SIAM REVIEW
Vol. 24, No. 4, October 1982

0036-1445/82/2404-0004 $01.00/0

UNDERSTANDING THE QR ALGORITHM*

DAVID S. WATKINS†

Abstract. The QR algorithm is currently the most popular method for finding all eigenvalues of a full matrix. While QR is now well understood by specialists in eigenvalue computations, this understanding is not being conveyed effectively to the mathematical public. Many accounts present Wilkinson's 1965 convergence proof. Others establish some of the connections between the QR algorithm, the power method and inverse iteration. Usually much emphasis is (rightly) placed on the refinements, such as shifts of origin, which are required to make the algorithm competitive. But practically all accounts fail to explain the basic meaning of QR iterations. As a consequence, the QR algorithm is widely thought to be difficult to understand. The purpose of this paper is to try to convince the reader that the opposite is true. In fact, the QR algorithm is neither more nor less than a clever implementation of simultaneous iteration, which is itself a natural, easily understood extension of the power method. This point of view deserves pre-eminence because it shows exactly what QR iterations are and evokes a clear geometric picture of the QR process. Furthermore, it provides a framework within which the rapid convergence associated with shifts of origin may be explained. No reference to inverse iteration is necessary. Inverse iteration has not, however, been banished from the paper—one section is devoted to an explanation of the interplay between inverse iteration, direct iteration and the QR algorithm. The key result of that section is a duality theorem which shows that whenever direct iteration takes place, inverse iteration automatically takes place at the same time.

1. Introduction. The QR algorithm is currently the most popular method for calculating the complete set of eigenvalues of a full (i.e., small) matrix. A descendant of Rutishauser's (1955), (1958) LR algorithm, it was discovered independently by Francis (1961), (1962) and Kublanovskaya (1961). The basic algorithm is as follows: Given a matrix A whose eigenvalues are desired, let $A_0 = A$. Then, given A_{m-1}, find unitary Q_m and upper triangular R_m such that $A_{m-1} = Q_m R_m$. Finally, define $A_m = R_m Q_m$. Thus

$$A_{m-1} = Q_m R_m, \qquad A_m = R_m Q_m.$$

One's first reaction on seeing this procedure is likely to be, "What does this have to do with eigenvalues?" or "What do these manipulations accomplish?" Most accounts answer these questions by presenting Wilkinson's (1965, p. 517) proof that, under suitable conditions, the sequence of (unitarily similar) matrices A_m converges to upper triangular form. That proof has its merits. For one, it is relatively brief and elementary. Also, it was the best available in the sixties. Unfortunately, it does not show what goes on in a QR iteration—the reader is shown that the method works, but is left wondering why. This author believes that the best way to explain what QR iterations are is to first introduce and discuss simultaneous iteration, an easily understood, multivector generalization of the power method, then show that the QR algorithm is just a clever way to do simultaneous iteration.

The connection between QR and simultaneous iteration has long been known. In fact, even before QR had come into being, Bauer (1958) showed that the LR algorithm is equivalent to a form of simultaneous iteration. The books of Faddeev and Faddeeva (1963) (LR case only), Householder (1964) and Wilkinson (1965, p. 608) all noted the connection, but no one seems to have appreciated at that time the appealing geometric picture of the QR algorithm which it evokes. The early convergence proofs made heavy use of determinants and were opaque and unwieldy. For an example see Wilkinson (1965, pp. 489–492). Wilkinson's (1965, p. 517) proof was a vast improvement, but still it did not

*Received by the editors August 4, 1981, and in revised form March 30, 1982. This work was supported in part by the National Science Foundation under grant MCS-8102382.

†Department of Pure and Applied Mathematics, Washington State University, Pullman, Washington 99164-2930.

reveal the meaning of QR iterations. Evidently the geometric point of view was first appreciated by Buurema (1970) and Parlett and Poole (1973). A decade has passed since those works appeared, yet the approach which they advocated is still not as widely understood as it deserves to be. This paper attempts to rectify that.

Among those specialists who understand the relationship between QR and simultaneous iteration there seems to be some reluctance to emphasize it. There is much more interest in the connection between QR and inverse iteration (i.e. the inverse power method), since this connection may be used to explain the rapid convergence of the *shifted* QR algorithm. The general feeling seems to be that this connection, rather than that with simultaneous iteration, should be regarded as primary, since the QR algorithm would be of no practical value if it did not converge swiftly. In response to that attitude we have adopted an approach in which the rapid convergence of the shifted QR algorithm is explained entirely within the context of simultaneous iteration, with no reference whatsoever to inverse iteration.

Another reason simultaneous iteration is not often connected with the QR algorithm is that, in its explicit form, simultaneous iteration is usually used only to find the few largest eigenvalues of large sparse matrices (c.f. Rutishauser (1969), (1970), Clint and Jennings (1970), (1971), Parlett (1980)), whereas QR is used mainly for small, full matrices. Thus, it appears that the classes of problems to which the two methods may be applied are nearly disjoint. This makes it easy to forget that QR is actually a form of simultaneous iteration.

The section contents are as follows. Section 2 introduces the power method and simultaneous iteration. Section 3 covers the QR algorithm. Inverse iteration has not been banished completely from the paper. An account of the beautiful relationship between it and QR is given in §4. Section 5 briefly discusses the related LR or LU and Cholesky algorithms and their connection with both simultaneous iteration and the QR algorithm.

The various implementations of QR—explicit and implicit QR, doubly shifted QR, rational QR, etc.—have not been covered. These have been documented in standard sources such as Wilkinson (1965), the *Handbook* of Wilkinson and Reinsch (1971), and the *EISPACK Guide* of Smith et al. (1976). Good implementations have long been available, much more widely available, in fact, than good explanations.

Notation. C^n denotes the space of n-tuples of complex numbers, and $\|\cdot\|_2$ denotes the Euclidean norm on C^n. Given a set of vectors $q_1, q_2, \cdots, q_k \in C^n$, $\langle q_1, q_2, \cdots, q_k\rangle$ will denote the subspace of C^n spanned by $q_1, q_2, \cdots, q_k$. Given a complex matrix M, the conjugate transpose of M will be denoted by M^*. Our object of study throughout this paper is a complex $n \times n$ matrix A with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. The eigenvalues will always be ordered so that $|\lambda_1| \geqq |\lambda_2| \geqq \cdots \geqq |\lambda_n|$. A may be real, in which case most of the algorithms discussed here may be carried out entirely in real arithmetic.

