State of the QR algorithm

\[ A \in \mathbb{R}^{n \times n} \]

\[ H = U_0^T A U_0 \quad \text{upper Hessenberg} \quad U_0^T U_0 = I \]

Suppose \( H \) is unreduced

For \( k = 1, 2, \ldots \)

\[
\begin{align*}
H - \mu I &= U R \quad \forall \\
H &= RU + \mu I
\end{align*}
\]

- If \( H \) is in fact reduced work on the unreduced blocks of \( H \).

- If \( \mu \) is an eigenvalue then * causes decoupling in the last row of \( H \). \((h_{n, n-1} = 0)\)

- There may be conjugate complex pairs of eigenvalues!

- In that case, how about a double shift?
- Suppose \( a_1 \) and \( a_2 \) are the (possibly conjugate complex) eigenvalues of

\[
\begin{bmatrix}
  b_{mm} & b_{mn} \\
  b_{nm} & b_{nn}
\end{bmatrix}, \quad m = n-1
\]

(the bottom right 2x2 block of \( H \))

- Consider the double shift:

\[
H - a_1 I = U_1 R_1, \tag{1}
\]

\[
H_1 = R_1 U_1 + a_1 I \tag{2}
\]

\[
H_1 - a_2 I = U_2 R_2 \tag{3}
\]

\[
H_2 = R_2 U_2 + a_2 I \tag{4}
\]

- But we don't want to use complex arithmetic.

- It turns out that

\[
H_2 = U_2^T U_1^T H U_1 U_2
\]

where \((U_1, U_2)(R_1, R_2) = (H - a_1 I)(H - a_2 I)\)

\[
= H^2 - (a_1 + a_2) I + a_1 a_2 I = M \text{ real}
\]

real \[\rightarrow\] real
- so we could do the double shift by computing $M$ and then the QR factorization of $M$, keeping everything real.

- why is (***) true?

\[
\begin{align*}
    H - a_2 I &= U_2 R_2 \\
    U_1 (H - a_2 I) R_1 &= U_1 U_2 R_2 R_1 \\
    U_1 H R_1 - a_2 U R_1 &= U_1 U_2 R_2 R_1
\end{align*}
\]

use $H_1 = R_1 U_1 + a_1 I$ and $U_1 R_1 = H - a_1 I$

\[
\begin{align*}
    U_1 (R_1 U_1 + a_1 I) R_1 - a_2 (H - a_1 I) &= U_1 U_2 R_2 R_1 \\
    U_1 R_1 (U_1 R_1 + a_1 I) - a_2 (H - a_1 I) &= U_1 U_2 R_2 R_1
\end{align*}
\]

use $U_1 R_1 = H - a_1 I$

\[
\begin{align*}
(H - a_1 I) (U_1 R_1 + a_1 I) - a_2 H + a_1 a_2 I &= U_1 U_2 R_2 R_1 \\
(H - a_1 I) H - a_2 H + a_1 a_2 I &= U_1 U_2 R_2 R_1 \\
(H - a_1 I) (H - a_2 I) &= U_1 U_2 R_2 R_1
\end{align*}
\]
- But computing $M$ may require $O(n^3)$ operations.

- $H^2$ is no longer upper Hessenberg.

- There is an additional non-zero line below the diagonal.
Implicit $Q$ Theorem (Golub/Van Loan 4th ed. p. 306)

Suppose $Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$
and $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$
are orthogonal matrices with the property that the matrices

\[
H = Q^T A Q \quad \text{and} \quad G_i = V^T A V
\]

are each upper Hessenberg.

- Suppose $H$ is unreduced.
- If $q_i = v_i$, then $q_i = \pm v_i$
  \[
  |b_i; i-1| = |g_i; i-1| \quad i = 2, \ldots, n
  \]

- For a more general statement covering the case that $H$ is reduced, see (Golub/Van Loan)

- Basically the IAT says that the first column of $Q$ determines the rest of $Q$. 
7.4.5 Important Hessenberg Matrix Properties

The Hessenberg decomposition is not unique. If $Z$ is any $n$-by-$n$ orthogonal matrix and we apply Algorithm 7.4.2 to $Z^T AZ$, then $Q^T AQ = H$ is upper Hessenberg where $Q = ZU_0$. However, $Qe_1 = Z(U_0 e_1) = Ze_1$ suggesting that $H$ is unique once the first column of $Q$ is specified. This is essentially the case provided $H$ has no zero subdiagonal entries. Hessenberg matrices with this property are said to be unreduced. Here is important theorem that clarifies these issues.

