These notes correspond to Sections 7.5 and 8.3 in the text.

**The Eigenvalue Problem: The Practical QR Algorithm**

**The Unsymmetric Eigenvalue Problem**

The efficiency of the QR Iteration for computing the eigenvalues of an \( n \times n \) matrix \( A \) is significantly improved by first reducing \( A \) to a Hessenberg matrix \( H \), so that only \( O(n^2) \) operations per iteration are required, instead of \( O(n^3) \). However, the iteration can still converge very slowly, so additional modifications are needed to make the QR Iteration a practical algorithm for computing the eigenvalues of a general matrix.

In general, the \( p \)th subdiagonal entry of \( H \) converges to zero at the rate

\[
\frac{\lambda_{p+1}}{\lambda_p},
\]

where \( \lambda_p \) is the \( p \)th largest eigenvalue of \( A \) in magnitude. It follows that convergence can be particularly slow if eigenvalues are very close to one another in magnitude. Suppose that we shift \( H \) by a scalar \( \mu \), meaning that we compute the QR factorization of \( H - \mu I \) instead of \( H \), and then update \( H \) to obtain a new Hessenberg \( \tilde{H} \) by multiplying the QR factors in reverse order as before, but then adding \( \mu I \). Then, we have

\[
\tilde{H} = RQ + \mu I = Q^T(H - \mu I)Q + \mu I = Q^THQ - \mu Q^TQ + \mu I = Q^THQ - \mu I + \mu I = Q^THQ.
\]

So, we are still performing an orthogonal similarity transformation of \( H \), but with a different \( Q \). Then, the convergence rate becomes \( |\lambda_{p+1} - \mu|/|\lambda_p - \mu| \). Then, if \( \mu \) is close to an eigenvalue, convergence of a particular subdiagonal entry will be much more rapid.

In fact, suppose \( H \) is unreduced, and that \( \mu \) happens to be an eigenvalue of \( H \). When we compute the QR factorization of \( H - \mu I \), which is now singular, then, because the first \( n - 1 \) columns of \( H - \mu I \) must be linearly independent, it follows that the first \( n - 1 \) columns of \( R \) must be linearly independent as well, and therefore the last row of \( R \) must be zero. Then, when we compute \( RQ \), which involves rotating columns of \( R \), it follows that the last row of \( RQ \) must also be zero. We then add \( \mu I \), but as this only changes the diagonal elements, we can conclude that \( \tilde{h}_{n,n-1} = 0 \). In other words, \( \tilde{H} \) is not an unreduced Hessenberg matrix, and deflation has occurred in one step.

If \( \mu \) is not an eigenvalue of \( H \), but is still close to an eigenvalue, then \( H - \mu I \) is nearly singular, which means that its columns are nearly linearly dependent. It follows that \( r_{nn} \) is small, and it can
be shown that \( h_{n,n-1} \) is also small, and \( h_{nn} \approx \mu \). Therefore, the problem is nearly decoupled, and \( \mu \) is revealed by the structure of \( \tilde{H} \) as an approximate eigenvalue of \( H \). This suggests using \( h_{nn} \) as the shift \( \mu \) during each iteration, because if \( h_{n,n-1} \) is small compared to \( h_{nn} \), then this choice of shift will drive \( h_{n,n-1} \) toward zero. In fact, it can be shown that this strategy generally causes \( h_{n,n-1} \) to converge to zero \textit{quadratically}, meaning that only a few similarity transformations are needed to achieve decoupling. This improvement over the linear convergence rate reported earlier is due to the changing of the shift during each step.

