ORIGINAL PAPER

The Niblack–Nigam iterative method revisited

J. P. Killingbeck · G. Jolicard

Abstract   The iterative algorithm of Niblack and Nigam for matrix eigenvalues is derived in a simple manner and is improved so as to permit reliable numerical calculations. The computational capabilities of the algorithm are tested on a bound state problem and it is then shown how to extract eigenvector information from numerical eigenvalues. A complexified form of the algorithm is shown to permit the accurate calculation of complex resonant state energies.

Keywords   Quantum physics · Matrix eigenvalues · Resonances

1 Introduction

In 1970 Niblack and Nigam [1] described an iterative algorithm for the calculation of matrix eigenvalues. They entitled their paper "an alternative to perturbation theory" and pointed out that, although their equations were derived in a manner resembling that used in the treatment of Brillouin–Wigner perturbation theory, the resulting calculation would work even for the case of strong perturbations (as measured by the magnitude of the matrix elements of the perturbing potential V in a basis set which is composed of the eigenfunctions of the unperturbed Hamiltonian). Although the final algorithm is quite simple to apply, the derivation of the algorithm given in [1] was lengthy and complicated and the only numerical application given was to the case of

J. P. Killingbeck · G. Jolicard
Institute UTINAM (CNRS UMR 6213), 41bis Avenue de l'Observatoire, BP 1615,
25010 Besancon cedex, France

J. P. Killingbeck (✉)
Centre for Mathematics, Loten Building, University of Hull, Hull HU6 7RX, UK
e-mail: j.p.killingbeck@hull.ac.uk

a harmonic oscillator with a trivial quadratic perturbation. These two features of [1] presumably explain why the method has not been taken up by other workers; a lengthy literature search did not reveal any subsequent use of the method in works dealing with either perturbation theory or matrix eigenvalues.

In the present work, we give a simplified derivation of that part of the algorithm which permits the calculation of matrix eigenvalues and point out that it can be used in different ways, whereas in [1] only one approach was envisaged. In [1] the case of real matrices and eigenvalues only was considered and the iterative procedure gave a sequence of increments to be added to an initial eigenvalue estimate. The final eigenvalue was thus found from the sum of these successive increments, so that questions of convergence could arise in principle, even though the method was more effective than a traditional perturbation approach. In a modern version of the method there are two obvious improvements which can be made. First, a complex variable version of the method can be constructed, making possible the calculation of the complex eigenvalues associated with resonance states. Second, a real or complex form of the Wynn algorithm can be used to deal with cases in which the computed sequence of increments does not converge rapidly to a finite sum.

In Sect. 2 we give an approach to the derivation of the eigenvalue algorithm which is more simple and direct than that of [1]. In Sect. 3 we outline some links between the Niblack–Nigam method and other techniques of eigenvalue calculation. Section 4 gives some bound state test calculations which explore the numerical performance of the algorithm. Section 5 shows that a direct search variant of the original method of [1] permits calculation of the eigenvalues even when the iterative form of the algorithm leads to a divergence. Section 6 explains how the use of an eigenvalue shift method can give selected components of a normalized eigenvector. Section 7 extends the iterative method by using a complex set of basis functions to study some complex resonant state energies. Section 8 gives a brief discussion.

2 Derivation of the algorithm

We start with the standard assumption that we have a set of normalized basis functions $|n)$ which are the eigenfunctions with energy E_n of an unperturbed Hamiltonian H_0. The matrix of the perturbed Hamiltonian $H = H_0 + V$ is then set up in this basis. We wish to find the perturbed eigenvalue ϵ which arises from the initial energy E_I of the unperturbed state $|I)$. The perturbed wave function Ψ is taken (as usual) to have an inner product of unity with the unperturbed wave function and so we can set $\Psi = \Phi + |I)$, where Φ is orthogonal to $|I)$. The perturbed eigenvalue equation can thus be written in the form

$$(H - \epsilon)\Phi + (H - \epsilon)|I) = 0 \tag{1}$$

Taking an inner product from the left with $|I)$ and remembering the orthogonality properties as well as the two terms included in H, we can re-arrange the resulting terms to obtain the standard equation