2. Direct iteration.

Basic power method. The basic (direct) power method consists of choosing a vector v and applying A to it repeatedly to form the sequence

$$v, Av, A^2v, A^3v, \cdots.$$

In practice one must rescale the vector at each step in order to avoid an eventual overflow or underflow, and to be able to judge whether the sequence is converging. Assuming a reasonable scaling strategy, the sequence of iterates will usually converge to an eigenvector of A. It is not hard to see why. Suppose A has eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ with $|\lambda_1| > |\lambda_2| \geqq \cdots \geqq |\lambda_n|$. We will assume for ease of exposition that A is simple; that is, A

has n linearly independent eigenvectors $v_1, v_2, \cdots, v_n$. This assumption is not critical, whereas the assumption $|\lambda_1| > |\lambda_2|$ is. We will order $v_1, v_2, \cdots, v_n$ so that v_i corresponds to λ_i. The starting vector v may be expressed uniquely as a linear combination of $v_1, v_2, \cdots, v_n$,

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

Applying A repeatedly we get

$$A^m v = c_1 \lambda_1^m v_1 + c_2 \lambda_2^m v_2 + \cdots + c_n \lambda_n^m v_n, \qquad m = 1, 2, 3, \cdots.$$

Since λ_1 dominates the other eigenvalues, the component in the direction of v_1 becomes relatively greater than the other components as m increases. If λ_1 were known in advance, one could rescale at each step by dividing by it to get

$$A^m v/(\lambda_1)^m = c_1 v_1 + c_2(\lambda_2/\lambda_1)^m v_2 + \cdots + c_n(\lambda_n/\lambda_1)^m,$$

which clearly converges to the eigenvector $c_1 v_1$, provided that c_1 is nonzero. Convergence is linear, with ratio of successive errors approximately $|\lambda_2/\lambda_1|$. This scaling strategy is unavailable in real problems, but the exact choice of scaling strategy is unimportant. The eigenvector is determined only up to a constant multiple: the direction is important, not the length.

The condition $c_1 \neq 0$ is equivalent to the condition $v \notin \langle v_2, \cdots, v_n \rangle$, where $\langle v_2, \cdots, v_n \rangle$ denotes the subspace spanned by $v_2, \cdots, v_n$. Any proper subspace is a very small subset (of measure zero, nowhere dense) of C^n. Therefore it is highly probable that a v chosen at random will not lie in $\langle v_2, \cdots, v_n \rangle$.

Subspace iteration. The eigenvector v_1 is merely a representative of the eigenspace $\langle v_1 \rangle$, which is the real object of interest. Likewise, in the sequence $v, Av, A^2v, A^3v, \cdots$, each of the iterates $A^m v$ may be viewed as a representative of the space $\langle A^m v \rangle$ which it spans. Rescaling a vector amounts to replacing one representative by a new representative of the same one-dimensional space. Thus the power method may be viewed as a process of iteration on subspaces: First a one-dimensional starting space $S = \langle v \rangle$ is chosen. Then iterates

$$S, AS, A^2S, A^3S, \cdots \tag{2.1}$$

are formed. This sequence converges linearly to the eigenspace $T = \langle v_1 \rangle$ in the sense that the angle between $A^m S$ and T converges to zero.

More generally one can choose a subspace S of any dimension, k, and form the sequence (2.1). It is not surprising that this sequence will generally converge to the invariant subspace spanned by the k leading eigenvectors. We will continue to assume that A is simple, with eigenvector basis $v_1, v_2, \cdots, v_n$. Let

$$T = \langle v_1, \cdots, v_k \rangle, \qquad U = \langle v_{k+1}, \cdots, v_n \rangle,$$

and assume that $|\lambda_k| > |\lambda_{k+1}|$. Both T and U are invariant under A, and are called *dominant* and *co-dominant* spaces, respectively. We shall see that the sequence (2.1) almost always converges to T.

In order to discuss convergence of subspaces we define a metric on the set of k-dimensional subspaces of C^n. A reasonable definition is

$$d(S, T) = \sup_{\substack{s \in S \\ \|s\|_2 = 1}} \inf_{t \in T} \|s - t\|_2$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The main result on convergence of subspace iteration is

THEOREM 2.1. *Let T and U be the dominant and co-dominant spaces defined above, and let S be a k-dimensional subspace of C^n such that $S \cap U = (0)$. Then there exists a constant C such that*

$$d(A^m S, T) \leqq C|\lambda_{k+1}/\lambda_k|^m$$

for all m. Thus $A^m S \rightarrow T$ linearly with ratio $|\lambda_{k+1}/\lambda_k|$.

We have opted to phrase Theorem 2.1 in terms of a metric because it is the easiest course—the metric can be defined in one line. A more natural notion is that of angle. The relative orientation of two k-dimensional subspaces is described by k *canonical angles*. The metric $d(S, T)$ is just the sine of the largest canonical angle between S and T. For more on angles see Björk and Golub (1973), Davis and Kahan (1970) and earlier references cited therein, and Stewart (1973b), (1977).

It is easy to argue the plausibility of Theorem 2.1. Let v be any nonzero vector in S. We will show that the iterates $A^m v$ lie (relatively) closer and closer to T as m increases. v may be expressed uniquely in the form

$$\begin{aligned} v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k & \qquad \text{(component in } T) \\ + c_{k+1} v_{k+1} + \cdots + c_n v_n & \qquad \text{(component in } U) \end{aligned}$$

in which v has been expressed as a sum of a component in T and a component in U. Since $v \notin U$, at least one of the coefficients $c_1, \cdots, c_k$ must be nonzero. Now

$$\begin{aligned} A^m v/(\lambda_k)^m = {} & c_1(\lambda_1/\lambda_k)^m v_1 + \cdots + c_{k-1}(\lambda_{k-1}/\lambda_k)^m v_{k-1} + c_k v_k \\ & + c_{k+1}(\lambda_{k+1}/\lambda_k)^m v_{k+1} + \cdots + c_n(\lambda_n/\lambda_k)^m v_n. \end{aligned}$$

Note that the nonzero coefficients of the component in T increase, or at least do not decrease, as m increases. At the same time the coefficients of the component in U tend to zero linearly with rate $|\lambda_{k+1}/\lambda_k|$ or better. Thus each sequence $(A^m v)$ converges to T at the stated rate, and consequently the limit of $(A^m S)$ lies in T. The limit cannot be a proper subspace of T because it has dimension k.

For a proof of Theorem 2.1 see Parlett and Poole (1973). Their treatment of the subject completely dispenses with the notions of metric and angle on the grounds that in the finite-dimensional setting any two reasonable notions of convergence are equivalent. The theorem still holds if A is not simple, except that the constant C must be replaced by a polynomial $C(m)$ if the eigenvalue λ_{k+1} is deficient. Parlett and Poole have also covered the case $\lambda_k = \lambda_{k+1}$. In that case convergence is too slow to be of any practical value.

