A convergence result for the QR method A convergence result is given for a large class of matrices, symmetric and nonsymmetric. Part of the proof will show the factor which determines the speed of convergence of the method. For a more general discussion of the method, see Golub-Van Loan (1983, §7.5, §8.2), Parlett (1968), Parlett (1980, Chap. 8), and Wilkinson (1965, Chap. 8).

Theorem 9.6 Let A be a real matrix of order n, and let its eigenvalues $\{\lambda_i\}$ satisfy

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0. \tag{9.5.13}$$

Then the iterates A_m of the QR method, defined in (9.5.1), will converge to an upper triangular matrix which contains the eigenvalues $\{\lambda_i\}$ in the diagonal positions. If A is symmetric, the sequence $\{A_m\}$ converges to a diagonal matrix.

Proof. The proof is fairly lengthy and involved, and the reader may wish to skip it and go onto the discussion following the proof. The main factor in determining the speed of convergence is contained in (9.5.19), and it is illustrated following the proof. This proof follows closely that of Wilkinson (1965. pp. 517-519).

Since A has distinct eigenvalues, there is a nonsingular matrix X for which

$$X^{-1}AX = D = \mathrm{diag}[\lambda_1, \ldots, \lambda_n]. \qquad (9.5.14)$$

Then

$$A^m = XD^mX^{-1}. \qquad (9.5.15)$$

Since A is real and all of its eigenvalues are of distinct magnitude, it cannot have any complex roots, as they would have to occur in conjugate pairs of equal magnitude.

The next few paragraphs will derive some alternative forms for A^m, based on modifying (9.5.15). Assume X^{-1} has the decomposition

$$X^{-1} = LU. \qquad (9.5.16)$$

For the appropriate modification to use when this is not possible and pivoting must be used, see Wilkinson (1965, p. 519). Combining (9.5.15) and (9.5.16),

$$A^m = X(D^mLD^{-m})D^mU. \qquad (9.5.17)$$

Recall that in the derivation of the decomposition (9.5.16) in §8.1 of Chapter 8, the diagonal elements of L could be chosen to be 1. Then the matrix D^mLD^{-m} is lower triangular with diagonal elements equal to 1, and

$$(D^mLD^{-m})_{ij} = \left[\frac{\lambda_i}{\lambda_j}\right]^m L_{ij}, \qquad 1 \le j < i \le n. \qquad (9.5.18)$$

Define E_m implicitly by

$$D^m L D^{-m} = I + E_m.$$

E_m is a lower triangular matrix which converges to zero: using (9.5.18) and (9.5.13),

$$\|E_m\|_\infty \le c\cdot \underset{1\le j\le n-1}{\text{Maximum}} \left|\frac{\lambda_{j+1}}{\lambda_j}\right|^m, \qquad m\ge 1, \tag{9.5.19}$$

for some constant $c>0$.

The matrix X can be factored.

$$X = QR,$$

for some orthogonal Q and nonsingular upper triangular R. Returning to (9.5.17), this leads to

$$\begin{aligned} A^m &= QR(I+E_m)D^m U \\ &= Q(I+RE_m R^{-1})RD^m U. \end{aligned} \tag{9.5.20}$$

Using another QR factorization,

$$I+RE_m R^{-1} = \tilde{Q}_m \tilde{R}_m. \tag{9.5.21}$$

We require the diagonal elements of R_m to be positive, which is possible from the construction for the factorization given in §8.3. Also see the discussion between (9.3.15) and (9.3.16), which shows that with this positivity assumption, the decomposition (9.5.21) is unique.

We can show that

$$\tilde{Q}_m, \tilde{R}_m \longrightarrow I \quad \text{as } m\to\infty. \tag{9.5.22}$$

Using (9.5.21) and (9.5.19), it is straightforward to show that

$$\tilde{R}_m^T \tilde{R}_m - I \longrightarrow 0 \qquad \text{as } m\to\infty.$$

A detailed examination of the coefficients of $\tilde{R}_m^T \tilde{R}_m$ will then show that $\tilde{R}_m \to I$, using the positivity of the diagonal elements. Using this result in (9.5.21) will then show $\tilde{Q}_m \to I$.

