An investigation of pseudo-convergence in an iterative method for calculating the low-lying eigenvectors of a Hermitian matrix

1979 J. Phys. A: Math. Gen. 12 1693

(http://iopscience.iop.org/0305-4470/12/10/015)

J. Phys. A: Math. Gen., Vol. 12, No. 10, 1979. Printed in Great Britain

An investigation of pseudo-convergence in an iterative method for calculating the low-lying eigenvectors of a Hermitian matrix

H G Miller and W A Berger

Institut für Theoretische Physik, Universität Frankfurt am Main, Robert-Mayer-Strasse 8-10, 6000 Frankfurt/Main 1, West Germany

Received 30 June 1978, in final form 10 January 1979

Abstract. In terms of a previously suggested iterative method for finding the lowest-lying eigenstate of hermitian matrices the problem of pseudo-convergence, i.e. the tendency to converge to an undesired eigenvalue, is discussed. A simple extension of this algorithm is presented which provides a means of identifying and circumventing pseudo-convergence in many cases. This extended scheme may also be used simultaneously to obtain excited states in a numerically economical manner.

1. Introduction

Under certain conditions most iterative diagonalisation schemes exhibit a disturbing tendency to converge to the wrong eigenvalue. This pseudo-convergence generally arises, as has recently been mentioned (Berger *et al* 1977), when the component of the desired eigenvector in the original arbitrarily chosen start vector is small (but non-vanishing) compared with the components of one or more of the neighbouring eigenvectors. In the case of the simple power method (Faddejew and Faddejewa 1964) or the more elegant Lanczos algorithm (Lanczos 1950), for example, the only known practical remedies for this situation are (Whitehead 1972, 1977): (*a*) perform additional iterations to ensure that convergence has really been achieved; or (*b*) choose another start vector and see if it converges to the same eigenvector. Although both of these prescriptions leave something to be desired, little else is available within the context of the aforementioned methods. In the present note, however, we wish to point out that within the framework of an algorithm recently suggested by us (Berger *et al* 1977) a simple extension provides under certain circumstances a rather straightforward means of alleviating this problem.

2. Pseudo-convergence in the old algorithm

We begin with an arbitrary start vector and perform the first two iterations of the Lanczos algorithm. After diagonalising the subsequent 2×2 matrix we obtain two mutually orthogonal vectors $|1, l\rangle$ and $|1, u\rangle$ with eigenvalues $e_{1,l}$ and $e_{1,u}$ such that $e_{1,l} < e_{1,u}$. Using $|1, l\rangle$ as the new start vector, a subsequent 2×2 matrix may be generated yielding upon diagonalisation two orthogonal vectors $|2, l\rangle$ and $|2, u\rangle$ which

0305-4470/79/101693+06$01.00

allows as before the choice of a new start vector. Previously we have proved that iterations performed in this manner are ultimately convergent to the eigenvector corresponding to the lowest eigenvalue which has a non-zero component in the original start vector. (We shall refer to this vector in the following as the ground-state vector.) This is mainly due to the fact that in each iteration step the component of the ground-state eigenvector increases.

At present, however, we are more interested in what happens to the components of the other eigenvectors in the iterated vector. Although these components must ultimately vanish when convergence is achieved they need not decrease monotonically in each iteration step. In fact if we expand (Berger *et al* 1977)

$$|k, l\rangle = \sum_{\lambda} b_k^{\lambda} |\lambda\rangle,$$

where $|\lambda\rangle$ are the eigenvectors of the matrix considered, we obtain

$$\frac{b_{k+1}^{\lambda}}{b_k^{\lambda}} = \frac{1 + (e_{k+1,l} - e_{k,l})(E_{\lambda} - e_{k,l}) V_k^{-2}}{[1 + (e_{k+1,l} - e_{k,l})^2 V_k^{-2}]^{1/2}} \tag{1}$$

where E_{λ} are the exact eigenvalues and V_k the off-diagonal matrix elements in the kth iteration. For all λ for which E_{λ} satisfies

$$e_{k+1,l} \geqslant E_{\kappa} \tag{2}$$

it can be shown that equation (1) gives $|b_{k+1}^{\lambda}| > |b_k^{\lambda}|$, in spite of the fact that all the b_k^{λ} for $\lambda \neq 0$ must ultimately vanish if convergence to the ground state is to be achieved. Thus in each iteration step only those components of the eigenvectors whose eigenvalues satisfy $e_{k+1,l} < E_{\lambda}$ may decrease in magnitude in the kth iteration, while the components of the eigenvectors with $E_{\lambda} \leqslant e_{k+1,l}$ will increase in magnitude. As can be seen from equation (1) the rate of increase is almost equal for components which are close in energy.