$$\in = E_I + (I \,|\, V \,|\, \Psi) \tag{2}$$

This equation, although of a familiar form, is NOT here associated with a perturbation expansion involving various orders of λ for a perturbation λV. A simple way to obtain the results of [1] is to postulate a tentative iterative form for Eq. (1):

$$(H - \in(n))\Phi + (H - \in(n+1)) \,|\, I) = 0 \tag{3}$$

We now encounter the technical problem which was solved by a more complicated route in [1] ; we wish to obtain an eigenvalue algorithm while avoiding the explicit calculation of the unknown wave function. We apply a linear operator A from the left and then take an inner product from the left with | I). The resulting equation is

$$(I \,|\, A \,(H - \in(n)) \,|\, \Phi) + (I \,|\, A \,(H - \in(n+1))|\, I) = 0 \tag{4}$$

We do not know Φ but we DO know that it must be some linear combination of the basis states | m) which specifically excludes | I). Thus, if the operator A renders zero all the off-diagonal matrix elements (I | A (H − ∈ (n))| m) then the first term in Eq. (4) must be zero. The second term must then also be zero. Remembering that it is the matrix of H − ∈ (n) which is being acted upon by A, we can rewrite the result that the second element on the right in (4) is zero as follows:

$$(I|A(H - \in(n)) \,|\, I) = (I \,|\, A(\in(n+1) - \in(n)) \,|\, I) \tag{5}$$

If the operator A also has the property that it leaves | I) invariant then the surviving diagonal element gives us the shift in the ∈ estimate and thus gives an operational meaning to our postulated iterative procedure. At first it might appear that finding an operator A with the desired properties might be difficult. However, on remembering the methods associated with Gaussian elimination, we see that the traditional process of subtracting a multiple of a row of the matrix H−∈(n) from other rows would be able to produce the required results. To eliminate the (I, J) element of the matrix (H−∈(n)) a multiple of row J is subtracted from row I; multiples of row J are also subtracted from the rows K with K > J so as to render the (K, J) element zero (thus ensuring that the use of later rows does not disturb the zero value of the (I, J) element. The row I is only acted UPON and does not participate in the subtraction process (this is the way in which the invariance of | I) is expressed in the row subtraction process). When all the off-diagonal elements (I, J) have been rendered zero by the row subtraction process then the remaining diagonal (I, I) element gives the shift ∈(n + 1) − ∈(n) (as shown by Eq. (5)) and the process is repeated for the matrix (H−∈(n + 1)). The idea behind the work of [1] is that the process will converge fairly quickly to give the correct value of the perturbed eigenvalue ∈. We undertake a numerical exploration of this idea later in this work.

3 Some links with other approaches

It is clear that convergence of the iterative process described above would produce a version of the matrix $H - \in$ in which ALL the elements in row I were zero, so that the determinant of the matrix would necessarily be zero. This is the traditional criterion which is used to find the eigenvalues of a matrix (the secular equation approach).The novelty of the Niblack–Nigam method is that the (I, I) element alone, when it is not zero, directly gives the numerical correction to the eigenvalue, whereas more traditional approaches would involve root-finding formulae, based on several values of the full determinant (of which the (I, I) element is only one factor).

Rendering zero the off-diagonal elements of a row in the matrix is reminiscent of the Bloch wave operator approach [2], which renders zero the off-diagonal elements in a column. Indeed, the basic element in a row operation is a matrix with just one element different from the unit matrix, as is the matrix which appears in the single cycle method associated with the wave operator approach [2]. However, the wave operator approach uses full (two sided) similarity transformations of the H matrix, while the present method uses an operator A which acts from the left on the varying matrix $H - \in(n)$. A similarity transformation leaves the matrix spectrum invariant, so that an isolated diagonal element in a row directly gives an eigenvalue. By contrast an operator acting from the left will not maintain the spectrum invariant but the row operations in A will leave the determinant of $H - \in(n)$ unchanged and a single factor in the determinant (that is, the (I, I) element) suffices to give the eigenvalue correction. The Niblack–Nigam method uses only linear operations, whereas the Bloch wave operator approach involves the more difficult solution of a nonlinear equation.

