0+\mathbf{PV}\psi_1=\Lambda\psi_0, \qquad \text{(i)}$$

$$[\mathbf{H}_0-E_0+\epsilon(\mathbf{1}-\mathbf{P})\mathbf{V}(\mathbf{1}-\mathbf{P})-\epsilon\Lambda]\psi_1=-\epsilon(\mathbf{1}-\mathbf{P})\mathbf{V}\psi_0. \qquad \text{(ii)}$$

Consider the case in which the zeroth order eigenvalue E_0 is nondegenerate, i.e., $E_n\neq E_0$ for $n\neq 0$. By a suitable choice of the normalization constant A, we can choose $\psi_0=\xi_0$. Then $\mathbf{PVP}\psi_0=(\xi_0,\mathbf{V}\xi_0)\xi_0$, and $\mathbf{PV}\psi_1=(\xi_0,\mathbf{V}\psi_1)\xi_0$. Equation (i) then reduces to

$$(\xi_0,\mathbf{V}\xi_0)+(\xi_0,\mathbf{V}\psi_1)=\Lambda. \qquad \text{(iii)}$$

Substituting (iii) into (ii) we get

$$(\mathbf{H}_0-E_0)\psi_1=\epsilon[(\xi_0,\mathbf{V}\xi_0)\psi_1-(\mathbf{1}-\mathbf{P})\mathbf{V}(\mathbf{1}-\mathbf{P})\psi_1+(\xi_0,\mathbf{V}\psi_1)\psi_1-(\mathbf{1}-\mathbf{P})\mathbf{V}\xi_0]. \qquad \text{(iv)}$$

Equations (iii) and (iv) are equivalent to (i) and (ii) or to (ii) and (iii). The perturbation theory we will develop in this paper reduces to Eqs. (iii) and (iv) when E_0 is nondegenerate. More specifically, Eqs. (14), (42), and (46) reduce to (iii), while Eq. (11) reduces to (iv). Substituting a power series expansion for ψ_1 into (iv) leads to a recursion relation for the coefficients in the power series which is a specialization of Eq. (34) for nondegenerate E_0.

The conclusion we must come to for nondegenerate E_0 is that, as far as obtaining the coefficients of the power series of ψ and E are concerned, the methods of this paper, as exmplified by Eqs. (iii) and (iv), are about equivalent to direct substitution into Eqs. (ii) and (iii), as far as simplicity in computing ψ and E to a given order are concerned. Of course, both methods lead to the same coefficients. However, the perturbation theory we will develop will be proven to be convergent under very general conditions, from which it follows that the conventional nondegenerate Schrödinger perturbation theory is convergent under these conditions. Furthermore, when the convergence criterion is satisfied, we can obtain bounds on the error which is made when ψ or E is replaced by its nth order perturbation approximation. Such bounds are very useful in the physical problems to which Schrödinger perturbation theory are applied. The remarks of the last two sentences apply equally well to the case of degenerate E_0, for the perturbation theory developed in this paper.

Before proceeding to the case of degenerate zeroth-order eigenvalue E_0, we give an alternative procedure for developing the nondegenerate Schrödinger perturbation theory. This procedure employs the methods used in the Brillouin-Wigner perturbation theory.[2] Multiplying (ii) by

$$(\mathbf{1}-\mathbf{P})\cdot[\mathbf{H}_0-E_0+\epsilon(\mathbf{1}-\mathbf{P})\mathbf{V}(\mathbf{1}-\mathbf{P})-\epsilon\Lambda]^{-1}=\sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\cdot\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\cdot\epsilon^n,$$

and noting that we can choose $\psi_0=\xi_0$, we get

$$\psi_1=\sum_{n=0}^{\infty}(-1)^{n+1}\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\cdot\epsilon^{n+1}\xi_0. \qquad \text{(v)}$$

Substitution of (v) into (iii) gives

$$\Lambda=(\xi_0,\mathbf{V}\xi_0)+\sum_{n=0}^{\infty}(-1)^{n+1}\cdot\left(\xi_0,\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)\cdot\epsilon^{n+1}. \qquad \text{(vi)}$$

If we substitute a power series expansion for Λ into (vi), we can obtain a recursion relation for the coefficients in the power series. Substitution of this power series for Λ into (v) will give a power series for ψ_1. However, because of the complicated implicit dependence of Eqs. (v) and (vi) on Λ, the procedure outlined above will be a much more complicated one for obtaining the coefficients in the power series for ψ and E than direct substitution into either Eqs. (ii) and (iii) or Eqs. (iii) and (iv).[3] Since the coefficients will be the same by any method, Eqs. (v) and (vi) are not the best starting points for the nondegenerate Schrödinger perturbation theory. The reason we consider (v) and (vi)

[2] L. Brillouin, J. phys. radium **3**, 373 (1932); E. P. Wigner, Math. u. naturw. Anz. ungar. Akad. Wiss. **50**, 475 (1935).

[3] Of course, if we want to express E_0 as a perturbation expansion in terms of E, just the converse is true. However, we are considering problems in which E_0 is given and E is to be found.

at all, is that they are readily generalized for the case of degenerate E_0, whereas we shall see that direct substitution of power series for ψ and Λ in (i) and (ii) can break down under conditions to which the convergence of both the theory developed in this paper and the theory based on generalizations of (v) and (vi) are insensitive.

We now turn to the case in which the zeroth-order eigenvalue E_0 is degenerate. Equations (iii) and (iv) are still valid if we substitute ψ_0 for ξ_0, but they provide no means for determining ψ_0. The perturbation theory to be developed in this paper provides generalizations of Eqs. (iii) and (iv) from which ψ_0, ψ_1 and E can be determined. Furthermore, we will now show that the method of direct substitution of power series for ψ and Λ in (i) and (ii) can break down under conditions to which the convergence of the theory developed in this paper are insensitive. We take as an example the case in which $E_1=E_0$, but $E_n\neq E_0$ for $n\geq 2$. Then the linear manifold M_{E_0} is spanned by the orthonormal vectors ξ_0 and ξ_1. These vectors can be chosen such that $(\xi_1,\mathbf{V}\xi_0)=0$, and we so choose them. We shall assume that $(\xi_1,\mathbf{V}\xi_1)\neq(\xi_0,\mathbf{V}\xi_0)$. By applying the methods given in Sec. VI, we obtain the following expression for the perturbed energy levels $E^{\pm}$:

$$2E^{\pm}=2E_0+\epsilon(a+b)-\epsilon^2(c+d)+\epsilon^3(f+g)+O(\epsilon^4)\pm\epsilon[(a-b)^2-2\epsilon(a-b)(c-d)+\epsilon^2\{(c-d)^2+2(a-b)(f-g)+4w^*w\}+O(\epsilon^3)]^{\frac{1}{2}}, \quad \text{(vii)}$$

where w is complex, a, b, c, d, f, g are real, and they are given by

$$a=(\xi_1,\mathbf{V}\xi_1), \quad b=(\xi_0,\mathbf{V}\xi_0),$$

$$c=\left(\xi_1, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_1\right), \quad d=\left(\xi_0, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right),$$

$$f=\left(\xi_1, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_1\right)-(\xi_1,\mathbf{V}\xi_1)\cdot\left(\xi_1, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\xi_1\right),$$

$$g=\left(\xi_0, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)-(\xi_0,\mathbf{V}\xi_0)\cdot\left(\xi_0, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\xi_0\right),$$

$$w=\left(\xi_1, \mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right),$$

where

$$\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}=\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\right]^2.$$

By neglecting terms of order ϵ^4 in (vii), we make an error whose size depends on the rapidity of convergence of the perturbation theory developed in this paper. The more rapid the convergence, the smaller the error. However, this error is insensitive to small changes in $(\xi_1,\mathbf{V}\xi_1)-(\xi_0,\mathbf{V}\xi_0)$, when $(\xi_1,\mathbf{V}\xi_1)+(\xi_0,\mathbf{V}\xi_0)$ and all matrix elements of $\mathbf{V}$ except $(\xi_0,\mathbf{V}\xi_0)$ and $(\xi_1,\mathbf{V}\xi_1)$ are held fixed. On the other hand, if we attempt to expand E^+ or E^- in a power series in ϵ, the expansion will break down for small enough but nonzero values of $(\xi_1,\mathbf{V}\xi_1)-(\xi_0,\mathbf{V}\xi_0)$, for values of ϵ for which the expansions used in (vii) are rapidly convergent, Specifically

$$E^+=E_0+\epsilon a-\epsilon^2c+\epsilon^3[f+w^*w/(a-b)]+O(\epsilon^4),$$
$$E^-=E_0+\epsilon b-\epsilon^2d+\epsilon^3[g-w^*w/(a-b)]+O(\epsilon^4).$$

Note that w is independent of the matrix elements $(\xi_1,\mathbf{V}\xi_1)$ and $(\xi_0,\mathbf{V}\xi_0)$, so that the term $w^*w/(a-b)$ will, in general, cause trouble when $a-b$ is small enough.