For example, consider now the case where the largest component in the start vector corresponds to the first excited eigenvector, i.e. the eigenvector whose eigenvalue E_1 is closest to the eigenvalue E_0 of the ground-state eigenvector, and the eigenvalues E_0 and E_1 are not widely separated. As long as equation (2) is satisfied for $\lambda = 0$ and $\lambda = 1$, b_k^0 and b_k^1 increase at almost equal rates. Because the norm of the iterated vectors is constant the overall contribution of the higher lying components to $|k, l\rangle$ decreases, and thus it may happen that the first excited state becomes the dominant contribution. This is exactly the behaviour exhibited in our calculations. Whenever pseudo-convergence to some E_{λ} occurred we found that the high-lying contributions were reduced, whereas the coefficient of the eigenvector $|\lambda\rangle$ had a maximum. Furthermore, the variance calculated with the iterated vectors showed a minimum (Kreuzer 1978).

3. The extended algorithm

We now propose the following. Instead of performing only ground-state iterations (which we shall refer to as *gs* iterations in the following) in the two-dimensional sub-spaces spanned by $|k, l\rangle$ and $|k, u\rangle$, a parallel set of 2×2 iterations (which we shall denote as p iterations) is performed simultaneously such that the start vector in each p iteration is orthogonal to a corresponding start vector in a *gs* iteration. For the first p iteration the normalised start vector $|0\rangle$ may be constructed quite simply in the following

manner:

$$|0) = N_0(1 - P_2)|, u\rangle$$

where

$$P_2 = |2, l\rangle\langle 2, l|$$

and N_0 is a normalisation constant. It is clear that this start vector is orthogonal to the normalised vector $|2, l\rangle$ which will be used as a start vector for the next *gs* iteration. In the subsequent p iterations the new start vectors may be constructed as follows:

$$|k) = N_k(1 - P_{k+2})|k-1, l)$$

$$P_{k+2} - |k+2, l\rangle\langle k+2, l|$$

where again N_k is a normalisation factor such that

$$(k|k) = 1$$

and where $|k-1, l)$ is the eigenvector corresponding to the lower eigenvalue of the 2×2 diagonalisation initiated with the start vector $|k-1)$. Again by construction, this start vector is orthogonal to $|k+2, l\rangle$ which is also the start vector for the next *gs* iteration.