The assumption $S \cap U = (0)$ corresponds to the assumption $c_1 \neq 0$ in the basic power method. It is important to realize that this assumption will be satisfied by virtually any subspaces S and U whose dimensions sum to n. This is most easily seen by analogy with the situation in R^3. There any two two-dimensional subspaces must intersect nontrivially because the sum of their dimensions exceeds three. By contrast, a two-dimensional subspace is not required to intersect nontrivially with a one-dimensional subspace, and it is obvious that it almost certainly will not. By the same token, since the sum of the dimensions of S and U does not exceed n, they are not required to intersect nontrivially, and they almost certainly will not.

Invariant subspaces are of interest to eigenvalue hunters partly because they allow one to *reduce* the problem. Indeed, let $Q = [Q_1 Q_2]$ be a unitary matrix whose first k

columns (Q_1) form an orthonormal basis for the invariant subspace T. Then

(2.2) $$Q^*AQ = \begin{bmatrix} Q_1^*AQ_1 & Q_1^*AQ_2 \\ Q_2^*AQ_1 & Q_2^*AQ_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where $Q_2^*AQ_1 = 0$ because T is invariant. Thus the eigenvalue problem for A has been divided into two smaller eigenvalue problems for A_{11} and A_{22}. The eigenvalues of A_{11} are exactly those of $A|_T$. In the case $k = 1$, A_{11} is 1×1, and its single entry is the eigenvalue λ_1.

In practice one never exactly attains an invariant subspace. Instead one has a subspace A^mS such that $d(A^mS, T)$ is small. Let $P = [P_1\ P_2]$ be a unitary matrix whose first k columns (P_1) span A^mS, and let

$$P^*AP = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

One would expect that, as $A^mS \rightarrow T$, B_{21} must converge to zero at the same rate. Indeed, it is not hard to show that this is so. The converse is true as well but harder to prove: if $B_{21} \rightarrow 0$, the linear span of the columns of P_1 approaches an invariant subspace of A at the same rate. See Stewart (1971), (1973b). In what follows it will become clear that these results are crucial to the convergence theory of the QR algorithm.

Simultaneous iteration is a practical means of carrying out subspace iteration. Since iteration on an entire subspace cannot be done in practice, one must instead choose a basis for S and iterate on the basis vectors simultaneously. Let S be a k-dimensional subspace of C^n such that $S \cap U = (0)$. Then S contains no null vectors of A, since all null vectors lie in U. From the discussion of Theorem 2.1 it is evident that $A^mS \cap U = (0)$ for all m, and therefore A^mS contains no null vectors. Let $q_1^0, \cdots, q_k^0$ be a basis of S. Then $A(q_1^0), \cdots, A(q_k^0)$ span AS. They are linearly independent as well, because S contains no null vectors, and they therefore form a basis for S. Likewise $A^m(q_1^0), \cdots, A^m(q_k^0)$ form a basis for A^mS, $m = 2, 3, 4 \cdots$. Thus, in theory at least, one can simply iterate on a basis of S to get bases for AS, A^2S, A^3S, $\cdots$. There are two reasons why it is not advisable to do this in practice: 1. The vectors will have to be rescaled in order to avoid overflow or underflow. 2. Each of the sequences $q_i^0, A(q_i^0), A^2(q_i^0), \cdots$ independently converges to the dominant subspace $\langle v_1 \rangle$. It follows that for m large the vectors $A^m(q_1^0), \cdots, A^m(q_k^0)$ all point in nearly the same direction. That is, the basis is ill-conditioned. Ill-conditioned bases can be avoided by replacing the basis gotten at each step by a well-conditioned basis for the same subspace. Probably the most effective way to do this is to orthonormalize. Thus, the following simultaneous iteration procedure is recommended: 1) Given $q_1^m, \cdots, q_k^m$, an orthonormal basis of A^mS, calculate $A(q_1^m), \cdots, A(q_k^m)$. 2) Orthonormalize $A(q_1^m), \cdots, A(q_k^m)$ from left to right to get $q_1^{m+1}, \cdots, q_k^{m+1}$, an orthonormal basis of $A^{m+1}S$.

Simultaneous iteration has the agreeable property of iterating on lower-dimensional subspaces at no extra cost. Let $S_i = \langle q_1^0, \cdots, q_i^0 \rangle$, $i = 1, \cdots, k$. Then $AS_i = \langle A(q_1^0), \cdots, A(q_i^0) \rangle = \langle q_1^1, \cdots, q_i^1 \rangle$ for all i, since the orthonormalization procedure preserves these subspaces. In general $A^mS_i = \langle q_1^m, \cdots, q_i^m \rangle$, $i = 1, \cdots, k$. Thus simultaneous iteration seeks not only an invariant subspace of dimension k, but also subspaces of dimensions $1, 2, \cdots, k - 1$ as well.

Now consider what happens when simultaneous iteration is applied to a complete set of orthonormal vectors $q_1^0, q_2^0, \cdots, q_n^0$. For $k = 1, 2, \cdots, n - 1$ let $S_k = \langle q_1^0, \cdots, q_k^0 \rangle$, $T_k = \langle v_1, \cdots, v_k \rangle$, $U_k = \langle v_{k+1}, \cdots, v_n \rangle$, and assume $S_k \cap U_k = (0)$ and $|\lambda_k| > |\lambda_{k+1}|$. Then $A^mS_k \rightarrow T_k$ linearly as $m \rightarrow \infty$. In terms of bases this means that $q_1^m, \cdots, q_n^m$ will converge (modulo factors of unit modulus) to an orthonormal basis $q_1, q_2, \cdots, q_n$ such that for each k, the first k vectors span the invariant subspace T_k. If we let $\hat{Q}_m$ denote the unitary matrix whose columns are $q_1^m, \cdots, q_n^m$, then the sequence of matrices $A_m = \hat{Q}_m^* A \hat{Q}_m$ converges to the block triangular form (2.2). But this holds for all k simultaneously, so the limiting form is upper triangular. The limiting main diagonal entries are $\lambda_1, \lambda_2, \cdots, \lambda_n$, in order. If some of the eigenvalues are equal in modulus, say $|\lambda_{i+1}| = |\lambda_{i+2}| = \cdots = |\lambda_{i+j}|$, then the limit will be block triangular with a $j \times j$ block in rows and columns $i+1$ through $i+j$. Of course the eigenvalues of the block are $\lambda_{i+1}, \cdots, \lambda_{i+j}$.

The conditions $S_k \cap U_k = (0)$ will be satisfied by practically any starting basis. If any of these conditions should be violated, the eigenvalues will still emerge on the main diagonal, but not in descending order (cf. Wilkinson (1965)). But this possibility is so remote that it is hardly worth thinking about, especially considering that a special relationship of the form $S_k \cap U_k \neq (0)$ will probably be destroyed by roundoff error in an actual computation. Furthermore, as we shall see, the subspace conditions are always satisfied by the QR algorithm on unreduced Hessenberg matrices.

