The Calculation of the Eigenvectors of Codiagonal Matrices

by J. H. Wilkinson

Summary: In the Givens method for calculating the eigenvalues and eigenvectors of a matrix, a collineatory transformation is constructed which reduces the matrix to codiagonal form. Givens (1954) has given a complete analysis of the problem of finding the eigenvalues and has described a very satisfactory practical procedure for evaluating them. No such analysis has been given for the eigenvectors, though Givens in an unpublished paper has described a procedure which, in his experience, has given accurate results. In this note an analysis of the problem is given, which explains why the straightforward use of the recursions often gives vectors which are catastrophically in error. A method of solution is described which has been used extensively for calculating the vectors on DEUCE. Much of what is written applies equally well to the codiagonal matrices produced by the method of Lanczos (1950), but because this method is usually programmed using floating-point arithmetic, there are one or two additional complications. These will be the subject of a later note.

STATEMENT OF THE PROBLEM

The codiagonal form produced by the Givens process will be denoted by $\boldsymbol{C}$ where we have

$$\boldsymbol{C} = \begin{bmatrix} \alpha_1 & \beta_2 & & & & \\ \beta_2 & \alpha_2 & \beta_3 & & & \\ & \beta_3 & \alpha_3 & \beta_4 & & \\ & \cdots & \cdots & \cdots & & \\ & & \cdots & \cdots & & \\ & & & \beta_{n-1} & \alpha_{n-1} & \beta_n \\ & & & & \beta_n & \alpha_n \end{bmatrix} \tag{1}$$

The process is usually programmed using fixed-point arithmetic, so that the α_i and β_i are all given to the same number of binary places.

We shall assume that some or all of the eigenvalues of the codiagonal form have been calculated to a high degree of accuracy (on the DEUCE, which has a 32 binary digit word, they will usually be correct to about 9 decimal places) and the problem is to find the corresponding eigenvectors. It is very desirable that the method should be completely automatic.

Corresponding to an exact eigenvalue, λ_1, of $\boldsymbol{C}$, there is a non-zero solution of the set of equations

$$\boldsymbol{C}\boldsymbol{x} = \lambda_1 \boldsymbol{x}, \tag{2}$$

determined apart from an arbitrary multiplier. We may solve any $(n-1)$ of the n homogeneous equations corresponding to (2) to obtain the ratios of the n components of $\boldsymbol{x}$. The remaining equation is then automatically satisfied. If λ is an approximation to λ_1 then the set of equations

$$\boldsymbol{C}\boldsymbol{x} = \lambda \boldsymbol{x} \tag{3}$$

has no non-zero solution, but by omitting each of the equations in turn we may determine n vectors

$$\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots \boldsymbol{x}^n$$

each of which is a function of λ, and we will expect that as λ tends to λ_1 each of these vectors will tend to $\boldsymbol{v}_1$, the eigenvector corresponding to λ_1. For a given approximation to an eigenvalue some of the $\boldsymbol{x}^i$ will be closer approximations to the eigenvector than others. In the next section the factors governing their rate of convergence are examined.

CONVERGENCE OF THE APPROXIMATE EIGENVECTORS

Equation (3) may be written

$$\begin{aligned} (\alpha_1 - \lambda)x_1 + \beta_2 x_2 &= 0 \\ \beta_2 x_1 + (\alpha_2 - \lambda)x_2 + \beta_3 x_3 &= 0 \\ \beta_3 x_2 + (\alpha_3 - \lambda)x_3 + \beta_4 x_4 &= 0 \\ \cdots\cdots\cdots\cdots\cdots & \\ \beta_{n-1} x_{n-2} + (\alpha_{n-1} - \lambda)x_{n-1} + \beta_n x_n &= 0 \\ \beta_n x_{n-1} + (\alpha_n - \lambda)x_n &= 0. \end{aligned} \tag{4}$$

The most convenient sets of $(n-1)$ equations to solve are the first $(n-1)$ and the last $(n-1)$ and it has frequently been suggested that one of these sets should be used for determining the eigenvectors. If the first $(n-1)$ equations are used we may take $x_1 = 1$ and use the first equation to determine x_2, the second to determine $x_3 \ldots$ the $(r-1)$th to determine $x_r \ldots$ and finally the $(n-1)$th to determine x_n. In general the nth equation will not be satisfied exactly by these values. If we omit an equation other than the first or the last, the situation is only slightly less simple. If we omit the ith equation the first $(i-1)$ equations give $x_2, x_3, \ldots x_i$ in terms of x_1 and the last $(n-i)$ give $x_{n-1}, x_{n-2}, \ldots x_i$ in terms of x_n and we can therefore obtain all the variables in terms of x_1. The solution obtained when omitting the ith equation satisfies

$$\begin{aligned} \beta_j x_{j-1} + (\alpha_j - \lambda)x_j + \beta_{j+1} x_{j+1} &= 0 \quad j \neq i \\ \beta_i x_{i-1} + (\alpha_i - \lambda)x_i + \beta_{i+1} x_{i+1} &= \delta \neq 0. \end{aligned} \tag{5}$$

