JOURNAL OF COMPUTATIONAL PHYSICS 2, 90-113 (1967)

Numerical Techniques in Matrix Mechanics[1]

CHARLES SCHWARTZ

Department of Physics, University of California, Berkeley, California 94720

ABSTRACT

We present, with numerical examples, several approximation techniques for solving some model problems in quantum mechanics. The emphasis is on matrix representations rather than differential or integral equations. Some of the methods are already known (extensions of the variational principle) and others are believed to be novel.

I. INTRODUCTION

Quantum mechanics is usually presented in two different forms. One is the Schrödinger differential (or integral) equation in which the independent variable x (or p) runs over the real line. The other is the Dirac-Heisenberg representation dealing with vectors in an infinite dimensional Hilbert space. The formal correspondence is established via the expansion

$$\psi(x) = \sum_n a_n u_n(x), \tag{1}$$

where the $u_n(x)$ are some complete set of functions in x, which provide a basis in Hilbert space.

When making approximate numerical calculations on some given problem which cannot be solved analytically, we most readily fall back on classical mathematical techniques which replace the differential or integral equation by some algebraic equations for the function on a finite set of mesh points. The variational method, with a linear superposition of trial functions, gives us the first connection with an approximation scheme which we can discuss in the Dirac-Heisenberg language. This is studied in some generalized forms in Section II. In Section III we present a couple of new methods contrived entirely within the matrix representation scheme; this is the most interesting part of our work, and we hope to see much more development in these directions.

[1] This research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, under Grant AF-AFOSR-130-66.

In all our examples we shall carry out computations at a sequence of orders of approximation $N = 1, 2, 3,...$ (N will be the number of mesh points or the number of trial functions or the number of dimensions in a restricted subspace of Hilbert space.) The "goodness" of any method is measured by the rapidity with which the desired answer E_N converges to the exact answer E as N increases.

Our first example, illustrating the classical method of discretizing the real line, is the simplest one we could think of which can be worked out exactly:

Example 1: *Calculate the lowest energy eigenvalue for a particle in a one-dimensional infinite square well potential.*

The differential equation and boundary conditions are

$$-\frac{1}{2}\frac{d^2}{dx^2}\psi(x) = E\psi(x), \qquad \psi(\pm 1) = 0. \tag{2}$$

We divide the segment $(-1, 1)$ of the x-axis into $2N$ equal intervals and then use the central difference formula,

$$\frac{d^2\psi(x)}{dx^2} \approx \frac{\psi(x+\delta) - 2\psi(x) + \psi(x-\delta)}{\delta^2}, \tag{3}$$

to convert (2) into the difference equation

$$\psi_{i+1} - 2\psi_i + \psi_{i-1} = -\frac{2E}{N^2}\psi_i, \qquad i = 0, \pm 1, \pm 2,..., \pm N$$

$$\psi_N = \psi_{-N} = 0. \tag{4}$$

This is readily solved to yield

$$E_N = N^2\left(1 - \cos\frac{\pi}{2N}\right) \xrightarrow[N\gg 1]{} \frac{\pi^2}{8}\left(1 - \frac{\pi^2}{48}N^{-2} + \cdots\right). \tag{5}$$

The error here decreases as N^{-2}, which we consider a very slow rate. For solving one-dimensional equations this is still probably the easiest method; for one can have a computer iterate over hundreds or thousands of points without work, worry, or cost of noticeable magnitudes. However, for multidimensional problems (the interesting ones) the situation is very different: in m dimensions, if we need N points per dimension to get the needed accuracy, then a total of N^m points will be involved in the computation. (The standard is often stated as, "Ten points per dimension, and nobody goes beyond three dimensions.") This emphasizes the need for more efficient numerical techniques and places a premium on faster

convergence; so we now turn to the variational methods, which have the reputation of being most powerful.[2]

II. Variational Methods

We now make the standard transition from the Schrödinger equation to matrix mechanics. Starting with

$$(H - E)\,\psi(x) = 0, \tag{6}$$

where H is a given Hermitian operator in the variable x, we introduce the (infinite) expansion (1) to obtain

$$(H - E)\sum_n a_n u_n(x) = 0. \tag{7}$$

Now we left-multiply by $u_m^*(x)$ and integrate over x to obtain the infinite set of algebraic equations

$$\sum_n (H_{mn} - EN_{mn})\,a_n = 0, \tag{8}$$

where

$$H_{mn} = \int u_m^*(x)\,Hu_n(x)\,dx \tag{9}$$

and

$$N_{mn} = \int u_m^*(x)\,u_n(x)\,dx. \tag{10}$$

If the basis functions u_n are orthonormal, then $N_{mn} = \delta_{mn}$; but this is not necessary.

If we could not solve exactly the continuous Eq. (6) then we do not expect to be able to solve the infinite discrete system (8) either. This simplest approximation scheme is to truncate (8) to N equations in N unknowns a_1 , a_2 ,..., a_N . The approximation to the energy E is then the eigenvalue E_N of an $N \times N$ matrix; this computation is readily carried out with the help of a computer once the basis is chosen.

Example 2: *The same problem as in Example* 1. *The exact ground state wave function is an even function of* x *with simple zeroes at* $x = \pm 1$; *so we choose the basis*

$$\begin{aligned} u_1 &= (1 - x^2) \\ u_2 &= (1 - x^2)\,x^2 \\ u_3 &= (1 - x^2)\,x^4, \quad \text{etc.} \end{aligned} \tag{11}$$

[2] Of course there do exist mesh point formulas more accurate than what we have used Eq. (3), and these give convergence at rates faster than that represented by Eq. (5). However, it is this authors impression that these techniques have not been sufficiently well refined for application to multidimensional problems.

The integrals (9) and (10) are elementary (one can also form these basis functions into orthogonal polynomials, but it does not affect the results), and the results are given in Table I. The convergence is fantastically rapid. The reason for this excellent

TABLE I

NUMERICAL RESULTS FOR EXAMPLE 2: LOWEST EIGENVALUE OF A SQUARE WELL

N	E_N	Error
1	1.25	2×10^{-2}
2	1.23372	2×10^{-5}
3	1.233700554	4×10^{-9}
4	1.2337005501365	3×10^{-13}
5	1.23370055013616984	1×10^{-17}
6	1.2337005501361698273545	2×10^{-22}
7	1.2337005501361698273543 11376	10^{-27}
Exact	$\pi^2/8 = 1.2337005501361698273543 11375$	

behavior is not only that the trial functions have the right overall shape, but also that they have appropriate analytic properties.[3] The exact solution is a function, all of whose derivatives, in the region of interest, are finite; and the same holds true for the basis we chose. By contrast the calculation of Example 1 may be described in terms of a set of trial functions which are nonzero only within the segments

$$\frac{i}{N} \leqslant x \leqslant \frac{i+1}{N}$$

and then matched so that value and slope are continuous. The lack of higher order differentiability then accounts for the slow (N^{-2}) convergence rate of that calculation.

