SIAM J. NUMER. ANAL.
Vol. 10, No. 2, April 1973

AN ALGORITHM FOR GENERALIZED MATRIX EIGENVALUE PROBLEMS*

C. B. MOLER† AND G. W. STEWART‡

Abstract. A new method, called the QZ algorithm, is presented for the solution of the matrix eigenvalue problem $Ax = \lambda Bx$ with general square matrices A and B. Particular attention is paid to the degeneracies which result when B is singular. No inversions of B or its submatrices are used. The algorithm is a generalization of the QR algorithm, and reduces to it when $B = I$. Problems involving higher powers of λ are also mentioned.

1. Introduction. We shall be concerned with the matrix eigenvalue problem of determining the nontrivial solutions of the equation

$$Ax = \lambda Bx,$$

where A and B are real matrices of order n. When B is nonsingular this problem is formally equivalent to the usual eigenvalue problem $B^{-1}Ax = \lambda x$.

When B is singular, however, such a reduction is not possible, and in fact the characteristic polynomial $\det(A - \lambda B)$ is of degree less than n, so that there is not a complete set of eigenvalues for the problem. In some cases the missing eigenvalues may be regarded as "infinite." In other cases the entire problem may be poorly posed. The term infinite eigenvalue is justified by the fact that if B is perturbed slightly so that it is no longer singular, there may appear a number of large eigenvalues that grow unboundedly as the perturbation is reduced to zero. However, if $\det(A - \lambda B)$ vanishes identically, say when A and B have a common null space, then any λ may be regarded as an eigenvalue. Such problems have unusually pathological features, and we refer to them as "ill-disposed" problems.

In numerical work the sharp distinction between singular and nonsingular matrices is blurred, and the pathological features associated with singular B carry over to the case of nearly singular B. The object of this paper is to describe an algorithm for computing the eigenvalues and corresponding eigenvectors that is unaffected by nearly singular B. The algorithm, the heart of which we call the QZ algorithm, is essentially an iterative method for computing the decomposition contained in the following theorem [10].

THEOREM. *There are unitary matrices Q and Z so that QAZ and QBZ are both upper triangular.*

We say that the eigenvalue problems $QAZy = \lambda QBZy$ and $Ax = \lambda Bx$ are unitarily equivalent. The two problems obviously have the same eigenvalues, and their eigenvectors are related by the equation $x = Zy$.

* Received by the editors October 19, 1971, and in revised form February 18, 1972.

† Department of Mathematics, University of New Mexico, Albuquerque, N.M. 87106. Most of the work of this author was done during a visit to Stanford University with support from the Computer Science Department, the Stanford Linear Accelerator Center and the National Science Foundation under Grant GJ-1158. Partial support was also obtained at the University of Michigan from the Office of Naval Research under Contract NR-044-377.

Cleve B. Moler received his Ph.D. in Mathematics in 1964 from Stanford University under the direction of Prof. Forsythe. After many years at the University of Michigan, he is now Professor of Mathematics at the University of New Mexico.

‡ Departments of Computer Science and Mathematics, Carnegie–Mellon University, Pittsburg, P.A. 15213. The work of this author was supported by the National Science Foundation under Grant GP-23655 and by the Office of Naval Research under Contract N00014-67-A-0216.

The algorithm proceeds in four stages. In the first, which is a generalization of the Householder reduction of a single matrix to Hessenberg form [4], [5], A is reduced to upper Hessenberg form and at the same time B is reduced to upper triangular form. In the second step, which is a generalization of the Francis implicit double shift QR algorithm [3], [8], A is reduced to quasi-triangular form while the triangular form of B is maintained. In the third stage the quasi-triangular matrix is effectively reduced to triangular form and the eigenvalues extracted. In the fourth stage the eigenvectors are obtained from the triangular matrices and then transformed back into the original coordinate system.

The transformations used in reducing A and B are applied in such a way that Wilkinson's general analysis of the roundoff errors in unitary transformations [11] shows that the computed matrices are exactly unitarily equivalent to slightly perturbed matrices $A + E$ and $B + F$. This means that the computed eigenvalues, which are the ratios of the diagonal elements of the final matrices, are the exact eigenvalues of the perturbed problem $(A + E)x = \lambda(B + F)x$. If an eigenvalue is well-conditioned in the sense that it is insensitive to small perturbations in A and B (see [10] for a detailed analysis), then it will be computed accurately. This accuracy is independent of the singularity or nonsingularity of B.

The use of unitary transformations in the reduction also simplifies the problem of convergence: a quantity may be set to zero if a perturbation of the same size can be tolerated in the original matrix.

Our computer program does not actually produce the eigenvalues λ_i but instead returns α_i and β_i, the diagonal elements of the triangular matrices QAZ and QBZ. The divisions in $\lambda_i = \alpha_i/\beta_i$ become the responsibility of the program's user. We emphasize this point because the α_i and β_i contain more information than the eigenvalues themselves.

Since our algorithm is an extension of the QR algorithm, the well-known properties of the QR algorithm apply to describe the behavior of our algorithm.

In their survey article [9], Peters and Wilkinson describe another approach for the case when B is nearly singular. In their method one computes an approximate null space for B and removes it from the problem. The technique is reapplied to the deflated problem, and so on until a well-conditioned problem is obtained. The method has the crucial drawback that one must determine the rank of B. If a wrong decision is reached, the well-conditioned eigenvalues may be seriously affected. A similar algorithm for rectangular matrices is given in [13].

The special case where A is symmetric and B is positive definite has been extensively treated. For the case of well-conditioned B the "Cholesky–Wilkinson" method [6] enjoys a well deserved popularity. A modification of this algorithm for band matrices is given by Crawford [1]. A variant of the Peters–Wilkinson method for nearly semidefinite B has been given by Fix and Heiberger [2]. Although our method does not preserve symmetry and is consequently more time consuming than these algorithms, its stability may make it preferable when B is nearly semidefinite.