In the process of row subtraction which represents the operator A it is clear that it is possible for some initially zero elements in row I to be rendered non-zero by the very subtractions which deliberately render zero the initially non-zero elements. Tracing the process through, it follows that all the rows in the matrix (except I) are in general required to participate in the subtraction process. However, this feature is a necessary part of ALL matrix eigenvalue techniques, since it ensures that the eigenvalues can change as the dimension of the matrix is increased.

One obvious practical feature in the numerical application of the Niblack–Nigam method is that the matrix H is only calculated once, with a copy C being used in the formation of the sequence of matrices $H - \in(n)$. In our calculations we allowed H to be a complex matrix, so that both bound states and resonances could be treated. Such a generalization was not, of course, envisaged in [1].

4 Some test bound state calculations

In [1] only a very simple numerical example was given and so we have used as a test problem the more modern one of a quartically perturbed oscillator. We set up the matrix of the Hamiltonian

$$H = -D^2 + x^2 + x^4 = [-D^2 + Wx^2] + x^4 - (W - 1)x^2 \tag{6}$$

Table 1 The first 5 even parity eigenvalues for the case W = 8 and a 50 × 50 matrix

I	E	NI
1	1.392351641530292	13
2	8.655049957759310	19
3	18.05755743630326	24
4	28.83533845950425	19
5	40.69038608210644	14

NI is the maximum number of iterations which suffices to find each eigenvalue as $\in(0)$ varies from 0 to 50. NI has only a very weak dependence on $\in(0)$

taking as a basis set the eigenfunctions of the oscillator Hamiltonian in the square bracket. For this "renormalized" problem the perturbation is relatively weak and so should be suitable to test an iterative approach. When setting up the program to carry out the calculation we have to provide for a variable choice of the unperturbed state | I), which we associate with row I in the matrix. Thus the rows involved in the subtraction process will skip over row I and the surviving (I, I) diagonal element will give the increment $\in(n + 1) - \in(n)$.

In the first calculation we used a 50 × 50 matrix with W = 8 and set the initial value $\in(0)$ to be the diagonal element H(I, I), letting I range from 1 to 5 and using an even parity basis set (see Table 1). For each I value the method converged in less than 15 iterations to give accurate eigenvalues. The calculation was repeated with $\in(0)$ varying between 0 and 50 for each I value, since the choice H(I, I) might appear to be copying standard perturbation theory too closely. The method performs equally well under these more extreme conditions, showing that (as commented in Sect. 2) the method is not reproducing a sequence of energy approximations which are those of a traditional perturbation series. The successive $\in$ estimates show an alternating convergence to the accurate eigenvalue, which would render the sequence suitable for treatment by the Wynn algorithm, although this is hardly necessary for these calculations. One feature was noteworthy: the eigenvalue obtained was determined by the value of the index I and not by the numerical value of $\in(0)$. This is presumably because the method finds the eigenvector with the greatest overlap with the reference state | I); the derivation in Sect. 2 assumed that this overlap is set at the value 1.

When the perturbation is made very strong by changing to a basis set with W = 1 then the calculation produces a reasonable convergence for the ground state but diverges for the higher states. The divergent sequence is not even alternating and so the Wynn algorithm is of little help. An improvement is obtained by using a relaxation parameter RE of 1/2 or 1/4 to multiply the predicted shift in $\in$ at each step. The lowest two eigenvalues then come out correctly for the indices I = 1 and I = 2, respectively, although a larger number of iterations is required. However, the higher eigenvalues come out sometimes in reversed order and sometimes are repeated as I is increased, with the further complication that the detailed behavior varies with the numerical value of the relaxation parameter RE.