Thus in the present method everything depends on the choice of basis functions u_n. One wants them to be sufficiently elaborate to represent the important structural details of the exact solution ψ; yet they must not be so complex that one cannot evaluate the integrals (9) and (10) required to set up the matrix. This bind is felt very keenly in the study of many-particle systems. The simplest trial functions to work with are just products of functions of the individual coordinates; yet one knows that there are important two-body correlations whose expansion in product functions is only slowly convergent. On the other hand, explicit correlation terms introduced into the trial function lead to very difficult multidimensional integrals for the evaluation of the matrix elements of H.

[3] A semiquantitative analysis of the convergence rate of this type of calculation has been attempted by C. Schwartz in "Methods in Computational Physics," Vol. 2, p. 241. Academic Press, New York (1963).

We shall now discuss a generalization of the standard matrix method presented above. Returning to Eq. (7) we left-multiply by some functions $v_m^*(x)$ and then integrate over x. The functions v_n need *not* be related to the functions u_n ; we merely require for each of these two sets of functions that their members be (internally) linearly independent and that they ultimately become complete as their number approaches infinity. We now solve the same matrix Eq. (8), but the definitions of the matrices H and N become

$$H_{mn} = \int v_m^*(x)\, Hu_n(x)\, dx, \tag{12}$$

$$N_{mn} = \int v_m^*(x)\, u_n(x)\, dx. \tag{13}$$

This general method of obtaining approximate solutions to the original Eq. (6) is called "the method of moments,"[4] and its application to atomic physics has recently been discussed by Szondy [1].

Before we go on to discuss the merits of this method of moments and present some examples, we shall first show its connection with the variational principle. Given the eigenvalue problem

$$(H - E) \mid \psi\rangle = 0 \tag{14}$$

and its adjoint problem (H need not be self-adjoint)

$$\langle\psi \mid (H - E) = 0, \tag{15}$$

we consider the quantity

$$J(\chi, \phi) = \langle\chi \mid H - E \mid \phi\rangle \tag{16}$$

for any two functions χ and ϕ. It is obvious that J is stationary with respect to arbitrary variations of the functions χ and ϕ independently when they are infinitesimally close to $\langle\psi \mid$ and $\mid \psi\rangle$, respectively. In practice we shall construct some class of functions over which ϕ can vary and do similarly for χ. Then variation within these classes will give us a best ϕ and a best χ as well as a best estimate for the eigenvalue E, which we shall call $E(\chi, \phi)$. Let us now measure the accuracy of such a calculation by setting

$$\begin{aligned} \chi &= \langle\psi \mid + \langle\Delta_\chi \mid \\ \phi &= \mid \psi\rangle + \mid \Delta_\phi\rangle \\ E(\chi, \phi) &= E + \delta E \end{aligned} \tag{17}$$

[4] See, for example, L. V. Kantorovich and V. I. Krylov, "Approximation Methods of Higher Analysis" (translated by C. D. Benster), p. 150, Wiley (Interscience), New York (1958); or L. Collatz, "Numerische und graphische Methoden" (Encyclopedia of Physics), Vol. ii, p. 438. Springer-Verlag Berlin (1955). The "method of moments" of F. R. Halpern [*Phys. Rev.* **107**, 1145 (1957)] is a special case of this general scheme wherein one chooses $u_n = v_n = H^n u_0$, $n = 0, 1, 2, \ldots$.

and substituting these into (16). We find

$$\delta E\langle\chi \mid \phi\rangle = -J(\chi, \phi) + \langle \Delta_\chi \mid \mathbf{H} - E \mid \Delta_\phi\rangle, \tag{18}$$

which tells us that the error in E is measured by the product of the errors in the two wavefunctions (the quantity J is known as a result of the calculation, and in the methods to be used will always be zero). In the familiar case, $\chi = \phi$, this is the usual statement that the error in the energy is of second order compared with the error in the wavefunction. In the general case we would only wish to know that the inner product $\langle\chi \mid \phi\rangle$ does not vanish, and this will "usually" be so; however, this may occasionally cause us trouble.

We shall use this variational principle with linear trial functions:

$$\phi = \sum_{n=1}^{N} a_n u_n \tag{19}$$

$$\chi = \sum_{m=1}^{N} b_m v_m \,. \tag{20}$$

Then the variation of J with respect to the N parameters b_m (or a_n) leads to the $N \times N$ matrix problem which we had constructed above, (8),

$$\det \mid H_{mn} - EN_{mn} \mid = 0, \tag{21}$$

where H_{mn} and N_{mn} are the same as (12) and (13). In this form, the value of J which goes into formula (18) is obviously zero, and we shall read formula (18) as follows. If we have two sets of N trial functions each, ϕ_1 and ϕ_2, then there are three variational calculations we could do:

$$\begin{aligned} &(\phi_1\,,\ \phi_1) \text{ which leads to an error in the energy of } \delta_{11}\,; \\ &(\phi_2\,,\ \phi_2) \text{ which leads to an error in the energy of } \delta_{22}\,; \\ &(\phi_1\,,\ \phi_2) \text{ which leads to an error in the energy of } \delta_{12}\,. \end{aligned} \tag{22}$$

These three quantities should be related approximately by

$$\delta_{12} \approx (\delta_{11}\delta_{22})^{1/2}. \tag{23}$$

Thus if ϕ_1 were an easy set of functions to work with but gave only modestly accurate results, and ϕ_2 were expected to give a much more accurate representation of the problem but matrix elements in this basis were too difficult to evaluate, then we could consider the mixed $(\phi_1\,,\ \phi_2)$ calculation as a way of getting intermediate accuracy with increased, but manageable, labor.

Another advantage of the method of moments over the usual statement of the variational principle is that it can readily be applied to nonself-adjoint equations. This point of view is in sharp contrast to the conclusion of an earlier work by the present author [2]. There, after several numerical examples, it was stated that a symmetric (i.e., self-adjoint) set of equations in an approximation scheme was much preferred to obtain good convergence. Where that conclusion pertained to the study of the New Tamm-Dancoff method, we have no new comment. However, in regard to the other simple examples adjoined to that earlier work, we must correct an error. Under Method I′ were presented two simple Schrödinger problems (a nonlinear oscillator and a Yukawa potential) which were solved after the equations were deliberately unsymmetrized. The results for the second of these two were just as good as for the usual symmetric variational method, but the first, the x^4 oscillator, showed terrible convergence behavior (see Figs. 2 and 3 of Ref. 2). We have now found that a programming error was entirely responsible for this result, and a corrected rerun of that example has given very nice convergence, in harmony with the results of the other new examples we shall now present.