Our algorithm can also be used to solve "λ-matrix" problems of the form

$$(\lambda^r C_r + \lambda^{r-1} C_{r-1} + \cdots + C_0)x = 0$$

by forming the generalized block companion matrices. For example, when $r = 3$,

$$A = \begin{pmatrix} C_2 & C_1 & C_0 \\ I & 0 & 0 \\ 0 & I & 0 \end{pmatrix}, \qquad B = -\begin{pmatrix} C_3 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Note that neither C_r nor C_0 is assumed to be nonsingular.

2. Reduction to Hessenberg-triangular form. In this section we shall give an algorithm whereby A is reduced to upper Hessenberg form and simultaneously B is reduced to triangular form. While a treatment of the reductions in this and the following sections can be given in terms of standard plane rotations and elementary Hermitian matrices, we find it convenient from a computational point of view to work exclusively with a modified form of the elementary Hermitians. Accordingly, we introduce the following notation.

By $\mathscr{H}_r(k)$ we mean the class of symmetric, orthogonal matrices of the form

$$I + vu^T,$$

where $v^Tu = -2$, v is a scalar multiple of u, only components $k, k+1, \cdots, k+r-1$ of u are nonzero, and $u_k = 1$. Given any vector x, it is easy to choose a member Q of $\mathscr{H}_r(k)$ so that

$$Qx = x + (u^Tx)v$$

has its $k+1, \cdots, k+r-1$ components equal to zero, its kth component changed and all other components unchanged. Since $u_k = 1$, the computation of Qy for any y requires only $2r-1$ multiplications and $2r-1$ additions. (In particular, use of a matrix in $\mathscr{H}_2$ requires only 3 multiplications instead of the 4 required by a standard plane rotation.)

For the most part, we shall use only matrices in $\mathscr{H}_2$ and $\mathscr{H}_3$. When a matrix Q in $\mathscr{H}_3(k)$ premultiplies a matrix A, only rows k, $k+1$, and $k+2$ in QA are changed. If the elements $k, k+1$, and $k+2$ in a column of A are zero, they remain zero in QA. Likewise, if $Z \in \mathscr{H}_3(k)$, only columns $k, k+1$, and $k+2$ are changed in AZ. If some row has elements $k, k+1$, and $k+2$ zero, then they remain zero in AZ. Similar considerations hold for the class $\mathscr{H}_2$.

All our transformations will be denoted by Q's and Z's with various subscripts. The Q's will always be premultipliers, that is, row operations. The Z's will always be postmultipliers, or column operations. The letter Q is being used in its traditional role to denote orthogonal matrices. The letter Z was chosen to denote orthogonal matrices which introduce zeros in strategic locations.

The first step in the reduction is to reduce B to upper triangular form by premultiplication by Householder reflections. The details of this reduction are well known (e.g., see [4], [11]) and we confine ourselves to a brief description to illustrate our notation. At the kth stage of the reduction (illustrated below for $k = 3$ and $n = 5$), the elements below the first $k-1$ diagonal elements of B are zero:

$$\begin{matrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & x^1 & x & x \\ 0 & 0 & x^1 & x & x \end{matrix}$$

Each x represents an arbitrary nonzero element. Each x^1 represents an element to be annihilated in the next step. A matrix $Q_k \in \mathscr{H}_{n-k+1}(k)$ is chosen to annihilate $b_{k+1,k}, b_{k+2,k}, \cdots, b_{n,k}$, and B is overwritten by Q_kB, giving a matrix of the form illustrated below.

$$\begin{matrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x^1 & x \end{matrix}$$

This process is repeated until $k = n - 1$. Of course A is overwritten by $Q_{n-1}Q_{n-2} \cdots Q_1A$.

After this reduction, A and B have the forms

$$A = \begin{matrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x^1 & x & x & x & x \end{matrix} \qquad B = \begin{matrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & x \end{matrix}$$

The problem now is to reduce A to upper Hessenberg form while preserving the triangularity of B. This is done as follows (for $k = 5$). First $Q \in \mathscr{H}_2(4)$ is determined to annihilate a_{51}. The matrices QA and QB, which overwrite A and B, then have the forms

$$\begin{matrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ 0 & x & x & x & x \end{matrix} \qquad \begin{matrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x^1 & x \end{matrix}$$

The transformation has introduced a nonzero element on the (5,4)-position of B. However, a $Z \in \mathscr{H}_2(4)$ can be used to restore the zero without disturbing the zero introduced in A. The elements of A can be annihilated in the following order:

$$\begin{matrix} x & x & x & x & x \\ x & x & x & x & x \\ x^3 & x & x & x & x \\ x^2 & x^5 & x & x & x \\ x^1 & x^4 & x^6 & x & x \end{matrix}$$

As each element of A is annihilated, it introduces a nonzero element on the subdiagonal of B, which is immediately annihilated by a suitably chosen Z. The entire algorithm, including the Householder triangularization of B, may be summed up as follows:

1. For $k = 1, 2, \cdots, n-1$,
   (i) choose $Q_k \in \mathscr{H}_{n-k+1}(k)$ to annihilate $b_{k+1,k}, b_{k+2,k}, \cdots, b_{n,k}$;
   (ii) $B \leftarrow Q_k B$, $A \leftarrow Q_k A$.
2. For $k = 1, 2, \cdots, n-2$,
   (i) for $l = n-1, n-2, \cdots, k+1$,
      (a) choose $Q_{kl} \in \mathscr{H}_2(l)$ to annihilate $a_{l+1,k}$;
      (b) $A \leftarrow Q_{kl} A$, $B \leftarrow Q_{kl} B$;
      (c) choose $Z_{kl} \in \mathscr{H}_2(l)$ to annihilate $b_{l+1,l}$;
      (d) $B \leftarrow B Z_{kl}$, $A \leftarrow A Z_{kl}$.