The rate of convergence to T_k is $|\lambda_{k+1}/\lambda_k|$, which will often be intolerably slow. To see how convergence might be accelerated, suppose we are able to find a number σ which is a very good approximation to λ_n. If we replace A by the *shifted* matrix $A - \sigma I$, the convergence rates will change to $|(\lambda_{k+1} - \sigma)/(\lambda_k - \sigma)|$, $k = 1, \cdots, n-1$. If σ is close enough to λ_n that $|\lambda_n - \sigma| \ll |\lambda_{n-1} - \sigma|$, convergence to the subspace T_{n-1} will be extremely fast. (If σ happens to exactly equal λ_n, convergence will take place in one iteration.) Better yet, suppose we are able to find a sequence of shifts σ_m such that $\sigma_m \to \lambda_n$, and on the mth step we replace A by $A - \sigma_m I$. Then convergence to T_{n-1} will be better than linear. The outward evidence of convergence to T_{n-1} is the convergence (up to a factor of unit modulus) of the last vector q_n, which is then orthogonal to T_{n-1}. Once this vector has converged, it can be dropped from the iterations, which can then be continued with A restricted to the subspace T_{n-1}. The shifts can then be chosen to approximate λ_{n-1}, the smallest eigenvalue of the restricted operator, causing rapid convergence to T_{n-2}. Continuing in this manner we can determine the eigenvalues in rapid succession.

We caution the reader that shifted simultaneous iteration, in the form just described, is numerically unstable because of the mode of deflation (deflation by restriction). A stable implementation is the QR algorithm, which employs deflation by similarity transformations. We shall see that the QR algorithm provides convenient sequences of shifts which converge to the eigenvalues quadratically.

Early references to simultaneous iteration are Bauer (1957), (1958) and Wilkinson (1965). The explicit formulations of simultaneous iteration alluded to in the introduction are actually much more sophisticated than the simple algorithm discussed in this section. All employ some form of Rayleigh–Ritz procedure to accelerate convergence to the eigenvalues. The reader is referred to Jennings (1967), Clint and Jennings (1970), (1971), Rutishauser (1969), (1970), Stewart (1969), (1976), and Parlett (1980).

3. The *QR* algorithm. The QR algorithm is based on the QR decomposition.

THEOREM 3.1. *Let A be a complex $n \times n$ matrix. Then there exist a unitary matrix Q and an upper triangular matrix R such that $A = QR$. If A is nonsingular, then R may be chosen so that all of its main diagonal entries are positive. In that case Q and R are uniquely determined.*

For a proof see, for example, Stewart (1973a). Not only do Q and R exist, but they can be constructed by a stable algorithm at a cost of about $2n^3/3$ multiplications. The QR decomposition is just a matrix realization of the Gram–Schmidt orthonormalization process. Indeed, suppose A is nonsingular, let $a_1, a_2, \cdots, a_n$ denote the columns of A, and let $q_1, q_2, \cdots, q_n$ denote the columns of Q. Then $a_1 = q_1 r_{11}$, $a_2 = q_1 r_{12} + q_2 r_{22}$, and in general

$$a_k = q_1 r_{1k} + q_2 r_{2k} + \cdots + q_k r_{kk}, \qquad r_{kk} > 0, \qquad k = 1, 2, \cdots, n.$$

It follows that $\langle a_1 \rangle = \langle q_1 \rangle$, $\langle a_1, a_2 \rangle = \langle q_1, q_2 \rangle$, and in general

$$\langle a_1, a_2, \cdots, a_k \rangle = \langle q_1, q_2, \cdots, q_k \rangle, \qquad k = 1, 2, \cdots, n.$$

That is, the columns of Q orthonormalize the columns of A.

In what follows we will assume that A is nonsingular, and thereby guarantee the uniqueness of all QR decompositions. The reader should not infer that the singular case is pathological. On the contrary, if A is singular, the zero eigenvalue will be disposed of in one iteration. This was already suggested by our discussion of simultaneous iteration. We will say more on this topic later in connection with Hessenberg matrices.

With the aid of the QR decomposition, we may express simultaneous iteration in matrix form as follows: Let $\hat{Q}_m$ be the matrix whose columns are $q_1^m, q_2^m, \cdots, q_n^m$, as in the previous section. If we let $D_{m+1} = A\hat{Q}_m$, then the columns of D_{m+1} are $Aq_1^m, Aq_2^m, \cdots, Aq_n^m$. These may be orthonormalized by a QR decomposition $D_{m+1} = \hat{Q}_{m+1} R_{m+1}$. To summarize,

$$D_{m+1} = A\hat{Q}_m, \qquad D_{m+1} = \hat{Q}_{m+1} R_{m+1}. \tag{3.1}$$

One way to check for convergence after m steps is to perform the similarity transformation

$$A_m = \hat{Q}_m^* A \hat{Q}_m \tag{3.2}$$

and check whether A_m is nearly upper triangular.

Suppose we start iterating with $\hat{Q}_0 = I$. That is, we start with the basis $e_1, e_2, \cdots, e_n$ of standard unit vectors. Then $D_1 = A$ and $A = D_1 = \hat{Q}_1 R_1$. Letting $Q_1 = \hat{Q}_1$ we have

$$A = Q_1 R_1. \tag{3.3}$$

If after one step we already wish to begin to monitor our convergence, we may do so by examining $A_1 = Q_1^* A Q_1$. Since $Q_1^* A = R_1$ by (3.3), A_1 may be gotten by

$$A_1 = R_1 Q_1. \tag{3.4}$$

Finding that A_1 is not upper triangular, we take another step. But now we have two matrices, A and A_1, which may be viewed as realizations of the same linear operator in two different coordinate systems. We can continue to operate on A, calculating $D_2 = A\hat{Q}_1$ and $D_2 = \hat{Q}_2 R_2$, or we can perform the equivalent operations on A_1. A vector which is represented by v in the A coordinate system is represented by $\hat{Q}_1^* v$ in the A_1 system. Therefore the vectors $q_1^1, \cdots, q_n^1$ in the A system become $e_1, \cdots, e_n$ in the A_1 system. Thus the equation $D_2 = A\hat{Q}_1$ is equivalent to $A_1 = A_1 I$, and the QR decomposition

$$D_2 = \hat{Q}_2 R_2$$

is equivalent to a QR decomposition of A_1:

$$A_1 = Q_2 R_2.$$

We have used the same symbol R_2 in both QR decompositions because it is the same matrix in both cases. This can be seen by noting that the equations $D_2 = AQ_1 = Q_1 A_1$ and $A_1 = Q_2 R_2$ may be combined to yield a second QR decomposition of D_2: $D_2 = (Q_1 Q_2) R_2$. From the uniqueness of the QR decomposition we find that R_2 is the same in both, and furthermore $\hat{\bar{Q}}_2 = Q_1 Q_2$. If we opt to operate with A_1 instead of A, we can check for convergence by calculating $A_2 = Q_2^* A_1 Q_2 = R_2 Q_2$. The equation $\hat{Q}_2 = Q_1 Q_2$ guarantees that this A_2 is the same as the one given by (3.2). We can continue this process to produce a sequence of matrices A_m, where

$$A_{m-1} = Q_m R_m, \qquad A_m = R_m Q_m. \tag{3.5}$$

This is the QR algorithm, and as we have just seen, it is equivalent to simultaneous iteration. The A_m produced by (3.5) are the same as those given by (3.2). The R_m of (3.5) are the same as those of (3.1), and the Q_m of (3.5) are related to the $\hat{Q}_m$ of (3.1) by

$$\hat{Q}_m = Q_1 Q_2 \cdots Q_m. \tag{3.6}$$

Q_m is the coordinate change at the mth step, whereas $\hat{Q}_m$ is the accumulated change of coordinates after m steps.

We have established the equivalence of simultaneous iteration and the QR algorithm by looking at the process one step at a time. Another way is to examine the cumulative effect of m steps. In this approach, Q_m, R_m and A_m are defined by (3.5), with $A_0 = A$, and $\hat{Q}_m$ is defined by (3.6). If, in addition, $\hat{R}_m$ is defined by $\hat{R}_m = R_m R_{m-1} \cdots R_1$, then

$$A = Q_1 R_1 = \hat{Q}_1 \hat{R}_1,$$
$$A^2 = Q_1 R_1 Q_1 R_1 = Q_1 Q_2 R_2 R_1 = \hat{Q}_2 \hat{R}_2,$$
$$A^3 = Q_1 R_1 Q_1 R_1 Q_1 R_1 = Q_1 Q_2 R_2 Q_2 R_2 R_1 = Q_1 Q_2 Q_3 R_3 R_2 R_1 = \hat{Q}_3 \hat{R}_3.$$

Clearly one could show by induction that

$$A^m = \hat{Q}_m \hat{R}_m, \qquad m = 1, 2, \cdots. \tag{3.7}$$

This equation has appeared repeatedly in the literature, and it has long been known to be central to the analysis of the QR algorithm. Unfortunately, in spite of its frequent appearance, its meaning is almost never explained. Equation (3.7) shows that $\hat{Q}_m$ and $\hat{R}_m$ are the QR factors of A^m. Recalling that the QR decomposition is an orthonormalization process, we conclude that, for all k, the first k columns of $\hat{Q}_m$ form an orthonormal basis for the space spanned by the first k columns of A^m. But what are the columns of A^m? The ith column of A^m is just $A^m e_i$. Thus

$$\langle A^m e_1, \cdots, A^m e_k \rangle = \langle q_1^m, \cdots, q_k^m \rangle, \qquad k = 1, \cdots, n, \quad m = 1, 2, \cdots.$$

That is, the columns of $\hat{Q}_m$ are just the result of m steps of simultaneous iteration, starting from the standard basis vectors $e_1, e_2, \cdots, e_n$.

Having established that QR is just simultaneous iteration starting with $e_1, e_2, \cdots, e_n$, we can conclude that the sequence A_m produced by QR converges to triangular (or at least block triangular) form, provided that the subspace conditions

$$\langle e_1, e_2, \cdots, e_k \rangle \cap \langle v_{k+1}, \cdots, v_n \rangle = (0), \qquad k = 1, \cdots, n-1, \tag{3.8}$$

are satisfied. The reader can easily verify that (3.8) is equivalent to the condition which is usually given—namely, that all leading principal minors of V^{-1} should be nonzero, where V is the matrix whose columns are the eigenvectors $v_1, \cdots, v_n$. It is the author's opinion that the geometric condition (3.8) is more illuminating than the equivalent condition on the minors of V^{-1}.

Refinements. The basic QR algorithm is too inefficient to be an effective tool, but two refinements suffice to make it competitive. 1) A preliminary reduction to Hessenberg form radically decreases the cost of each QR step. 2) The use of shifts of origin drastically reduces the total number of QR steps required to attain convergence.

Hessenberg form. A square matrix B is said to be in *upper Hessenberg form* if $b_{ij} = 0$ whenever $i > j + 1$. This means that B is nearly upper triangular, having all zeros in the lower triangle, except on the subdiagonal. Given any $n \times n$ matrix A, there exists an upper Hessenberg matrix B which is unitarily similar to A, which may be constructed from A at a cost of some $5n^3/3$ multiplications. (See e.g. Stewart (1973a).) If A is Hermitian, then B is tridiagonal and may be constructed in about $2n^3/3$ multiplications. (See Parlett (1980).) Hessenberg form is important to the QR algorithm because 1. it is preserved under QR iterations, and 2. the cost of a QR iteration for a Hessenberg matrix is $O(n^2)$ multiplications instead of $O(n^3)$. (To see that Hessenberg form is preserved, deduce from (3.5) that $A_m = R_m A_{m-1} R_m^{-1}$, and note that upper Hessenberg form is preserved upon pre- or postmultiplication by an upper triangular matrix.) For Hermitian matrices tridiagonal form is preserved, and the cost of a QR iteration is $O(n)$ multiplications.

An upper Hessenberg matrix B is in *unreduced* upper Hessenberg form if all of its subdiagonal entries are nonzero. If B is not unreduced, its eigenvalue problem may immediately be reduced to smaller eigenvalue problems involving unreduced upper Hessenberg matrices. Parlett (1968) has shown that for unreduced upper Hessenberg matrices the subspace relations (3.8) are always satisfied. The argument runs as follows: Given a nonzero $v \in \langle e_1, \cdots, e_k \rangle$, the special form of B guarantees that $v, Bv, B^2v, \cdots, B^{n-k}v$ are linearly independent. Thus the smallest invariant subspace containing v has dimension at least $n - k + 1$. Therefore v cannot lie in the $(n - k)$-dimensional invariant subspace $\langle v_{k+1}, \cdots, v_n \rangle$. It follows that the unshifted QR algorithm, started with an unreduced upper Hessenberg matrix, will always converge.

The same property guarantees that the QR algorithm will deal satisfactorily with singular matrices. Suppose B is a singular, unreduced, upper Hessenberg matrix, and consider a single QR step. Since the first $n - 1$ columns of B are linearly independent, the same must be true of R in the decomposition $B = QR$. Therefore $r_{kk} \neq 0$, $k = 1, \cdots, n - 1$. But the singularity of B requires that at least one r_{kk} be zero. We conclude that $r_{nn} = 0$. Therefore the iterate $B_1 = RQ$ has all zeros in the last row. That is, a zero eigenvalue has emerged. Future iterations may be applied to the deflated matrix gotten by deleting the last row and column of B_1.