Example 3: *Solve for the ground state of the hydrogen atom by using the basis functions*

$$r^{n-1}e^{-\alpha r}, \quad n = 1, 2, \ldots, N. \tag{24}$$

Of course with α equal to 1 we get the exact answer with the first term, so we fix α away from 1 and study the convergence of the Nth approximation to the energy toward the exact value. In Fig. 1 are plotted some results from the calculations using $\alpha = 4$ for the set ϕ_1 and $\alpha = 2$ for the set ϕ_2. The relation (23) is very well substantiated by the fact that the curve for the mixed calculation (2, 4) lies midway between the two curves (2, 2) and (4, 4). We can even take from the reference cited in footnote 3 the estimate for the convergence rate for this problem

$$| E_N - E_\infty | \sim \text{const } C_N(\alpha_1)\, C_N(\alpha_2)$$

$$C_N(\alpha) = N \left| \frac{\alpha - 1}{\alpha + 1} \right|^N, \tag{25}$$

and the curves of Fig. 1 are quite well fit by this formula.

Example 4: *The ground state of the helium atom. The general type of basis function we consider is that introduced by Hylleraas*:

$$e^{-ks/2} s^l u^m t^n$$

$$s = r_1 + r_2 \qquad t = r_1 - r_2 \qquad u = r_{12}. \tag{26}$$

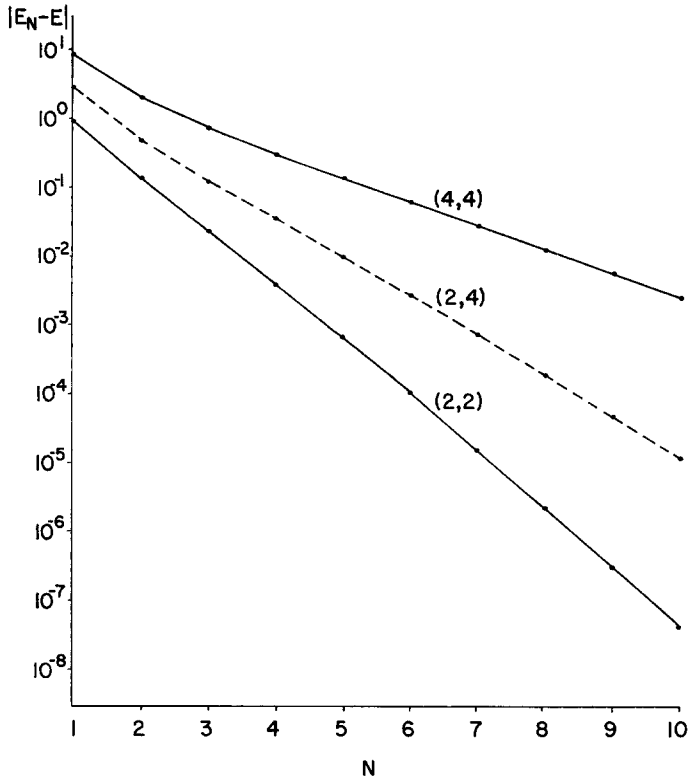

FIG. 1. Numerical results for example 3: convergence of a mixed basis calculation of the ground state energy of hydrogen.

For the ground state n is even, and the basis elements will be grouped and ordered according to the value of the sum $l + m + n$; the value of the scale factor k is kept fixed at 3.7.

The basis ϕ_1 will consist of the subsets of (26) for which the index m is an even integer. From the equation

$$u^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{12} \tag{27}$$

we see that this consists simply of product functions in the coordinates of the two electrons; this may be identified with the so-called configuration interaction basis. This is the simplest type of trial function to use in atomic structure problems, but we know that convergence will be slowed down by the lack of the odd powers of r_{12} ; we have in fact the exact condition

$$\Psi \xrightarrow[r_{12} \to 0]{} \text{const}[1 + \frac{1}{2} r_{12}]. \tag{28}$$

We take for our second basis ϕ_2 just the same set as for ϕ_1, but with each element multiplied by the factor $(1 + \frac{1}{2} r_{12})$ so that the property (28) will be exactly satisfied. The results are shown in Fig. 2. The curve (2, 2) shows results considerably more

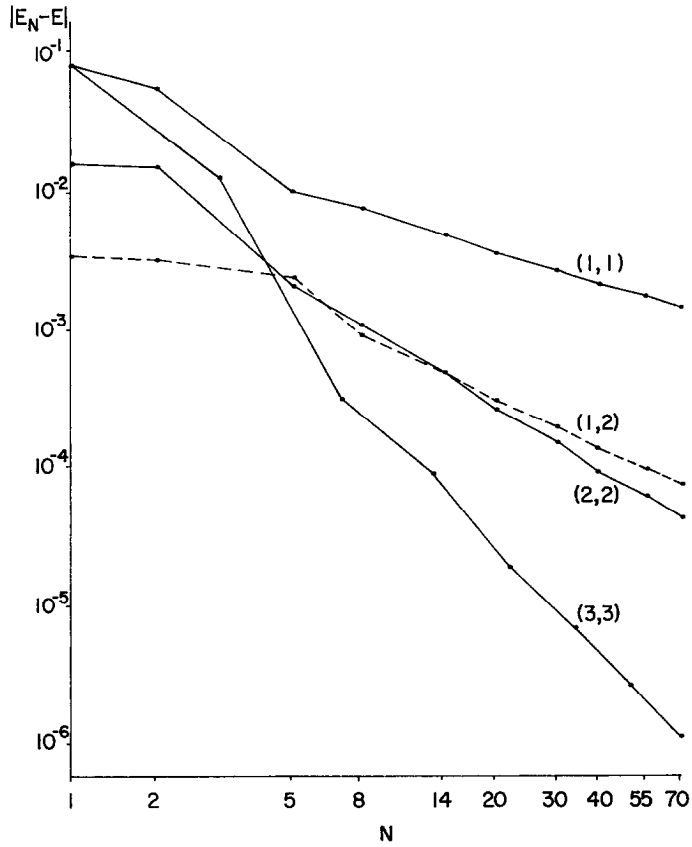

FIG. 2. Numerical results for example 4: convergence of a mixed basis calculation of the ground state of helium.

accurate than the curve (1, 1) as expected; furthermore the curve for the mixed (1, 2) calculation has a slope which is just about halfway between the slopes of the (1, 1) and (2, 2) curves, as predicted by (23). The fact that the (1, 2) curve lies lower than halfway between the other two is probably just an accident.

As far as the two-electron atom is concerned, there is nothing exciting about these results, since that problem has already been worked to death. However, for many-electron atoms we have here a way of putting some two-body correlation

into the wavefunction (on *one* side of the variational principle only) without making the computation of the required matrix elements too difficult. This will need further study.