The complete reduction requires about $\frac{17}{3}n^3$ multiplications, $\frac{17}{3}n^3$ additions and n^2 square roots. If eigenvectors are also to be computed, the product of the Z's must be accumulated. This requires an additional $\frac{3}{2}n^3$ multiplications and $\frac{3}{2}n^3$ additions. The product of the Q's is not required for the computation of eigenvectors.

3. The explicit QZ step. In this and the next section we assume that A is upper Hessenberg and B is upper triangular. In this section we shall propose an iterative technique for reducing A to upper triangular form while maintaining the triangularity of B. The idea of our approach is to pretend that B is nonsingular and examine the standard QR algorithm for $C = AB^{-1}$. The manipulations are then interpreted as unitary equivalences on A and B.

Specifically, suppose that one step of the QR algorithm with shift σ is applied to C. Then Q is determined as an orthogonal transformation such that the matrix

$$R = Q(C - \sigma I) \tag{3.1}$$

is upper triangular. The next iterate C' is defined as

$$C' = RQ^T + \sigma I \equiv QCQ^T$$

and is known to be upper Hessenberg. If we set

$$A' = QAZ \quad \text{and} \quad B' = QBZ,$$

where Z is any unitary matrix, then

$$A'B'^{-1} = QAZZ^T B^{-1} Q^T = QAB^{-1}Q^T = C'.$$

The matrix Z can be chosen so that B' is upper triangular. Then, since $A' = C'B'$ is the product of a Hessenberg and a triangular matrix, it is also upper Hessenberg. This insures that the nice distribution of zeros, introduced by the algorithm of § 2, is preserved by the QZ step. Thus a tentative form of our algorithm might read:

1. Determine Q so that QC is upper triangular.
2. Determine Z so that QAZ is upper Hessenberg and QBZ is upper triangular.
3. $A \leftarrow QAZ$, $B \leftarrow QBZ$.

The problem is then to give algorithms for computing Q and Z which do not explicitly require $C = AB^{-1}$.

The determination Q is relatively easy. For from (3.1) and the definition of C it follows that

$$Q(A - \sigma B) = RB \equiv S. \tag{3.2}$$

Since R and B are upper triangular, so is S. Thus Q is the unitary matrix that reduces $A - \sigma B$ to upper triangular form. Since $A - \sigma B$ is upper Hessenberg, Q can be expressed in the form

$$Q = Q_{n-1}Q_{n-2}\cdots Q_1, \quad \text{where } Q_k \in \mathscr{H}_2(k). \tag{3.3}$$

To calculate Z we apply Q in its factored form (3.3) to B and determine Z in a factored form so that B stays upper triangular. Specifically, Q_1B has the form ($n = 5$)

$$\begin{matrix} x & x & x & x & x \\ x^1 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & x \end{matrix}$$

If Q_1B is postmultiplied by a suitable $Z_1 \in \mathscr{H}_2(1)$, the nonzero element below the diagonal can be removed. Similarly, $Q_2Q_1BZ_1$ has the form

$$\begin{matrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & x^1 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & x \end{matrix}$$

and the offending nonzero element can be removed by a $Z_2 \in \mathscr{H}_2(2)$. Proceeding in this way, we construct Z in the form

$$Z = Z_1Z_2\cdots Z_{n-1}, \quad \text{where } Z_k \in \mathscr{H}_2(k).$$

Although QBZ is upper triangular, it is not at all clear that QAZ is upper Hessenberg. To see that it is, rewrite (3.2) in the form

$$QAZ = SZ + \sigma QBZ. \tag{3.4}$$

From the particular form of Z and the fact that S is upper triangular, it follows that SZ is upper Hessenberg. Thus (3.4) expresses QAZ as the sum of an upper Hessenberg and an upper triangular matrix. In fact, (3.4) represents a computationally convenient form for computing QAZ.

We summarize as follows.

1. Determine $Q = Q_{n-1}Q_{n-2}\cdots Q_1$ ($Q_k \in \mathscr{H}_2(k)$) so that $S = Q(A - \sigma B)$ is upper triangular.
2. Determine $Z = Z_1Z_2\cdots Z_{n-1}$ ($Z_k \in \mathscr{H}_2(k)$) so that $B' = QBZ$ is upper triangular.
3. $A' = SZ + \sigma B'$.

If this algorithm is applied iteratively with shifts $\sigma_1, \sigma_2, \cdots$, there result sequences of matrices $A_1, A_2, \cdots$, and $B_1, B_2, \cdots$ satisfying

$$A_{v+1} = Q_v A_v Z_v, \qquad B_{v+1} = Q_v B_v Z_v.$$

The matrices A_v are upper Hessenberg and the B_v are upper triangular. Moreover, if B_1 is nonsingular, then the matrices $C_v = A_v B_v^{-1}$ are the matrices which would have been obtained by applying the standard QR algorithm with shifts $\sigma_1, \sigma_2, \cdots$ to $C_1 = A_1 B_1^{-1}$. As C_v tends to upper triangular form, so must A_v, since B^{-1} is upper triangular.