In general, when E_0 is degenerate to any finite order N, the method of direct substitution of power series for ψ and Λ into (i) and (ii) will have poor convergence properties relative to that of the perturbation theory developed in this paper whenever a pair of eigenvalues of $\mathbf{PVP}$ within the linear manifold M_{E_0} are close to each other in value, yet are unequal. The difficulties will generally affect both the ψ's and Λ's corresponding to closely spaced eigenvalues. Similar remarks hold for the case of infinite degeneracy.

An alternative to direct substitution of power series for ψ and Λ in (i) and (ii) is to use the methods of the Brillouin-Wigner perturbation theory[2] to obtain equations analogous to (v) and (vi). Multiplying (ii) by

$$(\mathbf{1}-\mathbf{P})\cdot[\mathbf{H}_0-E_0+\epsilon(\mathbf{1}-\mathbf{P})\mathbf{V}(\mathbf{1}-\mathbf{P})-\epsilon\Lambda]^{-1}=\sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\epsilon^n,$$

we get

$$\psi_1=\sum_{n=0}^{\infty}(-1)^{n+1}\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\cdot\epsilon^{n+1}\psi_0. \quad \text{(viii)}$$

Substituting (viii) into (i), we get

$$\mathbf{H}'(\Lambda,\epsilon)\psi_0=\Lambda\psi_0,$$

$$\mathbf{H}'(\Lambda,\epsilon)\equiv\mathbf{PVP}+\sum_{n=0}^{\infty}(-1)^{n+1}\mathbf{PV}\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}(\mathbf{V}-\Lambda)\right]^n\times\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}\epsilon^{n+1}. \quad \text{(ix)}$$

Equations (viii) and (ix) provide a means of finding ψ and Λ for cases in which expansion of these quantities in power series in ϵ is not a valid procedure. In this sense it competes with the methods developed in this paper.

However, we already saw in the case of nondegenerate E_0 that development of the Schrödinger perturbation theory via the Brillouin-Wigner perturbation theory is a much more complicated procedure than using the methods developed in this paper. Let us now examine the relative merits of these two methods for the case of degenerate E_0. Before attempting to answer this question, we should consider the following: by appropriate formal manipulations, it is possible to transform the Brillouin-Wigner perturbation theory into the perturbation theory of this paper, and conversely. However, the spirit of the Brillouin-Wigner perturbation theory is to keep the operator expressions involved as simple as possible at the expense of an implicit dependence on the unknown eigenvalue Λ. On the other hand, the spirit of the perturbation theory developed in this paper is to avoid any implicit dependence on the eigenvalue at the expense of more complicated operators than appear in the Brillouin-Wigner perturbation theory. What we want to compare are the relative efficiencies of these two approaches to perturbation theory.

We first consider the case of finite degeneracy. In labeling our complete set of orthonormal eigenvectors of $\mathbf{H}_0$, ξ_n, we count the basis vectors with $E_n=E_0$ first. Thus $E_n=E_0$ for $N-1\geq n\geq 0$, and $E_n\neq E_0$ for $n\geq N$, where N is the order of the degeneracy of E_0.

In considering the relative merits of Eq. (ix) of the Brillouin-Wigner theory and Eq. (42) or (46) of the theory developed in this paper, we observe that the operator $\mathbf{H}'(\Lambda,\epsilon)$ of (ix) has a simpler perturbation expansion than either of the corresponding operators of (42) and (46), provided Λ is a known real constant. The fact that Λ is not known counterbalances this initial advantage of the Brillouin-Wigner perturbation theory in the following way: once the operator appearing on the left-hand side of Eq. (42) or (46) has been determined to a given order in ϵ, determining the N eigenvectors and eigenvalues of (42) or (46) is just a standard eigenvector-eigenvalue problem in an N-dimensional linear vector space. In particular, the eigenvalues are given by the roots of the standard secular equation. Equation (ix) is more difficult to solve in that, instead of the standard secular equation, we have the equation

$$\det\{H'_{ij}-\Lambda\delta_{ij}\}=0, \quad H'_{ij}\equiv(\xi_i,\mathbf{H}'(\Lambda,\epsilon)\xi_j) \quad \text{for } N-1\geq i,\ j\geq 0.$$

Because the H'_{ij}'s have an implicit dependence on Λ, the above equation is a more difficult one to solve for Λ than is the standard secular equation, if we wish to be accurate in $E=(E_0+\epsilon\Lambda)$ to a given order in ϵ beyond the third order. In general, if we wish to be accurate to an order r, $r\geq 4$, the above equation will be a polynomial equation of degree N^{r-2}. The N real roots of smallest absolute value will be the eigenvalues we seek, if ϵ is small enough. Assuming this to be the case, when one of these eigenvalues is substituted into Eq. (ix), this equation can be solved for the corresponding eigenvector or eigenvectors just as in the standard eigenvector-eigenvalue problem.

From the considerations of the preceding paragraph, we see that neither the Brillouin-Wigner perturbation theory nor the perturbation theory of this paper is clearly superior to the other for solving the eigenvector-eigenvalue problem. The situation is further complicated by the fact that the methods outlined in the previous paragraph are often replaced by more specialized procedures, but the qualitative features remain the same. It seems reasonable to assume that, for some eigenvector-eigenvalue problems, the methods developed in this paper will be more efficient, when carried out to some order in ϵ. Only experience can determine what the conditions are for this to be true.

We now consider some respects in which the perturbation theory developed in this paper is clearly superior to both the conventional Schrödinger perturbation theory and the Brillouin-Wigner perturbation theory. Suppose we want to compute the average eigenvalue of the perturbed eigenvectors, using either of the latter theories. What we would have to do is find the N eigenvalues Λ_1, Λ_2, $\cdots$, Λ_N and compute $\Lambda_{\text{Av}}=N^{-1}\sum_1^N\Lambda_n$. The average eigenvalue is then given by $E_0+\epsilon\Lambda_{\text{Av}}$. Using the perturbation theory developed in this paper, we need only compute the trace of $\mathbf{PVP}+\mathbf{PV}(1-\mathbf{P})\mathbf{KP}$ and divide by N to obtain Λ_{Av}. We do not have to find the individual eigenvalues. In general

$$(\Lambda^r)_{\text{Av}}=N^{-1}\sum_1^N(\Lambda_n)^r$$
$$=\operatorname{Tr}\{[\mathbf{PVP}+\mathbf{PV}(1-\mathbf{P})\mathbf{KP}]^r\}, \quad r=1,2,\cdots.$$

It is clear that it is much easier, for reasonable values of r, to compute $(\Lambda^r)_{\text{Av}}$ to a given order in ϵ by using the trace formula, than it is to compute $(\Lambda^r)_{\text{Av}}$ by obtaining each of the N eigenvalues Λ_n, to the same order in ϵ, and computing $N^{-1}\sum_1^N(\Lambda_n)^r$. In other words, the theory developed in this paper permits us to obtain some statistical information about the distribution of certain eigenvalues, with little effort compared to the work required to obtain this information by solving the appropriate eigenvalue problem.

Let us turn now to the case of infinite degeneracy. Many of the points discussed in connection with finite degeneracy are applicable to the case of infinite degeneracy, with perhaps some modifications or restrictions. We will not discuss such points here. When the zeroth-order eigenvalue E_0 is infinitely degenerate, we may not be particularly interested in solving the eigenvector-eigenvalue problem, but rather in eliminating from $\mathbf{H}_0+\epsilon\mathbf{V}$ the coupling between states in the closed linear manifold M_{E_0} and states outside M_{E_0}. This we we accomplish by means of a unitary transformation, and, in particular, we can use the unitary operator $\mathbf{S}$ defined in (7). In the latter case we obtain the Hermitean operator $\mathfrak{H}$ of Eqs. (9) and (16). The operator $\mathbf{P}\mathfrak{H}\mathbf{P}$ of Eq. (14) then represents the effect within M_{E_0} of $\mathbf{H}_0+\epsilon\mathbf{V}$ in the complete Hilbert space. An example of a case in which $\mathbf{P}\mathfrak{H}\mathbf{P}$ is of direct interest arises in connection with the nonrelativistic limit of the Dirac equation for an electron in a time-independent electromagnetic field. We start with the Dirac Hamiltonian operator

$$mc^2\{\boldsymbol{\beta}+(mc)^{-1}\sum_i \boldsymbol{\alpha}_i[\mathbf{p}_i-(e/c)A_i(\mathbf{x,y,z})] +(e/mc^2)\phi(\mathbf{x,y,z})\},$$