Hessenberg form makes testing for convergence easy. The subdiagonal block of dimension $(n - k) \times k$ has only one nonzero entry, which, in A_m, we will denote by $a^{(m)}_{k+1,k}$. This one number gives an indication of the distance of $\langle q_1^m, \cdots, q_k^m \rangle$ from $\langle v_1, \cdots, v_k \rangle$. If $|\lambda_k| > |\lambda_{k+1}|$, then $a^{(m)}_{k+1,k} \to 0$ linearly, with ratio $|\lambda_{k+1}/\lambda_k|$, as $m \to \infty$. Once $a^{(m)}_{k+1,k}$ is negligible, we can consider that the invariant subspace has been attained and reduce the problem.

Shifts of origin. As convergence approaches, the entry $a^{(m)}_{nn}$ will approach the eigenvalue λ_n, assuming $|\lambda_{n-1}| > |\lambda_n|$. Convergence is linear with ratio $|\lambda_n/\lambda_{n-1}|$. If $a^{(m)}_{nn}$ is sufficiently close to λ_n, the ratio $|(\lambda_n - a^{(m)}_{nn})/(\lambda_{n-1} - a^{(m)}_{nn})|$ will be substantially smaller than $|\lambda_n/\lambda_{n-1}|$. This suggests that we replace A_m by the shifted matrix $A_m - \sigma_m I$, where $\sigma_m = a^{(m)}_{nn}$, to arrive at the shifted QR algorithm:

$$A_{m-1} - \sigma_{m-1} I = Q_m R_m, \qquad A_m = R_m Q_m + \sigma_{m-1} I, \qquad m = 0, 1, 2, \cdots.$$

(In practice the shift may be restored after each iteration, as indicated here, or accumulated.) It is clear that convergence to λ_n will be better than linear because the convergence ratio $|(\lambda_n - \sigma_m)/(\lambda_{n-1} - \sigma_m)|$ tends to zero as $\sigma_m \to \lambda_n$. The positive feedback between the improving convergence ratio and the converging shifts results in quadratic (or better) convergence. Once λ_n has been found, the problem can be deflated, and attention can be turned to λ_{n-1}.

As we have presented it here, it appears that shifting should not begin until convergence is well under way. In fact it was discovered long ago that shifting may effectively be done right from the first iteration. The only consequence is that the eigenvalues no longer come out in order; typically the larger ones come out first. It was also found by Wilkinson (1965) that a better choice than $\sigma_m = a_{nn}^{(m)}$ is to take σ_m to be that eigenvalue of

$$\begin{bmatrix} a_{n-1,n-1}^{(m)} & a_{n-1,n}^{(m)} \\ a_{n,n-1}^{(m)} & a_{n,n}^{(m)} \end{bmatrix}$$

which is closer to $a_{nn}^{(m)}$. (In the real case, if this submatrix has complex conjugate eigenvalues, a double QR step using this complex conjugate pair of shifts is taken. See, for example, Stewart (1973a).)

Returning to the unshifted algorithm, we note that as we approach convergence we have estimates not only of λ_n, but of all other eigenvalues as well. Thus, we could choose a shift to approximate any one of them. However, it would be inappropriate to approximate any eigenvalue other than λ_n, as this would alter the ordering of the eigenvalues and cause them to attempt to converge in a new order. All of the progress toward convergence would be undone. It follows that in the general case in which shifting is done at each step, the shifts should always be determined by information from the lower right-hand corner of the matrix. That is, they should attempt to approximate whichever eigenvalue is due to emerge next at that corner of the matrix. In this way the ordering which emerges will be preserved, and therefore reinforced, throughout the iterations.

4. Inverse iteration and duality in the QR algorithm.

Inverse iteration. If A^{-1} exists, it has eigenvalues $(\lambda_n)^{-1}, (\lambda_{n-1})^{-1}, \cdots, (\lambda_1)^{-1}$ with eigenvectors $v_n, v_{n-1}, \cdots, v_1$. If also $|\lambda_{n-1}| > |\lambda_n|$, then the sequence

$$v, A^{-1}v, A^{-2}v, A^{-3}v, \cdots$$

will converge (if appropriately rescaled) to a multiple of v_n, provided that $c_n \neq 0$. The convergence is linear with ratio of successive errors roughly $|\lambda_n/\lambda_{n-1}|$. More generally, if σ is any non-eigenvalue, one can shift A by σ and form $(A - \sigma I)^{-1}$, whose eigenvalues are $(\lambda_1 - \sigma)^{-1}, (\lambda_2 - \sigma)^{-1}, \cdots, (\lambda_n - \sigma)^{-1}$. Suppose σ is a good approximation to some λ_i, good enough that $|\lambda_i - \sigma| \ll |\lambda_j - \sigma|$ for all $j \neq i$. Then the iterates

$$v, (A - \sigma I)^{-1}v, (A - \sigma I)^{-2}v, (A - \sigma I)^{-3}v, \cdots$$

(properly rescaled) will converge to a multiple of the eigenvector v_i, provided that $c_i \neq 0$. Convergence is linear with ratio of successive errors given by $r = \max(|(\lambda_i - \sigma)/(\lambda_j - \sigma)|)$. Since $r \ll 1$, convergence is fast.

Clearly A^{-1} or $(A - \sigma I)^{-1}$ may be applied to subspaces as well, with results analogous to those of §2.

Rayleigh quotient iteration is a variant of inverse iteration in which a different shift is used at each step. Suppose that after m steps we have the vector v^m, which approximates the eigenvector v_i. Then the *Rayleigh quotient*

$$\sigma_m = (v^{m*}Av^m)/(v^{m*}v^m)$$

is a good approximation to the corresponding eigenvalue λ_i. If it is a good enough approximation, then $|\lambda_i - \sigma_m|$ will be much smaller than $|\lambda_j - \sigma_m|$ for all $j \neq i$. We can then apply one step of inverse iteration to $A - \sigma_m I$ to get v^{m+1}, a much better approximation to v_i. The new Rayleigh quotient σ_{m+1} will then be much closer to λ_i. Ostrowski

(1958), (1959) established that (local) convergence is quadratic in general and cubic in the Hermitian case. The global convergence question is difficult because a different shift is used at each step. Kahan (Parlett and Kahan (1968)) showed that in the Hermitian case convergence takes place for almost all starting vectors. (The proof is also given in Parlett (1980).) The cases for which convergence does not occur are unstable under roundoff error, so in practice convergence is global. Parlett (1974) generalized the result to normal matrices. Chen (1977) showed some of the difficulties which occur in the nonnormal case.