Most of the properties of the QR algorithm carry over to the QZ algorithm. The eigenvalues will tend to appear in descending order as one proceeds along the diagonal. The convergence of $a_{n,n-1}^{(v)}$ to zero may be accelerated by employing one of the conventional shifting strategies. Once $a_{n,n-1}^{(v)}$ becomes negligible one can deflate the problem by working with the leading principal submatrices of order $n-1$. If some other subdiagonal element of A_v, say $a_{l,l-1}^{(v)}$, becomes negligible, one can effect a further savings by working with rows and columns l through n. Because we have used unitary transformations, an element of A_v or B_v can be regarded as negligible if a perturbation of the same size as the element can be tolerated in A_1 or B_1.

The algorithm given above is potentially unstable. If σ is large compared with A and B, the formula (3.4) will involve subtractive cancellation and A' will be computed inaccurately. Since the shift σ approximates the eigenvalue currently being found and the problem may have very large eigenvalues, there is a real possibility of encountering a large shift. Fortunately the large eigenvalues tend to be found last so that by the time a large shift emerges the small eigenvalues will have been computed stably. (The large eigenvalues are of course ill-conditioned and cannot be computed accurately.) To be safe one might perform the first few iterations with a zero shift in order to give the larger eigenvalues a chance to percolate to the top.

4. Implicit shifts. The potential instability in the explicit algorithm results from the fact that we have used formula (3.4) rather than unitary equivalences to compute A'. One way out of this difficulty is to generalize the implicit shift method for the QR algorithm to the QZ algorithm so that both A' and B' are computed by unitary equivalences. The implicit shift technique has the additional advantage that it can be adapted to perform two shifts at a time. For real matrices this means that a double shift in which the shifts are conjugate pairs can be performed in real arithmetic.

Since we are primarily interested in real matrices, we shall concentrate on double shifts. The method is based on the following observation. Suppose that A is upper Hessenberg and B is upper triangular and nonsingular. Then if Q and Z are unitary matrices such that QAZ is upper Hessenberg and QBZ is upper triangular, then Q is determined by its first row. In fact, AB^{-1} and $QAB^{-1}Q^H$ are both upper Hessenberg, so that, by the theorem in [11, p. 352], Q is determined by its first row.

Thus we must do two things. First, find the first row of Q. Second, determine Q and Z so that Q has the correct first row, QAZ is upper Hessenberg, and QBZ

is upper triangular. The first part is relatively easy. The first row that would be obtained from a double shifted QR applied to AB^{-1}. Since A is upper Hessenberg and B upper triangular, it is easy to calculate the first two columns of AB^{-1}. But these, along with the shifts, completely determine the first row of Q. Only nonsingularity of the upper 2×2 submatrix of B is actually required here. If either b_{11} or b_{22} is too small, so that this submatrix is nearly singular, a type of deflation can be carried out. We shall return to this point later.

The second part is a little more difficult, and is really the crux of the algorithm since it retains the Hessenberg and triangular forms. Only the first three elements of the first row of Q are nonzero. Thus, if Q_1 is a matrix in $\mathscr{H}_3(1)$ with the same first row of Q, then Q_1A and Q_1B have the following forms (when $n = 6$):

$$\begin{matrix} x & x & x & x & x & x \\ x & x & x & x & x & x \\ x & x & x & x & x & x \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \end{matrix} \qquad \begin{matrix} x & x & x & x & x & x \\ x^2 & x & x & x & x & x \\ x^1 & x^1 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & x \end{matrix}$$

As in the standard implicit shift QR algorithm, it is convenient to think of Q_1 as the reflection which annihilates two of the three nonzero elements in a fictitious "zeroth" column of A.

We must reduce Q_1A to upper Hessenberg and Q_1B to upper triangular by unitary equivalences. However, we may not premultiply by anything which affects the first row. This is done as follows. The matrix Q_1B has three nonzero elements outside the triangle. These can be annihilated by two Z's, a Z_1' in $\mathscr{H}_3(1)$ which annihilates (3, 1)- and (3, 2)-elements and then a Z_1'' in $\mathscr{H}_2(1)$ which annihilates the resulting (2, 1)-element. Let $Z_1 = Z_1'Z_1''$. Then Q_1BZ_1 is upper triangular. Applying Z_1 to Q_1A gives Q_1AZ_1 with the following form:

$$\begin{matrix} x & x & x & x & x & x \\ x & x & x & x & x & x \\ x^1 & x & x & x & x & x \\ x^1 & x & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \end{matrix}$$

This is multiplied by Q_2 in $\mathscr{H}_3(2)$ that annihilates the (3, 1)- and (4, 1)-elements. Then $Q_2Q_1AZ_1$ and $Q_2Q_1BZ_1$ have the forms

$$\begin{matrix} x & x & x & x & x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & 0 & x & x & x & x & x \\ 0 & x & x & x & x & x & 0 & x^2 & x & x & x & x \\ 0 & x & x & x & x & x & 0 & x^1 & x^1 & x & x & x \\ 0 & 0 & 0 & x & x & x & 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & x & x & 0 & 0 & 0 & 0 & 0 & x \end{matrix}$$

The first columns are now in the desired form. The nonzero elements outside the desired structure have been "chased" into the lower 5×5 submatrices.

Now, postmultiply by Z_2, a product of a matrix in $\mathscr{H}_3(2)$ and a matrix in $\mathscr{H}_2(2)$ that reduces the current B to triangular form. Then premultiply by Q_3 in $\mathscr{H}_3(3)$ to annihilate two elements outside the Hessenberg structure of the resulting A.

The process continues in a similar way, chasing the unwanted nonzero elements towards the lower, right-hand corners. It ends with a slightly simpler step which uses Q_{n-2} in $\mathscr{H}_2(n-1)$ to annihilate the $(n, n-2)$-element of the current A, thereby producing a Hessenberg matrix, and Z_{n-2} in $\mathscr{H}_2(n-1)$ which annihilates the $(n, n-1)$-element of the current B, producing a triangular B but not destroying the Hessenberg A.