Since we are not interested in arbitrary multipliers, this solution is the same as that to the set of equations

$$(\boldsymbol{C} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{e}_i \tag{6}$$

where $\boldsymbol{e}_i$ is the vector which has zero for all components except the ith, which is unity. If the true eigenvectors of $\boldsymbol{C}$ are $\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots \boldsymbol{v}_n$ corresponding to eigenvalues, $\lambda_1, \lambda_2, \ldots \lambda_n$ then we may express $\boldsymbol{e}_i$ in terms of the $\boldsymbol{v}_j$

$$\boldsymbol{e}_i = \sum_1^n k_{ij} \boldsymbol{v}_j. \tag{7}$$

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

The exact solution of the set of equations (6) is therefore

$$\boldsymbol{x} = \sum_{j=1}^{n} k_{ij}(\boldsymbol{C} - \lambda \boldsymbol{I})^{-1}\boldsymbol{v}_j. \tag{8}$$

Now $(\boldsymbol{C} - \lambda \boldsymbol{I})^{-1}$ has the same eigenvectors as $\boldsymbol{C}$, while its eigenvalues are $(\lambda_j - \lambda)^{-1}$, so that

$$\boldsymbol{x} = \sum_{j=1}^{n} k_{ij}\left(\frac{1}{\lambda_j - \lambda}\right)\boldsymbol{v}_j. \tag{9}$$

If $\lambda = \lambda_1 + \varepsilon$ we have

$$\boldsymbol{x} = k_{i1}\left(\frac{1}{-\varepsilon}\right)\boldsymbol{v}_1 + \sum_{j=2}^{n} k_{ij}\left(\frac{1}{\lambda_j - \lambda_1 - \varepsilon}\right)\boldsymbol{v}_j. \tag{10}$$

Provided k_{i1} is not zero we may deduce from (10) that

$$\boldsymbol{x} \to \boldsymbol{v}_1 \text{ as } \varepsilon \to 0$$

apart from a multiplicative constant. However, in practice, ε is always prescribed by the accuracy of the λ which we have calculated. Using single-length arithmetic, on most available machines ε will be at best of the order of 10^{-10}. If then k_{i1} happens to be "small," $\boldsymbol{x}$, as given by (10), will be by no means a good approximation to $\boldsymbol{v}_1$. If, for example, k_{i1} is itself of the order of 10^{-10} then $\boldsymbol{x}$ will contain components of other vectors which are as large as its components of $\boldsymbol{v}_1$.

In what follows the assumption will be made that $(\lambda_j - \lambda_1)^{-1}$ is not large, that is, that there are no roots very close to λ_1. It is reasonable in any case to ignore impurities in an approximation to $\boldsymbol{v}_1$ which are of the same order of magnitude as $10^{-10}(\lambda_j - \lambda_1)^{-1}$, for the following reason. The matrix $\boldsymbol{C}$ which we have regarded as exact from the point of view of this investigation will, in fact, have been derived by the Givens process and will therefore contain rounding errors. The errors in the calculated eigenvector $\boldsymbol{v}_1$ due to these rounding errors will contain the factors $(\lambda_j - \lambda)^{-1}$ and such errors are therefore inherent in the Givens process, or indeed in any process which derives the eigenvectors of $\boldsymbol{A}$ via a transformed matrix $\boldsymbol{B}$.

The latent vectors of Givens matrices are such that small k_{ij} are quite common, for the following reasons. If one of the values of β, say β_r, is exactly zero then the eigenvalues of $\boldsymbol{C}$ split into two groups, those which are the eigenvalues of the principal minor of order $(r-1)$ and those which are the eigenvalues of the matrix of order $(n-r+1)$ in the bottom right-hand corner. Eigenvectors corresponding to eigenvalues of the first group have zero components in each of the last $(n-r+1)$ positions. From equation (7) we have

$$k_{ij} = \frac{\boldsymbol{v}_j' \cdot \boldsymbol{e}_i}{\boldsymbol{v}_j' \cdot \boldsymbol{v}_j} \tag{11}$$

and therefore k_{ij} is zero for $i \geqslant r$ with $j < r$, and similarly also for $i < r$ with $j \geqslant r$. If β_r is very small then we will have eigenvectors with very small components in the last $(n-r+1)$ positions and these will give rise to small k_{ij}. A common situation with Givens matrices is that a number of the β_i are moderately small, and when this is true the roots of some principal minors of the codiagonal form are very close to roots of the complete matrix. Corresponding to these eigenvalues we find again that the eigenvectors have small components in the lower positions. From this it is clear that none of the $\boldsymbol{e}_i$ is likely to produce consistently accurate eigenvectors and, in particular, that $\boldsymbol{e}_n$ and $\boldsymbol{e}_1$ will, in general, be the least satisfactory. This means that the first $(n-1)$ equations of the n equations (4) will frequently produce inaccurate vectors.