**Theorem 7.4.2 (Implicit Q Theorem).** Suppose $Q = [q_1 \mid \cdots \mid q_n]$ and $V = [v_1 \mid \cdots \mid v_n]$ are orthogonal matrices with the property that the matrices $Q^T AQ = H$ and $V^T AV = G$ are each upper Hessenberg where $A \in \mathbb{R}^{n \times n}$. Let $k$ denote the smallest positive integer for which $h_{k+1,k} = 0$, with the convention that $k = n$ if $H$ is unreduced. If $q_1 = v_1$, then $q_i = \pm v_i$ and $|h_{i,i-1}| = |g_{i,i-1}|$ for $i = 2:k$. Moreover, if $k < n$, then $g_{k+1,k} = 0$.

**Proof.** Define the orthogonal matrix $W = [w_1 \mid \cdots \mid w_n] = VTQ$ and observe that $GW = WH$. By comparing column $i-1$ in this equation for $i = 2:k$ we see that

$$h_{i,i-1}w_i = Gw_{i-1} - \sum_{j=1}^{i-1} h_{j,i-1}w_j.$$ 

Since $w_1 = e_1$, it follows that $[w_1 \mid \cdots \mid w_n]$ is upper triangular and so for $i = 2:k$ we have $w_i = \pm I_i,i = \pm e_i$. Since $w_i = V^T q_i$ and $h_{i,i-1} = w_i^T Gw_{i-1}$ it follows that $v_i = \pm q_i$ and

$$|h_{i,i-1}| = |q_i^T A q_{i-1}| = |v_i^T A v_{i-1}| = |g_{i,i-1}|$$

for $i = 2:k$. If $k < n$, then

$$g_{k+1,k} = e_{k+1}^T G e_k = \pm e_{k+1}^T G W e_k = \pm e_{k+1}^T W H e_k$$

$$= \pm e_{k+1}^T \sum_{i=1}^{k} h_{ik} W e_i = \pm \sum_{i=1}^{k} h_{ik} e_{k+1} e_i = 0,$$

completing the proof of the theorem. \qed

The gist of the implicit $Q$ theorem is that if $Q^T AQ = H$ and $Z^T AZ = G$ are each unreduced upper Hessenberg matrices and $Q$ and $Z$ have the same first column, then $G$ and $H$ are “essentially equal” in the sense that $G = D^{-1} HD$ where $D = \text{diag}(\pm 1, \ldots , \pm 1)$.

Our next theorem involves a new type of matrix called a Krylov matrix. If $A \in \mathbb{R}^{n \times n}$ and $v \in \mathbb{R}^n$, then the Krylov matrix $K(A,v,j) \in \mathbb{R}^{n \times j}$ is defined by

$$K(A,v,j) = [v \mid Av \mid \cdots \mid A^{j-1} v].$$

It turns out that there is a connection between the Hessenberg reduction $Q^T AQ = H$ and the QR factorization of the Krylov matrix $K(A,Q(:,1),n)$.

**Theorem 7.4.3.** Suppose $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and $A \in \mathbb{R}^{n \times n}$. Then $Q^T AQ = H$ is an unreduced upper Hessenberg matrix if and only if $Q^T K(A,Q(:,1),n) = R$ is nonsingular and upper triangular.
and
\[ t = a_1 a_2 = h_{mn} h_{nn} - h_{nm} h_{mn} = \det(G) \in \mathbb{R}. \]
Thus, (7.5.7) is the QR factorization of a real matrix and we may choose \( U_1 \) and \( U_2 \) so that \( Z = U_1 U_2 \) is real orthogonal. It then follows that
\[ H_2 = U_2^H H_1 U_2 = U_2^H (U_1^H H U_1) U_2 = (U_1 U_2)^H H (U_1 U_2) = Z^T H Z \]
is real.
Unfortunately, roundoff error almost always prevents an exact return to the real field. A real \( H_2 \) could be guaranteed if we
- explicitly form the real matrix \( M = H^2 - sH + tI \),
- compute the real QR factorization \( M = ZR \), and
- set \( H_2 = Z^T H Z \).
But since the first of these steps requires \( O(n^3) \) flops, this is not a practical course of action.

### 7.5.5 The Double-Implicit-Shift Strategy

Fortunately, it turns out that we can implement the double-shift step with \( O(n^2) \) flops by appealing to the implicit Q theorem of §7.4.5. In particular we can effect the transition from \( H \) to \( H_2 \) in \( O(n^2) \) flops if we
- compute \( M e_1 \), the first column of \( M \);
- determine a Householder matrix \( P_0 \) such that \( P_0 (M e_1) \) is a multiple of \( e_1 \);
- compute Householder matrices \( P_1, \ldots, P_{n-2} \) such that if
\[ Z_1 = P_0 P_1 \cdots P_{n-2}, \]
then \( Z_1^T H Z_1 \) is upper Hessenberg and the first columns of \( Z \) and \( Z_1 \) are the same.
Under these circumstances, the implicit Q theorem permits us to conclude that, if \( Z^T H Z \) and \( Z_1^T H Z_1 \) are both unreduced upper Hessenberg matrices, then they are essentially equal. Note that if these Hessenberg matrices are not unreduced, then we can effect a decoupling and proceed with smaller unreduced subproblems.

