The QR algorithm

Tsung-Ming Huang

Department of Mathematics
National Taiwan Normal University, Taiwan

December 3, 2008

Outline

The power and inverse power methods

Let A be a nondefective matrix and (λ_i, x_i) for $i = 1, \cdots, n$ be a complete set of eigenpairs of A. That is $\{x_1, \cdots, x_n\}$ is linearly independent. Hence, for any $u_0 \neq 0$, $\exists\ \alpha_1, \cdots, \alpha_n$ such that

$$u_0 = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

Now $A^k x_i = \lambda_i^k x_i$, so that

$$A^k u_0 = \alpha_1 \lambda_1^k x_1 + \cdots + \alpha_n \lambda_n^k x_n. \tag{1}$$

If $|\lambda_1| > |\lambda_i|$ for $i \geq 2$ and $\alpha_1 \neq 0$, then

$$\frac{1}{\lambda_1^k} A^k u_0 = \alpha_1 x_1 + (\frac{\lambda_2}{\lambda_1})^k \alpha_2 x_2 + \cdots + \alpha_n (\frac{\lambda_n}{\lambda_1})^k x_n \rightarrow \alpha_1 x_1 \text{ as } k \rightarrow 0.$$

Theorem

Let A have a unique dominant eigenpair (λ_1, x_1) with $x_1^* x_1 = 1$ and $X = \begin{pmatrix} x_1 & X_2 \end{pmatrix}$ be a nonsingular matrix with $X_2^* X_2 = I$ such that

$$X^{-1}AX = \begin{pmatrix} \lambda_1 & 0 \\ 0 & M \end{pmatrix}.$$

Let $u_0 \neq 0$ be decomposed in $u_0 = r_1 x_1 + X_2 c_2$.
Then

$$\sin \angle(x_1, A^k u_0) \leq \frac{|\lambda_1|^{-k} \|M^k\|_2 \|c_2/r_1\|_2}{1 - |\lambda_1|^{-k} \|M^k\|_2 \|c_2/r_1\|_2}.$$

In particular $\forall\ \varepsilon > 0$, $\exists\ \sigma$ such that

$$\sin \angle(x_1, A^k u_0) \leq \frac{\sigma[\rho(M)/|\lambda_1| + \varepsilon]^k}{1 - \sigma[\rho(M)/|\lambda_1| + \varepsilon]^k},$$

where $\rho(M)$ is the spectral radius of M.

Proof: Since

$$u_0 = \alpha_1 x_1 + X_2 c_2 = \begin{pmatrix} x_1 & X_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ c_2 \end{pmatrix} = X \begin{pmatrix} \alpha_1 \\ c_2 \end{pmatrix},$$

it follows that

$$\begin{aligned} X^{-1}A^k u_0 &= X^{-1}A^k X \begin{pmatrix} \alpha_1 \\ c_2 \end{pmatrix} \\ &= (X^{-1}AX)(X^{-1}AX)\cdots(X^{-1}AX) \begin{pmatrix} \alpha_1 \\ c_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1^k & 0 \\ 0 & M^k \end{pmatrix} \begin{pmatrix} \alpha_1 \\ c_2 \end{pmatrix}. \end{aligned}$$

Hence,

$$A^k u_0 = X \begin{pmatrix} \lambda_1^k \alpha_1 \\ M^k c_2 \end{pmatrix} = \alpha_1 \lambda_1^k x_1 + X_2 M^k c_2.$$

Let the columns of Y form an orthonormal basis for the subspace orthogonal to x_1. By Lemma 3.12 in Chapter 1, we have

$$\sin\angle(x_1, A^k u_0) = \frac{\|Y^* A^k u_0\|_2}{\|A^k u_0\|_2} = \frac{\|Y^* X_2 M^k c_2\|_2}{\|\alpha_1\lambda_1^k x_1 + X_2 M^k c_2\|_2}.$$

But

$$\|Y^* X_2 M^k c_2\|_2 \le \|M^k\|_2 \|c_2\|_2$$

and

$$\|\alpha_1\lambda_1^k x_1 + X_2 M^k c_2\|_2 \ge |\alpha_1||\lambda_1^k| - \|M^k\|_2\|c_2\|_2,$$

we get

$$\sin\angle(x_1, A^k u_0) \le \frac{|\lambda_1|^{-k}\|M^k\|_2\|c_2/\alpha_1\|_2}{1 - |\lambda_1|^{-k}\|M^k\|_2\|c_2/\alpha_1\|_2}.$$

By Theorem 2.9 in Chapter 1, $\forall\, \varepsilon > 0$, $\exists\, \hat{\sigma}$ such that

$$\|M^k\|_2 \le \hat{\sigma}(\rho(M) + \varepsilon)^k.$$

Take $\sigma = \hat{\sigma}\|c_2/\alpha_1\|_2$. Then $\forall\ \varepsilon > 0$,

$$\sin \angle(x_1, A^k u_0) \leq \frac{\sigma[\rho(M)/|\lambda_1| + \varepsilon]^k}{1 - \sigma[\rho(M)/|\lambda_1| + \varepsilon]^k}.$$

■

- The error in the eigenvector approximation converges to zero at an asymptotic rate of $[\rho(M)/|\lambda_1|]^k$.
- If A has a complete system of eigenvectors with $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$, then the convergence is as $|\lambda_2/\lambda_1|^k$.

Algorithm (Power Method with 2-norm)

Choose an initial $u \neq 0$ *with* $\|u\|_2 = 1$.
Iterate until convergence
  Compute $v = Au;\ k = \|v\|_2;\ u := v/k$

Theorem

The sequence defined by Algorithm 1 is satisfied

$$\lim_{i\to\infty} k_i = |\lambda_1|$$

$$\lim_{i\to\infty} \varepsilon^i u_i = \frac{x_1}{\|x_1\|}\frac{\alpha_1}{|\alpha_1|}, \quad \text{where } \varepsilon = \frac{|\lambda_1|}{\lambda_1}$$

Proof: It is obvious that

$$u_s = A^s u_0 / \|A^s u_0\|, \quad k_s = \|A^s u_0\| / \|A^{s-1} u_0\|. \tag{2}$$

This follows from ${\lambda_1}^{-s} A^s u_0 \longrightarrow \alpha_1 x_1$ that

$$|\lambda_1|^{-s} \|A^s u_0\| \longrightarrow |\alpha_1| \|x_1\|$$
$$|\lambda_1|^{-s+1} \|A^{s-1} u_0\| \longrightarrow |\alpha_1| \|x_1\|$$

and then

$$|\lambda_1|^{-1} \|A^s u_0\| / \|A^{s-1} u_0\| = |\lambda_1|^{-1} k_s \longrightarrow 1.$$

From (1) follows now for $s \to \infty$

$$\begin{aligned} \varepsilon^s u_s &= \varepsilon^s \frac{A^s u_0}{\|A^s u_0\|} = \frac{\alpha_1 x_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^s x_i}{\|\alpha_1 x_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^s x_i\|} \\ &\to \frac{\alpha_1 x_1}{\|\alpha_1 x_1\|} = \frac{x_1}{\|x_1\|} \frac{\alpha_1}{|\alpha_1|}. \end{aligned}$$

Algorithm (Power Method with Linear Function)

Choose an initial $u \neq 0$.
Iterate until convergence
Compute $v = Au; \ k = \ell(v); \ u := v/k$
where $\ell(v)$, *e.g.* $e_1(v)$ *or* $e_n(v)$, *is a linear functional.*