\textbf{Example} Consider the 2 \( \times \) 2 matrix

\[
H = \begin{bmatrix} a & b \\ \epsilon & 0 \end{bmatrix}, \quad \epsilon > 0,
\]

that arises naturally when using \( h_{nn} \) as a shift. To compute its QR factorization of \( H \), we perform a single Givens rotation to obtain \( H = GR \), where

\[
G = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \quad c = \frac{a}{\sqrt{a^2 + \epsilon^2}}, \quad s = \frac{\epsilon}{\sqrt{a^2 + \epsilon^2}}.
\]

Performing the similarity transformation \( \tilde{H} = G^T H G \) yields

\[
\tilde{H} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a & b \\ \epsilon & 0 \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} ac + bs & bc - \epsilon s \\ \epsilon c & -\epsilon s \end{bmatrix} = \begin{bmatrix} ac^2 + bcs + \epsilon cs & bc^2 - acs - \epsilon s^2 \\ -acs - bs^2 + \epsilon c^2 & -bcs + as^2 - \epsilon cs \end{bmatrix} = \begin{bmatrix} a + bcs & bc^2 - \epsilon \\ -bs^2 & -bcs \end{bmatrix}.
\]

We see that the one subdiagonal element is

\[
-bs^2 = -b \frac{\epsilon^2}{\epsilon^2 + a^2},
\]

cmpared to the original element \( \epsilon \). It follows that if \( \epsilon \) is small compared to \( a \) and \( b \), then subsequent \( QR \) steps will cause the subdiagonal element to converge to zero \textit{quadratically}. For example, if

\[
H = \begin{bmatrix} 0.6324 & 0.2785 \\ 0.0975 & 0.5469 \end{bmatrix},
\]

then the value of \( h_{21} \) after each of the first three \( QR \) steps is 0.1575, \(-0.0037\), and \( 2.0876 \times 10^{-5} \). \(\square\)
This shifting strategy is called the *single shift strategy*. Unfortunately, it is not very effective if \( H \) has complex eigenvalues. An alternative is the *double shift strategy*, which is used if the two eigenvalues, \( \mu_1 \) and \( \mu_2 \), of the lower-right \( 2 \times 2 \) block of \( H \) are complex. Then, these two eigenvalues are used as shifts in consecutive iterations to achieve quadratic convergence in the complex case as well. That is, we compute

\[
\begin{align*}
H - \mu_1 I &= U_1 R_1 \\
H_1 &= R_1 U_1 + \mu_1 I \\
H_1 - \mu_2 I &= U_2 R_2 \\
H_2 &= R_2 U_2 + \mu_2 I.
\end{align*}
\]

To avoid complex arithmetic when using complex shifts, the *double implicit shift strategy* is used. We first note that

\[
U_1 U_2 R_2 R_1 = U_1 (H_1 - \mu_2 I) R_1 \\
= U_1 H_1 R_1 - \mu_2 U_1 R_1 \\
= U_1 (R_1 U_1 + \mu_1 I) R_1 - \mu_2 (H - \mu_1 I) \\
= U_1 R_1 U_1 R_1 + \mu_1 U_1 R_1 - \mu_2 (H - \mu_1 I) \\
= (H - \mu_1 I)^2 + \mu_1 (H - \mu_1 I) - \mu_2 (H - \mu_1 I) \\
= H^2 - 2 \mu_1 H + \mu_1^2 I + \mu_1 H - \mu_1^2 I - \mu_2 I + \mu_1 \mu_2 I \\
= H^2 - (\mu_1 + \mu_2) H + \mu_1 \mu_2 I.
\]

Since \( \mu_1 = a + bi \) and \( \mu_2 = a - bi \) are a complex-conjugate pair, it follows that \( \mu_1 + \mu_2 = ab \) and \( \mu_1 \mu_2 = a^2 + b^2 \) are real. Therefore, \( U_1 U_2 R_2 R_1 = (U_1 U_2)(R_2 R_1) \) represents the QR factorization of a real matrix.