We should complete this example by considering the third basis ϕ_3 consisting of all the Hylleraas functions (26). The curve (3, 3) in Fig. 2 shows extremely rapid convergence compared with our other curves. However, when we attempted the mixed calculation (1, 3) we encountered our first failure. The results were so erratic that we could not sensibly plot them in Fig. 2. It seems that the trouble lay in the term $\langle \chi \mid \phi \rangle$ on the left side of Eq. (18); that is, the matrix N_{mn} was singular. For the usual symmetrical case this matrix is always positive, and for the non-symmetrical examples we have done up until now we could show again that N was positive; but in this (1, 3) example we were able to see that N had zero eigenvalues, and this apparently not only allows trouble to develop but forces it to occur. Some further study to learn what to do in these cases is needed (see Appendix V).

III. Some Purely Matrix Techniques

In matrix mechanics we are given the Hamiltonian H as some specified function of x and p; x, p, and H are all infinite matrices. Our objective is to invent ways of constructing finite matrices $\langle x \rangle$, $\langle p \rangle$, and $\langle H \rangle$ which approximate these, and then we can easily calculate the eigenvalues and eigenvectors by mechanical means. There is almost nothing to be found in the literature[5] on such approximation techniques, and we should feel free to invent many games. The most obvious hurdle to be overcome is the canonical commutator condition

$$xp - px = iI; \tag{29}$$

this cannot ever be satisfied by finite matrices, and we shall have to learn how to approximate this equation.

Instead of trying to discuss many possibilities which we were not able to carry out, we shall present below a couple of successes which we have found. Our first step (and a rather conservative one) is to choose some known representation for the matrices $\langle a \rangle$, $\langle b \rangle$, etc., where a, b, etc. are the variables in the problem; then we simply carry out algebraic manipulations on these finite (truncated from the infinite) dimensional matrices:

$$\langle ab \rangle \approx \langle a \rangle \langle b \rangle. \tag{30}$$

[5] A recent paper by D. I. Fivel [*Phys. Rev.* **142**, 1219 (1966)] stands as an exception to this statement.

Example 5: *The nonlinear oscillator*

$$H = \frac{p^2}{2} + \frac{x^4}{4}. \tag{31}$$

To define our basis we consider the known linear oscillator problem

$$H_0 = \frac{p^2}{2} + \frac{\omega^2 x^2}{2} \tag{32}$$

with some natural frequency ω chosen for convenience. In the basis of the eigenvectors of H_0 we can construct the well-known infinite matrices

$$x, p, x^2, p^2, x^4 \tag{33}$$

and then truncate these at dimension N to get

$$\langle x\rangle, \langle p\rangle, \langle x^2\rangle, \langle p^2\rangle, \langle x^4\rangle. \tag{34}$$

For our first game we simply take

$$\langle H\rangle = \tfrac{1}{2}\langle p^2\rangle + \tfrac{1}{4}\langle x^4\rangle \tag{35}$$

and then compute the eigenvalue E_N. It should be obvious that this is not a new game, but it is exactly the same as the variational method of Section II. Numerical results showing very rapid convergence for the ground state are shown in Fig. 3. (We have used $\omega = (3/2)^{1/3}$, and these results were already given in Ref. 2.)

For our second game we use instead of $\langle x^4\rangle$ the result of the finite matrix multiplication

$$\langle x^2\rangle \cdot \langle x^2\rangle \tag{36}$$

to get $\langle H\rangle$. The resulting eigenvalues are also shown in Fig. 3, and we see that the convergence is just as good as that of the first game above.

Finally, for the third game, we go all the way with

$$\langle H\rangle = \tfrac{1}{2}\langle p\rangle \cdot \langle p\rangle + \tfrac{1}{4}\langle x\rangle \cdot \langle x\rangle \cdot \langle x\rangle \cdot \langle x\rangle. \tag{37}$$

These results, also shown in Fig. 3, again converge just about as well as those of the first game.

Games two and three may be read as further approximations made within the $N \times N$ approximation of the usual variational method (game one). It is delightful to see that nothing essential has been lost in this way, and maybe much flexibility has been gained. (The new results do oscillate about the exact answer while the old method gave strictly an upper bound, but we consider this an unimportant

aspect of the convergence study.) After having found these results we can justify them as follows. The difference between say $\langle x^2 \rangle$ and $\langle x \rangle \langle x \rangle$ lies only in the farthest corner of the $N \times N$ matrix. The eigenvector which we are constructing has only

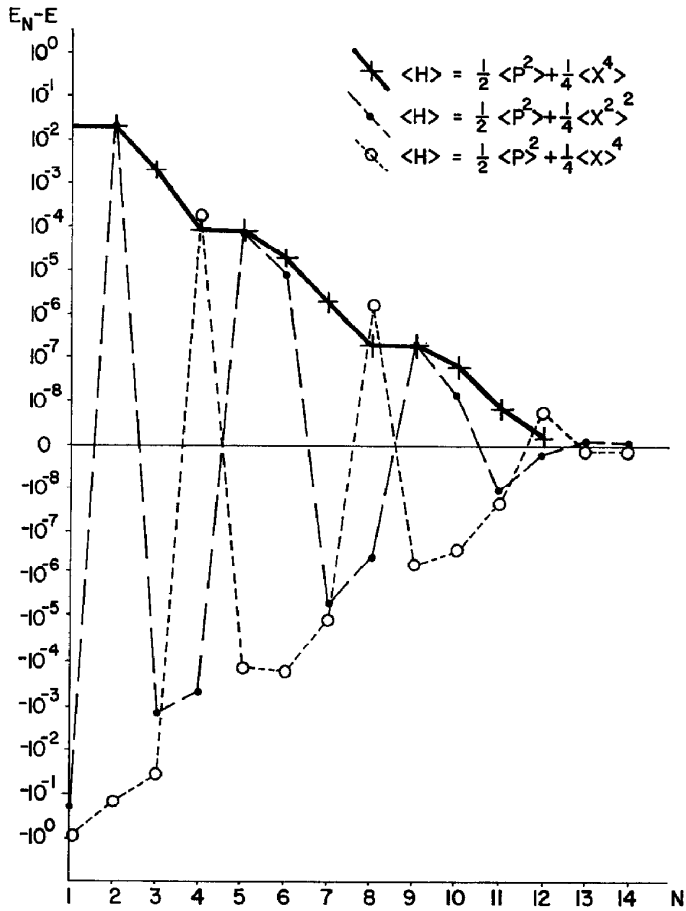

FIG. 3. Numerical results for example 5: convergence of three different matrix representations for the ground state of the nonlinear oscillator (31).

a very small component in this last element of the space (this we know because the first method was converging rapidly), and so the additional error is quite insignificant.