Theorem

Suppose $\ell(x_1) \neq 0$ *and* $\ell(v_i) \neq 0, i = 1, 2, \ldots,$ *then*

$$\lim_{i\to\infty} k_i = \lambda_1$$

$$\lim_{i\to\infty} u_i = \frac{x_1}{\ell(x_1)}.$$

Proof: As above we show that

$$u_i = A^i u_0 / \ell(A^i u_0), \quad k_i = \ell(A^i u_0)/\ell(A^{i-1} u_0).$$

From (1) we get for $s \to \infty$

$$\lambda_1{}^{-s} \ell(A^s u_0) \longrightarrow \alpha_1 \ell(x_1),$$

$$\lambda_1{}^{-s+1} \ell(A^{s-1} u_0) \longrightarrow \alpha_1 \ell(x_1),$$

thus

$$\lambda_1{}^{-1} k_s \longrightarrow 1.$$

Similarly for $i \longrightarrow \infty$,

$$u_i = \frac{A^i u_0}{\ell(A^i u_0)} = \frac{\alpha_1 x_1 + \sum_{j=2}^n \alpha_j (\frac{\lambda_j}{\lambda_1})^i x_j}{\ell(\alpha_1 x_1 + \sum_{j=2}^n \alpha_j (\frac{\lambda_j}{\lambda_1})^i x_j)} \longrightarrow \frac{\alpha_1 x_1}{\alpha_1 \ell(x_1)}$$

- Note that:

$$\begin{aligned} k_s &= \frac{\ell(A^s u_0)}{\ell(A^{s-1}u_0)} = \lambda_1 \frac{\alpha_1 \ell(x_1) + \sum_{j=2}^n \alpha_j (\frac{\lambda_j}{\lambda_1})^s \ell(x_j)}{\alpha_1 \ell(x_1) + \sum_{j=2}^n \alpha_j (\frac{\lambda_j}{\lambda_1})^{s-1} \ell(x_j)} \\ &= \lambda_1 + O\left(\left| \frac{\lambda_2}{\lambda_1} \right|^{s-1} \right). \end{aligned}$$

That is the convergent rate is $\left|\frac{\lambda_2}{\lambda_1}\right|$.

Theorem

Let $u \neq 0$ and for any μ set $r_\mu = Au - \mu u$. Then $\|r_\mu\|_2$ is minimized when

$$\mu = u^* A u / u^* u.$$

In this case $r_\mu \perp u$.

Proof: W.L.O.G. assume $\|u\|_2 = 1$. Let $\begin{pmatrix} u & U \end{pmatrix}$ be unitary and set

$$\begin{pmatrix} u^* \\ U^* \end{pmatrix} A \begin{pmatrix} u & U \end{pmatrix} \equiv \begin{pmatrix} \nu & h^* \\ g & B \end{pmatrix} = \begin{pmatrix} u^* A u & u^* A U \\ U^* A u & U^* A U \end{pmatrix}.$$

Then

$$\begin{aligned}
\begin{pmatrix} u^* \\ U^* \end{pmatrix} r_\mu &= \begin{pmatrix} u^* \\ U^* \end{pmatrix} Au - \mu \begin{pmatrix} u^* \\ U^* \end{pmatrix} u \\
&= \begin{pmatrix} u^* \\ U^* \end{pmatrix} A \begin{pmatrix} u & U \end{pmatrix} \begin{pmatrix} u^* \\ U^* \end{pmatrix} u - \mu \begin{pmatrix} u^* \\ U^* \end{pmatrix} u \\
&= \begin{pmatrix} \nu & h^* \\ g & B \end{pmatrix} \begin{pmatrix} u^* \\ U^* \end{pmatrix} u - \mu \begin{pmatrix} u^* \\ U^* \end{pmatrix} u \\
&= \begin{pmatrix} \nu & h^* \\ g & B \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \mu \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \nu - \mu \\ g \end{pmatrix}.
\end{aligned}$$

It follows that

$$\|r_\mu\|_2^2 = \| \begin{pmatrix} u^* \\ U^* \end{pmatrix} r_\mu \|_2^2 = \| \begin{pmatrix} \nu - \mu \\ g \end{pmatrix} \|_2^2 = |\nu - \mu|^2 + \|g\|_2^2.$$

Hence

$$\min_{\mu} \|r_\mu\|_2 = \|g\|_2 = \|r_\nu\|_2.$$

That is $\mu = \nu = u^* A u$. On the other hand, since

$$u^* r_\mu = u^*(Au - \mu u) = u^* A u - \mu = 0,$$

it implies that $r_\mu \perp u$. ■

Definition (Rayleigh quotient)

Let u and v be vectors with $v^* u \neq 0$. Then $v^* A u / v^* u$ is called a Rayleigh quotient.

If u or v is an eigenvector corresponding to an eigenvalue λ of A, then

$$\frac{v^* A u}{v^* u} = \lambda \frac{v^* u}{v^* u} = \lambda.$$

Therefore, $u_k^* A u_k / u_k^* u_k$ provide a sequence of approximation to λ in the power method.

Inverse power method

Goal

Find the eigenvalue of A that is in a given region or closest to a certain scalar σ and the corresponding eigenvector.

Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of A. Suppose λ_1 is simple and $\sigma \approx \lambda_1$. Then

$$\mu_1 = \frac{1}{\lambda_1 - \sigma}, \mu_2 = \frac{1}{\lambda_2 - \sigma}, \cdots, \mu_n = \frac{1}{\lambda_n - \sigma}$$

are eigenvalues of $(A - \sigma I)^{-1}$ and $\mu_1 \to \infty$ as $\sigma \to \lambda_1$. Thus we transform λ_1 into a dominant eigenvalue μ_1.

The inverse power method is simply the power method applied to $(A - \sigma I)^{-1}$.

Let

$$y = (A - \sigma I)^{-1} x \text{ and } \hat{x} = y / \|y\|_2.$$

It holds that

$$(A - \sigma I)\hat{x} = \frac{x}{\|y\|_2} \equiv w.$$

Set

$$\rho = \hat{x}^*(A - \sigma I)\hat{x} = \hat{x}^* w.$$

Then

$$r = [A - (\sigma + \rho)I]\hat{x} = (A - \sigma I)\hat{x} - \rho\hat{x} = w - \rho\hat{x}.$$

Algorithm (Inverse power method with a fixed shift)

Choose an initial $u_0 \neq 0$.
For $i = 0, 1, 2, \ldots$
  Compute $v_{i+1} = (A - \sigma I)^{-1} u_i$ *and* $k_{i+1} = \ell(v_{i+1})$.
  Set $u_{i+1} = v_{i+1} / k_{i+1}$

- The convergence of Algorithm 3 is $|\frac{\lambda_1-\sigma}{\lambda_2-\sigma}|$ whenever λ_1 and λ_2 are the closest and the second closest eigenvalues to σ.
- Algorithm 3 is linearly convergent.

Algorithm (Inverse power method with variant shifts)

Choose an initial $u_0 \neq 0$.
Given $\sigma_0 = \sigma$.
For $i = 0, 1, 2, \ldots$
  Compute $v_{i+1} = (A - \sigma_i I)^{-1} u_i$ *and* $k_{i+1} = \ell(v_{i+1})$.
  Set $u_{i+1} = v_{i+1}/k_{i+1}$ *and* $\sigma_{i+1} = \sigma_i + 1/k_{i+1}$.

- Above algorithm is locally quadratic convergent.