where m is the electron mass, e the electron charge, c the velocity of light, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_x$, $\boldsymbol{\alpha}_y$, $\boldsymbol{\alpha}_z$ are Dirac matrices, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ coordinate operators, $\mathbf{p}_x$, $\mathbf{p}_y$, $\mathbf{p}_z$, the momentum operators, and $\phi(x,y,z)$, $A_i(x,y,z)$ the electromagnetic potentials at positition (x,y,z). We set $\mathbf{H}_0=\boldsymbol{\beta}$, $\epsilon=(mc)^{-1}$ and $\mathbf{V}=\sum_i\boldsymbol{\alpha}_i\{\mathbf{p}_i-(e/c)A_i(\mathbf{x,y,z})\}+(e/c)\phi(\mathbf{x,y,z})$. Since $\boldsymbol{\beta}^2=\mathbf{1}$, $\mathbf{H}_0$ has eigenvalues ± 1; we choose $E_0=1$. Then we easily obtain the following:

$$\mathbf{P}=\tfrac{1}{2}(\mathbf{1}+\boldsymbol{\beta}),\quad \mathbf{1}-\mathbf{P}=\tfrac{1}{2}(\mathbf{1}-\boldsymbol{\beta}),$$

$$\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^r}=\left[\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\right]^r=-(-\tfrac{1}{2})^{r+1}\cdot(\mathbf{1}-\boldsymbol{\beta}),$$

$$\mathbf{PVP}=(e/2c)(\mathbf{1}+\boldsymbol{\beta})\cdot\phi(\mathbf{x,y,z}),$$

$$(\mathbf{1}-\mathbf{P})\mathbf{V}(\mathbf{1}-\mathbf{P})=(e/2c)(\mathbf{1}-\boldsymbol{\beta})\cdot\phi(\mathbf{x,y,z}),$$

$$\mathbf{PV}(\mathbf{1}-\mathbf{P})=\tfrac{1}{2}(\mathbf{1}+\boldsymbol{\beta})\sum_i\boldsymbol{\alpha}_i\{\mathbf{p}_i-(e/c)A_i(\mathbf{x,y,z})\},$$

$$(\mathbf{1}-\mathbf{P})\mathbf{VP}=\tfrac{1}{2}(\mathbf{1}-\boldsymbol{\beta})\sum_i\boldsymbol{\alpha}_i\{\mathbf{p}_i-(e/c)A_i(\mathbf{x,y,z})\}.$$

When the above formulas are substituted into Eq. (54), $mc^2\mathbf{P}\mathfrak{H}\mathbf{P}$ will represent the equivalent nonrelativistic Hamiltonian operator for the electron, plus the higher order correction terms proportional to $(mc)^{-2}$ and $(mc)^{-3}$, among states for which $\boldsymbol{\beta}$ has the eigenvalue plus one.

It is worth noting that, in practice, if the Hermitean operator $\mathbf{H}_0$ has a continuous spectrum, the physical problem can be modified in a trivial manner, e.g., by box normalization in quantum mechanics, so as to give $\mathbf{H}_0$ a purely discrete spectrum. Then $\mathbf{H}_0$ will possess a complete orthonormal set of eigenvectors ξ_0, ξ_1, $\cdots$, ξ_n, $\cdots$, as we have been assuming. Furthermore, the modifications that have to be made for the case of a finite dimensional Hilbert space, in results obtained on the assumption that ξ_0, ξ_1, $\cdots$, ξ_n, $\cdots$ formed an infinite set, are obvious.

The remainder of this paper is organized as follows: in Sec. II we consider some concepts related to Hilbert space operators which will be of use later on. In Sec. III we state all the assumptions necessary to prove the perturbation theory has a nonzero radius of convergence. We also give a less restrictive form of Schrödinger perturbation theory, which differs from the conventional form. However, since the two forms do have a lot in common, we see no point in burdening the reader by introducing a new terminology to distinguish between them. This less restrictive form is developed in Sec. IV into one simple operator equation which is the starting point for the power series solutions considered in Sec. V. Section V contains the sufficient conditions for convergence of these power series. In Sec. VI the results of Secs. IV and V are applied to the eigenvector problem. In Sec. VII we discuss how the general theory developed in Secs. III, IV, V, and VI can be used in applications. Finally, in Sec. VIII we give the low-order terms in the power series expansions of quantities used in applications.

II. SOME CONCEPTS RELATED TO HILBERT SPACE OPERATORS

We wish to consider here the concepts of the norm of a Hilbert space operator and of a resolution of the identity. The results that follow will be used implicitly in the remainder of this paper.

The norm of an operator $\mathbf{A}$, $\|\mathbf{A}\|$, is defined as the supremum or least upper bound of $[(\mathbf{A}\psi,\mathbf{A}\psi)/(\psi,\psi)]^{\frac{1}{2}}$ for all nonzero vectors ψ in the Hilbert space. $\|\mathbf{A}\|$ is also equal to the supremum of $|(\chi,\mathbf{A}\psi)|/[(\chi,\chi)(\psi,\psi)]^{\frac{1}{2}}$ for all nonzero vectors ψ and χ in the Hilbert space. If $\mathbf{A}$ is Hermitean, then $\|\mathbf{A}\|$ is equal to the supremum of $|(\psi,\mathbf{A}\psi)|/(\psi,\psi)$.[4] It follows from the above results that,[5] for any operator $\mathbf{A}$, $\|\mathbf{A}\|\geq 0$, with $\|\mathbf{A}\|=0$ if and only if $\mathbf{A}=\mathbf{0}$; $\|\mathbf{A}^\dagger\|=\|\mathbf{A}\|$; and $\|\mathbf{A}^\dagger\mathbf{A}\|=\|\mathbf{A}\|^2$. If c is any complex number, then $\|c\mathbf{A}\|=|c|\cdot\|\mathbf{A}\|$. Also, for any operators $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{A}+\mathbf{B}\|\leq\|\mathbf{A}\|+\|\mathbf{B}\|$ and $\|\mathbf{AB}\|\leq\|\mathbf{A}\|\cdot\|\mathbf{B}\|$.

A family of projection operators $\mathbf{E}(\lambda)$ of the real variable λ is called a resolution of the identity if and only if it has the following properties:

$$\lim_{\lambda\to-\infty}\mathbf{E}(\lambda)=\mathbf{0},\quad \lim_{\lambda\to\infty}\mathbf{E}(\lambda)=\mathbf{1},\quad \lim_{\lambda\to\lambda_0+}\mathbf{E}(\lambda)=\mathbf{E}(\lambda_0),$$
$$\mathbf{E}(\lambda'')\geq\mathbf{E}(\lambda')\quad\text{for }\lambda''\geq\lambda'.$$

A resolution of the identity $\mathbf{E}(\lambda)$ belongs to a Hermitean operator $\mathbf{H}$ if and only if[6]

$$\mathbf{H}=\int_{-\infty}^{\infty}\lambda d\mathbf{E}(\lambda).$$

The following relations exist between any Hermitean operator and possible resolutions of the identity belonging to it:

(i) There exists a unique resolution of the identity belonging to each Hermitean operator of finite norm.[7]

(ii) There exists none or exactly one resolution of the identity belonging to a given Hermitean operator of infinite norm.[8]

[4] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955), p. 99. The constant C of theorem 18 is chosen to be $\|\mathbf{A}\|$.

[5] We use the notation $\mathbf{A}^\dagger$ to denote the adjoint or Hermitian conjugate of an operator $\mathbf{A}$.

[6] Reference 4, pp. 118, 119.

[7] Reference 4, pp. 150, 99.

[8] Reference 4, p. 154. We include the maximal property in our definition of a Hermitean operator.

If $\mathbf{H}$ is Hermitean and there exists a resolution of the identity $\mathbf{E}(\lambda)$, belonging to it, then this resolution is unique. If $f(x)$ is a real function of the real variable x, we can define

$$f(\mathbf{H})\equiv\int_{-\infty}^{\infty} f(\lambda)d\mathbf{E}(\lambda),$$

whenever the right-hand integral exists. If $f_3(x)=f_1(x)+f_2(x)$, then $f_3(\mathbf{H})=f_1(\mathbf{H})+f_2(\mathbf{H})$. If $f_3(x)=f_1(x)\cdot f_2(x)$, then $f_3(\mathbf{H})=f_1(\mathbf{H})\cdot f_2(\mathbf{H})$. Thus we can manipulate such functions of $\mathbf{H}$ as if $\mathbf{H}$ were a real variable.[9]

The reader may wonder why we introduce the concept of a resolution of the identity belonging to a given Hermitean operator, instead of the more familiar concept of a complete set of eigenvectors. The reason for this is that we want to use the result on the existence of a unique resolution of the identity belonging to any Hermitean operator of finite norm. This result is used to obtain Eq. (7) under more general conditions than would otherwise be possible.