In order to emphasize that it is not necessary for any of the β_i to be pathologically small, we may consider the matrix $\boldsymbol{C}$ of order 21:

$$\boldsymbol{C} = \begin{bmatrix} 10 & 1 & & & & & & & & & \\ 1 & 9 & 1 & & & & & & & & \\ & 1 & 8 & 1 & & & & & & & \\ & & 1 & \ldots\ldots\ldots & & & & & & & \\ & & & & 1 & -8 & 1 & & & & \\ & & & & & 1 & 9 & 1 & & & \\ & & & & & & 1 & -10 & & & \end{bmatrix}$$

The largest root of this matrix is clearly greater than 10. If we denote its *exact* value by λ, then, using this value, we may obtain the eigenvector from any $(n-1)$ equations. Suppose we use the last $(n-1)$ and put $x_{21} = 1$. We have

$$\begin{aligned} x_{20} &= (\lambda - 10)x_{21} \\ x_{19} &= (\lambda - 9)x_{20} - x_{21} \\ x_{18} &= (\lambda - 8)x_{19} - x_{20} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

giving

$$\begin{aligned} x_{20} &= (\lambda - 10) \\ x_{19} &= (\lambda - 9)(\lambda - 10) - 1 \\ x_{18} &= (\lambda - 8)[(\lambda - 9)(\lambda - 10) - 1] - (\lambda - 10) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Since λ is greater than 10 it is clear that the terms in this sequence increase rapidly. Quite a crude estimate of the rate of growth will suffice for our purposes. If $x_{r-1} > x_{r-2}$ then we have

$$\begin{aligned} x_r &= (\lambda + r - 10)x_{r-1} - x_{r-2} \\ &> r\,x_{r-1} - x_{r-2}, \text{ since } \lambda > 10 \\ &> r\,x_{r-1} - x_{r-1} \\ &= (r-1)x_{r-1}. \end{aligned}$$

From this we may deduce that the x_i certainly increase up to x_2 and that

$$\begin{aligned} x_2 &> x_3 \\ &> 2 \cdot x_4 \\ &> 2 \cdot 3 \cdot x_5 \\ &\quad\ldots\ldots\ldots \\ &> 19!\,x_{21} \\ &= 19! > 10^{17}. \end{aligned}$$

The 21st component of the normalized eigenvector is therefore less than 10^{-17}. The eigenvector, correct to eight decimals, is given below. From the above argument it is clear that k_{n1} is less than 10^{-17}, so that the eigenvector obtained by solving *exactly* the first 20 equations, using a value of λ with an error as small as 10^{-9}, would be hopelessly inaccurate.

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

Eigenvector obtained by solving first 20 equations	*Accurate eigenvector*
1·00000 00000	+1·00000 000
0·74619 42000	0·74619 419
0·30299 99841	0·30299 994
0·08590 25989	0·08590 250
0·01880 78337	0·01880 748
0·00336 30323	0·00336 146
0·00051 68030	0·00050 815
0·00012 34211	0·00006 659
0·00043 92408	0·00000 771
0·00371 82642	0·00000 080
−0·03579 96842	0·00000 007
Components now increase rapidly until $x_{21} \simeq 10^8$.	0·00000 001
	0·00000 000
	0·00000 000
	0·00000 000
	0·00000 000
	0·00000 000
	0·00000 000
	0·00000 000
	0·00000 000
	−0·00000 000

$\lambda = 10 \cdot 74619\ 42$ correct to 9 significant figures.

This last point cannot be too strongly emphasized since it is sometimes thought that the inaccuracy arises from the rounding errors inherent in calculating the eigenvector from the first $(n-1)$ equations. If one takes the value 10·74619 42 for λ (this is correct to 9 significant figures) and solves the first 20 equations, keeping 9 figures in the calculation, then the solution obtained does differ substantially from the exact solution of the first 20 equations, and this difference is, of course, due to the effect of rounding errors. However, there is no 9-figure approximation to λ_1 for which the *exact* normalized solution of the first 20 equations bears any relation to $\boldsymbol{v}_1$. The result is somewhat surprising since the eigenvector of $\boldsymbol{C}$ is given by

$$\begin{matrix} 1 \\ -p_1(\lambda_1) \\ +p_2(\lambda_1) \\ -p_3(\lambda_1) \\ \cdots\cdots \\ +p_{20}(\lambda_1) \end{matrix}$$

where $p_r(\lambda)$ is the determinant of the rth principal minor of $(\boldsymbol{C} - \lambda \boldsymbol{I})$. It is precisely this sequence of values, calculated for a succession of approximations to the eigenvalue, which is used in the Givens process for the determination of the eigenvalue itself, and this it does extremely accurately.

The above example is useful because it emphasizes that the phenomenon is not dependent upon the matrix having close roots (the roots of this matrix are very well separated), or upon the existence of very small β_i. However, the presence of close roots in a matrix or the existence of small β_i does increase the probability of the occurrence of the phenomenon. If, in the matrix $\boldsymbol{C}$ above, the values of β are taken to be 0·01 instead of 1·0 then the difficulty arises in a more extreme form. With $\beta = 1$ the first 20 equations do not determine the first eigenvector correct to r decimals until we have a value of λ_1 which is correct to about $(r + 17)$ decimals. With $\beta = 0 \cdot 01$ we need a value of λ correct to about $(50 + r)$ decimals.