Connection with Newton method

Consider the nonlinear equations:

$$F\left(\begin{bmatrix} u \\ \lambda \end{bmatrix}\right) \equiv \begin{bmatrix} Au - \lambda u \\ \ell^T u - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (3)$$

Newton method for (3): for $i = 0, 1, 2, \ldots$

$$\begin{bmatrix} u_{i+1} \\ \lambda_{i+1} \end{bmatrix} = \begin{bmatrix} u_i \\ \lambda_i \end{bmatrix} - \left[F'\left(\begin{bmatrix} u_i \\ \lambda_i \end{bmatrix}\right)\right]^{-1} F\left(\begin{bmatrix} u_i \\ \lambda_i \end{bmatrix}\right).$$

Since

$$F'\left(\begin{bmatrix} u \\ \lambda \end{bmatrix}\right) = \begin{bmatrix} A - \lambda I & -u \\ \ell^T & 0 \end{bmatrix},$$

the Newton method can be rewritten by component-wise

$$(A - \lambda_i)u_{i+1} = (\lambda_{i+1} - \lambda_i)u_i \quad (4)$$

$$\ell^T u_{i+1} = 1. \quad (5)$$

Let

$$v_{i+1} = \frac{u_{i+1}}{\lambda_{i+1} - \lambda_i}.$$

Substituting v_{i+1} into (4), we get

$$(A - \lambda_i I)v_{i+1} = u_i.$$

By equation (5), we have

$$k_{i+1} = \ell(v_{i+1}) = \frac{\ell(u_{i+1})}{\lambda_{i+1} - \lambda_i} = \frac{1}{\lambda_{i+1} - \lambda_i}.$$

It follows that

$$\lambda_{i+1} = \lambda_i + \frac{1}{k_{i+1}}.$$

Hence the Newton's iterations (4) and (5) are identified with Algorithm 4.

Algorithm (Inverse power method with Rayleigh Quotient)

Choose an initial $u_0 \neq 0$ *with* $\|u_0\|_2 = 1$.
Compute $\sigma_0 = u_0^T A u_0$.
For $i = 0, 1, 2, \ldots$
  Compute $v_{i+1} = (A - \sigma_i I)^{-1} u_i$.
  Set $u_{i+1} = v_{i+1}/\|v_{i+1}\|_2$ *and* $\sigma_{i+1} = u_{i+1}^T A u_{i+1}$.

- For symmetric A, Algorithm 5 is cubically convergent.

The explicitly shift QR algorithm

The QR algorithm is an iterative method for reducing a matrix A to triangular form by unitary similarity transformations.

Algorithm (explicitly shift QR algorithm)

Set $A_0 = A$.
For $k = 0, 1, 2, \cdots$
    Choose a shift σ_k*;*
    Factor $A_k - \sigma_k I = Q_k R_k$, *where* Q_k *is orthogonal and* R_k *is upper triangular;*
    $A_{k+1} = R_k Q_k + \sigma_k I$*;*
end for

Since

$$A_k - \sigma_k I = Q_k R_k \Longrightarrow R_k = Q_k^*(A_k - \sigma_k I),$$

it holds that

$$\begin{aligned} A_{k+1} &= R_k Q_k + \sigma_k I \\ &= Q_k^*(A_k - \sigma_k I) Q_k + \sigma_k I \\ &= Q_k^* A_k Q_k \end{aligned}$$

The algorithm is a variant of the power method.

Let $Q = \begin{pmatrix} \hat{Q} & q \end{pmatrix}$ be unitary and write

$$Q^*AQ = \begin{pmatrix} \hat{Q}^*A\hat{Q} & \hat{Q}^*Aq \\ q^*A\hat{Q} & q^*Aq \end{pmatrix} \equiv \begin{pmatrix} \hat{B} & \hat{h} \\ \hat{g}^* & \hat{\mu} \end{pmatrix}.$$

If (λ, q) is a left eigenpair of A, then

$$\hat{g}^* = q^*A\hat{Q} = \lambda q^*\hat{Q} = 0 \text{ and } \hat{\mu} = q^*Aq = \lambda q^*q = \lambda.$$

That is

$$Q^*AQ = \begin{pmatrix} \hat{B} & \hat{h} \\ 0 & \lambda \end{pmatrix}.$$

But it is not an effective computational procedure because it requires q is an eigenvector of A.

Let q be an approximate left eigenvector of A with

$$q^* q = 1, \ \hat{\mu} = q^* A q \ \text{ and } \ r = q^* A - \hat{\mu} q^*.$$

Then

$$\begin{aligned} r \begin{pmatrix} \hat{Q} & q \end{pmatrix} &= (q^* A - \hat{\mu} q^*) \begin{pmatrix} \hat{Q} & q \end{pmatrix} \\ &= \begin{pmatrix} q^* A \hat{Q} - \hat{\mu} q^* \hat{Q} & q^* A q - \hat{\mu} q^* q \end{pmatrix} \\ &= \begin{pmatrix} q^* A \hat{Q} & 0 \end{pmatrix} = \begin{pmatrix} \hat{g}^* & 0 \end{pmatrix}. \end{aligned}$$

Therefore,

$$\|\hat{g}^*\|_2 = \|r \begin{pmatrix} \hat{Q} & q \end{pmatrix}\|_2 = \|r\|_2.$$

The QR algorithm implicitly chooses q to be a vector produced by the inverse power method with shift σ.

Write the QR factorization of $A - \sigma I$ as

$$\begin{pmatrix} \hat{Q}^* \\ q^* \end{pmatrix} (A - \sigma I) = R \equiv \begin{pmatrix} \hat{R}^* \\ r^* \end{pmatrix}.$$

It holds that

$$q^*(A - \sigma I) = r^* = r_{nn} e_n^T \Rightarrow q^* = r_{nn} e_n^T (A - \sigma I)^{-1} \tag{6}$$

Hence, the last column of Q generated by the QR algorithm is the result of the inverse power method with shift σ applied to e_n^T.

Question

How to choose shift σ?

Let

$$A = \begin{pmatrix} B & h \\ g^* & \mu \end{pmatrix}.$$

Then

$$e_n^T A e_n = \mu \text{ and } e_n^T A - \mu e_n = \begin{pmatrix} g^* & \mu \end{pmatrix} - \mu e_n = \begin{pmatrix} g^* & 0 \end{pmatrix}.$$

- If we take (μ, e_n) to be an approximate left eigenvector of A, then the corresponding residual norm is $\|g\|_2$.
- If g is small, then μ should approximate an eigenvalue of A and choose $\sigma = \mu = e_n^T A e_n$ (Rayleigh quotient shift).

Question

Why the QR algorithm converges?

Let

$$\begin{aligned} A - \sigma I &\equiv \begin{pmatrix} B - \sigma I & h \\ g^* & \mu - \sigma \end{pmatrix} \\ &= QR \equiv \begin{pmatrix} P & f \\ e^* & \pi \end{pmatrix} \begin{pmatrix} S & r \\ 0 & \rho \end{pmatrix} \end{aligned} \quad (7)$$

be the QR factorization of $A - \sigma I$. Take

$$\hat{A} \equiv \begin{pmatrix} \hat{B} & \hat{h} \\ \hat{g}^* & \hat{\mu} \end{pmatrix} = RQ + \sigma I. \quad (8)$$

Since Q is unitary, we have

$$\|e\|_2^2 + \pi^2 = \|f\|_2^2 + \pi^2 = 1$$

which implies that

$$\|e\|_2 = \|f\|_2 \text{ and } |\pi| \leq 1.$$

From (7), we have

$$g^* = e^* S.$$

Assume S is nonsingular and $\kappa = \|S^{-1}\|_2$, then

$$\|e\|_2 \leq \kappa \|g\|_2.$$

Since

$$R \equiv \begin{pmatrix} S & r \\ 0 & \rho \end{pmatrix} = Q^*(A - \sigma I) \equiv \begin{pmatrix} P^* & e \\ f^* & \bar{\pi} \end{pmatrix} \begin{pmatrix} B - \sigma I & h \\ g^* & \mu - \sigma \end{pmatrix},$$

it implies that

$$\rho = f^* h + \bar{\pi}(\mu - \sigma)$$

and then

$$\begin{aligned} |\rho| &\leq \|f\| \|h\| + |\pi| |\mu - \sigma| = \|e\|_2 \|h\|_2 + |\pi| |\mu - \sigma| \\ &\leq \kappa \|g\|_2 \|h\|_2 + |\mu - \sigma|. \end{aligned}$$