Table 2 The approximate roots of $F(\epsilon) = 0$ as found by the scan described in Sect. 5, for the case $W = 1$, using a fixed I value and varying ϵ in steps of 0.001

I = 1		I = 2	
E	F	E	F
1.392352	−6.7(−4)	1.392351	−2.1(−2)
		1.687662	4.0(3)
8.482640	1.9(3)	8.655050	−1.5(−2)
8.655050	−4.0(−2)		
17.952521	3.3(3)	15.868195	1.0(5)
18.057560	−6.9(−2)	18.057558	−2.4(−3)
28.791091	1.3(4)	27.271703	3.8(4)
28.835344	−4.1(−1)	28.835339	−8.2(−3)
40.671960	7.4(4)	39.691045	5.9(5)
40.690398	−1.2(0)	40.690386	−1.7(−2)

Note the very large $F(\epsilon + d\epsilon)$ associated with the partner roots and the negative sign of $F(\epsilon + d\epsilon)$ for the true eqiured roots. The root associated with the chosen I does not have a partner root. The numerical values shown have been severely truncated.

5 A Direct search approach

In the $W = 1$ calculation we deliberately pushed the iterative method to breaking point by using a very strong perturbation; for a weaker perturbation λx^4 with $\lambda = 0, 2$ (for example) the iterative method still works satisfactorily. In [1] the authors simply commented that after a few iterations "it becomes clear whether the iterative method is converging or not" but did not suggest any remedy for the case of non-convergence. The relaxation parameter approach is only a partial and inconsistent aid for the difficult cases and so we tried a scanning approach. The trial ϵ value was varied in steps of 0.001 from 0 to 50.00 for several fixed I values, with the computer program adapted to find those points at which $F(\epsilon) = [H - \epsilon] (I, I)$ changes sign between ϵ and $\epsilon + d\epsilon$ and to calculate the linearly interpolated ϵ value for which $F(\epsilon) = 0$. The output of the program is the interpolated ϵ value and also $F(\epsilon + d\epsilon)$. Table 2 shows some results for this calculation. Even for a fixed I the direct scan gives all the required eigenvalues in the scanned region. There is an unexpected side effect in that those eigenvalues not arising from the chosen I value have a partner zero with a very large value of $F(\epsilon + d\epsilon)$. The true approximate eigenvalues can be isolated, since they have a negative (and very small) value of $F(\epsilon + d\epsilon)$. This scanning calculation shows that the original approach of [1] can be modified to yield the eigenvalues even when the iterative form of the calculation (the only one envisaged in [1]) fails to converge satisfactorily. The simple scanning program can, of course, be augmented by adding a secant or Newton's method roorfinder to it to refine the eigenvalues to high precision.

6 Using energies to analyze the eigenvector

Since we noted in the calculation with $W = 8$ that the I value determines the resulting eigenvalue for a wide range of $\epsilon(0)$ values, the question naturally arises of whether the erratic behavior for the case $W = 1$ indicates that for that case the I-th eigenvector actually does NOT have the I-th basis vector as its dominant component. Since we

Table 3 The values of $A(J)^2$ as found by the energy shift method of Sect. 6, for the case W=1 and for the even parity states with I = 3, 4, 5

	$A(3)^2$	$A(4)^2$	$A(5)^2$
3	0.1076	0.3337	0.2415
4	0.1558	0.0335	0.0497
5	0.1469	0.0017	0.0976

have deliberately obtained a simple eigenvalue algorithm by carefully eliminating the unknown eigenvector from the calculation, it might seem necessary to construct it by inverse iteration in order to check this point. However, we can check the coefficients of basis vectors 3, 4 and 5 in the eigenvectors by remembering some simple principles of first order perturbation theory. If we add a small increment ß to the (J, J) element of a matrix then the shift produced in the K-th eigenvalue is given by an expectation value which can quickly be seen to be ß $A(K, J)^2$, where A (K, J) is the coefficient of the basis vector J in the expansion of the NORMALIZED K-th eigenvector. We thus have an energy based method for obtaining numerical estimates of selected eigenvector components. A value of ß = 0.0001 is quite small enough to check the question raised in this section. The results of the calculation, shown in Table 3, show that our conjecture was correct and reveals how strongly the basis states are coupled by the strong perturbation associated with the case W = 1.