It is rather difficult to develop a feeling for how to measure the "smallness" of the error made in some general approximation method which might be suggested,

and this is the biggest drawback to progress in these searches. Thus, because of the formal equation (29), we were very hesitant about ever even attempting game three. One can easily see the error made here:

$$[\langle x\rangle, \langle p\rangle] - i\langle I\rangle = -i \begin{bmatrix} 0 & 0 & \cdots & & 0 \\ 0 & 0 & & & \\ \vdots & & \ddots & & \\ & & & 0 & 0 \\ 0 & & & 0 & N \end{bmatrix}, \tag{38}$$

and the matrix on the right does not look small (cf. Appendix II).

This exercise with the oscillator has been amusing and possibly stimulating, but it is not a problem of practical interest. We turn now to another example of these games which may be somewhat practical.

Example 6: *In a recent work* [3] *on scattering calculations we were led to rewrite the standard equation for the T-matrix,*

$$T = V + VGT, \tag{39}$$

where

$$G = (E - H_0 + i\epsilon)^{-1}, \tag{40}$$

in the form,

$$V^{-1}T = I + GT \quad \text{[see footnote 6]}; \tag{41}$$

and then construct the symmetric variational principle

$$[T] = 2T + TGT - TV^{-1}T. \tag{42}$$

Trial functions were then introduced for T, and the integrals could be easily carried out for a simple potential $V(r)$. However if V were some more complicated operator we wondered how one could effectively evaluate the matrix elements of $\langle V^{-1}\rangle$, and our suggestion is now to try to use $\langle V\rangle^{-1}$ instead, in the same spirit as we played games with the oscillator problem. However, since we shall be dealing with a nonorthonormal basis for these matrices a little formal preparation is necessary.

Given two complete sets of functions, $u_n(x)$ and $v_m(x)$, we construct the general resolution of the identity:

$$\delta(x - y) = \sum_{n,m} M_x \mid u_n(x)\rangle\, W_{nm}\langle v_m(y)\mid N_y\,, \tag{43}$$

[6] If one rewrites this as $T^{-1} = V^{-1} + (H_0 - E)^{-1}$, an amusing analogy with electric circuit theory is suggested.

where M and N are any two nonsingular operators and the numbers W_{nm} will now be determined. If we multiply (43) by $\langle v_k(x) \mid N_x$ and integrate over x we get

$$\langle v_k(y)| N_y = \sum_{n,m} \langle v_k \mid NM \mid u_n\rangle W_{nm}\langle v_m(y)| N_y ; \tag{44}$$

thus from the assumed linear independence of the functions v_m we can conclude

$$\sum_n \langle v_k \mid NM \mid u_n\rangle W_{nm} = \delta_{km} . \tag{45}$$

Now using (43) we can expand the general matrix element of a product as

$$\langle v_k \mid AB \mid u_l\rangle = \sum_{n,m} \langle v_k \mid AM \mid u_n\rangle W_{nm}\langle v_m \mid NB \mid u_\ell\rangle. \tag{46}$$

This is so far rigorous, but we now use the notation $\langle X\rangle$ to mean the finite $(N \times N)$ matrix of elements $\langle v_k \mid X \mid u_\ell\rangle k$, $\ell = 1, 2,..., N$, and we have our approximate matrix multiplication formula

$$\langle AB\rangle \approx \langle AM\rangle\langle NM\rangle^{-1}\langle NB\rangle. \tag{47}$$

For the case $M = N = I$, and $\langle v_k \mid u_\ell\rangle = \delta_{k\ell}$, this is just the formula

$$\langle AB\rangle \approx \langle A\rangle \cdot \langle B\rangle$$

which we used above, Eq. (30). For the present problem we choose $A = KS$, $B = S^{-1}L$, and then manipulate (47) to get

$$\langle NS^{-1}L\rangle \approx \langle NM\rangle\langle KSM\rangle^{-1}\langle KL\rangle. \tag{48}$$

Now, getting back to the scattering calculation, consider the potential

$$V = e^{-r}/r. \tag{49}$$

The trial functions for T (which is $V\psi$) behave like $1/r$ near the origin so that the matrix elements $\int TVT$ are not defined. We therefore make the following break-up to use formula (48):

$$K = N = r, \qquad S = e^{-r}, \qquad L = M = 1; \tag{50}$$

$$\left\langle \left(\frac{e^{-r}}{r}\right)^{-1} \right\rangle \approx \langle r\rangle\langle re^{-r}\rangle^{-1}\langle r\rangle. \tag{51}$$

The matrix elements in (51) are now easy to calculate, and so we repeated the phase shift calculation given in Ref. 3 by using the result of the matrix mani-

pulation of Eq. (51) to replace $\langle V^{-1}\rangle$. The results were not very good; with sufficiently careful handling of the data one could see some convergence to the right answer, but the situation was not clear and at best the convergence was quite slow. The following discussion seems to be a reasonable explanation of this poor result.

We should expect that for the truncated Eq. (48) to be accurate the finite basis that we have kept must be able to give a good representation of the operators S and S^{-1}. In the above example S and S^{-1} are two very different operators in the region of large values of the coordinate r, and the large r region is probably not negligible in the scattering problem (as it might be in a bound state problem). However, the trial functions used went as $r^n e^{-\alpha r}$, and these functions are not very efficient at building up behavior at large r. Ideally we should arrange to invert an operator S which is as close to the identity as we can manage. The choice $K = N = V^{-1}$, $S = 1$, $L = M = 1$ leaves us with

$$\langle V^{-1}\rangle \approx \langle V^{-1}\rangle\langle V^{-1}\rangle^{-1}\langle V^{-1}\rangle, \tag{52}$$

which is an identity and not interesting.

Now we are led to consider a more complex situation:

$$V_2 = \frac{e^{-r}}{r} + \sigma\frac{e^{-2r}}{r}, \tag{53}$$

TABLE II

NUMERICAL RESULTS OF CALCULATIONS OF THE SCATTERING LENGTHS ($a = -\tan\delta/k$ at $k = 0$) FOR THE POTENTIAL $V = -2(e^{-r}/r)(1 + \sigma e^{-r})$[a]

N	a for $\sigma = 0$	a for $\sigma = 2$	a for $\sigma = -2$
1	8.0938697	1.9231311	0.0529962
2	8.0086267	0.3977105	−1.1506452
3	7.9122814	0.5427499	−1.4369220
4	7.9117816	0.2816501	−1.4874346
5	7.9117804	0.2947919	−1.5078973
6	7.9114674	0.2687904	−1.5101818
7	7.9114205	0.2692808	−1.5111852
8	7.9113885	0.2668049	−1.5113125
9	7.9113829	0.2667269	−1.5113533
10	7.9113807	0.2664709	−1.5113611
11	7.9113804	0.2664439	−1.5113621
12	7.9113802	0.2664083	−1.5113624