Furthermore,

\[
H_2 = R_2 U_2 + \mu_2 I = U_2^T H_1 U_2 = U_2^T (R_1 U_1 + \mu_1 I) U_2 = U_2^T U_1^T H U_1 U_2.
\]

That is, \( U_1 U_2 \) is the orthogonal matrix that implements the similarity transformation of \( H \) to obtain \( H_2 \). Therefore, we could use exclusively real arithmetic by forming \( M = H^2 - (\mu_1 + \mu_2) H + \mu_1 \mu_2 I \), compute its QR factorization to obtain \( M = Z R \), and then compute \( H_2 = Z^T H Z \), since \( Z = U_1 U_2 \), in view of the uniqueness of the QR decomposition. However, because \( M \) is computed by squaring \( H \), which requires \( O(n^3) \) operations. Therefore, this is not a practical approach.

We can work around this difficulty using the Implicit Q Theorem. Instead of forming \( M \) in its entirety, we only form its first column, which, being a second-degree polynomial of a Hessenberg matrix, has only three nonzero entries. We compute a Householder transformation \( P_0 \) that makes this first column a multiple of \( e_1 \). Then, we compute \( P_0 H P_0 \), which is no longer Hessenberg, because it operates on the first three rows and columns of \( H \). Finally, we apply a series of Householder
reflections $P_1, P_2, \ldots, P_{n-2}$ that restore Hessenberg form. Because these reflections are not applied to the first row or column, it follows that if we define $\tilde{Z} = P_0 P_1 P_2 \cdots P_{n-2}$, then $Z$ and $\tilde{Z}$ have the same first column. Since both matrices implement similarity transformations that preserve the Hessenberg form of $H$, it follows from the Implicit Q Theorem that $Z$ and $\tilde{Z}$ are essentially equal, and that they essentially produce the same updated matrix $H_2$. This variation of a Hessenberg QR step is called a Francis QR step.

A Francis QR step requires $10n^2$ operations, with an additional $10n^2$ operations if orthogonal transformations are being accumulated to obtain the entire real Schur decomposition. Generally, the entire QR algorithm, including the initial reduction to Hessenberg form, requires about $10n^3$ operations, with an additional $15n^3$ operations to compute the orthogonal matrix $Q$ such that $A = QTQ^T$ is the real Schur decomposition of $A$.

The Symmetric Eigenvalue Problem

In the symmetric case, there is no need for a double-shift strategy, because the eigenvalues are real. However, the Implicit Q Theorem can be used for a different purpose: computing the similarity transformation to be used during each iteration without explicitly computing $T - \mu I$, where $T$ is the tridiagonal matrix that is to be reduced to diagonal form. Instead, the first column of $T - \mu I$ can be computed, and then a Householder transformation to make it a multiple of $e_1$. This can then be applied directly to $T$, followed by a series of Givens rotations to restore tridiagonal form. By the Implicit Q Theorem, this accomplishes the same effect as computing the QR factorization $UR = T - \mu I$ and then computing $\tilde{T} = RU + \mu I$.

While the shift $\mu = t_{nn}$ can always be used, it is actually more effective to use the Wilkinson shift, which is given by

$$\mu = t_{nn} + d - \text{sign}(d) \sqrt{d^2 + t_{n,n-1}^2}, \quad d = \frac{t_{n-1,n-1} - t_{nn}}{2}.$$  

This expression yields the eigenvalue of the lower $2 \times 2$ block of $T$ that is closer to $t_{nn}$. It can be shown that this choice of shift leads to cubic convergence of $t_{n,n-1}$ to zero.

The symmetric QR algorithm is much faster than the unsymmetric QR algorithm. A single QR step requires about $30n$ operations, because it operates on a tridiagonal matrix rather than a Hessenberg matrix, with an additional $6n^2$ operations for accumulating orthogonal transformations. The overall symmetric QR algorithm requires $4n^3/3$ operations to compute only the eigenvalues, and approximately $8n^3$ additional operations to accumulate transformations. Because a symmetric matrix is unitarily diagonalizable, then the columns of the orthogonal matrix $Q$ such that $Q^T A Q$ is diagonal contains the eigenvectors of $A$. 