From (8), we have

$$\hat{g}^* = \rho e^*$$

which implies that

$$\|\hat{g}\|_2 \leq |\rho|\|e\|_2 \leq |\rho|\kappa\|g\|_2 \leq \kappa^2\|h\|_2\|g\|_2^2 + \kappa|\mu - \sigma|\|g\|_2.$$

Consequently,

$$\|g_{j+1}\|_2 \leq \kappa_j^2\|h_j\|_2\|g_j\|_2^2 + \kappa_j|\mu_j - \sigma_j|\|g_j\|_2.$$

If g_0 is sufficiently small and μ_0 is sufficiently near a simple eigenvalue λ, then $g_j \to 0$ and $\mu_j \to \lambda$.
Assume $\exists\ \eta$ and κ such that

$$\|h_j\|_2 \leq \eta \ \text{ and } \ \kappa_j = \|S_j^{-1}\|_2 \leq \kappa.$$

Take the Rayleigh quotient shift $\sigma_j = \mu_j$. Then

$$\|g_{j+1}\|_2 \leq \kappa^2 \eta \|g_j\|_2^2,$$

which means that $\|g_j\|_2$ converges at least quadratically to zero. If A_0 is Hermitian, then A_k is also Hermitian. It holds that

$$h_j = g_j$$

and then

$$\|g_{j+1}\|_2 \leq \kappa^2 \|g_j\|_2^3.$$

Therefore, the convergent rate is cubic.

The unshifted QR algorithm

QR algorithm

$$A_{k+1} = Q_k^* A_k Q_k$$

or

$$A_{k+1} = Q_k^* Q_{k-1}^* \cdots Q_0 A_0 Q_0 \cdots Q_{k-1} Q_k$$

for $k = 0, 1, 2, \cdots$.

Let

$$\hat{Q}_k = Q_0 \cdots Q_{k-1} Q_k.$$

Then

$$A_{k+1} = \hat{Q}_k^* A_0 \hat{Q}_k.$$

Theorem

Let $Q_0, \cdots, Q_k$ *and* $R_0, \cdots, R_k$ *be the orthogonal and triangular matrices generated by the QR algorithm with shifts* $\sigma_0, \cdots, \sigma_k$ *starting with* A*. Let*

$$\hat{Q}_k = Q_0 \cdots Q_k \text{ and } \hat{R}_k = R_0 \cdots R_k.$$

Then

$$\hat{Q}_k \hat{R}_k = (A - \sigma_0 I) \cdots (A - \sigma_k I).$$

Proof: Since

$$\begin{aligned} R_k &= (A_{k+1} - \sigma_k I) Q_k^* \\ &= \hat{Q}_k^* (A - \sigma_k I) \hat{Q}_k Q_k^* \\ &= \hat{Q}_k^* (A - \sigma_k I) \hat{Q}_{k-1}, \end{aligned}$$

it follows that

$$\hat{R}_k = R_k \hat{R}_{k-1} = \hat{Q}_k^* (A - \sigma_k I) \hat{Q}_{k-1} \hat{R}_{k-1}$$

and

$$\hat{Q}_k \hat{R}_k = (A - \sigma_k I) \hat{Q}_{k-1} \hat{R}_{k-1}.$$

By induction on $\hat{Q}_{k-1} \hat{R}_{k-1}$, we have

$$\hat{Q}_k \hat{R}_k = (A - \sigma_k I) \cdots (A - \sigma_0 I).$$

■

If $\sigma_k = 0$ for $k = 0, 1, 2, \cdots$, then $\hat{Q}_k \hat{R}_k = A^{k+1}$ and

$$\hat{r}_{11}^{(k)} \hat{q}_1^{(k)} = \hat{Q}_k \hat{R}_k e_1 = A^{k+1} e_1.$$

This implies that the first column of $\hat{Q}_k$ is the normalized result of applying $k+1$ iterations of the power method to e_1.

Hence, $\hat{q}_1^{(k)}$ approaches the dominant eigenvector of A, i.e., if

$$A_k = \hat{Q}_k^* A Q_k = \begin{pmatrix} \mu_k & h_k^* \\ g_k & B_k \end{pmatrix},$$

then $g_k \to 0$ and $\mu_k \to \lambda_1$, where λ_1 is the dominant eigenvalue of A.

Theorem

Let

$$X^{-1} A X = \Lambda \equiv diag(\lambda_1, \cdots, \lambda_n)$$

where $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$. *Suppose* X^{-1} *has an LU factorization* $X^{-1} = LU$, *where* L *is unit lower triangular, and let* $X = QR$ *be the QR factorization of* X. *If* A^k *has the QR factorization* $A^k = \hat{Q}_k \hat{R}_k$, *then* $\exists$ *diagonal matrices* D_k *with* $|D_k| = I$ *such that* $\hat{Q}_k D_k \longrightarrow Q$.

Proof: By the assumptions, we get

$$A^k = X\Lambda^k X^{-1} = QR\Lambda^k LU = QR(\Lambda^k L\Lambda^{-k})(\Lambda^k U).$$

Since

$$(\Lambda^k L\Lambda^{-k})_{ij} = \ell_{ij}(\lambda_i/\lambda_j)^k \to 0 \;\text{ for }\; i > j,$$

it holds that

$$\Lambda^k L\Lambda^{-k} \to I \;\text{ as }\; k \to \infty.$$

Let

$$\Lambda^k L\Lambda^{-k} = I + E_k,$$

where $E_k \to 0$ as $k \to \infty$. Then

$$A^k = QR(I + E_k)(\Lambda^k U) = Q(I + RE_kR^{-1})(R\Lambda^k U).$$

Let

$$I + RE_kR^{-1} = \bar{Q}_k\bar{R}_k$$

be the QR factorization of $I + RE_kR^{-1}$. Then

$$A^k = (Q\bar{Q}_k)(\bar{R}_kR\Lambda^kU).$$

Since

$$I + RE_kR^{-1} \to I \text{ as } k \to \infty,$$

we have

$$\bar{Q}_k \to I \text{ as } k \to \infty.$$

Let the diagonals of $\bar{R}_kR\Lambda^kU$ be $\delta_1, \cdots, \delta_m$ and set

$$D_k = diag(\bar{\delta}_1/|\delta_1|, \cdots, \bar{\delta}_n/\delta_n).$$

Then $A^k = (Q\bar{Q}_kD_k^{-1})(D_k\bar{R}_kR\Lambda^kU) = \hat{Q}_k\hat{R}_k$.

Since the diagonals of $D_k\bar{R}_k R\Lambda^k U$ and $\hat{R}_k$ are positive, by the uniqueness of the QR factorization

$$\hat{Q}_k = Q\bar{Q}_k D_k^{-1},$$

which implies that

$$\hat{Q}_k D_k = Q\bar{Q}_k \to Q \text{ as } k \to \infty.$$

■

Remark:

(i) Since $X^{-1}AX = \Lambda \equiv diag(\lambda_1, \cdots, \lambda_n)$, we have

$$A = X\Lambda X^{-1} = (QR)\Lambda(QR)^{-1} = Q(R\Lambda R^{-1})Q^* \equiv QTQ^*$$

which is a Schur decomposition of A. Therefore, the column of $\hat{Q}_k D_k$ converge to the Schur vector of A and $A_k = \hat{Q}_k^* A\hat{Q}_k$ converges to the triangular factor of the Schur decomposition of A.