It might be imagined that trouble of this kind would occur comparatively infrequently, but this has not been the author's experience. For every matrix of the first batch of 30 on which the experiment was tried, the use of the first $(n-1)$ recursions produced some completely inaccurate vectors. This experience was confirmed by Professor Givens on matrices arising from a different source.

PRACTICAL PROCEDURE FOR CALCULATING EIGENVECTORS

Instead of solving the equations

$$(\boldsymbol{C} - \lambda \boldsymbol{I})x = \boldsymbol{e}_i \tag{12}$$

for some value of i we analyse the solution of

$$(\boldsymbol{C} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{b} \tag{13}$$

for any vector $\boldsymbol{b}$. If $\boldsymbol{b}$ is expressed in terms of the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_n$

$$\boldsymbol{b} = \sum_1^n \gamma_i \boldsymbol{v}_i \tag{14}$$

then we have

$$\boldsymbol{x} = \sum_1^n \gamma_i \left(\frac{1}{\lambda_i - \lambda}\right)\boldsymbol{v}_i. \tag{15}$$

If λ is an accurate approximation to a root λ_j then it is clear from (15) that the corresponding $\boldsymbol{x}$ will be an accurate approximation to $\boldsymbol{v}_j$ provided $\boldsymbol{b}$ is chosen so that it is not particularly defective in $\boldsymbol{v}_j$.

If the largest component of $\boldsymbol{v}_k$ is in the jth position, then it is clear that $\boldsymbol{b} = \boldsymbol{e}_j$ would be a satisfactory choice for the determination of $\boldsymbol{v}_k$. This means that solving the recursions omitting the jth equation would give an accurate vector $\boldsymbol{v}_k$. However, we do not know *a priori* which will be the largest component of $\boldsymbol{v}_k$.

In the procedure programmed for DEUCE, $\boldsymbol{b}$ is chosen in a way which makes it unlikely that it is particularly defective in any of the $\boldsymbol{v}_i$ and at the same time so as to minimise the computation involved in calculating $\boldsymbol{x}$ from (13). The equations may be written in full

$$\begin{aligned} (\alpha_1 - \lambda)x_1 + \beta_2 x_2 &= b_1 \\ \beta_2 x_1 + (\alpha_2 - \lambda)x_2 + \beta_3 x_3 &= b_2 \\ \beta_3 x_2 + (\alpha_3 - \lambda)x_3 + \beta_4 x_4 &= b_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \beta_{n-1} x_{n-2} + (\alpha_{n-1} - \lambda)x_{n-1} + \beta_n x_n &= b_{n-1} \\ \beta_n x_{n-1} + (\alpha_n - \lambda)x_n &= b_n. \end{aligned} \tag{16}$$

These are solved by the successive elimination of the variables $x_1, x_2, \ldots x_{n-1}$ in their natural order but taking as "pivotal" row at each stage that equation which has the largest coefficient of the variable which is

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

being eliminated. At each stage there will be only two equations containing that variable. Because the equations are not necessarily used in their natural order the resulting equations must be written:

$$\begin{aligned} &p_1x_1 + q_1x_2 + r_1x_3 = c_1 \\ &p_2x_2 + q_2x_3 + r_2x_4 = c_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &p_{n-2}x_{n-2} + q_{n-2}x_{n-1} + r_{n-2}x_n = c_{n-2} \\ &p_{n-1}x_{n-1} + q_{n-1}x_n = c_{n-1} \\ &p_n x_n = c_n \end{aligned} \tag{17}$$

though usually many of the r_i will be zero. The c_i are derived from the b_i, which are still at our disposal. We may assume that the b_i were chosen so that $c_i = 1$ for all values of i, and in this way we avoid any necessity for computation of the right-hand sides of (17). We therefore solve the set of equations (17) with $c_i = 1$. The reasons for expecting that the vector $\boldsymbol{b}$ corresponding to this choice for $\boldsymbol{c}$ will not be deficient in the eigenvector which we are attempting to calculate, are explained at the end of this section. The process has been used in exactly the above form on a large number of matrices, many of which have produced codiagonal forms with very small β_i. More than 400 matrices of orders from 10 to 128 have so far been solved and the vectors have been consistently accurate. The time taken to find the eigenvectors of the codiagonal form by this method is, however, quite a small percentage of the total time taken for the whole problem; so that, if it were found on a more extended test that it was not satisfactory, the following alternative, which is somewhat longer, could be used.