III. FUNDAMENTAL ASSUMPTIONS

We make the following assumptions:

> $\mathbf{H}_0$ and $\mathbf{V}$ are Hermitean operators in a Hilbert space; there exists a resolution of the identity belonging to $\mathbf{H}_0$; E_0 is an isolated eigenvalue of $\mathbf{H}_0$; $\mathbf{H}_0\neq E_0\mathbf{1}$; and ϵ is a real parameter. (1)

Let $\mathbf{E}(\lambda)$ be a resolution of the identity belonging to $\mathbf{H}_0$. Then $\mathbf{E}(\lambda)$ is unique and

$$\mathbf{H}_0=\int_{-\infty}^{\infty}\lambda d\mathbf{E}(\lambda).$$

What we mean by E_0 being an isolated eigenvalue of $\mathbf{H}_0$ is that we can find a $\delta>0$ such that $d\mathbf{E}(\lambda)/d\lambda=\mathbf{0}$ for $0<|\lambda-E_0|\leq\delta$. $\mathbf{H}_0\neq E_0\mathbf{1}$ implies that

$$\mathbf{E}(E_0+\delta)-\mathbf{E}(E_0-\delta)\neq\mathbf{1}.$$

Note that we are only requiring that the one eigenvalue E_0 be isolated.

Let $\mathbf{P}$ be the projection operator onto the closed linear manifold M_{E_0} of all solutions ψ_0 of $\mathbf{H}_0\psi_0=E_0\psi_0$. Let Δ be the superemum or least upper bound of all possible δ's of the previous paragraph. Then $0<\Delta<\infty$, and

$$\mathbf{H}_0=E_0\mathbf{P}+\int_{|\lambda-E_0|\geq\Delta-}\lambda d\mathbf{E}(\lambda).$$

We can thus define

$$\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\equiv\int_{|\lambda-E_0|\geq\Delta-}(\lambda-E_0)^{-\frac{1}{2}}d\mathbf{E}(\lambda),$$

and

$$\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\equiv\int_{|\lambda-E_0|\geq\Delta-}(\lambda-E_0)^{-1}d\mathbf{E}(\lambda),$$

the right-hand integrals being absolutely convergent. Then

$$0<\left\|\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|=\Delta^{-\frac{1}{2}}<\infty,\quad 0<\left\|\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\right\|=\Delta^{-1}<\infty. \tag{2}$$

Some further properties are[5]:

$$E_0^*=E_0,\quad \mathbf{P}\neq\mathbf{0},\quad \mathbf{P}\neq\mathbf{1},\quad \mathbf{P}^\dagger=\mathbf{P},\quad \mathbf{P}^2=\mathbf{P},\quad \mathbf{H}_0\mathbf{P}=\mathbf{P}\mathbf{H}_0=E_0\mathbf{P}. \tag{3}$$

In addition, we assume that

$$\|\mathbf{PVP}\|<\infty,\quad \left\|\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|<\infty,\quad \left\|\mathbf{PV}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|<\infty. \tag{4}$$

Note that $\|\mathbf{V}\|<\infty$ is a sufficient condition for (4) to be satisfied, but not a necessary one if $\|\mathbf{H}_0\|=\infty$.

Let us see what form the assumptions we have made take for the case in which $\mathbf{H}_0$ possesses a complete orthonormal set of eigenvectors $\xi_0, \xi_1, \cdots, \xi_n, \cdots$ with eigenvalues $E_0, E_1, \cdots, E_n, \cdots$, respectively. The assumptions that $\mathbf{H}_0$ and $\mathbf{V}$ are Hermitean reduce to the assumptions that E_n is real for all n, and $(\xi_m,\mathbf{V}\xi_n)=(\xi_n,\mathbf{V}\xi_m)^*$ for all m and n, respectively. The assumption of the existence of a resolution of the identity belonging to $\mathbf{H}_0$ is satisfied by the following unique choice of $\mathbf{E}(\lambda)$:

$$\mathbf{E}(\lambda)\xi_n \begin{cases} =\xi_n & \text{for } E_n\leq\lambda \\ =0 & \text{for } E_n>\lambda. \end{cases}$$

The assumptions that E_0 is an isolated eigenvalue of $\mathbf{H}_0$, and that $\mathbf{H}_0\neq E_0\mathbf{1}$ reduce to the following: If $E_n\neq E_0$, then $|E_n-E_0|>\delta>0$ for some fixed δ; and, there exists an $E_n\neq E_0$, respectively.

The definitions of the operators $\mathbf{P}$ and

$$(\mathbf{1}-\mathbf{P})/(\mathbf{H}_0-E_0)$$

reduce to the definitions given in section I. The definition of $(\mathbf{1}-\mathbf{P})/(\mathbf{H}_0-E_0)^{\frac{1}{2}}$ reduces to the following:

$$\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\xi_n \begin{cases} =(E_n-E_0)^{-\frac{1}{2}}\xi_n & \text{for } E_n\neq E_0 \\ =0 & \text{for } E_n=E_0. \end{cases}$$

Equations (2) and (3) follows easily from these definitions, and the operators appearing in Eq. (4) are just products of operators already defined.

In Sec. I we have already noted differences between the perturbation theory developed in this paper and the conventional Schrödinger perturbation theory, for the case of degenerate zeroth order eigenvalue E_0. However,

[9] Reference 4, pp. 141–145.

the differences are not really marked enough to justify the introduction of a new terminology to distinguish between them. Instead, we use the descriptive phrase "Schrödinger perturbation theory in its least restrictive form" to describe the perturbation problem with which Secs. IV and V will be concerned, and which is described below.

Schrödinger perturbation theory in its least restrictive form is concerned with the problem of finding a unitary operator $\mathbf{U}(\epsilon)$ such that

$$(1-\mathbf{P})\mathbf{U}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{U}\mathbf{P}=\mathbf{P}\mathbf{U}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{U}(1-\mathbf{P})=0,$$

and such that $\mathbf{U}(\epsilon)$ and $\mathbf{U}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{U}$ possess derivatives with respect to ϵ of all orders at $\epsilon=0$. Questions as to the convergence of the power series expansions for $\mathbf{U}(\epsilon)$ and $\mathbf{U}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{U}$, and as to the equality of these operators to their power series expansions, are left unanswered. Furthermore, we cannot speak of *the* Schrödinger perturbation theory because $\mathbf{U}(\epsilon)$ is not unique. If $\mathbf{T}(\epsilon)$ is any unitary operator that commutes with $\mathbf{P}$, then $\mathbf{UT}$ is unitary and

$$(1-\mathbf{P})(\mathbf{UT})^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{UTP} =\mathbf{T}^\dagger(1-\mathbf{P})\mathbf{U}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{UPT}=0.$$

Similarly

$$\mathbf{P}(\mathbf{UT})^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{UT}(1-\mathbf{P})=0.$$

Thus $\mathbf{U}(\epsilon)\mathbf{T}(\epsilon)$ leads to a Schrödinger perturbation theory if $\mathbf{T}(\epsilon)$ possesses derivatives with respect to ϵ of all orders at $\epsilon=0$.

The assumptions contained in (1) and (4), however, are sufficient to enable us to exhibit a unitary operator $\mathbf{S}(\epsilon)$ for ϵ in some finite region about the origin, such that $(1-\mathbf{P})\mathbf{S}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{SP}=\mathbf{PS}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{S}(1-\mathbf{P})=0$, and such that $\mathbf{S}(\epsilon)$ and $\mathbf{S}^\dagger(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{S}$ are equal to power series in ϵ with nonzero radii of convergence.