(ii) Write

$$R(\Lambda^k L \Lambda^{-k}) = \begin{pmatrix} R_{11} & r_{1,i} & R_{1,i+1} \\ 0 & r_{ii} & r^*_{i,i+1} \\ 0 & 0 & R_{i+1,i+1} \end{pmatrix} \begin{pmatrix} L_{11}^{(k)} & 0 & 0 \\ \ell_{i,1}^{(k)*} & 1 & 0 \\ L_{i+1,1}^{(k)} & \ell_{i+1,i}^{(k)} & L_{i+1,i+1}^{(k)} \end{pmatrix}.$$

If $\ell_{i,1}^{(k)*}$, $L_{i+1,1}^{(k)}$ and $\ell_{i+1,i}^{(k)}$ are zeros, then

$$R(\Lambda^k L \Lambda^{-k}) = \begin{pmatrix} R_{11}L_{11}^{(k)} & r_{1,i} & R_{1,i+1}L_{i+1,i+1} \\ 0 & r_{i,i} & r^*_{i,i+1}L_{i+1,i+1} \\ 0 & 0 & R_{i+1,i+1}L_{i+1,i+1} \end{pmatrix}$$

and

$$\begin{aligned} I + RE_kR^{-1} &= R(I+E_k)R^{-1} = R(\Lambda^k L \Lambda^{-k})R^{-1} \\ &= \begin{pmatrix} G_{11} & g_{1,i} & G_{1,i+1} \\ 0 & g_{ii} & g^*_{i,i+1} \\ 0 & 0 & G_{i+1,i+1} \end{pmatrix} \\ &= \bar{Q}_k \bar{R}_k \quad \sim \quad \text{QR factorization} \end{aligned}$$

which implies that

$$\bar{Q}_k = diag(\bar{Q}_{11}^k, w, \bar{Q}_{i+1,i+1}^k)$$

and

$$\begin{aligned} A_k &= \hat{Q}_k^* A \hat{Q}_k = \bar{Q}_k^* Q^* A Q \bar{Q}_k = \bar{Q}_k^* T \bar{Q}_k \\ &= \begin{pmatrix} A_{11}^{(k)} & a_{1,i}^{(k)} & A_{1,i+1}^{(k)} \\ 0 & \lambda_i & A_{i,i+1}^{(k)} \\ 0 & 0 & A_{i+1,i+1}^{(k)} \end{pmatrix}. \end{aligned}$$

Therefore, A_k decouples at its ith diagonal element.
The rate of convergence is at least as fast as the approach of $\max\{|\lambda_i/\lambda_{i-1}|, |\lambda_{i+1}/\lambda_i|\}^k$ to zero.

Definition

A Householder transformation or elementary reflector is a matrix of

$$H = I - uu^*$$

where $\|u\|_2 = \sqrt{2}$.

Note that H is Hermitian and unitary.

Theorem

Let x be a vector such that $\|x\|_2 = 1$ and x_1 is real and nonnegative. Let

$$u = (x + e_1)/\sqrt{1 + x_1}.$$

Then

$$Hx = (I - uu^*)x = -e_1.$$

Proof:

$$
\begin{aligned}
I - uu^*x &= x - (u^*x)u = x - \frac{x^*x + x_1}{\sqrt{1+x_1}} \cdot \frac{x+e_1}{\sqrt{1+x_1}} \\
&= x - (x + e_1) = -e_1
\end{aligned}
$$

■

Theorem

Let x be a vector with $x_1 \neq 0$. Let

$$
u = \frac{\rho \frac{x}{\|x\|_2} + e_1}{\sqrt{1 + \rho \frac{x_1}{\|x\|_2}}},
$$

where $\rho = \bar{x}_1/|x_1|$. Then

$$
Hx = -\bar{\rho}\|x\|_2 e_1.
$$

Proof: Since

$$\begin{aligned} & [\bar{\rho}x^*/\|x\|_2 + e_1^T][\rho x/\|x\|_2 + e_1] \\ = & \bar{\rho}\rho + \rho x_1/\|x\|_2 + \bar{\rho}\bar{x}_1/\|x\|_2 + 1 \\ = & 2[1 + \rho x_1/\|x\|_2], \end{aligned}$$

it follows that

$$u^*u = 2 \quad \Rightarrow \quad \|u\|_2 = \sqrt{2}$$

and

$$u^*x = \frac{\bar{\rho}\|x\|_2 + x_1}{\sqrt{1 + \rho\frac{x_1}{\|x\|_2}}}.$$

Hence,

$$
\begin{aligned}
Hx &= x - (u^*x)u = x - \frac{\bar{\rho}\|x\|_2 + x_1}{\sqrt{1+\rho\frac{x_1}{\|x\|_2}}} \frac{\rho\frac{x}{\|x\|_2} + e_1}{\sqrt{1+\rho\frac{x_1}{\|x\|_2}}} \\
&= \left[1 - \frac{(\bar{\rho}\|x\|_2 + x_1)\frac{\rho}{\|x\|_2}}{1+\rho\frac{x_1}{\|x\|_2}}\right] x - \frac{\bar{\rho}\|x\|_2 + x_1}{1+\rho\frac{x_1}{\|x\|_2}} e_1 \\
&= -\frac{\bar{\rho}\|x\|_2 + x_1}{1+\rho\frac{x_1}{\|x\|_2}} e_1 \\
&= -\bar{\rho}\|x\|_2 e_1.
\end{aligned}
$$

■

- Reduction to Hessenberg form

Take

$$A = \begin{pmatrix} \alpha_{11} & a_{12}^* \\ a_{21} & A_{22} \end{pmatrix}.$$

Let $\hat{H}_1$ be a Householder transformation such that

$$\hat{H}_1 a_{21} = v_1 e_1.$$

Set $H_1 = diag(1, \hat{H}_1)$. Then

$$H_1 A H_1 = \begin{pmatrix} \alpha_{11} & a_{12}^* \hat{H}_1 \\ \hat{H}_1 a_{21} & \hat{H}_1 A_{22} \hat{H}_1 \end{pmatrix} = \begin{pmatrix} \alpha_{11} & a_{12}^* \hat{H}_1 \\ v_1 e_1 & \hat{H}_1 A_{22} \hat{H}_1 \end{pmatrix}$$

For the general step, suppose $H_1, \cdots, H_{k-1}$ are Householder transformation such that

$$H_{k-1} \cdots H_1 A H_1 \cdots H_{k-1} = \begin{pmatrix} A_{11} & a_{1,k} & A_{1,k+1} \\ 0 & \alpha_{kk} & a_{k,k+1}^* \\ 0 & a_{k+1,k} & A_{k+1,k+1} \end{pmatrix},$$

where A_{11} is a Hessenberg matrix of order $k-1$. Let $\hat{H}_k$ be a Householder transformation such that

$$\hat{H}_k a_{k+1,k} = v_k e_1.$$

Set $H_k = diag(I_k, \hat{H}_k)$, then

$$H_k H_{k-1} \cdots H_1 A H_1 \cdots H_{k-1} H_k = \begin{pmatrix} A_{11} & a_{1,k} & A_{1,k+1}\hat{H}_k \\ 0 & \alpha_{kk} & a_{k,k+1}^* \hat{H}_k \\ 0 & v_k e_1 & \hat{H}_k A_{k+1,k+1} \hat{H}_k \end{pmatrix}.$$

Hessenberg form

$$
\begin{aligned}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
{\color{red}\times} & \times & \times & \times & \times \\
{\color{red}\times} & \times & \times & \times & \times \\
{\color{red}\times} & \times & \times & \times & \times \\
{\color{red}\times} & \times & \times & \times & \times
\end{pmatrix}
& \xrightarrow{H_1}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & {\color{red}\times} & \times & \times & \times \\
0 & {\color{red}\times} & \times & \times & \times \\
0 & {\color{red}\times} & \times & \times & \times
\end{pmatrix} \\
& \xrightarrow{H_2}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & {\color{red}\times} & \times & \times \\
0 & 0 & {\color{red}\times} & \times & \times
\end{pmatrix} \\
& \xrightarrow{H_3}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
\end{aligned}
$$

Definition (Givens rotation)

A plane rotation (also called a Givens rotation) is a matrix of the form

$$P = \begin{pmatrix} c & s \\ -\bar{s} & \bar{c} \end{pmatrix}$$

where $|c|^2 + |s|^2 = 1$.