By the above process we may calculate $\boldsymbol{x}_1$ corresponding to the known approximation λ. This vector $\boldsymbol{x}_1$ may now be used as the vector $\boldsymbol{b}$ in the sets of equations (16). Even if $\boldsymbol{x}_1$ is not a sufficiently good approximation to $\boldsymbol{v}_1$ it will certainly contain a substantial component of $\boldsymbol{v}_1$ so that the vector $\boldsymbol{x}_2$ obtained with $\boldsymbol{b} = \boldsymbol{x}_1$ will be a very accurate eigenvector. There is about twice as much work in this process as in that which has been described previously, but even with this refinement the time taken to calculate the eigenvectors of the codiagonal form is very short. When deriving $\boldsymbol{x}_2$ from $\boldsymbol{x}_1$ there is no need to recalculate the p_i, q_i and r_i since they are the same as before. If a copy is retained of the multiplying factors used in the initial reduction, the amount of work involved in calculating $\boldsymbol{x}_2$ is less than that in calculating $\boldsymbol{x}_1$. It is not possible to obtain unlimited accuracy in the vectors of $\boldsymbol{C}$ (regarded for this purpose as an exact matrix) by repeating the above process to obtain vectors $\boldsymbol{x}_3, \boldsymbol{x}_4, \ldots$, etc., but it is quite easy to obtain a vector of more than single-length accuracy without resorting to double-length arithmetic, while still using the original approximate value of λ and the original values of p_i, q_i and r_i. The process described above is repeated until two successive values of $\boldsymbol{x}_i$ (when normalised in some convenient way) agree to within three or four binary places. This will seldom mean producing anything further than $\boldsymbol{x}_3$. We then form the vector $\boldsymbol{r}$ defined by

$$\boldsymbol{r} = \boldsymbol{x}_{i-1} - (\boldsymbol{C} - \lambda\boldsymbol{I})\boldsymbol{x}_i$$

and use $\boldsymbol{r}$ as our vector $\boldsymbol{b}$ in the equations (16). The vector $\boldsymbol{r}$ will be small compared with $\boldsymbol{x}_{i-1}$ and we can therefore take more binary places in $\boldsymbol{r}$ than in $\boldsymbol{x}_{i-1}$. The figures in these extra binary positions will be determined exactly provided all the figures involved in the multiplication of $\boldsymbol{x}_i$ by $(\boldsymbol{C} - \lambda\boldsymbol{I})$ are retained. If $(\boldsymbol{C} - \lambda\boldsymbol{I})\boldsymbol{y} = \boldsymbol{r}$ then $(\boldsymbol{x}_i - \boldsymbol{y})$ is a more accurate vector.

Since we have the relation

$$|\boldsymbol{C} - \lambda\boldsymbol{I}| = \pm p_1p_2 \ldots p_n \tag{18}$$

where the sign depends upon the number of interchanges used during the reduction and

$$|\boldsymbol{C} - \lambda\boldsymbol{I}| = \prod_{i=1}^{n}(\lambda_i - \lambda) = 0 \tag{19}$$

for an exact eigenvalue, we might expect one of the p_i to be small. In fact it has been the author's experience with general matrices that when Gaussian elimination with interchange has been performed on $(\boldsymbol{A} - \lambda\boldsymbol{I})$ with a value of λ which was an eigenvalue (correct to the number of working figures), then the last pivotal value has been, in general, almost zero to working accuracy. For the codiagonal forms arising from the Givens transformations it is quite common for *none* of the pivotal elements to be small. For the matrix of order 21 given above and $\lambda = 10{\cdot}74619\,42$ (correct to 9 significant figures) the pivotal elements p_i are given below.

Pivotal elements for $\lambda = 10{\cdot}74619\,42$

$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$1{\cdot}0$
$-4{\cdot}07784\,23$
$-12{\cdot}66037\,37$
$13{\cdot}66720\,76$
$-14{\cdot}67302\,64$
$15{\cdot}67804\,19$
$-16{\cdot}68241\,07$
$17{\cdot}68625\,08$
$-18{\cdot}68965\,31$
$19{\cdot}69268\,87$
$-20{\cdot}69541\,39$

For values of λ which are in error in the 9th significant figure the 12th pivot is extremely sensitive to variations in λ. Changing the last figure of λ alters all figures of this pivotal value. For values of λ of lower accuracy, say $10{\cdot}7460$, $10{\cdot}7461$, $10{\cdot}7463$, it is the 8th pivot which

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

is sensitive, as is shown below. (10·7462 is not included because this value is much more accurate than the other three values.)

Pivotal elements for various values of λ

$\lambda = 10\cdot7460$	$\lambda = 10\cdot7461$	$\lambda = 10\cdot7463$
1·0	1·0	1·0
1·0	1·0	1·0
1·0	1·0	1·0
1·0	1·0	1·0
1·0	1·0	1·0
1·0	1·0	1·0
1·0	1·0	1·0
−4·90289 29	−2·37850 29	2·67335 13
−8·61432 48	−8·61444 08	−8·61459 23
−9·62991 43	9·63001 59	−9·63021 79
−10·64215 69	−10·64225 80	−10·64246 02
−11·65203 41	−11·65213 50	−11·65233 68
12·66017 81	−12·66027 88	−12·66048 03
−13·66701 22	13·66711 28	−13·66731 41
14·67283 11	−14·67293 17	−14·67313 27
−15·67784 68	−15·67794 73	−15·67814 82
16·68221 57	−16·68231 61	−16·68251 70
−17·68605 59	17·68615 63	−17·68635 70
18·68945 83	−18·68955 86	−18·68975 92
−19·69249 39	19·69259 42	−19·69279 48
20·69521 92	−20·69531 95	−20·69552 00