The fictitious zeroth column of A is determined in part by the shifts. In analogy with the implicit double shift algorithm, we take the shifts σ_1 and σ_2 to be the two zeros of the 2×2 problem

$$\det(\bar{A} - \sigma\bar{B}) = 0,$$

where

$$\bar{A} = \begin{pmatrix} a_{n-1,n-1} & a_{n-1,n} \\ a_{n,n-1} & a_{n,n} \end{pmatrix}, \qquad \bar{B} = \begin{pmatrix} b_{n-1,n-1} & b_{n-1,n} \\ 0 & b_{n,n} \end{pmatrix}.$$

It is not desirable to compute σ_1 and σ_2 explicitly, or even to find the coefficients in the quadratic polynomial $\det(\bar{A} - \sigma\bar{B})$. Instead, following the techniques used in "*hqr2*" [8], we obtain ratios of the three nonzero elements of the first column of $(AB^{-1} - \sigma_1 I)(AB^{-1} - \sigma_2 I)$ directly from formulas which involve only the differences of diagonal elements. This insures that small, but nonnegligible, off-diagonal elements are not lost in the shift calculation. The formulas are $(m = n - 1)$

$$a_{10} = \left[\left(\frac{a_{mm}}{b_{mm}} - \frac{a_{11}}{b_{11}}\right)\left(\frac{a_{nn}}{b_{nn}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{mn}}{b_{nn}}\right)\left(\frac{a_{nm}}{b_{mm}}\right) + \left(\frac{a_{nm}}{b_{mm}}\right)\left(\frac{b_{mn}}{b_{nn}}\right)\left(\frac{a_{11}}{b_{11}}\right)\right] \cdot \left(\frac{b_{11}}{a_{21}}\right) + \frac{a_{12}}{b_{22}} - \left(\frac{a_{11}}{b_{11}}\right)\left(\frac{b_{12}}{b_{22}}\right),$$

$$a_{20} = \left(\frac{a_{22}}{b_{22}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{21}}{b_{11}}\right)\left(\frac{b_{12}}{b_{22}}\right) - \left(\frac{a_{mm}}{b_{mm}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{nn}}{b_{nn}} - \frac{a_{11}}{b_{11}}\right) + \left(\frac{a_{nm}}{b_{mm}}\right)\left(\frac{b_{mn}}{b_{nn}}\right), \tag{4.1}$$

$$a_{30} = \frac{a_{32}}{b_{22}}.$$

We are now in a position to summarize the double implicit shift method. It is understood that A and B are to be overwritten by the transformed matrices as they are generated.

1. Compute a_{10}, a_{20}, and a_{30} by (4.1).
2. For $k = 1, 2, \cdots, n-2$,
   (a) determine $Q_k \in \mathscr{H}_3(k)$ to annihilate $a_{k+1,k-1}$ and $a_{k+2,k-1}$;
   (b) determine $Z'_k \in \mathscr{H}_3(k)$ to annihilate $b_{k+2,k+1}$ and $b_{k+2,k}$;
   (c) determine $Z''_k \in \mathscr{H}_2(k)$ to annihilate $b_{k+1,k}$.
3. Determine $Q_{n-1} \in \mathscr{H}_2(n-1)$ to annihilate $a_{n,n-2}$.
4. Determine $Z_{n-1} \in \mathscr{H}_2(n-1)$ to annihilate $b_{n,n-1}$.

For each k, determination of Q_k requires a few multiplications and one square root. Application of Q_k to both A and B requires about $10(n-k)$ multiplications. The work involved with each Z'_k is the same. Application of Z''_k requires only about $6(n-k)$ multiplications. The number of additions is about the same. Summing these for k from 1 to $n-1$ gives a total of about $13n^2$ multiplications, $13n^2$ additions and $3n$ square roots per double iteration.

By way of comparison, for the double shift QR algorithm as implemented in "*hqr*," Z'_k becomes simply Q_k^T and Z''_k is not used. Furthermore, the transformations are carried out on only one matrix. Consequently, each double iteration requires about $5n^2$ multiplications, $5n^2$ additions and n square roots. Thus the QZ algorithm applied on two matrices can be expected to require roughly 2.6 times as much work per iteration as the QR algorithm on a single matrix.

In order to obtain eigenvectors, the Q's are ignored and the Z's accumulated. This requires about $8n^2$ more multiplications and $8n^2$ more additions per double iteration.

There is one difficulty. The formulas for a_{10}, a_{20}, and a_{30} are not defined when b_{11} and b_{22} are zero. Moreover, if b_{11} and b_{22} are small the terms that determine the shift (terms involving a_{nn}, b_{nn}, etc.) become negligible compared to the other terms, so that the effect of the shift is felt only weakly.

Part of the solution to this difficulty is to deflate from the top. If b_{11} is negligible, it may be set to zero to give the forms for A and B ($n = 4$):

$$\begin{matrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \end{matrix} \qquad \begin{matrix} 0 & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & 0 & x \end{matrix}$$

A Q in $\mathscr{H}_2(1)$ can then be used to annihilate the (2, 1)-element of A, which deflates the problem.

The rest of the solution lies in recognizing that there is not much of a problem. If b_{11} and b_{22} are small, then the problem has large eigenvalues. We have already observed that the larger eigenvalues tend to emerge at the upper left, and the larger the eigenvalue, the swifter its emergence. Moreover, the speed will not be affected by a small shift. This means that whenever the implicit shift is diluted by a small b_{11} or b_{22}, the algorithm is none the less profitably employed in finding a large eigenvalue.