Given $a \neq 0$ and b, set

$$v = \sqrt{|a|^2 + |b|^2}, \; c = |a|/v \;\text{ and }\; s = \frac{a}{|a|} \cdot \frac{\bar{b}}{v},$$

then

$$\begin{pmatrix} c & s \\ -\bar{s} & \bar{c} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} v\dfrac{a}{|a|} \\ 0 \end{pmatrix}.$$

Let

$$P_{ij} = \begin{pmatrix} I_{i-1} & & & & \\ & c & & s & \\ & & I_{j-i-1} & & \\ & -\bar{s} & & \bar{c} & \\ & & & & I_{n-j} \end{pmatrix}.$$

$$\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} \xrightarrow{P_{12}} \begin{pmatrix} + & + & + & + \\ 0 & + & + & + \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} \xrightarrow{P_{13}} \begin{pmatrix} + & + & + & + \\ 0 & \times & \times & \times \\ 0 & + & + & + \\ \times & \times & \times & \times \end{pmatrix}$$

$$\xrightarrow{P_{14}} \begin{pmatrix} + & + & + & + \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & + & + & + \end{pmatrix}$$

Hessenberg form

$$
\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix}
\xrightarrow{P_{12}}
\begin{pmatrix} + & 0 & \times & \times \\ + & + & \times & \times \\ + & + & \times & \times \\ + & + & \times & \times \end{pmatrix}
$$

$$
\xrightarrow{P_{13}}
\begin{pmatrix} + & 0 & 0 & \times \\ + & \times & + & \times \\ + & \times & + & \times \\ + & \times & + & \times \end{pmatrix}
$$

$$
\xrightarrow{P_{14}}
\begin{pmatrix} + & 0 & 0 & 0 \\ + & \times & \times & + \\ + & \times & \times & + \\ + & \times & \times & + \end{pmatrix}
$$

(i) Reduce a matrix to Hessenberg form by QR factorization.

$$
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times
\end{pmatrix}
\xrightarrow{Q_1 A Q_1^*}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & \times & \times & \times & \times
\end{pmatrix}
$$

$$
\xrightarrow{Q_2 A Q_2^*}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & \times & \times & \times
\end{pmatrix}
$$

$$
\xrightarrow{Q_3 A Q_3^*}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
$$

: upper He

(ii) Reduce upper Hessenberg matrix to upper triangular form by Givens rotations

$$
\begin{pmatrix}
\times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
\xrightarrow{P_{12}A_1}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
$$

$$
\xrightarrow{P_{23}A_2}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
\xrightarrow{P_{34}A_3}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times \\
0 & 0 & 0 & \times & \times
\end{pmatrix}
$$

$$
\xrightarrow{P_{45}A_4}
\begin{pmatrix}
\times & \times & \times & \times & \times \\
0 & \times & \times & \times & \times \\
0 & 0 & \times & \times & \times \\
0 & 0 & 0 & \times & \times \\
0 & 0 & 0 & 0 & \times
\end{pmatrix}
= T \quad \text{(upper triangular)}
$$

Hessenberg form

$$\begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{pmatrix} \xrightarrow{A_1 P_{12}^*} \begin{pmatrix} + & + & \times & \times & \times \\ + & + & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{pmatrix}$$

$$\xrightarrow{A_2 P_{23}^*} \begin{pmatrix} \times & + & + & \times & \times \\ \times & + & + & \times & \times \\ 0 & + & + & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{pmatrix} \xrightarrow{A_3 P_{34}^*} \begin{pmatrix} \times & \times & + & + & \times \\ \times & \times & + & + & \times \\ 0 & \times & + & + & \times \\ 0 & 0 & + & + & \times \\ 0 & 0 & 0 & 0 & \times \end{pmatrix}$$

$$\xrightarrow{A_4 P_{45}^*} \begin{pmatrix} \times & \times & \times & + & + \\ \times & \times & \times & + & + \\ 0 & \times & \times & + & + \\ 0 & 0 & \times & + & + \\ 0 & 0 & 0 & + & + \end{pmatrix} = H \quad \text{(upper Hessenberg)}$$

A practical algorithm for reducing an upper Hessenberg matrix H to Schur form:

1. If the shifted QR algorithm is applied to H, then $h_{n,n-1}$ will tend rapidly to zero and other subdiagonal elements may also tend to zero, slowly.
2. If $h_{i,i-1} \approx 0$, then deflate the matrix to save computation.
   - How to decide $h_{i,i-1}$ to be negligible?
   - If
     $$|h_{i+1,i}| \leq \varepsilon \|A\|_F$$
     for a small number ε, then $h_{i+1,i}$ is negligible.
   - Let Q be an orthogonal matrix such that
     $$H = Q^* A Q \equiv [h_{ij}]$$
     is upper Hessenberg. Let
     $$\tilde{H} = H - h_{i+1,i} e_{i+1} e_i^T \qquad \sim \text{ deflated matrix}$$

Set

$$E = Q(h_{i+1,i}e_{i+1}e_i^T)Q^*.$$

Then

$$\tilde{H} = Q^*(A - E)Q.$$

If $|h_{i+1,i}| \leq \varepsilon\|A\|_F$, then

$$\|E\|_F = \|Q(h_{i+1,i}e_{i+1}e_i^T)Q^*\|_F = |h_{i+1,i}| \leq \varepsilon\|A\|_F$$

or

$$\frac{\|E\|_F}{\|A\|_F} \leq \varepsilon.$$

When ε equals the rounding unit ε_M, the perturbation E is of a size with the perturbation due to rounding the elements of A.

The Wilkinson shift

1. The Rayleigh-quotient shift $\sigma = h_{n,n}$
   $\Rightarrow$ local quadratic convergence to simple
2. If H is real
   $\Rightarrow$ Rayleigh-quotient shift is also real
   $\Rightarrow$ can not approximate a complex eigenvalue
3. The Wilkinson shift μ :
   If λ_1, λ_2 are eigenvalues of $\begin{pmatrix} h_{n-1,n-1} & h_{n-1,n} \\ h_{n,n-1} & h_{n,n} \end{pmatrix}$ with
   $|\lambda_1 - h_{n,n}| \leq |\lambda_2 - h_{n,n}|$, then $\mu = \lambda_1$.

Algorithm

do $k = 1, 2, \cdots$

compute Wilkinson shift μ_k

Reduce upper Hessenberg $H_k - \mu_k I$ to upper triangular T_k :

$$P_{n-1,n}^{(k)} \cdots P_{12}^{(k)} (H_k - \mu_k I) = T_k;$$

compute

$$H_{k+1} = T_k P_{12}^{(k)*} \cdots P_{n-1,n}^{(k)*} + \mu_k I;$$

end do

$\Rightarrow$ Schur form of A $\Rightarrow$ eigenvalues of A.