If we were to take a succession of values of λ correct to 4, 5, 6 . . . decimals then it would be found that the number of p_i equal to unity would increase, until for a very accurate value of λ all p_i would be unity except p_{21}. Further increase in accuracy would result in $p_1, p_2, \ldots p_{20}$ remaining at 1 while p_{21} would progressively diminish. In fact

$$p_{21} \to 0 \text{ as } \lambda \to \lambda_1. \qquad (20)$$

Again we have the phenomenon that the relation (20) does not become evident until λ has approached very closely to λ_1. It is not surprising that the two phenomena arise together since they both have the same root cause. Choosing a matrix of the simple form of this example makes it easier to observe what is happening.

Some comment on the choice of right-hand side, $\boldsymbol{b}$, may now be made. If we have an exact value of λ for a matrix $\boldsymbol{A}$ and we perform the reduction process exactly, the last pivotal value will be zero. The last pivotal row is a linear combination,

$$k_1\boldsymbol{r}_1 + k_2\boldsymbol{r}_2 + \ldots + k_n\boldsymbol{r}_n \text{ (say)},$$

of the rows, $\boldsymbol{r}_i$, of $(\boldsymbol{A} - \lambda\boldsymbol{I})$ and the only linear combination of these rows which is identically zero is that for which the k_i are proportional to the elements of the eigenvector of $\boldsymbol{A}'$ corresponding to the eigenvalue λ. We shall denote this eigenvector by $\boldsymbol{v}^*$. If we append any right-hand side, $\boldsymbol{b}$, to give the equations

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{b}$$

then the right-hand side of the last pivotal equation will be

$$\boldsymbol{c}_n = k_1b_1 + k_2b_2 + \ldots + k_nb_n = \boldsymbol{b}'\boldsymbol{v}^*.$$

Since $\boldsymbol{v}^*$ is orthogonal to all eigenvectors of $\boldsymbol{A}$ other than that corresponding to λ a good choice of $\boldsymbol{b}$ is one for which $\boldsymbol{c}_n$ is not small. This argument would suggest that for a normal matrix, that is, one for which the last pivot would be small, the choice

$$\boldsymbol{c} = \boldsymbol{e}_n,$$

would be sound. (Because of possible interchanges this does not imply $\boldsymbol{b} = \boldsymbol{e}_n$). For matrices which exhibit the peculiarity described above we would not expect $\boldsymbol{c} = \boldsymbol{e}_n$ to be a good choice, but it is reasonable to assume that $\boldsymbol{c} = \boldsymbol{e}_r$, where r is such that the rth principal minor gives a small last pivot, will be a good choice. Taking $\boldsymbol{c} = \boldsymbol{e} = \sum_{r=1}^{n} \boldsymbol{e}_r$ is likely, therefore, to be very safe.

INFINITE MATRICES OF CODIAGONAL FORM

An interesting application of the above method arose recently in connection with a study of the Jahn–Teller effect (Longuet-Higgins, Öpik, Pryce, and Sack, 1958). Eigenvalues and eigenvectors of the two-parameter system of symmetric codiagonal matrices given below were required.

$$\boldsymbol{C}_{m,k}$$

$$\begin{bmatrix} m+1 & k\sqrt{m-1} & & & & \\ k\sqrt{m-1} & m-2 & k\sqrt{1} & & & \\ & k\sqrt{1} & m-3 & k\sqrt{m-2} & & \\ & & k\sqrt{m+2} & m+4 & k\sqrt{2} & \\ & & & k\sqrt{2} & m-5 & \\ & & & \ldots & \ldots & \ldots \end{bmatrix} \qquad (21)$$

If we consider the principal minor of order n then it has n eigenvalues of which the rth in order of increasing magnitude may be denoted by $\lambda_r(n)$. For fixed r we have

$$\lambda_r(n) \to \lambda_r \text{ as } n \to \infty$$

where λ_r is the rth eigenvalue of the infinite matrix. If $\boldsymbol{v}_r$ is the rth eigenvector of the infinite matrix normalized so that the sum of the squares of its components, v_{ri}, is unity, then

$$v_{ri} \to 0 \text{ as } i \to \infty$$

so that for a given accuracy all components of an eigenvector are zero from a certain point onwards. For an exact value of λ_r an eigenvector could be found by taking $v_{ri} = 1$ and solving exactly the equations

$$(m+1-\lambda_r) + k\sqrt{m+1}\, v_{r2} = 0$$

$$k\sqrt{m+1} + (m-2-\lambda_r)v_{r2} - k\sqrt{1}\, v_{r3} = 0$$

$$k\sqrt{1}\, v_{r2} - (m+3-\lambda_r)v_{r3} - k\sqrt{m-2}\, v_{r4} = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

successively to give $v_{r,2}, v_{r,3}, v_{r,4} \ldots$ For a good approximation to λ_r the exact solution of these equations always