7 Some resonant state calculations

For the first test calculations we studied the real bound state energies associated with the use of a positive λ value in the perturbation λx^4. Giving λ a small negative value is also of interest from a modern point of view. The potential $x^2 - \lambda x^4$ has a central well with an outer region in which the potential descends downwards to minus infinity. In a time-dependent approach a wave packet initially localized in the central well would leak out into the outer region in a quantum mechanical tunneling process. This tunneling processes arising in a time-dependent approach shows up in the more easy time-independent approach via complex eigenvalues of form ER + i EI, with EI being proportional to the inverse of the tunneling lifetime. These complex eigenvalues can be studied by a process of complexification, in which the basis functions are taken to be the eigenfunctions of the Hamiltonian $-D^2 + (WR + iWI)x^2$. This is achieved by simply using this complex W value in every formula for the matrix elements in which the original real W appeared in the bound state calculation. The row subtraction process is also carried out in complex arithmetic. The calculation turns out to work very well in the complex case and leads to complex resonant state energies which are in accord with those found by the accurate Riccati–Pade approach [3] and also by the hypervirial perturbation theory approach [4]. Table 4 shows some typical results which were obtained by using the simple fixed choice W = 1 + i. Since the λ values used were of relatively small magnitude the iterative method converges quickly to give the complex eigenvalue for each I value.

Table 4 The I = 1 and I = 2 resonant state energies for the Hamiltonian $-D^2 + x^2 - \lambda x^4$ Obtained using the choice W = 1 + i and for a basis of 50 even parity complex basis functions as explained in Sect. 7 of the text

λ	ER	EI
0.02	0.9844276697652552	5.111(−14)
	0.4786335047926524	2.8426265(−9)
0.04	0.9674512352369713	5.965307361(−7)
	0.4504482756443305	4.670427976358(−3)
0.06	0.9483297726717095	1.11911527151015(−4)
	0.4124540696226597	1.41422243854514(−1)
0.08	0.9259424610731429	1.5440221243205(−3)
	0.3841131497773914	4.32096228521483(−1)

8 Discussion

The Niblack–Nigam algorithm, when modified as shown in the present work, turns out to be a simple and effective method for finding both real and complex eigenvalues. Since the original derivation of the algorithm given in [1] was unduly complicated we have shown that a much shorter derivation is possible. Although our unorthodox proof might at first appear to have some elements of black magic about it (since the unknown wave function appears in the equations) it is clear that the essential element is the separating out of the term Φ which must be orthogonal to the unperturbed state | I). A purist might well argue that, since in the standard Eq. (2) the choices Є(n + 1) and Ψ (n) would be the obvious (indeed, necessary) ones to set up an iterative process, then in setting up an iterative version of Eq. (1) we should use the term Φ(n) instead of Φ in Eq. (3), thus allowing the wave function to evolve as well as the energy. The fortunate consequence of the approach used here is that the use of such an alternative form of Eq. (3) makes no difference to the resulting method for calculating the eigenvalue! In Sect. 6 we have shown that we are still able to obtain useful information about the normalized eigenvectors by using purely numerical calculations involving the eigenvalues. It is, of course, possible to use inverse iteration to find a full eigenvector once the corresponding eigenvalue has been found; in another work we have actually used complex inverse iteration to study a variety of resonant states [5].

The numerical results show that, although the method described uses some of the traditional language of perturbation theory, the sequence of energy estimates obtained as Є(0) varies is not that of any of the traditional perturbation techniques. In the present work, we only cite a few references, since there are very few to cite. This work appears to be the first one to have made an analysis or application of the Niblack–Nigam algorithm since its original appearance in 1970.

References

1. W.K. Niblack, B.P. Nigam, Am. J. Phys **38**, 101 (1970)
2. J.P. Killingbeck, G. Jolicard, J. Phys. A Math. Gen. **36**, R105 (2003)
3. F.M. Fernandez, R.H. Tipping, J. Mol. Struct (Theochem) **488**, 157 (1999)
4. J.P. Killingbeck, A. Grosjean, G. Jolicard, J. Phys. A Math. Gen. **37**, 7971 (2004)
5. J.P. Killingbeck, A. Grosjean arXiv:0906.4420v2 [quant-ph] (2009)