Question

How to get eigenvectors of A?

If $A = QTQ^*$ is the Schur decomposition of A and X is the matrix of right eigenvectors of T, then QX is the matrix of right eigenvalues of A.

If

$$T = \begin{pmatrix} T_{11} & t_{1,k} & t_{1,k+1} \\ 0 & \tau_{kk} & t^*_{k,k+1} \\ 0 & 0 & T_{k+1,k+1} \end{pmatrix}$$

and τ_{kk} is a simple eigenvalue of T, then

$$\begin{pmatrix} -(T_{11} - \tau_{kk} I)^{-1} t_{1,k} \\ 1 \\ 0 \end{pmatrix}$$

is an eigenvector of T and

$$\begin{pmatrix} 0 & 1 & -t^*_{k,k+1}(T_{k+1,k+1} - \tau_{kk} I)^{-1} \end{pmatrix}$$

is a left eigenvector of T corresponding to τ_{kk}.

The implicity shifted QR algorithm

Theorem (Real Schur form)

Let A be real of order n. Then $\exists$ an orthogonal matrix U such that

$$U^T A U = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1k} \\ 0 & T_{22} & \cdots & T_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & T_{kk} \end{pmatrix} \sim \textit{quasi-triangular}$$

The diagonal blocks of T are of order one or two. The blocks of order one contain the real eigenvalue of A. The block of order two contain the pairs of complex conjugate eigenvalue of A. The blocks can be made to appear in any order.

Proof: Let (λ, x) be a complex eigenpair with $\lambda = \mu + i\nu$ and $x = y + iz$. That is

$$2y = x + \bar{x}, \quad 2zi = x - \bar{x}$$

and

$$\begin{aligned} Ay &= \frac{1}{2}[\lambda x + \bar{\lambda}\bar{x}] \\ &= \frac{1}{2}[(\mu y - \nu z) + i(\mu z + \nu y) + (\mu y - \nu z) - i(\nu y + \mu z)] \\ &= \mu y - \nu z. \end{aligned} \tag{9}$$

Similarly,

$$Az = \frac{1}{2i}[\lambda x - \bar{\lambda}\bar{x}] = \nu y + \mu z. \tag{10}$$

From (9) and (10), we have

$$\begin{aligned} A \begin{pmatrix} y & z \end{pmatrix} &= \begin{pmatrix} \mu y - \nu z & \nu y + \mu z \end{pmatrix} \\ &= \begin{pmatrix} y & z \end{pmatrix} \begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix} = \begin{pmatrix} y & z \end{pmatrix} L. \end{aligned}$$

Let

$$\begin{pmatrix} y & z \end{pmatrix} = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} = X_1 R$$

be a QR factorization of $\begin{pmatrix} y & z \end{pmatrix}$. Since y and z are linearly independent, it holds that R is nonsingular and

$$X_1 = \begin{pmatrix} y & z \end{pmatrix} R^{-1}.$$

Consequently,

$$AX_1 = A \begin{pmatrix} y & z \end{pmatrix} R^{-1} = \begin{pmatrix} y & z \end{pmatrix} LR^{-1} = X_1 RLR^{-1}.$$

Using this result and $(X_1 \; X_2)$ is unitary, we have

$$\begin{aligned} \begin{pmatrix} X_1^T \\ X_2^T \end{pmatrix} A \begin{pmatrix} X_1 & X_2 \end{pmatrix} &= \begin{pmatrix} X_1^T A X_1 & X_1^T A X_2 \\ X_2^T A X_1 & X_2^T A X_2 \end{pmatrix} \\ &= \begin{pmatrix} RLR^{-1} & X_1^T A X_2 \\ 0 & X_2^T A X_2 \end{pmatrix}. \quad (11) \end{aligned}$$

Since λ and $\bar{\lambda}$ are eigenvalues of L and RLR^{-1} is similar to L, (11) completes the deflation of the complex conjugate pair λ and $\bar{\lambda}$. ∎

Remark

$AX_1 = X_1(RLR^{-1})$
$\Rightarrow$ *A maps the column space of X_1 into itself*
$\Rightarrow$ *$span(X_1)$ is called an eigenspace or invariant subspace.*

- Francis double shift

1. If the Wilkinson shift σ is complex, then $\bar{\sigma}$ is also a candidate for a shift.
2. Apply two steps of the QR algorithm, one with shift σ and the other with shift $\bar{\sigma}$ to yield a matrix $\hat{H}$.

Let

$$\hat{Q}\hat{R} = (H - \sigma I)(H - \hat{\sigma} I)$$

be the QR factorization of $(H - \sigma I)(H - \hat{\sigma} I)$, then

$$\hat{H} = \hat{Q}^* H \hat{Q}.$$

Since

$$(H - \sigma I)(H - \hat{\sigma} I) = H^2 - 2Re(\sigma)H + |\sigma|^2 I \in \mathbb{R}^{n\times n},$$

we have that $\hat{Q} \in \mathbb{R}^{n\times n}$ and $\hat{H} \in \mathbb{R}^{n\times n}$. Therefore, the QR algorithm with two complex conjugate shifts preserves reality.

Francis double shift strategy

1. Compute the Wilkinson shift σ;
2. From the matrix $H^2 - 2Re(\sigma)H + |\sigma|^2 I := \tilde{H} \sim O(n^3)$ operations;
3. Compute QR factorization of $\tilde{H}$: $\tilde{H} = \hat{Q}\hat{R}$;
4. Compute $\hat{H} = \hat{Q}^* H \hat{Q}$.

- The uniqueness of Hessenberg reduction

Definition

Let H be upper Hessenberg of order n. Then H is unreduced if $h_{i+1,i} \neq 0$ for $i = 1, \cdots, n-1$.

Theorem (Implicit Q theorem)

Suppose $Q = \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix}$ *and* $V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ *are unitary matrices with*

$$Q^* A Q = H \quad \text{and} \quad V^* A V = G$$

being upper Hessenberg. Let k *denote the smallest positive integer for which* $h_{k+1,k} = 0$, *with the convection that* $k = n$ *if* H *is unreduced. If* $v_1 = q_1$, *then* $v_i = \pm q_i$ *and* $|h_{i,i-1}| = |g_{i,i-1}|$ *for* $i = 2, \cdots, k$. *Moreover, if* $k < n$, *then* $g_{k+1,k} = 0$.