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

MATRIX 1

0·25000	0·06675	0·04000	−0·02475	−0·07050	−0·06375	−0·06925	−0·02050	−0·03600	0·01025	0·00175	−0·02750	−0·02300	−0·00260
	0·25000	0·10400	0·07475	0·03625	0·11675	0·11050	0·06225	0·05100	−0·03250	−0·02400	0·03600	0·06350	0·05300
		−0·25000	0·14575	0·03725	0·07175	0·07800	0·12200	0·11275	0·09375	0·10175	0·09600	0·14300	0·11550
			−0·25000	0·05375	0·07000	0·05225	0·12800	0·12475	0·10550	0·13000	0·14575	0·13975	0·13375
				−0·25000	0·04575	0·05750	0·05700	0·05050	0·01475	0·04500	0·07150	0·05300	0·01600
					−0·25000	0·08625	0·08800	0·07150	0·04850	0·03200	0·04475	0·03300	0·04500
						−0·25000	0·08725	0·06950	0·03725	0·04025	0·04300	0·04075	0·01450
							−0·25000	0·14100	0·13275	0·15550	0·13050	0·11825	0·09125
								0·25000	0·07425	0·10750	0·09175	0·10725	0·08225
									0·25000	0·15500	0·09625	0·09950	0·09425
										−0·25000	0·13350	0·14850	0·13050
											0·25000	0·11100	0·10075
												0·25000	0·14325
													−0·25000

gives a good approximation to the earlier component of r_r, but the later components tend to infinity instead of to zero. Accurate eigenvectors were found for a large number of values of k and m, using the method described in this paper. Its automatic application gave consistently accurate eigenvectors. For some of the values of k and m which were of interest it was necessary to consider matrices of orders up to 128. An example is given at the end of this paper.

NUMERICAL EXAMPLES

Three examples are given which illustrate the above points. The first is derived from the matrix of order 14 given by Brooker and Sumner (1956) (see Matrix 1).

GIVENS CODIAGONAL FORM AND ITS ROOTS

α_i	β_i	λ_i
0·25000 0000	0·00000 0000	1·33403 4844
0·76849 1173	0·15366 0746	0·46276 6202
0·91955 6756	0·46726 0328	0·26773 3297
0·23093 8895	0·11925 6498	0·23163 9484
0·13305 3788	0·08076 3539	0·17735 6337
0·22254 9575	0·03394 7196	0·17130 7560
0·11612 7856	0·03609 0904	0·16632 4602
0·12033 9373	0·03502 2375	0·14342 2880
0·12371 9912	0·02915 7561	0·12278 7524
0·12856 1407	0·03745 3705	0·10321 5761
0·10776 8089	0·01609 0599	0·09720 9219
0·13703 9203	0·02382 6467	0·08422 5269
0·13805 7030	0·02946 8449	0·07359 7119
0·10379 6943	0·00764 6394	−0·06437 9910

FIRST FOUR EIGENVECTORS OF CODIAGONAL FORM CALCULATED BY DEUCE PROGRAM

−0·12180 621	0·72220 468	−0·49474 263	−0·34701 603
−0·85930 971	1·00000 000	−0·05709 601	−0·04146 403
−1·00000 000	0·89179 298	−0·22388 762	+0·06647 838
0·10864 611	−0·50225 957	1·00000 000	−0·54593 352
0·00731 255	−0·12487 537	−0·78617 724	−0·10289 858
0·00022 356	−0·01793 071	−0·73992 634	1·00000 000
0·00000 663	−0·00188 653	0·18686 459	0·34864 816
0·00000 019	−0·00019 439	−0·04640 116	0·11940 974
0·00000 001	−0·00001 692	−0·01011 081	−0·03703 434
0·00000 000	−0·00000 190	−0·00275 399	−0·01375 120
−0·00000 000	−0·00000 009	−0·00028 519	−0·00188 770
+0·00000 000	−0·00000 000	+0·00005 483	−0·00052 743
+0·00000 000	−0·00000 000	−0·00001 250	+0·00016 690
−0·00000 000	−0·00000 000	−0·00000 059	−0·00000 998

The eigenvectors of the codiagonal matrix illustrate the phenomenon referred to. The later vectors do not provide any points of interest and are quite normal in form. All vectors were correct to within 2 units of the seventh decimal, the largest errors being in the vectors corresponding to the smallest eigenvalues. For the four vectors quoted above the maximum error is 6 units in the eighth decimal place. In each of the first two vectors the last component is so small that we would expect the solution of the first $(n \quad 1)$ equations to be quite valueless. This is illustrated below where the vector obtained by inserting the approximate value of λ_1 in the first $(n - 1)$ equations is given. To emphasize that the result is not due to solving these equations inaccurately with this value of λ_1 the $(n - 1)$ equations were solved to double-length accuracy. The resulting vector is given below in normalized form.

−0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 000
0·00000 004
0·00000 306
0·00015 755
0·00639 731
−1·00000 000

As a second example we consider the matrix given below which arose in the theory of molecular orbitals.