[a] The basis functions for the $N \times N$ matrices are those given in Ref. 3 with $\alpha = 1.8$; and the matrix of V^{-1} is constructed according to (48) and (54).

and make the decomposition

$$K = N = \left(\frac{e^{-r}}{r}\right)^{-1}, \qquad S = (1 + \sigma e^{-r}), \qquad L = M = 1. \tag{54}$$

Here the quantity S which our matrix must invert varies very little from unity, and the envelope of the function V_2 is handled exactly by the N term. We carried out numerical calculations with the potential $-2V_2$, and Table II shows the results. For $\sigma = 0$ this is just a repeat of the old method, which converges very nicely; the convergence for the cases $\sigma = +2$ and $\sigma = -2$ appears to be just about as good, and this is very satisfying. The situation for $\sigma = -2$ is particularly delightful since this potential has a node and the original matrix elements $\langle V^{-1} \rangle$ would have been somewhat tricky to handle, but in this matrix inversion game apparently everything took care of itself automatically.

Summary

Several new or expanded techniques for approximation schemes have been discussed and illustrated. It is expected that these will add to our flexibility and power to attack hard problems in the future. It is also hoped that further invention in the realm of matrix games will be stimulated by our novel successes of Section III. Our mathematical discussions about the theoretical soundness of the methods are admittedly very weak; most of our work has been really "experimental numeracity."

The appendices present several discussions attemuting to explore further various peripheral questions rvised by our studies.

Appendix I. Eigenvalues of Non-Hermitian Matrices

In Section II we considered non-Hermitian matrices representing the Hamiltonian of some problems and then computed eigenvalues. We are familiar with the theorem that any eigenvalue of a Hermitian matrix must be real. We now ask the question: Should we be surprised that an eigenvalue of a non-Hermitian matrix comes out to have zero imaginary part? It is a fact that in all the numerical examples presented here, we had no difficulty in finding purely real eigenvalues when we looked for them.

Imagine that we can choose a basis in which the matrix in question is real (this is so in all our examples). Now start with a symmetric matrix which is somehow close to the given one, and then continuously vary the matrix elements until the actual matrix is reached. The eigenvalues of the starting (symmetric) matrix all lie on the real axis, and then they all move continuously as the matrix is varied. Since the secular equation has purely real coefficients, any nonreal roots must

occur in complex conjugate pairs. Thus the only way in which any eigenvalue can move off the real axis is for two eigenvalues to meet on the real line and then go off into the complex plane as a reflected pair.

The above discussion is intended to show that it is not completely easy for eigenvalues to leave the real axis; we can add only a plausibility argument to say why in our examples they in fact did not become complex. If the two sets of functions u_n and v_m introduced in Section II do indeed give a good representation of the actual wavefunction ψ, then we may expect the emerging eigenvalue spectrum to be a good approximation to the true spectrum. The true spectrum consists of well-separated points on the real line, and so we may not be surprised that there has been no opportunity for the approximate eigenvalues to meet and then become complex.

Appendix II. The Commutator Problem

In Section III it was noted that the canonical commutator condition, Eq. (29), could never be satisfied by finite matrices trying to represent the operators x and p, and furthermore it was not clear how to measure the amount of error one may tolerate in approximating this condition. We present here another attempt which, by its failure, further illustrates this situation.

Assume that we are working in the space of $N \times N$ matrices, we have some Hermitian matrices x and p, and we want to study the error matrix

$$\Delta = [x, p] - iI. \tag{A1}$$

We would like Δ to be small in some sense, so let us consider the error function $\mathfrak{E}$ defined by

$$\mathfrak{E} = \text{trace } \Delta^+\Delta = \sum_{\alpha,\beta=1}^{N} |\Delta_{\alpha\beta}|^2. \tag{A2}$$

In order to proceed it is convenient to change basis; since x is Hermitian there is a unitary matrix U which will diagonalize it.

$$\begin{gathered} x = U\tilde{x}U^+; \qquad U^+U = UU^+ = I; \qquad p = U\tilde{p}U^+ \\ \tilde{x}_{\alpha\beta} = \tilde{x}_{\alpha\alpha}\delta_{\alpha\beta}\,. \end{gathered} \tag{A3}$$

We now calculate the error function,

$$\begin{aligned} \mathfrak{E} &= \sum_{\alpha,\beta=1}^{N} |\tilde{\Delta}_{\alpha\beta}|^2 = \sum_{\alpha,\beta=1}^{N} |\tilde{p}_{\alpha\beta}(\tilde{x}_{\alpha\alpha} - \tilde{x}_{\beta\beta}) - i\delta_{\alpha\beta}|^2 \\ &= \sum_{\alpha,\beta=1}^{N} |\tilde{p}_{\alpha\beta}|^2 (\tilde{x}_{\alpha\alpha} - \tilde{x}_{\beta\beta})^2 + N; \end{aligned} \tag{A4}$$

and find that its minimum occurs for

$$|\tilde{p}_{\alpha\beta}| = 0 \qquad \text{if} \quad \tilde{x}_{\alpha\alpha} \neq \tilde{x}_{\beta\beta}\,, \tag{A5}$$

which is precisely the condition that p and x commute. This is a ridiculus situation since, for example, it implies that the Hamiltonian will be diagonal when x is. By contrast, the approximation made in Example 5 [Eq. (38)] gives an error $\mathfrak{E} = N^2$ compared with $\mathfrak{E} = N$ above, but that approximation worked very well.

Appendix III. Accuracy of the Matrix Approximations

In Section III we approximated infinite sums by finite sums in order to play some new matrix approximation games. Let us look at the error made in going from the exact Eq. (46) to the approximate equation (47); the terms lost can be separated as

$$\sum_{n\leqslant N}\sum_{m>N} \tag{A6}$$

$$\sum_{n>N}\sum_{m\leqslant N} \tag{A7}$$

$$\sum_{n>N}\sum_{m>N} \tag{A8}$$

where the summand in each case is

$$\langle v_k \mid AM \mid u_n\rangle W_{nm}\langle v_m \mid NB \mid u_\ell\rangle. \tag{A9}$$

We now wish to show that the first two terms, (A6) and (A7), are effectively zero, and thus the error (A8) may be said to be small *of second order*. Consider the set of functions u_n for $n > N$; the calculation in our N-dimensional subspace should be unaffected by a redefinition of the basis outside this subspace, and so we choose to make these "outside" u_n orthogonal to all the "inside" $v_m (m \leqslant N)$, via the weighting operator NM. We redefine similarly the outside v_m to be orthogonal, via NM to the inside u_n. Thus the infinite matrix NM reduces to an $N \times N$ block and an $(\infty - N) \times (\infty - N)$ block with no connecting elements. The matrix W is just the inverse of NM, and so it too has no elements connecting the "in" to the "out" bases. There may be some conditions necessary for the above analysis to be correct (e.g., if the set u_n is identical with the set v_m then we would like NM to be Hermitian, or at least normal) but we expect they can easily be satisfied.