Duality in the QR algorithm. The following duality theorem provides the link between the QR algorithm and inverse iteration. It shows that whenever direct (subspace) iteration takes place, inverse (subspace) iteration also takes place automatically.

THEOREM 4.1. *Suppose A is nonsingular, and let S and $S^\perp$ be orthogonal, complementary subspaces of C^n. Then, for all integers m, A^mS and $(A^*)^{-m}S^\perp$ are also orthogonal complements.*

Proof. Let $x, y \in C^n$. Then $(x, y) = (Ax, (A^*)^{-1}y)$, etc. □

Thus the sequences

$$S, AS, A^2S, \cdots$$
$$S^\perp, (A^*)^{-1}S^\perp, (A^*)^{-2}S^\perp, \cdots$$

are equivalent in that they yield orthogonal complements. That is, subspace iteration by A on S is equivalent to subspace iteration by $(A^*)^{-1}$ on $S^\perp$. How is this reflected in the QR algorithm? The starting subspaces for QR are $\langle e_1, \cdots, e_k\rangle$, $k = 1, \cdots, n$, so it must be that iteration by $(A^*)^{-1}$ is also tacitly taking place on the subspaces $\langle e_{k+1}, \cdots, e_n\rangle$. Let $q_1^m, \cdots, q_n^m$ be as in §§2 and 3. Then since

$$\langle q_1^m, \cdots, q_k^m\rangle = \langle A^m e_1, \cdots, A^m e_k\rangle, \qquad k = 1, \cdots, n = 1,$$

it follows from Theorem 4.1 that

$$\langle q_{k+1}^m, \cdots, q_n^m\rangle = \langle (A^*)^{-m} e_{k+1}, \cdots, (A^*)^{-m} e_n\rangle. \tag{4.1}$$

These equations can also be derived from the basic equations of the QR algorithm, (3.5) and (3.7). Taking conjugate transposes and inverting each of these equations we get

$$(A_{m-1}^*)^{-1} = Q_m L_m, \qquad (A_m^*)^{-1} = L_m Q_m, \tag{4.2}$$

$$(A^*)^{-m} = \hat{Q}_m \hat{L}_m \tag{4.3}$$

where $L_m = (R_m^*)^{-1}$ and $\hat{L}_m = (\hat{R}_m^*)^{-1} = L_m L_{m-1} \cdots L_1$. L_m and $\hat{L}_m$ are lower triangular. Equations (4.2) show that the QR algorithm on A is equivalent to a QL algorithm on $(A^*)^{-1}$. The QL algorithm is based on the QL decomposition, for which there is a theorem analogous to Theorem 3.1. The QL decomposition is also an orthonormalization procedure, but with the last column orthonormalized first. That is, if $B = QL$, then $\langle b_n\rangle = \langle q_n\rangle$, $\langle b_{n-1}, b_n\rangle = \langle q_{n-1}, q_n\rangle$, and so on. Just as (3.7) connects the QR algorithm with simultaneous iteration by A, (4.3) establishes the connection with simultaneous iteration by $(A^*)^{-1}$. Specifically, the equations (4.1) follow immediately.

Now consider the case $k = n - 1$ in (4.1). Since $\langle q_n^m\rangle = \langle (A^*)^{-m} e_n\rangle$, the last column of $\hat{Q}_m$ represents the effect of inverse iteration by A^*, with starting vector e_n. Therefore q_n^m should converge to the eigenvector of A^* corresponding to its smallest eigenvalue $\bar{\lambda}_n$. Convergence can be accelerated by subtracting a shift $\bar{\sigma}_m$ which approximates $\bar{\lambda}_n$. A reasonable choice of shift is the Rayleigh quotient,

$$\bar{\sigma}_m = (q_n^m)^* A^* (q_n^m).$$

Then

$$\sigma_m = \overline{(q_n^m)^* A^* (q_n^m)} = (q_n^m)^* A (q_n^m) = a_{nn}^{(m)},$$

by (3.2). Thus the Rayleigh quotient is just the shift suggested originally in §3.

From the almost global convergence of Rayleigh quotient iteration it follows that the shifted QR algorithm for Hermitian matrices almost always converges. If the Wilkinson shift (introduced in §3) is used, convergence always takes place. See Wilkinson (1968) or Parlett (1980). In the non-Hermitian case, shifted QR is thought to converge almost always, but no proof has been found. A proof of (almost) global convergence of Rayleigh quotient iteration is required.

5. Variants. The QR algorithm was preceded by the LR or LU algorithm of Rutishauser (1955), (1958). The LU algorithm, as we shall call it, is based on successive LU decompositions, where L is lower triangular with 1's on the main diagonal and U is upper triangular. Thus, the unshifted algorithm has the form

$$B_{m-1} = L_m U_m, \qquad B_m = U_m L_m, \tag{5.1}$$

where $B_0 = A$. Not every matrix has an LU decomposition, so this procedure cannot always be carried out. We will not concern ourselves with that. If A is Hermitian and positive definite one can use the Cholesky decomposition $A = GG^*$, where G is lower triangular with positive main diagonal entries. From (5.1) we have

$$B_m = L_m^{-1} B_{m-1} L_m = U_m B_{m-1} U_m^{-1}, \tag{5.2}$$

from which

$$B_m = \hat{L}_m^{-1} A \hat{L}_m = \hat{U}_m A \hat{U}_m^{-1}, \tag{5.3}$$

where $\hat{L}_m = L_1 L_2 \cdots L_m$ and $\hat{U}_m = U_m U_{m-1} \cdots U_1$. Also, in analogy with (3.7),

$$A^m = \hat{L}_m \hat{U}_m. \tag{5.4}$$

Like the QR decomposition, the LU decomposition is a normalization procedure. If $A = LU$, then the columns of A and L are related by $\langle a_1, \cdots, a_k \rangle = \langle l_1, \cdots, l_k \rangle$, $k = 1, \cdots, n$. This follows from the fact that U is upper triangular. (Since L is also triangular, the *rows* of A and U satisfy a similar relationship.) The columns of L are not orthonormal. Instead they are normalized so that the ith column has $i - 1$ initial zeroes followed by a one. This can be thought of as a cheap alternative to orthonormalization.