IV. DEVELOPMENT OF THE THEORY

Consider any operator $\mathbf{K}$ such that

$$\|(1-\mathbf{P})\mathbf{KP}\|<\infty. \tag{5}$$

Then

$$\|\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\|=\|[(1-\mathbf{P})\mathbf{KP}]^\dagger(1-\mathbf{P})\mathbf{KP}\| =\|(1-\mathbf{P})\mathbf{KP}\|^2<\infty.$$

Thus there exists a unique resolution of the identity $\mathbf{E}(\lambda)$ belonging to $\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}$, i.e.,

$$\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}=\int_{-\infty}^{\infty}\lambda d\mathbf{E}(\lambda).$$

But for any vector ψ in the Hilbert space,

$$0\leq(\psi,\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\psi)\leq\|(1-\mathbf{P})\mathbf{KP}\|^2(\psi,\psi).$$

This implies $\mathbf{E}(\lambda)=0$ for $\lambda<0$ and $\mathbf{E}(\lambda)=1$ for $\lambda>\|(1-\mathbf{P})\mathbf{KP}\|^2$, so that

$$\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}=\int_{0-}^{\|(1-\mathbf{P})\mathbf{KP}\|^2+}\lambda d\mathbf{E}(\lambda).$$

We can thus define the operators

$$1/[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\equiv\int_{0-}^{\|(1-\mathbf{P})\mathbf{KP}\|^2+}(1+\lambda)^{-\frac{1}{2}}d\mathbf{E}(\lambda),$$

and

$$[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\equiv\int_{0-}^{\|(1-\mathbf{P})\mathbf{KP}\|^2+}(1+\lambda)^{\frac{1}{2}}d\mathbf{E}(\lambda).$$

Similarly we can define the operators

$$1/[1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}}$$

and

$$[1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}}.$$

Thus the operators

$$1/[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}},\quad[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}},\quad 1/[1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}},\quad [1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}} \tag{6}$$

exist. We can therefore define the operator $\mathbf{S}$ in terms of $(1-\mathbf{P})\mathbf{KP}$ as follows:

$$\mathbf{S}\equiv[1+(1-\mathbf{P})\mathbf{KP}]\frac{\mathbf{P}}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}} +[1-\mathbf{PK}^\dagger(1-\mathbf{P})]\frac{(1-\mathbf{P})}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}. \tag{7}$$

Then

$$\mathbf{S}^\dagger\mathbf{S}=\left[\frac{\mathbf{P}}{[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}}\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\}+\frac{(1-\mathbf{P})}{[1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}}}\{1-(1-\mathbf{P})\mathbf{KP}\}\right] \cdot\left[\{1+(1-\mathbf{P})\mathbf{KP}\}\frac{\mathbf{P}}{[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}}+\{1-\mathbf{PK}^\dagger(1-\mathbf{P})\}\frac{(1-\mathbf{P})}{[1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]^{\frac{1}{2}}}\right]=1.$$

$$\mathbf{SS}^\dagger=\{1+(1-\mathbf{P})\mathbf{KP}\}\frac{\mathbf{P}}{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}}\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\} +\{1-\mathbf{PK}^\dagger(1-\mathbf{P})\}\frac{(1-\mathbf{P})}{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})}\{1-(1-\mathbf{P})\mathbf{KP}\};$$

$$[1+\mathbf{PK}^\dagger(1-\mathbf{P})-(1-\mathbf{P})\mathbf{KP}][\mathbf{SS}^\dagger-1]=[\mathbf{P}\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\}+(1-\mathbf{P})\{1-(1-\mathbf{P})\mathbf{KP}\}][\mathbf{SS}^\dagger-1]=0.$$

Let ψ be any vector in the Hilbert space, and let $\chi=[\mathbf{SS}^\dagger-1]\psi$. Then

$$0=([1+\mathbf{PK}^\dagger(1-\mathbf{P})-(1-\mathbf{P})\mathbf{KP}]\chi,\ \times[1+\mathbf{PK}^\dagger(1-\mathbf{P})-(1-\mathbf{P})\mathbf{KP}]\chi)$$
$$=(\chi,[1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})]\chi)\geq(\chi,\chi).$$

Thus $\chi=0$ for arbitrary ψ, so that we must have $\mathbf{SS}^\dagger=\mathbf{1}$. Thus $\mathbf{S}$ is unitary, i.e.,

$$\mathbf{S}^\dagger=\mathbf{S}^{-1}. \tag{8}$$

We now let $(1-\mathbf{P})\mathbf{KP}$, and thus $\mathbf{S}$, be functions of ϵ. We define the Hermitean operator $\mathfrak{H}(\epsilon)$ as follows:

$$\mathfrak{H}(\epsilon)\equiv\mathbf{S}^\dagger(\epsilon)[\mathbf{H}_0+\epsilon\mathbf{V}]\mathbf{S}(\epsilon). \tag{9}$$

Substituting (7) in (9) we get

$$(1-\mathbf{P})\mathfrak{H}\mathbf{P}=\frac{(1-\mathbf{P})}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}\times[1-(1-\mathbf{P})\mathbf{KP}]\cdot[\mathbf{H}_0+\epsilon\mathbf{V}]\times[1+(1-\mathbf{P})\mathbf{KP}]\frac{\mathbf{P}}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}$$

which simplifies to

$$(1-\mathbf{P})\mathfrak{H}\mathbf{P}=\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}\times[(\mathbf{H}_0-E_0)(1-\mathbf{P})\mathbf{KP}-\epsilon\{(1-\mathbf{P})\mathbf{KPVP}-(1-\mathbf{P})\mathbf{V}(1-\mathbf{P})\mathbf{KP}+(1-\mathbf{P})\mathbf{KPV}(1-\mathbf{P})\mathbf{KP}-(1-\mathbf{P})\mathbf{VP}\}]\frac{1}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}. \tag{10}$$

Combining (10) with the relation

$$\mathbf{P}\mathfrak{H}(1-\mathbf{P})=[(1-\mathbf{P})\mathfrak{H}\mathbf{P}]^\dagger,$$

we obtain the following:

A sufficient condition for the relations $(1-\mathbf{P})\mathfrak{H}\mathbf{P}=\mathbf{P}\mathfrak{H}(1-\mathbf{P})=\mathbf{0}$ to hold is that

$$(\mathbf{H}_0-E_0)(1-\mathbf{P})\mathbf{KP}=\epsilon[(1-\mathbf{P})\mathbf{KPVP}-(1-\mathbf{P})\mathbf{V}(1-\mathbf{P})\mathbf{KP}+(1-\mathbf{P})\mathbf{KPV}(1-\mathbf{P})\mathbf{KP}-(1-\mathbf{P})\mathbf{VP}]. \tag{11}$$

In Sec. V we will find a solution of (11) as a power series in ϵ with a nonzero radius of convergence. For the present we will assume we have an operator $(1-\mathbf{P})\mathbf{K}(\epsilon)\mathbf{P}$ satisfying (11) and evaluate $\mathfrak{H}(\epsilon)$. We immediately have

$$(1-\mathbf{P})\mathfrak{H}\mathbf{P}=\mathbf{P}\mathfrak{H}(1-\mathbf{P})=\mathbf{0}. \tag{12}$$

Multiplying (9) on the left with $\mathbf{S}$ and using (8), we get

$$\mathbf{S}\mathfrak{H}=[\mathbf{H}_0+\epsilon\mathbf{V}]\mathbf{S}. \tag{13}$$

Combining (7), (12), and (13), we get

$$\mathbf{PS}\mathfrak{H}\mathbf{P}=\mathbf{PSP}\mathfrak{H}\mathbf{P}=\frac{1}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}\mathbf{P}\mathfrak{H}\mathbf{P}=\mathbf{P}[\mathbf{H}_0+\epsilon\mathbf{V}]\mathbf{SP}$$
$$=[E_0\mathbf{P}+\epsilon\mathbf{PVP}+\epsilon\mathbf{PV}(1-\mathbf{P})\mathbf{KP}]\frac{1}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}.$$

It then follows that

$$\mathbf{P}\mathfrak{H}\mathbf{P}=E_0\mathbf{P}+\epsilon\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}[\mathbf{PVP}+\mathbf{PV}(1-\mathbf{P})\mathbf{KP}]\frac{1}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}. \tag{14}$$

By similar reasoning

$$(1-\mathbf{P})\mathbf{S}\mathfrak{H}(1-\mathbf{P})=(1-\mathbf{P})\mathbf{S}(1-\mathbf{P})\mathfrak{H}\mathbf{P}=\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}(1-\mathbf{P})\mathfrak{H}(1-\mathbf{P})=(1-\mathbf{P})[\mathfrak{H}_0+\epsilon\mathbf{V}]\mathbf{S}(1-\mathbf{P})$$
$$=[\mathbf{H}_0(1-\mathbf{P})+\epsilon(1-\mathbf{P})\mathbf{V}(1-\mathbf{P})-\epsilon(1-\mathbf{P})\mathbf{VPK}^\dagger(1-\mathbf{P})]\cdot\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}.$$

Thus

$$(1-\mathbf{P})\mathfrak{H}(1-\mathbf{P})=\mathfrak{H}_0(1-\mathbf{P})+\epsilon\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}[(1-\mathbf{P})\mathbf{V}(1-\mathbf{P})-(1-\mathbf{P})\mathbf{PVK}^\dagger(1-\mathbf{P})]\times\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}. \tag{15}$$

Combining (12), (14), and (15), we get

$$\mathfrak{H}=\mathbf{H}_0+\epsilon\Bigg\{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}[\mathbf{PVP}+\mathbf{PV}(1-\mathbf{P})\mathbf{KP}]\frac{1}{\{1+\mathbf{PK}^\dagger(1-\mathbf{P})\mathbf{KP}\}^{\frac{1}{2}}}+\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}[(1-\mathbf{P})\mathbf{V}(1-\mathbf{P})-(1-\mathbf{P})\mathbf{VPK}^\dagger(1-\mathbf{P})]\frac{1}{\{1+(1-\mathbf{P})\mathbf{KPK}^\dagger(1-\mathbf{P})\}^{\frac{1}{2}}}\Bigg\}. \tag{16}$$