This second order smallness for the error in our matrix manipulations is similar to the second-order smallness in the error in the eigenvalue as given by the varia-

tional principle, and this may explain why we get comparably accurate results with our new games as with the older methods.

Appendix IV. Recursion Methods

One of the oldest methods for solving differential equations is to make a power series expansion, collect terms of a given power, and thus get recursion formulas for the expansion coefficients. We wish to see now how this general method may be compared with the general method of moments of Section II. Let us start with the general equation

$$\mathfrak{L}\psi = 0, \tag{A10}$$

where $\mathfrak{L}$ is some linear operator. We postulate the approximate expansion

$$\psi = \sum_{n=1}^{N} a_n u_n , \tag{A11}$$

and the assume that we can find some recursion formula which gives the result of $\mathfrak{L}$ acting on each u_n:

$$\mathfrak{L}u_n = \sum_{\ell} B_{\ell n} w_\ell . \tag{A12}$$

In general the two complete sets u_n and w_n need not be identical, and the sum (A12) may contain an infinite number of terms. The appropriate equation now reads

$$\sum_{n=1}^{N} a_n \sum_{\ell} B_{\ell n} w_\ell \approx 0, \tag{A13}$$

and since there are N unknowns a_n, we set separately to zero the coefficients of the first N (linearly independent) functions w:

$$\sum_{n=1}^{N} B_{\ell n} a_n = 0 \qquad \ell = 1, 2, \ldots, N. \tag{A14}$$

The first question we wish to ask about this recursion method (A14) is: When is this equivalent to some form of the method of moments? Starting with (A10) and (A11) the method of moments gave the approximate solution as:

$$\sum_{n=1}^{N} \mathfrak{L}_{mn} a_n = 0 \qquad m = 1, 2, \ldots, N, \tag{A15}$$

where

$$\mathfrak{L}_{mn} = \langle v_m \mid \mathfrak{L} \mid u_n \rangle, \tag{A16}$$

and the functions v_m have been chosen somehow. By using (A12) we can write this matrix element as

$$\begin{aligned} \mathfrak{Q}_{mn} &= \sum_{\ell} \langle v_m \mid w_\ell \rangle \, B_{\ell n} \\ &\equiv \mathfrak{Q}^{<}_{mn} + \mathfrak{Q}^{>}_{mn} , \end{aligned} \tag{A17}$$

where the sign $<$ indicates those terms of the sum (A17) for which $\ell \leqslant N$ and the sign $>$ indicates those for which $\ell > N$. Now if the term $\mathfrak{Q}^{>}$ vanishes, then we can see that (A15) is equivalent to (A14), and thus the two methods are the same: we need simply multiply the equations (A15) from the left by the inverse of the matrix

$$S_{mk} = \langle v_m \mid w_k \rangle \quad m, k = 1, 2, \ldots, N, \tag{A18}$$

and we are assured that this inverse exists.

The easiest way to assure the vanishing of $\mathfrak{Q}^{>}$ is to require

$$\langle v_m \mid w_\ell \rangle = 0 \quad \text{for all} \quad m = 1, 2, \ldots, N \quad \text{and all} \quad \ell = N + 1, N + 2, \ldots . \tag{A19}$$

This (A19) is thus a sufficient condition for the equivalence of the two methods. We cannot prove that it is also a necessary condition, but would guess that it probably is. The simplest way in which to satisfy (A19) would be to have an orthogonality condition,

$$\langle v_m \mid w_\ell \rangle = \delta_{m\ell} . \tag{A20}$$

[The following argument—patterned after what we did in Appendix III—might seem to make the above equivalence more general. Suppose, for simplicity, we had $w_n = u_n$. Let us define a new set of functions u'_n as

$$\begin{aligned} u'_n &= u_n \qquad n = 1, 2, \ldots, N \\ u'_n &= u_n + \sum_{m=1}^{N} \Lambda_{nm} u_m \qquad n = N + 1, N + 2, \ldots . \end{aligned} \tag{A21}$$

We can choose the coefficients Λ so as to satisfy condition (A19) by the Schmidt orthogonalization procedure. Since the functions u_n for $n \leqslant N$ have not been changed, one might think that we have now demonstrated a more general equivalence of the two methods. This is a false conclusion since now the matrix B actually has been changed. For example, suppose that contained in the original recursion formula was

$$\mathfrak{Q} u_N = B_{N+1,N} u_{N+1} + \text{other things.} \tag{A22}$$

Our truncation procedure would tell us to drop this u_{N+1} term. However, after the redefinition (A21) the right side of (A22) would contain terms

$$B_{N+1,N} \left(\sum_{m=1}^{N} \Lambda_{N+1,m} u'_m \right), \tag{A23}$$

which we must now keep.]

We would next like to try to answer the question: If the recursion method is not equivalent to a variational method, does it have the same general order of accuracy, or is it expected to be worse? It is our guess that in general the recursion method is accurate only to first order and thus is weaker than the variational methods. For example, consider again the infinite square well and take as the expanded wave function

$$\phi_N = \sum_{\ell=1}^{N} C_\ell (1 - x^2)^\ell. \tag{A24}$$

Substituting this into the wave equation and collecting powers of $(1 - x^2)$ we get the algebraic system

$$4\ell(\ell + 1)\, C_{\ell+1} - 2\ell(2\ell - 1)\, C_\ell + 2EC_{\ell-1} = 0 \qquad \ell = 1, 2, \ldots, N \tag{A25}$$

with the constraints

$$C_0 = C_{N+1} = 0. \tag{A26}$$

The eigenvalues E_N are readily computed and the results converge at just half the rate of those from the variational method shown in Table I. (Thus at 15 terms of the present development we get an eigenvalue which is as accurate as that obtained with 7 terms in the previous scheme.) This is the first-order accuracy we claim will generally result from these schemes. Obviously if we reorganize the basis (A24) we will get different results because the matrix B will be changed, and it appears that the best reorganization would be to get some appropriately orthogonal polynomials, for then—by the above discussion—we would be back at the variational method with its second order accuracy.