5. Further reduction of the quasi-triangular form. The result of the algorithm described so far is in an upper triangular matrix B and a quasi-upper triangular matrix A in which no two consecutive subdiagonal elements are nonzero. This means that the original problem decomposes into 1×1 and 2×2 subproblems. The eigenvalues of the 1×1 problems are the ratios of the corresponding diagonal elements of A and B. The eigenvalues of the 2×2 problems might be calculated as the roots of a quadratic equation, and may be complex even for real A and B.

There are two good reasons for not using the quadratic directly, but instead reducing the 2×2 problems. First, when A and B are real, the calculation of eigenvectors is greatly facilitated if all the real eigenvalues are contained in 1×1 problems. A more important second reason is that the 1×1 problems contain more information than the eigenvalues alone. For example, if a_{11} and b_{11} are small, then the eigenvalue $\lambda_1 = a_{11}/b_{11}$ is ill-conditioned, however reasonable it may appear. This reason obviously applies to complex eigenvalues as well as real ones. Accordingly, we recommend that the 2×2 problems be reduced to 1×1 problems and that the diagonal elements, rather than the eigenvalues, be reported.

Without loss of generality we may consider the problem of reducing 2×2 matrices A and B simultaneously to upper triangular form by unitary equivalences. For our purposes we may assume that B is upper triangular.

Two special cases may be disposed of immediately. If b_{11} is zero, then a $Q \in \mathcal{H}_2(1)$ may be chosen to reduce a_{21} to zero. The zero elements of QB are not disturbed. Similarly, if b_{22} is zero, a $Z \in \mathcal{H}_2(1)$ may be chosen to reduce a_{21} to zero without disturbing b_{21}.

In the general 2×2 case, it is not difficult to write down formulas for the elements of $A' = QAZ$ and $B' = QBZ$ for any Q and Z. Moreover, these formulas can be arranged so that numerically one of a'_{21} or b'_{21} is effectively zero. It is not obvious, however, that the other element is numerically zero, and the effect of assuming that it is by setting it to zero could be disastrous. Consequently, we must consider a somewhat more complicated procedure.

The theoretical procedure for reducing A to triangular form may be described as follows. Let λ be an eigenvalue of the problem and form the matrix $E = A - \lambda B$. Choose a $Z \in \mathcal{H}_2(1)$ to annihilate either e_{11} or e_{21}. Since the rows of E are parallel, it follows that whichever of e_{11} or e_{21} is annihilated the other must also be annihilated. Now choose $Q \in \mathcal{H}_2(1)$ so that either QAZ or QBZ is upper triangular. Since the first column of QEZ is zero and $QEZ = QAZ - \lambda QBZ$, it follows that, however Q is chosen, both QAZ and QBZ must be upper triangular.

In the presence of rounding error the method of computing λ and the choice of Z and Q are critical to the stability of the process. A rigorous rounding error analysis will show that, under a reasonable assumption concerning the computed λ, the process described below is stable. However, to avoid excessive detail, we only outline the analysis. We assume that all computations are done in floating-point arithmetic with t base β digits and that the problem has been so scaled that underflows and overflows do not occur. We further assume that a_{21} is not negligible in the sense that $|a_{21}| < \beta^{-t}\|A\|$, where $\|\cdot\|$ denotes, say, the row sum norm.

The algorithm for computing λ amounts to making an appropriate origin shift and computing an eigenvalue from the characteristic equation. It goes as follows.

$$\mu = a_{11}/b_{11},$$
$$\bar{a}_{12} = a_{12} - \mu b_{12},$$
$$\bar{a}_{22} = a_{22} - \mu b_{22},$$
$$p = \frac{1}{2}\left(\frac{\bar{a}_{22}}{b_{22}} - \frac{b_{12}a_{21}}{b_{11}b_{22}}\right),$$
$$q = \frac{a_{21}\bar{a}_{12}}{b_{11}b_{22}},$$
$$r = p^2 + q,$$
$$\lambda = \mu + p + \operatorname{sgn}(p)\cdot\sqrt{r} \quad \text{(complex if } r < 0\text{)}. \tag{5.1}$$

We must now assume that the computed λ satisfies the equation

$$\det(A' - \lambda B') = 0,$$

where $\|A - A'\| \leqq \sigma_A\|A\|$ and $\|B - B'\| \leqq \sigma_B\|B\|$ with σ_A and σ_B small constants of order β^{-t}. Define

$$E' = A' - \lambda B'$$

and let E denote the computed value

$$E = \mathrm{fl}(A - \lambda B).$$

Then

$$E' = E + H$$

with $\|H\| \leqq \sigma \max\{\|A\|, |\lambda|\,\|B\|\}$ with σ of order β^{-t}.

We claim that, approximately,

$$\|E\| \geqq \beta^{-t}\max\{\|A\|, |\lambda|\,\|B\|\}. \tag{5.2}$$

First we note that

$$\|E\| \geqq |e_{21}| = |a_{21}| \geqq \beta^{-t}\|A\|, \tag{5.3}$$

by the assumption that a_{21} is significant. Now assume that $\|E\| < \beta^{-t}|\lambda|\,\|B\|$. Then subtractive cancellation must occur in the computation of e_{11}, e_{12}, and e_{22}. Thus $a_{11} \approx \lambda b_{11}$, $a_{12} \approx \lambda b_{12}$ and $a_{22} \approx \lambda b_{22}$. Hence we have $\|A\| \geqq |\lambda|\,\|B\|$, and, from (5.3), $\|E\| \geqq \beta^{-t}|\lambda|\,\|B\|$, a contradiction.