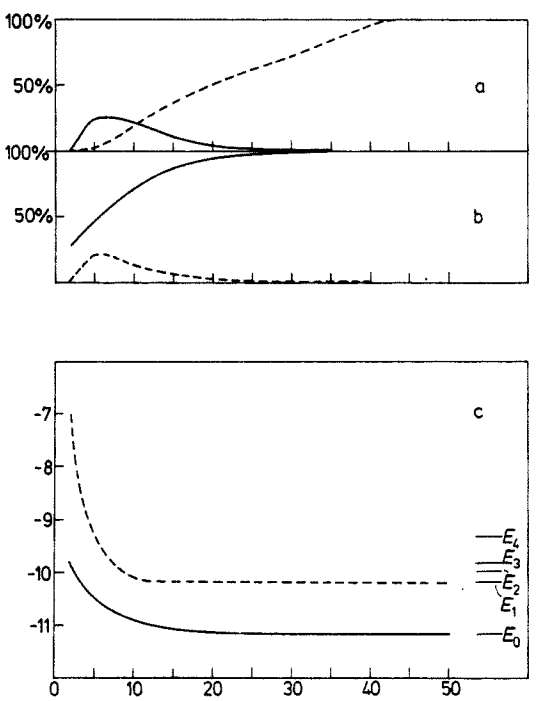

Figure 1. Convergence pattern of the two lowest-lying eigenvectors of a random 100×100 symmetric matrix: (a) probabilities of the ground state (full curve) and the first excited state (broken curve) in the iterated p vectors; (b) same as (a) but for iterated *gs* vectors; (c) expectation values of the vectors of the *gs* iterations (full curve) and the vectors of the p iterations (broken curve). On the right-hand side the four lowest-lying eigenvalues of the matrix are given. The probabilities of the ground and first excited states in the original start vector are 10% each.

In the case where the iterated *gs* vectors (i.e. the vectors of the *gs* iterations) converge monotonically to the ground state, the orthogonalisation condition is quite simply a convenient means of projecting the converging ground-state eigenvector out of 2×2 iterations involving the *p* vectors since

$$|k, l\rangle \xrightarrow[k\to\infty]{} |k=0\rangle \qquad P_k \xrightarrow[k\to\infty]{} P^0|\lambda=0\rangle\langle\lambda=0|$$

where $|\lambda=0\rangle$ is the ground-state eigenvector of the matrix to be diagonalised. Hence by removing the component of this vector, the iterated *p* vectors must ultimately converge to the eigenvector corresponding to the first excited state, assuming that its component in the original start vector is non-zero and that it was not accidentally removed during the iterations. Furthermore, as will be discussed later, the parallel iterations require only a small amount of additional numerical effort, so that the proposed method is a reasonably efficient means of calculating the first excited state.

On the other hand, what interests us is what happens when pseudo-convergence occurs in the *gs* iterations. In this case the expectation values of the iterated *gs* vectors, $e_{k,l}$, remain almost constant for a number of iterations. At the same time the eigenvalues $\epsilon_{k,l}$ of the iterated *p* vectors may continue to decrease in each iteration step. The orthogonality condition, however, prevents the two iterated vectors from becoming the same vector. Hence if the pseudo-convergence in *gs* iterations persists for a sufficient number of iterations it may eventually happen that

$$\epsilon_{k,l} < e_{k+2,l}$$

i.e. that the expectation values of the *p* and *gs* iterations may cross.

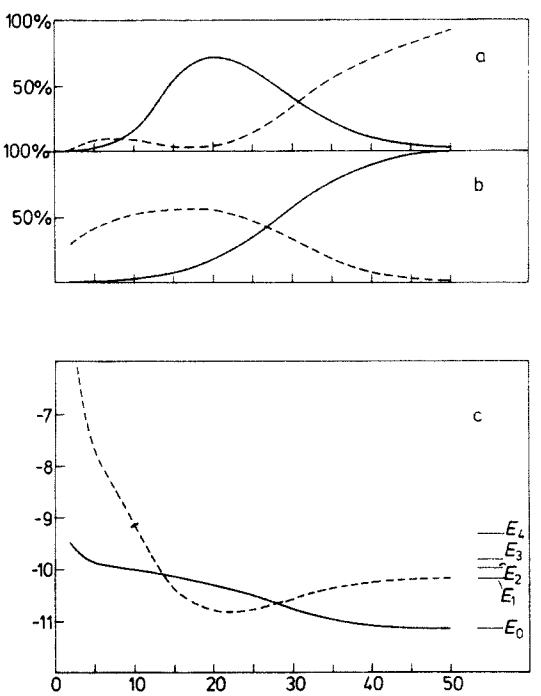

Figure 2. As figure 1 with the probabilities of ground and first excited states in the original start vector equal to 0·1% and 10% respectively.

From the previous discussion we know that the component of the eigenstate $|\lambda\rangle$ becomes dominant in the iterated *gs* vectors if pseudo-convergence to E_λ occurs. Because of the orthogonalisation there is a good chance that in the *p* vectors the corresponding component will be suppressed and if crossing occurs the component of the ground-state will become dominant.