V. POWER SERIES SOLUTION AND CONVERGENCE

We will now find a solution of (11) of the form

$$(1-\mathbf{P})\mathbf{K}\mathbf{P}=\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}\mathbf{P},\quad \|\mathfrak{K}\|<\infty. \tag{17}$$

Equations (2) and (17) imply that

$$\|(1-\mathbf{P})\mathbf{K}\mathbf{P}\|=\left\|\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}\mathbf{P}\right\| \leq\left\|\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|\cdot\|\mathfrak{K}\|<\infty.$$

Thus (17) implies (5). Substitution of (17) into (11) leads to the following:

A sufficient condition for (11) to be satisfied is for the operator $\mathfrak{K}$ to satisfy the equation

$$\mathfrak{K}=\epsilon\left[\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathfrak{K}\mathbf{P}\mathbf{V}\mathbf{P}-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}\mathbf{P} +\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathfrak{K}\mathbf{P}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}\mathbf{P}-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\mathbf{P}\right]. \tag{18}$$

We seek a solution of (18) of the form

$$\mathfrak{K}(\epsilon)=\sum_{n=1}^{\infty}\mathfrak{K}_n\epsilon^n. \tag{19}$$

The condition $\|\mathfrak{K}\|<\infty$ will be guaranteed if we require that

$$\sum_{n=1}^{\infty}\|\mathfrak{K}_n\|\cdot|\epsilon|^n<\infty. \tag{20}$$

By substituting (19) into (18) and equating coefficients of like powers of ϵ on both sides of the resulting equation we obtain

$$\mathfrak{K}_1=-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\mathbf{P},$$

$$\mathfrak{K}_n=\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathfrak{K}_{n-1}\mathbf{P}\mathbf{V}\mathbf{P}-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}_{n-1}\mathbf{P} +\sum_{k=1}^{n-2}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathfrak{K}_k\mathbf{P}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}_{n-k-1}\mathbf{P}\quad\text{for } n\geq 2. \tag{21}$$

If the expressions for $\mathfrak{K}_n$, given by (21) satisfy (20), then (19) is a solution of (18), which implies

$$(1-\mathbf{P})\mathbf{K}(\epsilon)\mathbf{P}=\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}(\epsilon)\mathbf{P}$$

is a solution of (11). We will now obtain a sufficient condition for this to be the case.

We first note that, for any ψ in the Hilbert space,

$$\left(\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\mathbf{P}\psi,\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\mathbf{P}\psi\right) =\left(\left[\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right]^{\dagger}\mathbf{V}\mathbf{P}\psi,\left[\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right]^{\dagger}\mathbf{V}\mathbf{P}\psi\right);$$

so that

$$\left\|\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\mathbf{P}\right\|=\left\|\left[\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right]^{\dagger}\mathbf{V}\mathbf{P}\right\| =\left\|\mathbf{P}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|. \tag{22}$$

(2) and (4) enable us to define

$$a\equiv\Delta^{-1}\|\mathbf{P}\mathbf{V}\mathbf{P}\|+\left\|\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|,$$

$$b\equiv\Delta^{-\frac{1}{2}}\left\|\mathbf{P}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|. \tag{23}$$

Applying (22), (23), and some results of Sec. II to (21) we get

$$\|\mathfrak{K}_1\|=\Delta^{\frac{1}{2}}b,$$

$$\|\mathfrak{K}_n\|\leq a\|\mathfrak{K}_{n-1}\|+b\Delta^{-\frac{1}{2}}\sum_{k=1}^{n-2}\|\mathfrak{K}_k\|\cdot\|\mathfrak{K}_{n-k-1}\|\quad\text{for } n\geq 2. \tag{24}$$

We next define a function $G(z)$ of the complex variable z as follows:

$$G(z)\equiv(\Delta^{\frac{1}{2}}/2bz)[1-az-\{(1-az)^2-4b^2z^2\}^{\frac{1}{2}}]. \tag{25}$$

It is easy to show that $G(z)$ has the following properties:

$$G(z)=z[aG(z)+b\Delta^{-\frac{1}{2}}G^2(z)+\Delta^{\frac{1}{2}}b]. \tag{26}$$

$G(z)$ is analytic in z for $|z|<1/(a+2b)$ and continuous in z for $|z|\leq 1/(a+2b)$.

We can define M_n as follows:

$$M_n\equiv n!^{-1}[d^nG(z)/dz^n]_{z=0}. \tag{27}$$

It can then be shown that

$$G(z)=\sum_{n=1}^{\infty}M_nz^n\quad\text{for } |z|\leq 1/(a+2b). \tag{28}$$

Substituting (28) in (26) and equating coefficients of like powers of z on both sides of the resulting equation, we get

$$M_1=\Delta^{\frac{1}{2}}b,$$
$$M_n=aM_{n-1}+b\Delta^{-\frac{1}{2}}\sum_{k=1}^{n-2}M_kM_{n-k-1}\quad\text{for } n\geq 2. \tag{29}$$

Note the similarity in form of (24) and (29). A simple induction suffices to prove that[10]

$$\|\mathfrak{R}_n\|\leq M_n \quad \text{for } n\geq 1. \tag{30}$$

Using (28) and (30) we get for $|\epsilon|\leq 1/(a+2b)$ that

$$\sum_{n=1}^{\infty}\|\mathfrak{R}_n\|\cdot|\epsilon|^n\leq \sum_{n=1}^{\infty}M_n|\epsilon|^n=G(|\epsilon|)$$
$$\leq G(1/[a+2b])=\Delta^{\frac{1}{2}}<\infty.$$

Thus (20) is satisfied. We can now state the following:

$$\mathfrak{R}(\epsilon)=\sum_{n=1}^{\infty}\mathfrak{R}_n\epsilon^n \tag{31}$$

is a solution of (18) for $|\epsilon|\leq 1/(a+2b)$,

$$\|\mathfrak{R}(\epsilon)\|\leq G(|\epsilon|)<\Delta^{\frac{1}{2}} \quad \text{for } |\epsilon|<1/(a+2b). \tag{32}$$

Let us define

$$\mathbf{K}_n\equiv\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{R}_n\mathbf{P} \quad \text{for } n\geq 1. \tag{33}$$

Then, from (17)–(21), (23), and (30)–(33), we get

$$\mathbf{K}_1=-\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\mathbf{P},$$

$$\mathbf{K}_n=\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{K}_{n-1}\mathbf{P}\mathbf{V}\mathbf{P}-\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}(\mathbf{1}-\mathbf{P})\mathbf{K}_{n-1}\mathbf{P}$$
$$+\sum_{k=1}^{n-2}\frac{\mathbf{1}-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{K}_k\mathbf{P}\mathbf{V}(\mathbf{1}-\mathbf{P})\mathbf{K}_{n-k-1}\mathbf{P} \quad \text{for } n\geq 2. \tag{34}$$

$$\|\mathbf{K}_n\|\leq\Delta^{-\frac{1}{2}}M_n \quad \text{for } n\geq 1. \tag{35}$$

$$(\mathbf{1}-\mathbf{P})\mathbf{K}(\epsilon)\mathbf{P}=\sum_{n=1}^{\infty}\mathbf{K}_n\epsilon^n \tag{36}$$

is a solution of (11) for $|\epsilon|\leq 1/(a+2b)$.

$$\|(\mathbf{1}-\mathbf{P})\mathbf{K}(\epsilon)\mathbf{P}\|\leq\Delta^{-\frac{1}{2}}G(|\epsilon|)<1$$
$$\text{for } |\epsilon|<1/(a+2b). \tag{37}$$

[10] Since $\|\mathfrak{R}_1\|=\Delta^{\frac{1}{2}}b=M_1$, (30) is true for $n=1$. Assume (30) is true for $N\geq n\geq 1$. Then, from (30) with $N\geq n\geq 1$, (24) and (29) we get

$$\|\mathfrak{R}_{N+1}\|\leq a\|\mathfrak{R}_N\|+b\Delta^{-\frac{1}{2}}\sum_{k=1}^{N-1}\|\mathfrak{R}_k\|\cdot\|\mathfrak{R}_{N-k}\|\leq aM_N$$
$$+b\Delta^{-\frac{1}{2}}\sum_{k=1}^{N-1}M_kM_{N-k}=M_{N+1}.$$

This imples that (30) is true for $n=N+1$. By mathematical induction, (30) is thus true for all $n\geq 1$.