Using (9.5.21) in (9.5.20),

$$A^m = (Q\tilde{Q}_m)(\tilde{R}_m R D^m U). \qquad (9.5.23)$$

Clearly $Q\tilde{Q}_m$ is orthogonal. And because $\tilde{R}_m$, R, U are upper triangular and D^m is diagonal, we have their product is upper triangular. Thus (9.5.23) is a QR factorization of A^m. Returning to (9.5.7), we have the second QR factorization

$$A^m = P_m U_m$$

Comparing these results and using the uniqueness of the QR factorization expressed in (9.3.15) and (9.3.16), we have

$$P_m = (Q\tilde{Q}_m)\tilde{D}_m, \qquad U_m = \tilde{D}_m(\tilde{R}_m R D^m U), \qquad (9.5.24)$$

for some diagonal matrix $\tilde{D}_m$ with

$$\tilde{D}_m^2 = I, \qquad m \geq 1. \qquad (9.5.25)$$

We now examine the behavior of the sequence $\{A_m\}$ as $m \to \infty$. From (9.5.5) and (9.5.24),

$$\begin{aligned} A_{m+1} &= P_m^T A P_m \\ &= \tilde{D}_m \tilde{Q}_m^T Q^T A Q \tilde{Q}_m \tilde{D}_m \end{aligned}$$

From earlier $X=QR$, and

$$Q = XR^{-1}$$

$$Q^T = Q^{-1} = RX^{-1}.$$

Substituting above,

$$\begin{aligned} A_{m+1} &= \tilde{D}_m^T \tilde{Q}_m^T R X^{-1} A X R^{-1} \tilde{Q}_m \tilde{D}_m \\ &= \tilde{D}_m^T \tilde{Q}_m^T R D R^{-1} \tilde{Q}_m \tilde{D}_m \end{aligned} \qquad (9.5.26)$$

Consider just the diagonal elements of A_{m+1} since they are the main point of interest. The matrix RDR^{-1} is upper triangular and its diagonal elements are just $\{\lambda_1, \ldots, \lambda_n\}$. Using (9.5.22) and (9.5.25), that $\tilde{Q}_m \to I$ and $\tilde{D}_m^2 = I$, we will then have the diagonal elements of A_{m+1} will converge to the eigenvalues of A, ordered from largest to smallest in magnitude. In addition, since RDR^{-1} is upper triangular, the elements below the diagonal in A_{m+1} will converge to zero. The speed of convergence will depend completely on the speed of convergence of $\tilde{Q}_m$ to I; and this depends on the bound in (9.5.19).

If A is symmetric, then the iterates A_m are also symmetric. Since the lower triangular part of A_m converges to zero, the same is true of the part above the diagonal. This proves that for a symmetric matrix satisfying (9.5.13), A_m converges to a diagonal matrix. This completes the proof. ■

As pointed out in the proof, the critical factor in determining the speed of convergence are the ratios λ_{j+1}/λ_j, $1 \le j \le n-1$. Thus there is a geometric rate of convergence, and this can be very slow. In the example (9.5.8), the ratios of successive eigenvalues are

$$\frac{\lambda_2}{\lambda_1} \doteq 0.63, \qquad \frac{\lambda_3}{\lambda_2} \doteq 0.42.$$

And if the off-diagonal elements are observed, we see that they decrease with about these ratios.

For matrices whose eigenvalues do not satisfy (9.5.13), the iterates A_m may not converge to a triangular matrix. For A symmetric, the sequence $\{A_m\}$ will converge to a block diagonal matrix

$$A_m \longrightarrow D = \begin{bmatrix} B_1 & & & 0 \\ & B_2 & & \\ & & \ddots & \\ 0 & & & B_r \end{bmatrix}, \qquad (9.5.27)$$

in which all blocks B_i have order 1 or 2. Thus the eigenvalues of A can be easily computed from those of D. If A is real and nonsymmetric, the situation is more complicated, but acceptable. For a discussion, see Wilkinson (1965, Chap. 8) and Parlett (1968).

To see that $\{A_m\}$ does not always converge to a diagonal matrix, consider the simple symmetric example

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Its eigenvalues are $\lambda=\pm 1$. Since A is orthogonal, we have

$$A = Q_1R_1 \qquad \text{with } Q_1 = A, \quad R_1 = I.$$

And thus

$$A_2 = R_1Q_1 = A,$$

and all iterates $A_m=A$. The sequence $\{A_m\}$ does not converge to a diagonal matrix.