Let us work out the details. Observe first that \( P_0 \) can be determined in \( O(1) \) flops since \( M e_1 = [x, y, z, 0, \ldots, 0]^T \) where
\[ x = h_{11}^2 + h_{12} h_{21} - sh_{11} + t, \]
\[ y = h_{21} (h_{11} + h_{22} - s), \]
\[ z = h_{22} h_{32}. \]
Since a similarity transformation with \( P_0 \) only changes rows and columns 1, 2, and 3, we see that
The Practical QR Algorithm

\[
P_0H_0P_0 = \begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\]

Now the mission of the Householder matrices \( P_1, \ldots, P_{n-2} \) is to restore this matrix to upper Hessenberg form. The calculation proceeds as follows:

\[
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\xrightarrow{P_1}
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\xrightarrow{P_2}
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\xrightarrow{P_3}
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\xrightarrow{P_4}
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  x & x & x & x & x \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & x \\
\end{bmatrix}
\]

Each \( P_k \) is the identity with a 3-by-3 or 2-by-2 Householder somewhere along its diagonal, e.g.,

\[
P_1 = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  0 & x & x & x & 0 \\
  0 & x & x & x & 0 \\
  0 & 0 & 0 & 0 & 1 \\
  0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}, \\
P_2 = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 \\
  0 & 0 & x & x & 0 \\
  0 & 0 & x & x & 0 \\
  0 & 0 & 0 & 0 & 1 \\
\end{bmatrix},
\]

\[
P_3 = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & x & x \\
\end{bmatrix}, \\
P_4 = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 \\
  0 & 0 & 0 & x & x \\
  0 & 0 & 0 & x & x \\
\end{bmatrix}
\]

The applicability of Theorem 7.4.3 (the implicit Q theorem) follows from the observation that \( P_k e_1 = e_1 \) for \( k = 1:n - 2 \) and that \( P_0 \) and \( Z \) have the same first column. Hence, \( Z_1 e_1 = Z e_1 \), and we can assert that \( Z_1 \) essentially equals \( Z \) provided that the upper Hessenberg matrices \( Z^T H Z \) and \( Z_1^T H Z_1 \) are each unreduced.
The implicit determination of $H_2$ from $H$ outlined above was first described by Francis (1961) and we refer to it as a Francis QR step. The complete Francis step is summarized as follows:

**Algorithm 7.5.1 (Francis QR step)** Given the unreduced upper Hessenberg matrix $H \in \mathbb{R}^{n \times n}$ whose trailing 2-by-2 principal submatrix has eigenvalues $a_1$ and $a_2$, this algorithm overwrites $H$ with $Z^T H Z$, where $Z$ is a product of Householder matrices and $Z^T (H - a_1 I)(H - a_2 I)$ is upper triangular.

\[
m = n - 1
\]

\{Compute first column of $(H - a_1 I)(H - a_2 I)$\}

\[
s = H(m, m) + H(n, n)
\]

\[
t = H(m, m) \cdot H(n, n) - H(m, n) \cdot H(n, m)
\]

\[
x = H(1, 1) \cdot H(1, 1) + H(1, 2) \cdot H(2, 1) - s \cdot H(1, 1) + t
\]

\[
y = H(2, 1) \cdot H(1, 1) + H(2, 2) - s
\]

\[
z = H(2, 1) \cdot H(3, 2)
\]

**for** $k = 0:n - 3$

\[
[v, \beta] = \text{house}([x \ y \ z]^T)
\]

\[
q = \max\{1, k\}.
\]

\[
H(k + 1:k + 3, q:n) = (I - \beta vv^T) \cdot H(k + 1:k + 3, q:n)
\]

\[
r = \min\{k + 4, n\}
\]

\[
H(1:r, k + 1:k + 3) = H(1:r, k + 1:k + 3) \cdot (I - \beta vv^T)
\]

\[
x = H(k + 2, k + 1)
\]

\[
y = H(k + 3, k + 1)
\]

**if** $k < n - 3$

\[
z = H(k + 4, k + 1)
\]

**end**

\[
[v, \beta] = \text{house}([x \ y]^T)
\]

\[
H(n - 1:n, n - 2:n) = (I - \beta vv^T) \cdot H(n - 1:n, n - 2:n)
\]

\[
H(1:n, n - 1:n) = H(1:n, n - 1:n) \cdot (I - \beta vv^T)
\]

This algorithm requires $10n^2$ flops. If $Z$ is accumulated into a given orthogonal matrix, an additional $10n^2$ flops are necessary.