Because of (37) we can use the expansions

$$(\mathbf{1}+\mathbf{A})^{\frac{1}{2}}=\mathbf{1}+\sum_{n=0}^{\infty}\frac{(-1)^n(2n)!}{(n+1)2^{2n+1}(n!)^2}\mathbf{A}^{n+1}, \quad \|\mathbf{A}\|<1,$$
$$\frac{1}{(\mathbf{1}+\mathbf{A})^{\frac{1}{2}}}=\sum_{n=0}^{\infty}\frac{(-1)^n(2n)!}{2^n(n!)^2}\mathbf{A}^n, \quad \|\mathbf{A}\|<1, \tag{38}$$

with[11]

$$\mathbf{A}=\mathbf{P}\mathbf{K}^\dagger(\mathbf{1}-\mathbf{P})\mathbf{K}\mathbf{P} \quad \text{and} \quad \mathbf{A}=(\mathbf{1}-\mathbf{P})\mathbf{K}\mathbf{P}\mathbf{K}^\dagger(\mathbf{1}-\mathbf{P}),$$
$$\text{for } |\epsilon|<1/(a+2b).$$

(38) may be used to express $\mathbf{S}(\epsilon)$ and $\mathfrak{H}(\epsilon)$ as power series in ϵ, using (7), (16), (34), and (36), which will be convergent for $|\epsilon|<1/(a+2b)$, i.e.,

$$\mathbf{S}(\epsilon)=\mathbf{1}+\sum_{n=1}^{\infty}n!^{-1}[d^n\mathbf{S}(\epsilon)/d\epsilon^n]_{\epsilon=0}\cdot\epsilon^n$$
$$\text{for } |\epsilon|<1/(a+2b), \tag{39}$$

$$\mathfrak{H}(\epsilon)=\mathbf{H}_0+\sum_{n=1}^{\infty}n!^{-1}[d^n\mathfrak{H}(\epsilon)/d\epsilon^n]_{\epsilon=0}\cdot\epsilon^n$$
$$\text{for } |\epsilon|<1/(a+2b). \tag{40}$$

This completes the proof of the statements made in the last paragraph of Sec. III.

VI. APPLICATION TO THE EIGENVECTOR PROBLEM

We want to apply the results of Secs. IV and V to the eigenvector problem

$$(\mathbf{H}_0+\epsilon\mathbf{V})\psi=E\psi. \tag{41}$$

Suppose we have a vector χ_0 in the closed linear manifold M_{E_0} of all solutions ψ_0 of the equation $\mathbf{H}_0\psi_0=E_0\psi_0$ such that

$$(\mathbf{P}\mathfrak{H}\mathbf{P})\chi_0=E\chi_0. \tag{42}$$

Then, using (12) and the definition of $\mathbf{P}$, we see that χ_0 also satisfies the equations

$$\mathbf{P}\chi_0=\chi_0, \quad \mathfrak{H}\chi_0=E\chi_0, \tag{43}$$

and, using (13), (42) and (43), we get

$$(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{S}\mathbf{P}\chi_0=\mathbf{S}\mathfrak{H}\chi_0=E\mathbf{S}\mathbf{P}\chi_0, \tag{44}$$

i.e. $\psi=\mathbf{S}\mathbf{P}\chi_0$ is a solution of (41). Thus, we have reduced the problem of finding certain eigenvectors of the Hermitean operator $\mathbf{H}_0+\epsilon\mathbf{V}$ in the full Hilbert space to that of finding the eigenvectors of the Hermitean operator $\mathbf{P}\mathfrak{H}\mathbf{P}$ in M_{E_0}, and then operating on the latter eigenvectors with $\mathbf{SP}$. The eigenvectors we obtain by

[11] $\|\mathbf{P}\mathbf{K}^\dagger(\mathbf{1}-\mathbf{P})\mathbf{K}\mathbf{P}\|=\|(\mathbf{1}-\mathbf{P})\mathbf{K}\mathbf{P}\mathbf{K}^\dagger(\mathbf{1}-\mathbf{P})\|=\|(\mathbf{1}-\mathbf{P})\mathbf{K}\mathbf{P}\|^2<1.$

this method are namely those satisfying the relation:

$$\mathbf{SPS}^{\dagger}\psi=\psi. \tag{45}$$

Unfortunately, the perturbation expansions of the operators $\mathbf{P\mathfrak{H}P}$ and $\mathbf{SP}$ are more complex than that of $(\mathbf{1}-\mathbf{P})\mathbf{KP}$. If we are willing to find the eigenvectors of a non-Hermitian operator in M_{E_0}, then we can find operators corresponding to $\mathbf{P\mathfrak{H}P}$ and $\mathbf{SP}$ which are so simply related to $(\mathbf{1}-\mathbf{P})\mathbf{KP}$, that their perturbation expansions can be written down by inspection from the perturbation expansion of $(\mathbf{1}-\mathbf{P})\mathbf{KP}$. We proceed as follows: suppose we have a vector Ω_0 in M_{E_0} such that

$$\epsilon[\mathbf{PVP}+\mathbf{PV}(\mathbf{1}-\mathbf{P})\mathbf{KP}]\Omega_0=(E-E_0)\Omega_0. \tag{46}$$

Then

$$\mathbf{P}\Omega_0=\Omega_0, \tag{47}$$

and using (7), (12), (13), (14), (46), and (47), we get

$$\begin{aligned}(\mathbf{H}_0+\epsilon\mathbf{V})[\mathbf{1}+(\mathbf{1}-\mathbf{P})\mathbf{KP}]\Omega_0
&=(\mathbf{H}_0+\epsilon\mathbf{V})\mathbf{SP}[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\Omega_0\\
&=\mathbf{SP\mathfrak{H}P}[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\Omega_0\\
&=\mathbf{SP}[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\\
&\qquad\times\{E_0+\epsilon[\mathbf{PVP}+\mathbf{PV}(\mathbf{1}-\mathbf{P})\mathbf{KP}]\}\Omega_0\\
&=E\mathbf{SP}[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}\Omega_0\\
&=E[\mathbf{1}+(\mathbf{1}-\mathbf{P})\mathbf{KP}]\Omega_0;\end{aligned}$$

that is,

$$\psi=[\mathbf{1}+(\mathbf{1}-\mathbf{P})\mathbf{KP}]\Omega_0 \tag{48}$$

is a solution of (41). The relation between (42) and (46) is as follows: From any Ω_0 we can obtain a χ_0 by multiplication by $[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}$; conversely, from any χ_0 we can obtain an Ω_0 by multiplication by $1/[\mathbf{1}+\mathbf{PK}^{\dagger}(\mathbf{1}-\mathbf{P})\mathbf{KP}]^{\frac{1}{2}}$. $\epsilon[\mathbf{PVP}+\mathbf{PV}(\mathbf{1}-\mathbf{P})\mathbf{KP}]$ and $[\mathbf{1}+(\mathbf{1}-\mathbf{P})\mathbf{KP}]$ are the counterparts of $\mathbf{P\mathfrak{H}P}$ and $\mathbf{SP}$, respectively, and possess the properties we anticipated.

If $|\epsilon|\leq 1/(a+2b)$, what can we say about the possible range of E? Utilizing (17), (23), (32), and (46), we get

$$|E-E_0|\leq\left\|\epsilon\left[\mathbf{PVP}+\mathbf{PV}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathfrak{K}\mathbf{P}\right]\right\|\leq\left[\|\mathbf{PVP}\|+\Delta^{\frac{1}{2}}\left\|\mathbf{PV}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|\right]\Bigg/\left[\Delta^{-1}\|\mathbf{PVP}\|+\left\|\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|+2\Delta^{-\frac{1}{2}}\left\|\mathbf{PV}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|\right].$$

Thus $|E-E_0|/\Delta\leq 1$. A stronger result, applicable under quite general circumstances, but not in all cases, is the following:

$$|E-E_0|/\Delta<\tfrac{1}{2}\quad\text{for } |\epsilon|\leq 1/(a+2b). \tag{49}$$

A sufficient condition for (49) to be true is that

$$\left\|\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{\mathbf{1}-\mathbf{P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|>\Delta^{-1}\|\mathbf{PVP}\|. \tag{50}$$

VII. DISCUSSION

Although all the results that are necessary in applying the general theory presented in Secs. III, IV, V, and VI are contained in these sections, they are presented in logical order rather than in a way most convenient for applications. The purpose of the remarks that follow is to show how the general theory can be used in applications.

The formal expressions for the coefficients of ϵ^n in the perturbation expansions of $(\mathbf{1}-\mathbf{P})\mathbf{KP}$, $\mathbf{S}$ and $\mathfrak{H}$ do not depend on whether or not our convergence criteria are satisfied, or even on whether the perturbation expansions converge, but only on whether the operator products that appear in the coefficients exist. The formal expressions for these coefficients are given directly or can be derived from the following equations: (7), (16), (34), (36), and (38).