In order to demonstrate this we have performed the following numerical calculations. A 100×100 symmetric random matrix was generated. In the first case (see figure 1) the components of the ground and first excited states in the original start vector were of equal magnitude. As expected the *gs* iterations converge monotonically to the ground state, and the iterated *p* vectors converge ultimately to the first excited state. Adjusting the start vector so that the dominant component is that of the first excited state (see figure 2) yielded pseudo-convergence in the iterated *gs* vector. Furthermore the expectation values obtained from the iterated *gs* and *p* vectors crossed, the expectation value of the iterated *p* vector becoming less than that of the *gs* vector. Iterating further produced a second crossing as the two vectors converged ultimately to ground and first excited eigenvectors. Realising that the crossing of the iterated expectation values is indicative of pseudo-convergence, the problem may be remedied quite simply at this point by interchanging the two vectors (see figure 3), i.e. taking $|k-2, l\rangle$ as new start vector for the *gs* iterations and $|k, l\rangle$ as new start vector for the *p* iterations. This, needless to say, greatly improves the convergence rate for the (new) *gs* vector, since its dominant component is now that of the ground-state eigenvector. Furthermore, as far as we know this is the only way to detect pseudo-convergence directly during the iteration procedure without either having to perform additional iterations or to begin again with a different start vector.

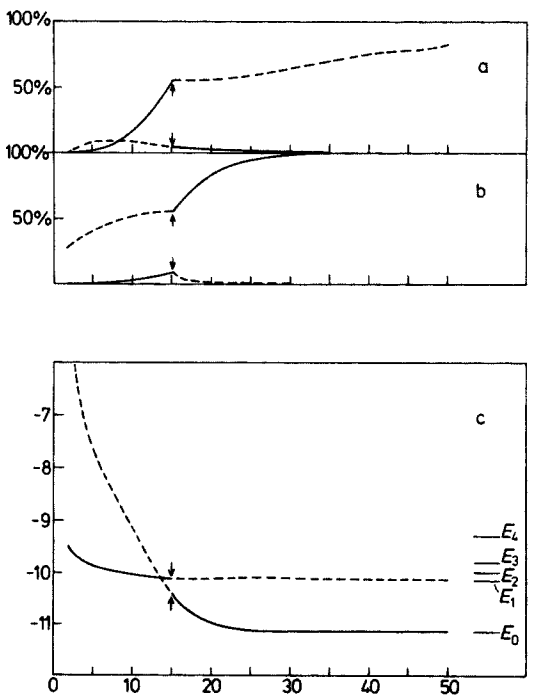

Figure 3. Same as figure 2 with interchange of the vectors of the *gs* iteration and *p* iteration at the point indicated by the arrows.

We also wish to point out that the set of parallel iterations requires very little additional numerical work since the iterated vectors in *both cases* may always be represented as a sum of terms involving different powers of the matrix to be diagonalised acting on the original start vector. These must in any case always be calculated for the *gs* iterations and may therefore be stored for use in the *p* iterations.

Furthermore, extending the present procedure to include more than just one parallel set of 2×2 iterations, may well provide a reasonable means of obtaining the lower-lying eigenstates. For an arbitrary original start vector the convergence rates of the respective states may be increased in cases of pseudo-convergence by an appropriate re-ordering of the start vectors for the subsequent iterations.

Acknowledgments

We wish to acknowledge useful discussions with K G Kreuzer, Dr P Chattopadhyay and Professor R M Dreizler. We also acknowledge the use of the computer facilities of the HRZ of the University of Frankfurt.

References

Berger W A, Miller H G, Kreuzer K-G and Dreizler R M 1977 *J. Phys. A: Math. Gen.* **10** 1089

Faddejew D K and Faddejewa W N 1964 *Numerische Methoden der Linearen Algebra* (München: Oldenbourg)

Kreuzer K-G 1978 *Diplomarbeit* University of Frankfurt

Lanczos C 1950 *J. Res. Natn. Bur. Stand.* **45** 255

Whitehead R R 1972 *Nucl. Phys.* A **182** 290

Whitehead R R, Watt A, Cole B J and Morrison I 1977 *Advances in Nuclear Physics* ed M Baranger and E Vogt (New York: Plenum) vol 9 123–76