Now

$$0 = \det(E') = \det(E) + (e_{11} + h_{11})h_{22} - (e_{12} + h_{12})h_{21} + h_{11}e_{22} - h_{12}e_{21}.$$

Hence

$$|\det(E)| \leqq \rho_1\|E\|\max\{\|A\|, |\lambda|\,\|B\|\} + \rho_2[\max\{\|A\|, |\lambda|\,\|B\|\}]^2,$$

where ρ_1 and ρ_2 are of order β^{-t}. From (5.2) it then follows that

$$|\det(E)| \leqq \rho\|E\|\max\{\|A\|, |\lambda|\,\|B\|\},$$

where ρ is of order β^{-t}.

Now consider the determination of Z. Assume that the second row of E is larger than the first. Then $Z \in \mathscr{H}_2(1)$ is chosen to annihilate e_{21}. Let $F = EZ$. Then f_{21} is essentially zero. Furthermore, since Z is unitary,

$$|f_{11}f_{22}| = |\det(E)| \leqq \rho\|E\|\max\{\|A\|, |\lambda|\,\|B\|\}.$$

But $|f_{22}| \cong \|e_2\|$ and, since e_2 was assumed to be the larger row, $\|e_2\| = \|E\|$. Hence we have approximately

$$|f_{11}| \leqq \rho \max \{\|A\|, |\lambda| \, \|B\|\}.$$

To choose Q, let

$$C = AZ, \qquad D = BZ,$$

and let f_1, c_1, and d_1 be the first columns of F, C, and D. Let q_2^T denote the second row of Q. If $\|A\| \geqq |\lambda| \, \|B\|$, we choose Q to annihilate d_{21}. Numerically this means that

$$|q_2^T d_1| \leqq \sigma \|B\|,$$

where σ is a constant of the order of β^{-t}. We must show that $q_2^T c_1$ is negligible. But

$$\begin{aligned} |q_2^T c_1| &= |q_2^T f_1 + \lambda q_c^T d_1| \\ &\leqq \|f_1\| + |\lambda| \, \|q_c^T d_1\| \\ &\leqq \rho \max \{\|A\|, |\lambda| \, \|B\|\} + \sigma |\lambda| \, \|B\| \\ &\leqq (\rho + \sigma) \|A\|. \end{aligned}$$

If, on the other hand, $|\lambda| \, \|B\| > \|A\|$, we choose Q so that

$$|q_2^T c_1| \leqq \sigma \|A\|.$$

It then follows that

$$\begin{aligned} |q_2^T d_1| &= |q_1^T f - q_2^T c_1| / |\lambda| \\ &\leqq \rho |\lambda|^{-1} \max \{\|A\|, |\lambda| \, \|B\|\} + \sigma |\lambda|^{-1} \|A\| \\ &\leqq (\rho + \sigma) \|B\|. \end{aligned}$$

In summary, λ is computed using (5.1), Z is chosen to annihilate the first element of the larger of the two rows of $A - \lambda B$ and Q is chosen to annihilate the (2, 1)-element of the smaller of the two matrices AZ and λBZ. In this way, we can be sure that the computed (2, 1)-elements of both QAZ and QBZ are negligible.

In practice with matrices of any order, if the transformations are real, they are applied to the entire matrices. If the transformations are complex, they are used to compute the diagonal elements that would result, but are not actually applied. We thus obtain a quasi-triangular problem in which each 2×2 block is known to correspond to a pair of complex eigenvalues.

The generalized eigenvectors of this reduced problem can be found by a back-substitution process which is a straightforward extension of the method used in "*hqr*2" [8]. The vectors of the original problem are then found by applying the accumulated Z's.

6. Some numerical results. The entire process described above has been implemented in a FORTRAN program [7]. There are four main subroutines: the initial reduction to Hessenberg-triangular form, the iteration itself, the computation of the final diagonal elements, and the computation of the eigenvectors. The complete program contains about 600 FORTRAN statements, although this could be reduced somewhat at the expense of some clarity.

The numerical properties observed experimentally are consistent with the use of unitary transformations. The eigenvalues are always found to whatever accuracy

is justified by their condition. If an eigenvalue and eigenvector are not too "ill-disposed," then they produce a small relative residual.

Similar numerical properties cannot generally be expected from any algorithm which inverts B or any submatrix of B. This is even true of 2×2 submatrices, as illustrated by the following example due to Wilkinson:

$$A = \begin{pmatrix} .1 & .2 \\ .3 & .4 \end{pmatrix}, \qquad B = \begin{pmatrix} .1 & .1 \\ 0 & \mu \end{pmatrix}.$$

Here μ is about the square root of the machine precision, that is, μ is not negligible compared to 1, but μ^2 is. There is one eigenvalue near -2. Small relative changes in the elements of the matrices cause only small relative changes in this eigenvalue. The other eigenvalue becomes infinite as μ approaches zero. Great care must be taken in solving this problem so that the mild instability of the one eigenvalue does not cause an inaccurate result for the other, stable eigenvalue.

Of course, the use of unitary transformations makes our technique somewhat slower than others which might be considered. But the added cost is not very great. In testing our program, we solve problems of order 50 regularly. A few problems of orders greater than 100 have been run, but these become somewhat expensive when they are merely tests.

One typical example of order 50 requires 45 seconds on Stanford's IBM 360 model 67. Of this, 13 seconds are spent in the initial reduction, 29 seconds are used for the 61 double iterations required, and 3 seconds are needed for the diagonal elements and eigenvectors. If the eigenvectors are not needed and so the transformations not saved, the total time is reduced to 27 seconds. By way of comparison, formation of $B^{-1}A$ à la Peters and Wilkinson [9] and use of FORTRAN versions [12] of "orthes" [5] and "*hqr2*" [8] requires a total of 27 seconds for this example. (All of these times are for code generated by the IBM FORTRAN IV compiler, *H* level, with the optimization parameter set to 2.)