MATRIX 2

0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	1	0	0	0
0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
0	0	0	0	0	0	1	0	0	0	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	1

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022

The eigenvalues produced by the DEUCE program include the three values

$$\lambda_5 = 1 + 2^{-26}$$
$$\lambda_6 = 1 \text{ exactly}$$
$$\lambda_7 = 1 - 2^{-27}$$

which were the approximations produced for the triple root $\lambda = 1$. The corresponding normalized vectors were

v_5	v_6	v_7
0·1108 773	−0·4540 761	−0·4116 218
−0·3033 805	+0·1960 617	+0·0402 348
0·1925 032	0·2580 145	−0·4518 566
+0·1108 773	−0·4540 761	+0·4116 218
−0·3033 805	−0·1960 617	−0·0402 348
−0·0000 000	−0·0000 000	+0·0000 000
0·3033 805	−0·1960 617	+0·0402 348
−0·1530 415	+0·3486 117	−0·0720 252
+0·4564 220	−0·1525 500	−0·0317 904
−0·3033 805	−0·1960 617	−0·0402 348
−0·1530 414	−0·3486 117	−0·0720 253
−0·4564 220	−0·1525 500	0·0317 905
+0·0000 000	−0·0000 000	−0·0000 000
−0·1925 032	0·2580 145	0·4518 566
0·2639 188	−0·1054 644	−0·4836 470

These vectors were produced entirely automatically and it may easily be verified that they are accurate eigenvectors. They are also independent and therefore form a complete set of vectors corresponding to $\lambda = 1$. That this should happen is not surprising when we consider the meaning of equation (15) for coincident roots. An interesting feature of this example is that we may easily guess three eigenvectors of A from any one of the vectors v_5, v_6, or v_7.

For example, v_5 is of the form shown in Table 1,

$$\text{where } a = 0{\cdot}1108773$$
$$b = 0{\cdot}3033805$$
$$c = 0{\cdot}1530415$$

Since the a, b and c do not appear to be simply related we might expect the vector v_5 to split into

$$a\boldsymbol{x} - b\boldsymbol{y} + c\boldsymbol{z}$$

where $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ were eigenvectors as shown in Table 1. It may readily be verified that $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent vectors of the original matrix. The above device has frequently been used with "exact" matrices with "exact" coincident roots.

TABLE 1

v_5	x	y	z
$-a$	−1	0	0
$-b$	0	−1	0
$(a-b)$	1	−1	0
a	1	0	0
b	0	1	0
0	0	0	0
$-b$	0	−1	0
c	0	0	1
$(b-c)$	0	1	1
b	0	1	0
$-c$	0	0	−1
$(-b-c)$	0	−1	−1
0	0	0	0
$(-a+b)$	−1	1	0
$(a-c)$	1	0	1

As a final example we give the 20th root and vector of the infinite matrix discussed earlier with $k = 1$, $m = 0$.

$\lambda_{20} = 19{\cdot}4880$

v_{20} (first 35 components. All later components are zero to 4 decimal places.)

Components 1 to 12	*Components 13 to 24*	*Components 25 to 35*
−0·0000	−0·1083	0·3618
−0·0000	−0·2290	−0·3167
0·0000	−0·3668	0·2103
−0·0000	−0·3678	−0·1169
0·0000	−0·0867	0·0557
−0·0000	−0·2748	−0·0239
0·0000	−0·2231	0·0092
−0·0001	−0·2263	−0·0033
0·0007	−0·1864	0·0011
−0·0030	−0·3007	−0·0003
0·0122	−0·0414	0·0001
−0·0394	−0·2460	

This vector, which is a correct infinite vector to four decimal places, was calculated automatically by the DEUCE program and it is significant that a value of λ correct to 4 decimal places gave the full accuracy in the vector.

ACKNOWLEDGEMENT

The work described here has been carried out as part of the Research Programme of the National Physical Laboratory and this paper is published by permission of the Director of the Laboratory.

REFERENCES

BROOKER, R. A., and SUMNER, F. H. (1956). "The Method of Lanczos for Calculating the Characteristic Roots and Vectors of a Real Symmetric Matrix," *Proc. I.E.E.*, Vol. 103, Part B, Supplement No. 1, p. 114.

GIVENS, W. (1954). "Numerical Computation of the Characteristic Values of a Real Symmetric Matrix," *Oak Ridge National Laboratory Report*, No. 1574.

LANCZOS, C. (1950). "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators," *Journal of Research of the National Bureau of Standards*, Vol. 45, p. 255.

LONGUET-HIGGINS, H. C., ÖPIK, U., PRYCE, M. H. L., and SACK, R. A. (1958). "Studies of the Jahn–Teller Effect. II. The Dynamical Problem," *Proceedings of the Royal Society*. A. Vol. 244, p. 1.

NATIONAL PHYSICAL LABORATORY (1957). *Modern Computing Methods*. Notes on Applied Science No. 16. London: H.M.S.O.

Downloaded from https://academic.oup.com/comjnl/article/1/2/90/425260 by guest on 21 August 2022