We have developed two alternative procedures for finding solutions of Eq. (41) which also satisfy (45). One method involves simpler expressions for the coefficients of ϵ^n in the perturbation expansions of the relevant operators, but involves finding eigenvectors of a non-Hermitean operator in the closed linear manifold M_{E_0} of all solutions ψ_0 of the equation $\mathbf{H}_0\psi_0=E_0\psi_0$. The equations needed for applying this method are (34), (36), (46) and (48). The other method involves more complex expressions for the coefficients of ϵ^n, but involves finding eigenvectors of a Hermitean operator in M_{E_0}. The equations needed for applying this method are (7), (14), (34), (36), (38), (42), and (44). The former method is simpler when the dimensionality of M_{E_0} is a small finite number, and certainly when M_{E_0} is one dimensional, i.e., E_0 is nondegenerate. The latter method is simpler when the dimensionality of M_{E_0} is infinite or finite but large; the reason for this is that the operator in M_{E_0} whose eigenvectors we are seeking is Hermitean, so that we can, for example, write it in the form $\mathbf{H}_0'+\epsilon\mathbf{V}'$ and apply the methods of this paper to it, whereas we could not do this with a non-Hermitean operator.

Let us now consider the sufficient conditions for convergence that have been developed. If we only desire convergence for ϵ small enough but not zero, then the criteria we are interested in are given in Eqs. (1) and (4). If we want to use the general theory to prove convergence for a particular value of ϵ, then we must have $|\epsilon|\leq 1/(a+2b)$ for the expansion of the operator $(\mathbf{1}-\mathbf{P})\mathbf{KP}$, or $|\epsilon|<1/(a+2b)$ for the expansion of the operator $\mathbf{S}$ or $\mathfrak{H}$, where a and b are defined in Eq. (23).

In practice we can always find Δ exactly and can often find

$$\|\mathbf{PVP}\| \quad \text{and} \quad \left\|\mathbf{PV}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|$$

exactly, but usually can only find bounds on

$$\left\|\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^{\frac{1}{2}}}\right\|.$$

In cases where $\|\mathbf{A}\|$ cannot be found exactly, the definition of $\|\mathbf{A}\|$ as the supremum or least upper bound of $[(\mathbf{A}\psi,\mathbf{A}\psi)/(\psi,\psi)]^{\frac{1}{2}}$ for all nonzero vectors ψ in the Hilbert space provides a natural and convenient variational principle for obtaining a good approximation to $\|\mathbf{A}\|$ from below. The smaller the value of $|\epsilon|(a+2b)<1$, the smaller, approximately, is the upper bound provided by the general theory to the error in neglecting all terms in the perturbation expansion of the operator $(\mathbf{1-P})\mathbf{KP}$, $\mathbf{S}$ or $\mathfrak{H}$ beyond a certain order. We have not studied these bounds, but they can be derived, using Eqs. (25), (28), (29), (35), and (36).[12] This suggests that, if we have a choice of $\mathbf{H}_0$ out of a group of possibilities, we should choose that $\mathbf{H}_0$ which minimizes $|\epsilon|(a+2b)$.

It is often the case that a good estimate of the eigenvalue E for an eigenvector we are interested in can be obtained either theoretically or experimentally. If $|E-E_0|/\Delta\geq\frac{1}{2}$, then we know from Eqs. (49) and (50) that $|\epsilon|>1/(a+2b)$ under quite general circumstances, so that we cannot use the general theory to establish convergence of the perturbation expansion of $(I-\mathbf{P})\mathbf{KP}$, $\mathbf{S}$ or $\mathfrak{H}$. For example, if we wish to treat the electron-electron repulsion in the neutral helium atom as a perturbation in computing the ground state energy, then, since $|E-E_0|/\Delta\approx 11/15>\frac{1}{2}$, we know that almost certainly $|\epsilon|>1/(a+2b)$.

For additional applications and a more detailed discussion see Sec. I.

VIII. LOW-ORDER PERTURBATION TERMS

From (34) we obtain the following expressions for $\mathbf{K}_1$, $\mathbf{K}_2$, and $\mathbf{K}_3$:

$$\mathbf{K}_1=-\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP},\quad \mathbf{K}_2=\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}-\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP},$$

$$\begin{aligned}\mathbf{K}_3=&\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VPVP}+\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}+\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}\\&-\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VPVPVP}-\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}.\end{aligned}\tag{51}$$

For the case in which M_{E_0} is one-dimensional, i.e. E_0 is nondegenerate, let ξ_0 be a normal basis vector in M_{E_0}. Then, substituting (51) in (46), we get[13]

$$\begin{aligned}E=&E_0+\epsilon(\xi_0,\mathbf{V}\xi_0)-\epsilon^2\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)+\epsilon^3\left[\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)-(\xi_0,\mathbf{V}\xi_0)\cdot\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\xi_0\right)\right]\\&+\epsilon^4\Bigg[(\xi_0,\mathbf{V}\xi_0)\cdot\left\{\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)+\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\xi_0\right)\right\}+\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)\\&\cdot\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\xi_0\right)-(\xi_0,\mathbf{V}\xi_0)^2\cdot\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^3}\mathbf{V}\xi_0\right)-\left(\xi_0,\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\xi_0\right)\Bigg]+O(\epsilon^5).\end{aligned}\tag{52}$$

For the more general case, we get

$$\begin{aligned}\epsilon[\mathbf{PVP}+\mathbf{PV}(\mathbf{1-P})\mathbf{KP}]=&\epsilon\mathbf{PVP}-\epsilon^2\mathbf{PV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}+\epsilon^3\left[\mathbf{PV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}-\mathbf{PV}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}\right]\\&+\epsilon^4\Bigg[\mathbf{PV}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VPVP}+\mathbf{PV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}+\mathbf{PV}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}\\&-\mathbf{PV}\frac{\mathbf{1-P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VPVPVP}-\mathbf{PV}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{\mathbf{1-P}}{\mathbf{H}_0-E_0}\mathbf{VP}\Bigg]+O(\epsilon^5).\end{aligned}\tag{53}$$

---

[12] $\|(\mathbf{1-P})(\mathbf{K}(\epsilon)\mathbf{P}-\sum_{n=1}^{N}\mathbf{K}_n\epsilon^n\|\leq\Delta^{-\frac{1}{2}}[G(|\epsilon|)-\sum_{n=1}^{N}M_n|\epsilon|^n]$, for example.

[13] This formula is equivalent to the formula obtained by K. F. Niessen, Phys. Rev. 34, 263 (1929).

The expression for $\mathbf{P}\mathfrak{H}\mathbf{P}$ obtained by substituting (51) in (14) and using (38) is

$$\mathbf{P}\mathfrak{H}\mathbf{P}=E_0\mathbf{P}+\epsilon\mathbf{PVP}-\epsilon^2\mathbf{PV}\frac{1-\mathbf{P}}{\mathfrak{H}_0-E_0}\mathbf{VP}+\epsilon^3\left[\mathbf{PV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}-\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}+\mathbf{PVPV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}\right\}\right]$$
$$+\epsilon^4\left[\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VPVP}+\mathbf{PVPV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}\right\}\right.$$
$$+\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}+\mathbf{PVPV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}\right\}$$
$$+\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}+\mathbf{PV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VPV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}\right\}-\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VPVPVP}\right.$$
$$\left.\left.+\mathbf{PVPVPV}\frac{1+\mathbf{P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VP}\right\}-\mathbf{PV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}\right]+O(\epsilon^5). \quad (54)$$

The expression for $\mathbf{SP}$ obtained by substituting (51) in (7) and using (38) is

$$\mathbf{SP}=\mathbf{P}-\epsilon\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}+\epsilon^2\left[\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}-\tfrac{1}{2}\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}\right]$$
$$+\epsilon^3\left[\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VPVP}+\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPVP}+\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VPV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}\right.$$
$$-\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VPVPVP}-\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}+\frac{1}{2}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VPV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}$$
$$+\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{V}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{VP}+\mathbf{PV}\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\mathbf{V}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^2}\mathbf{VP}\right\}$$
$$\left.-\frac{1}{2}\left\{\mathbf{PV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VPVP}+\mathbf{PVPV}\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^3}\mathbf{VP}\right\}\right]+O(\epsilon^4). \quad (55)$$

For the reader who is unfamiliar with functions of operators, it should be pointed out that

$$\frac{1-\mathbf{P}}{(\mathbf{H}_0-E_0)^r}=\left[\frac{1-\mathbf{P}}{\mathbf{H}_0-E_0}\right]^r, \quad r=1,2,\cdots.$$

Thus, Eqs. (51)–(55) are just algebraic combinations of operators discussed in Sec. I, so that the reader should be able to write these equations in matrix form when $\mathbf{H}_0$ possesses a complete orthonormal set of eigenvectors ξ_0, ξ_1, $\cdots$, ξ_n, $\cdots$. For other cases the reader should note the next to last paragraph of Sec. I.