In the examples we have seen so far, the total number of double iterations required is usually about 1.2 or 1.3 times the order of the matrices. This figure is fairly constant, although it is not difficult to find examples which require many fewer or many more iterations. As a rule of thumb, for a matrix of order n the time required on the model 67 is about $.36\,n^3$ milliseconds if vectors are computed, $.22\,n^3$ milliseconds if they are not.

The example in Table 1 is not typical, but it does illustrate several interesting points. It was generated by applying nonorthogonal rank one modifications of the identity to direct sums of companion matrices. The companion matrices were chosen so that the resulting problem has three double roots,

$$\lambda_1 = \lambda_2 = \infty,$$
$$\lambda_3 = \lambda_5 = \frac{1}{2} + \frac{\sqrt{3}}{2}i,$$
$$\lambda_4 = \lambda_6 = \frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

TABLE 1

$$A = \begin{matrix} 50 & -60 & 50 & -27 & 6 & 6 \\ 38 & -28 & 27 & -17 & 5 & 5 \\ 27 & -17 & 27 & -17 & 5 & 5 \\ 27 & -28 & 38 & -17 & 5 & 5 \\ 27 & -28 & 27 & -17 & 16 & 5 \\ 27 & -28 & 27 & -17 & 5 & 16 \end{matrix} \qquad B = \begin{matrix} 16 & 5 & 5 & 5 & -6 & 5 \\ 5 & 16 & 5 & 5 & -6 & 5 \\ 5 & 5 & 16 & 5 & -6 & 5 \\ 5 & 5 & 5 & 16 & -6 & 5 \\ 5 & 5 & 5 & 5 & -6 & 16 \\ 6 & 6 & 6 & 6 & -5 & 6 \end{matrix}$$

α	β	α/β
25.768670843143	$.2637605112.10^{-6}$	$0.976972281.10^{8}$
-12.821841071323	$.1312405807.10^{-6}$	$-0.976972290.10^{8}$
$5.814535434181 + 10.071071345641\,i$	11.629071028730	$0.49999999310489 + 0.86602543924271\,i$
$5.800765071150 - 10.047220375909\,i$	11.601530302268	$0.49999999310489 - 0.86602543924271\,i$
$5.736511506410 + 9.935928843473\,i$	11.473022854605	$0.50000000689511 + 0.86602536832617\,i$
$5.510879468089 - 9.545122710676\,i$	11.021758784186	$0.50000000689511 - 0.86602536832617\,i$

The double root at ∞ results from the fact that B has a double zero eigenvalue. All three roots are associated with quadratic elementary divisors; i.e., each root has only one corresponding eigenvector. The computed diagonals of the triangularized matrices are given in the table. Note that the four finite eigenvalues are obtained with a relative accuracy of about 10^{-8}. This is about the square root of the machine precision and is the expected behavior for eigenvalues with quadratic elementary divisors. The singularity of B does not cause any further deterioration in their accuracy. Furthermore, the infinite eigenvalues are obtained from the reciprocals of quantities which are roughly the square root of the machine precision times the norm of B. Consequently we are somewhat justified if we claim to have computed the square root of infinity.[1]

Acknowledgments. W. Kahan and J. H. Wilkinson made several helpful comments. Linda Kaufman of Stanford has recently shown how to carry out the corresponding generalization of the LR algorithm and has written a program that accepts general complex matrices. Charles Van Loan of the University of Michigan has done a detailed roundoff error analysis of the process described in section 5.

[1] This prompts us to recall the limerick which introduces George Gamow's *One, Two, Three, Infinity*:

There was a young fellow from Trinity
Who tried $\sqrt{\infty}$
But the number of digits
Gave him such fidgets
That he gave up Math for Divinity.

REFERENCES

[1] CHARLES CRAWFORD, *The numerical solution of the generalized eigenvalue problem*, Doctoral thesis, University of Michigan, Ann Arbor, 1970; Comm. ACM., 16 (1973), pp. 41–44.

[2] G. FIX AND R. HEIBERGER, *An algorithm for the ill-conditioned generalized eigenvalue problem*, Numer. Math., this Journal, 9 (1972), pp. 78–88.

[3] J. G. F. FRANCIS, *The QR transformation—a unitary analogue to the LR transformation*, Comput. J., 4 (1961–62), pp. 265–271, 332–345.

[4] A. S. HOUSEHOLDER, *Unitary triangularization of a nonsymmetric matrix*, J. Assoc. Comput. Mach., 5 (1958), pp. 339–342.

[5] R. S. MARTIN AND J. H. WILKINSON, *Similarity reduction of a general matrix to Hessenberg form*, Numer. Math., 12 (1968), pp. 349–368.

[6] ———, *Reduction of the symmetric eigenproblem* $Ax = \lambda Bx$ *and related problems to standard form*, Ibid., 11 (1968), pp. 99–110.

[7] C. B. MOLER AND G. W. STEWART, *The QZ algorithm for* $Ax = \lambda Bx$, this Journal, 9 (1972), pp. 669–686.

[8] G. PETERS AND J. H. WILKINSON, *Eigenvectors of real and complex matrices by LR and QR triangularizations*, Numer. Math., 16 (1970), pp. 181–204.

[9] ———, $Ax = \lambda Bx$ *and the generalized eigenproblem*, this Journal, 7 (1970), pp. 479–492.

[10] G. W. STEWART, *On the sensitivity of the eigenvalue problem* $Ax = \lambda Bx$, submitted for publication.

[11] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.

[12] FORTRAN translations of the ALGOL procedures in [5] and [8] may be obtained from the NATS Project, Applied Mathematics Division, Argonne National Laboratories, Argonne, Illinois.

[13] A. M. DELL, R. L. WEIL AND G. L. THOMPSON, *Roots of Matrix Pencils*, Comm. ACM., 14 (1971), pp. 113–117.