From (5.4) one sees that $\hat{L}_m$ and $\hat{U}_m$ are the LU factors of A^m. Thus

$$\langle A^m e_1, \cdots, A^m e_k \rangle = \langle l_1^m, \cdots, l_k^m \rangle, \qquad m = 1, 2, \cdots,$$

where $l_1^m, \cdots, l_k^m$ are the columns of $\hat{L}_m$. This shows that the first k columns of $\hat{L}_m$ span the space gotten by m steps of subspace iteration on $\langle e_1, \cdots, e_k \rangle$, which is the same space as is spanned by the first k columns of $\hat{Q}_m$ in the QR algorithm. This equality of subspaces was recognized early and later reemphasized by Parlett and Poole (1973). However, equality of the underlying subspaces does not imply that the two algorithms give the same results. The major focus of the algorithms is not on subspaces, but on sequences of matrices $A_0, A_1, A_2, \cdots$ and $B_0, B_1, B_2, \cdots$. The two methods generate different sequences:

$$A_m = \hat{Q}_m^* A \hat{Q}_m, \qquad B_m = \hat{L}_m^{-1} A \hat{L}_m.$$

The difference in these two sequences is dramatized by the fact (cf. Wilkinson (1965, p. 545)) that if the Cholesky variant is used, then $B_{2m} = A_m$. That is, one QR step equals two

Cholesky *LU* steps. In the general case the relationship is not so clear, but practice has shown that *QR* usually converges faster than *LU*. In addition, the unitary matrices $\hat{Q}_m$ are amenable to analysis, whereas the $\hat{L}_m$ are not so convenient analytically. The entries of $\hat{L}_m$ and $\hat{L}_m^{-1}$ may grow with m, as may the entries of B_m. As a consequence, convergence of B_m to triangular form cannot be deduced from convergence of the subspaces. By contrast the entries of $\hat{Q}_m$ and $\hat{Q}_m^*$ are bounded by 1, those of A_m are bounded by the spectral norm of A, and convergence of the subspaces implies convergence of A_m to triangular form. Nevertheless, a recent paper of Dax and Kaniel (1981) suggests that the *LU* algorithm may not yet be dead.

Acknowledgment. The author thanks Beresford Parlett for making many helpful suggestions and providing historical perspective which the author could not possibly possess.

REFERENCES

F. L. BAUER (1957), *Das Verfahren der Treppeniteration und verwandte Verfahren zur Losung algebraischer Eigenwertprobleme*, Z. Angew. Math. Phys., 8, pp. 214–235.

——— (1958), *On modern matrix iteration processes of Bernoulli and Graeffe type*, J. Assoc. Comput. Mach., 5, pp. 246–257.

A. BJÖRK AND G. H. GOLUB (1973), *Numerical methods for calculating angles between subspaces*, Math. Comp., 27, pp. 579–594.

H. J. BUUREMA (1970), *A geometric proof of convergence for the QR method*, thesis, Rijksuniversiteit te Groningen.

N.-F. CHEN (1977), *Inverse iteration on defective matrices*, Math. Comp., 31, pp. 726–732.

M. CLINT AND A. JENNINGS (1970), *The evaluation of eigenvalues and eigenvectors of real symmetric matrices by simultaneous iteration*, Comput. J., 13, pp. 76–80.

——— (1971), *A simultaneous iteration method for the unsymmetric eigenvalue problem*, J. Inst. Math. Appl., 8, pp. 111–121.

CH. DAVIS AND W. M. KAHAN (1970), *The rotation of eigenvectors by a perturbation*. III, SIAM J. Numer. Anal., 7, pp. 1–46.

A. DAX AND S. KANIEL (1981), *The ELR method for computing the eigenvalues of a general matrix*, SIAM J. Numer. Anal., 18, pp. 597–605.

D. K. FADDEEV AND V. N. FADDEEVA (1963), *Computational Methods of Linear Algebra*, Freeman, San Francisco.

J. G. F. FRANCIS (1961), (1962), *The QR transformation* I, II, Comput. J., 4, pp. 265–271, 332–345.

A. S. HOUSEHOLDER (1964), *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York.

A. JENNINGS (1967), *A direct iteration method of obtaining latent roots and vectors of a symmetric matrix*, Proc. Cambridge Philos. Soc., 63, pp. 755–765.

V. N. KUBLANOVSKAYA (1961), *On some algorithms for the solution of the complete eigenvalue problem*, USSR Comput. Math. Math. Phys., 3, pp. 637–657.

A. M. OSTROWSKI (1958), (1959), *On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors*, I–VI, Arch. Rat. Mech. Anal., 1, pp. 233–241; 2, pp. 423–428; 3, pp. 325–340, pp. 341–347, pp. 472–481; 4, pp. 153–165.

B. N. PARLETT (1968), *Global convergence of the basic QR algorithm on Hessenberg matrices*, Math. Comp., 22, pp. 803–817.

——— (1974), *The Rayleigh quotient iteration and some generalizations for non-normal matrices*, Math. Comp., 28, pp. 679–693.

——— (1980), *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ.

B. N. PARLETT AND W. KAHAN (1968), *On the convergence of a practical QR algorithm*, Proc. IFIP Congress, 1968, pp. A25–A30.

B. N. PARLETT AND W. G. POOLE, JR. (1973), *A geometric theory for the QR, LU, and power iterations*, SIAM J. Numer. Anal., 8, pp. 389–412.

H. RUTISHAUSER (1955), *Une méthode pour la détermination des valeurs propres d'une matrice*, Comptes Rendus Acad. Sci. Paris, 240, pp. 34–36.

——— (1958), *Solution of eigenvalue problems with the LR transformation,* Nat. Bur. Stand. Appl. Math. Ser., 49, pp. 47–81.

——— (1969), *Computational aspects of F. L. Bauer's simultaneous iteration method,* Numer. Math., 13, pp. 4–13.

——— (1970), *Simultaneous iteration method for symmetric matrices,* Numer. Math., 16, pp. 205–223. Contribution II/9 of Wilkinson and Reinsch (1971).

B. T. SMITH ET AL. (1976), *Matrix Eigensystem Routines—EISPACK Guide,* 2nd ed., Lecture Notes in Computer Science, 6, Springer-Verlag, New York.

G. W. STEWART (1969), *Accelerating the orthogonal iteration for the eigenvalues of a Hermitian matrix,* Numer. Math., 13, pp. 362–376.

——— (1971), *Error bounds for approximate invariant subspaces of closed linear operators,* SIAM J. Numer. Anal., 8, pp. 796–808.

——— (1973a), *Introduction to Matrix Computations,* Academic Press, New York.

——— (1973b), *Error and perturbation bounds for subspaces associated with certain eigenvalue problems,* this Review, 15, pp. 727–764.

——— (1976), *Simultaneous iteration for computing invariant subspaces of non-Hermitian matrices,* Numer. Math., 25, pp. 123–136.

——— (1977), *On the perturbation of pseudo-inverses, projections, and linear least squares problems,* this Review, 19, pp. 634–662.

J. H. WILKINSON (1965), *The Algebraic Eigenvalue Problem,* Clarendon Press, Oxford.

——— (1968), *Global convergence of the QR algorithm,* Proc. IFIP Congress, pp. A22–A24.

J. H. WILKINSON AND C. REINSCH (1971), *Handbook for Automatic Computation, vol. 2, Linear Algebra,* Springer-Verlag, New York.