In an attempt at a more general analysis of the accuracy (first or second order) of the recursion scheme we can proceed as follows. From the general problem

$$\mathfrak{L}\psi = 0 \tag{A27}$$

with the expansion

$$\psi = \sum c_n u_n, \tag{A28}$$

we get by our recursion (or any other) method the infinite matrix problem

$$B \cdot c = 0 \tag{A29}$$

which exactly represents the original equation. The Nth approximation amounts to truncating this to an $N \times N$ matrix problem

$$\tilde{B} \cdot \tilde{c} = 0, \tag{A30}$$

and if λ is an eigenvalue in (A29) then $\tilde{\lambda}$ is the corresponding eigenvalue in (A30). We can postulate the existence of dual vectors d and $\tilde{d}$ which solve

$$d \cdot B = 0 \tag{A31}$$

$$\tilde{d} \cdot \tilde{B} = 0. \tag{A32}$$

If we write, for concreteness, $B = K - \lambda M$ then we can do the standard error analysis: (A33)

$$\begin{aligned}\langle \tilde{d} \mid B(\tilde{\lambda}) \mid \tilde{c} \rangle &= \langle \tilde{d} \mid B(\lambda) \mid \tilde{c} \rangle + (\lambda - \tilde{\lambda}) \langle \tilde{d} \mid M \mid \tilde{c} \rangle \\ &= \langle d \mid B \mid c \rangle + \langle \tilde{d} - d \mid B \mid c \rangle + \langle d \mid N \mid \tilde{c} - c \rangle \\ &\quad + \langle \tilde{d} - d \mid B \mid \tilde{c} - c \rangle + (\lambda - \tilde{\lambda}) \langle \tilde{d} \mid M \mid \tilde{c} \rangle. \end{aligned} \tag{A34}$$

The left-hand term is zero by (A30) or (A32), and the first three terms on the right are zero by (A29) and (A31); and so we have

$$(\tilde{\lambda} - \lambda) \langle \tilde{d} \mid M \mid \tilde{c} \rangle = \langle \tilde{d} - d \mid B \mid \tilde{c} - c \rangle, \tag{A35}$$

which is just like Eq. (18) and gives us the mixed second-order accuracy. The problem is to be able to estimate the error associated with the approximation (N term truncation) of the dual vector d (we presumably start with some estimate of the accuracy of the c-vector approximation; this defines our "first order error"). If the overall accuracy in the eigenvalue is no better than that of the c-vector (i.e., first order), then we would guess that the accuracy of the d-vector approximation was essentially nil (i.e., order of 100% error). We shall shortly try to see this in some example. On the other hand, if we start with some given c-vector approximation sequence but have at our disposal the various methods of building the matrix B, we might be drawn to try to select a method that gives a self-adjoint matrix B, for then the d-vector is the same as the c-vector and we could expect the better second order accuracy. Thus the reliance on symmetric equations (as promulgated in Ref. 2) represents not really a great ideal, but rather a reasonable guide when any further analysis is absent. This discussion suggests why the improved calculation of the NTD method in Ref. 2 worked well, but it also suggests that there may be other (even better or simpler) ways of gaining the objective of better convergence for that technique.

(It should not be inferred here that the recursion methods are intrinsically inferior to the variational methods. Given a good set of expansion functions, the u_n, one can reduce the accuracy of the variational method by using some very poor set of weight functions v_m—the dual set here. However, having in mind from the outset that one must choose two sets, one will probably choose them both wisely and get "second-order good" results. On the other hand, if one approaches the recursion method without realizing the role played by the dual vector, a great deal of accuracy may be needlessly lost.)

To illustrate the importance of the dual vector in the recursion method, we return to the square well problem and the expansion (A24). The dual of the recursion formula (A25) is

$$4\ell(\ell - 1)\, D_{\ell-1} - 2\ell(2\ell - 1)\, D_\ell + 2ED_{\ell+1} = 0, \qquad \ell = 1, 2,..., \tag{A36}$$

and if we interpret this as the direct recursion formula for an expansion of the original wavefunction as

$$\chi = \sum_{\ell=1} D_\ell v_\ell\,, \tag{A37}$$

we find the behavior

$$v_\ell \xrightarrow[x\to\pm 1]{} \text{const.}\,(1 - x^2)^{a-\ell}. \tag{A38}$$

This represents an extremely ill-chosen basis and explains why this dual vector does nothing to raise the convergence rate above the "first-order" rate which we observed.

This sort of result might be seen by the following alternative approach. We showed above that the recursion method was equivalent to a moments method with the functions v_m chosen so that

$$\langle v_m \mid w_n \rangle = \delta_{mn}\,. \tag{A39}$$

If, as above, the basis functions u_n are simply powers of a single variable

$$u_n = y^n,$$

then the appropriately orthogonal functions v_m may be represented as

$$v_m = \frac{1}{m!}\left(\frac{-d}{dy}\right)^m \delta(y). \tag{A40}$$

This is a very singular function [as is (A38) in the above example where $y = (1 - x^2)$], and again we would guess that this dual basis adds nothing to the first-order convergence rate.

One might expect that as the original u_n basis was gradually modified from simple powers to orthogonal polynomials, one would see the corresponding smooth improvement in the nature of the dual basis and smooth improvement of the convergence rate. However, we have no example to demonstrate this.

Appendix V. The Ill-Conditioned Problem in the Method of Moments

In our analysis of the method of moments we noted that $\langle \chi \mid \phi \rangle$ should not vanish if we wished to obtain second-order accuracy for the eigenvalue, and in

reporting the failure of the (1, 3) computation on the helium atom we thought this was the source of trouble. We are not at all sure about this for the following reasons. It seems that all that occurred in that example was the vanishing of some eigenvalue of the metric matrix N; we would expect that the functions χ and ϕ should be close to the true function ψ and thus their inner product would not vanish. Consider the following 2×2 matrix problem

$$\det \left\| \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} - E \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} \right\| = 0. \tag{A41}$$

This seems to typify the point in question; we have assumed a basis in which, for simplicity, the metric matrix is diagonal and we have shown one of its eigenvalues as being variable (we shall let it approach zero); the matrix elements of H may be assumed all finite. Now solving the quadratic and letting ϵ get small, we have the eigenvalues

$$E_1 \xrightarrow[\epsilon \to 0]{} H_{11} - \frac{H_{12}H_{21}}{H_{22}} + 0(\epsilon) \tag{A42}$$

$$E_2 \xrightarrow[\epsilon \to 0]{} \frac{H_{22}}{\epsilon} \qquad + 0(1). \tag{A43}$$

Thus the eigenvalue E_2 behaves badly, but E_1 seems to be quite reasonable, and this is the answer we are interested in if indeed the 1×1 approximation was close. The indicated conclusion is that the vanishing eigenvalue of the metric matrix does not completely wreck the calculation, but we have not been able to see why we did in our (1, 3) example lose the second-order accuracy which was expected.

References

1. T. Szondy, *Acta Phys. Hung.* **17**, 303 (1964); E. Szondy and T. Szondy, *Acta Phys. Hung.* **20**, 253 (1966).
2. C. Schwartz, *Ann. Phys. (N.Y.)* **32**, 277 (1965).
3. C. Schwartz, *Phys. Rev.* **141**, 1468 (1966).