**7.5.6 The Overall Process**

Reduction of $A$ to Hessenberg form using Algorithm 7.4.2 and then iteration with Algorithm 7.5.1 to produce the real Schur form is the standard means by which the dense unsymmetric eigenproblem is solved. During the iteration it is necessary to monitor the subdiagonal elements in $H$ in order to spot any possible decoupling. How this is done is illustrated in the following algorithm:
Algorithm 7.5.2 (QR Algorithm) Given \( A \in \mathbb{R}^{n \times n} \) and a tolerance \( \text{tol} \) greater than the unit roundoff, this algorithm computes the real Schur canonical form \( Q^T AQ = T \). If \( Q \) and \( T \) are desired, then \( T \) is stored in \( H \). If only the eigenvalues are desired, then diagonal blocks in \( T \) are stored in the corresponding positions in \( H \).

Use Algorithm 7.4.2 to compute the Hessenberg reduction

\[
H = U_0^T A U_0 \text{ where } U_0 = P_1 \cdots P_{n-2}.
\]

If \( Q \) is desired form \( Q = P_1 \cdots P_{n-2} \). (See §5.1.6.)

until \( q = n \)

Set to zero all subdiagonal elements that satisfy:

\[
|h_{i,i-1}| \leq \text{tol} \cdot (|h_{ii}| + |h_{i-1,i-1}|).
\]

Find the largest nonnegative \( q \) and the smallest non-negative \( p \) such that

\[
H = \begin{bmatrix}
H_{11} & H_{12} & H_{13} \\
0 & H_{22} & H_{23} \\
0 & 0 & H_{33}
\end{bmatrix}
\]

where \( H_{33} \) is upper quasi-triangular and \( H_{22} \) is unreduced.

if \( q < n \)

Perform a Francis QR step on \( H_{22} \): \( H_{22} = Z^T H_{22} Z \).

if \( Q \) is required

\[
Q = Q \cdot \text{diag}(I_p, Z, I_q)
\]

\[
H_{12} = H_{12} Z
\]

\[
H_{23} = Z^T H_{23}
\]

end

end

Upper triangularize all 2-by-2 diagonal blocks in \( H \) that have real eigenvalues and accumulate the transformations (if necessary).

This algorithm requires \( 25n^3 \) flops if \( Q \) and \( T \) are computed. If only the eigenvalues are desired, then \( 10n^3 \) flops are necessary. These flops counts are very approximate and are based on the empirical observation that on average only two Francis iterations are required before the lower 1-by-1 or 2-by-2 decouples.

The roundoff properties of the QR algorithm are what one would expect of any orthogonal matrix technique. The computed real Schur form \( \hat{T} \) is orthogonally similar to a matrix near to \( A \), i.e.,

\[
Q^T (A + E) Q = \hat{T}
\]

where \( Q^T Q = I \) and \( \| E \|_2 \approx u \| A \|_2 \). The computed \( \hat{Q} \) is almost orthogonal in the sense that \( \hat{Q}^T \hat{Q} = I + F \) where \( \| F \|_2 \approx u \).

The order of the eigenvalues along \( \hat{T} \) is somewhat arbitrary. But as we discuss in §7.6, any ordering can be achieved by using a simple procedure for swapping two adjacent diagonal entries.
John Francis and 50 years of QR

John Francis submitted his first QR paper almost 50 years ago in October 1959. By 1962 he had left the NA field. When his algorithm was judged one of the top ten algorithms of the 20th century in 2000 by Jack Dongarra and Francis Sullivan, nobody alive in the mathematics community had ever seen John Francis or knew where or if he lived. Gene Golub and Frank Uhlig independently tracked John Francis down, joined forces, and visited and interviewed him over the last couple of years.

When first contacted, John Francis had no idea about QR’s impact. He is 74 years old now and well. Re QR, he remembers his math and computational work of 50 years ago clearly. John Francis will be the lead-off speaker at a mini symposium, held in his honor, at the 23rd Biennial Conference on Numerical Analysis, June 23rd - 26th 2009 in Glasgow to which everyone is cordially invited.

For a detailed list of invited speakers, see http://www.maths.strath.ac.uk/naconf/minisymposia
- Lessons:
  - we saw how something very sophisticated grew out of something very humble
    
    power method → QR algorithm

- Issues:
  - rapid convergence (use shifts of origin to decouple u. h.)
  - preprocessing: get as close to triangular form as possible.
  - Efficiency in each step:
    $O(n^2)$ instead of $O(n^3)$
  - Keep Arithmetic Real
  - conceptually accomplish everything by multiplying with orthogonal matrices
  - Finding eigenvalues and eigenvectors is vastly (75 times) more expensive than solving a linear system.
    - but same order: $O(n^3)$
Additional Eigenvalue Problems that we will skip:

- $Ax = 2Bx$
  generalized eigenvalue problem
- Finding subsets of evinced vectors
- Symmetric Matrices
- Complex Matrices
- Sparsity