Proof: Define $W \equiv \begin{pmatrix} w_1 & \cdots & w_n \end{pmatrix} = V^* Q$. Then

$$GW = GV^*Q = V^*AQ = V^*QH = WH$$

which implies that

$$h_{i,i-1} w_i = G w_{i-1} - \sum_{j=1}^{i-1} h_{j,i-1} w_j \;\text{ for }\; i = 2, \cdots, k.$$

Since $v_1 = q_1$, it holds that

$$\begin{aligned} w_1 &= e_1, \\ h_{21} w_2 &= G w_1 - h_{11} w_1 = \alpha_{21} e_1 + \alpha_{22} e_2. \end{aligned}$$

Assume

$$w_{i-1} = \alpha_{i-1,1} e_1 + \cdots + \alpha_{i-1,i-1} e_{i-1}.$$

Then

$$\begin{aligned} h_{i,i-1} w_i &= G[\alpha_{i-1,1} e_1 + \cdots + \alpha_{i-1,i-1} e_{i-1}] - \sum_{j=1}^{i-1} \beta_{i,j} e_j \\ &= \bar{\alpha}_{i,1} e_1 + \cdots + \bar{\alpha}_{i,i} e_i. \end{aligned}$$

By induction, $\begin{pmatrix} w_1 & \cdots & w_k \end{pmatrix}$ is upper triangular. Since V and Q are unitary, $W = V * Q$ is also unitary and then

$$w_1^* w_j = 0, \text{ for } j = 2, \cdots, k.$$

That is

$$w_{1j} = 0, \text{ for } j = 2, \cdots, k$$

which implies that

$$w_2 = \pm e_2.$$

Similarly, by

$$w_2^* w_j = 0, \text{ for } j = 3, \cdots, k,$$

i.e.,

$$w_{2j} = 0, \quad \text{for} \quad j = 3, \cdots, k.$$

We get $w_3 = \pm e_3$. By induction,

$$w_i = \pm e_i, \quad \text{for} \quad i = 2, \cdots, k.$$

Since $w_i = V^* q_i$ and $h_{i,i-1} = w_i^* G w_{i-1}$, we have

$$v_i = V e_i = \pm V w_i = \pm q_i$$

and

$$|h_{i,i-1}| = |g_{i,i-1}| \quad \text{for} \quad i = 2, \cdots, k.$$

If $h_{k+1,k} = 0$, then

$$\begin{aligned} g_{k+1,k} &= e_{k+1}^T G e_k = \pm e_{k+1}^T G W e_k = \pm e_{k+1}^T W H e_k \\ &= \pm e_{k+1}^T \sum_{i=1}^{k} h_{ik} w_i = \pm \sum_{i=1}^{k} h_{ik} e_{k+1}^T e_i = 0 \end{aligned}$$

General algorithm

1. Determine the first column c_1 of $C = H^2 - 2Re(\sigma)H + |\sigma|^2 I$.
2. Let Q_0 be a Householder transformation such that $Q_0^* c_1 = \sigma e_1$.
3. Set $H_1 = Q_0^* H Q_0$.
4. Use Householder transformation Q_1 to reduce H_1 to upper Hessenberg form $\hat{H}$.
5. Set $\hat{Q} = Q_0 Q_1$.

Question

General algorithm= the Francis double shift QR algorithm ?

Answer:

(I) Let

$$C = \begin{pmatrix} c_1 & C_* \end{pmatrix} = \hat{Q}\hat{R} = \begin{pmatrix} \hat{q} & \hat{Q}_* \end{pmatrix} \begin{pmatrix} \rho & r^* \\ 0 & R_* \end{pmatrix}$$

be the QR factorization of C. Then $c_1 = \rho\hat{q}$. Partition $Q_0 \equiv \begin{pmatrix} q_0 & Q_*^{(0)} \end{pmatrix}$, then $c_1 = \sigma Q_0 e_1 = \sigma q_0$ which implies that $\hat{q}$ and q_0 are proportional to c_1.

(II) Since $\hat{H} = Q_1^* H_1 Q_1$ is upper Hessenberg, we have

$$Q_1 e_1 = e_1.$$

Hence,

$$(Q_0 Q_1) e_1 = Q_0 e_1 = q_0$$

which implies that the first column of $Q_0 Q_1$ is proportional to $\hat{q}$.

(III) Since $(Q_0Q_1)^*H(Q_0Q_1)$ is upper Hessenberg and the first column of Q_0Q_1 is proportional to $\hat{q}$, by the implicit Q Theorem, if $\hat{H}$ is unreduced, then $\hat{Q} = Q_0Q_1$ and $\hat{H} = (Q_0Q_1)^*H(Q_0Q_1)$.

• Computation of the first column of $C = H^2 - 2Re(\sigma)H + |\sigma|^2 I$: Let

$$
\begin{aligned}
t &\equiv 2Re(\sigma) = \text{trace}\begin{pmatrix} h_{n-1,n-1} & h_{n-1,n} \\ h_{n,n-1} & h_{n,n} \end{pmatrix}, \\
d &\equiv |\sigma|^2 = \det\begin{pmatrix} h_{n-1,n-1} & h_{n-1,n} \\ h_{n,n-1} & h_{n,n} \end{pmatrix}.
\end{aligned}
$$

Since H is upper Hessenberg, it holds that the first column of H^2 is

$$
\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ & h_{32} \end{pmatrix}\begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix} = \begin{pmatrix} h_{11}^2 + h_{12}h_{21} \\ h_{21}(h_{11} + h_{22}) \\ h_{21}h_{32} \end{pmatrix}.
$$

Thus, the first three components of the first column of C are

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} h_{11}^2 + h_{12}h_{21} - t \cdot h_{11} + d \\ h_{21}(h_{11} + h_{22}) - t \cdot h_{21} \\ h_{21}h_{32} \end{pmatrix}$$

$$= \quad h_{21} \begin{pmatrix} (h_{nn} - h_{11})(h_{n-1,n-1} - h_{11}) - h_{n,n-1}h_{n-1,n}/h_{21} + h_{12} \\ (h_{22} - h_{11}) - (h_{nn} - h_{11}) - (h_{n-1,n-1} - h_{11}) \\ h_{32} \end{pmatrix}$$

which requires $O(1)$ operations.

The implicit double shift

$$\begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix} \xrightarrow{Q_0 H} \begin{pmatrix} + & + & + & + & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix}$$

$$\xrightarrow{H Q_0} \begin{pmatrix} + & + & + & \times & \times & \times \\ + & + & + & \times & \times & \times \\ + & + & + & \times & \times & \times \\ + & + & + & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix} \xrightarrow{Q_1 H Q_1} \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & + & \times & \times & \times & \times \\ 0 & + & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix}$$

$$\xrightarrow{Q_2 H Q_2} \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & + & \times & \times & \times \\ 0 & 0 & + & 0 & \times & \times \end{pmatrix} \xrightarrow{Q_3 H Q_3} \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & + & \times & \times \end{pmatrix}$$

$$\xrightarrow{Q_4 H Q_4} \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix} \quad \sim O(n^2) \text{ operations}$$

- Deflation:
  1. If the eigenvalues of $\begin{pmatrix} h_{n-1,n-1} & h_{n-1,n} \\ h_{n,n-1} & h_{n,n} \end{pmatrix}$ are complex and nondefective, then $h_{n-1,n-2}$ converges quadratically to zero.
  2. If the eigenvalues are real and nondefective, both the $h_{n-1,n-2}$ converge quadratically to zero. The subdiagonal elements other than $h_{n-1,n-2}$ and $h_{n,n-1}$ may show a slow convergent to zero.
  3. Deflate matrix to a middle size of matrix.

1. Converge to a block upper triangular with order one or two diagonal blocks. $i, e.$ converge to real Schur form.

- Eigenvector:

Suppose

$$T = \begin{pmatrix} T_{11} & t_{12} & t_{13} \\ 0 & \tau_{22} & \tau_{23} \\ 0 & 0 & \tau_{33} \end{pmatrix}$$

and $\begin{pmatrix} x_1^T & \xi_2 & 1 \end{pmatrix}^T$ is the eigenvector corresponding to eigenvalue $\lambda = \tau_{33}$. Then

$$\begin{pmatrix} T_{11} & t_{12} & t_{13} \\ 0 & \tau_{22} & \tau_{23} \\ 0 & 0 & \tau_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ \xi_2 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ \xi_2 \\ 1 \end{pmatrix}.$$

That is

$$\begin{cases} \tau_{22}\xi_2 - \lambda\xi_2 = -\tau_{23}, \\ T_{11}x_1 - \lambda x_1 = -t_{13} - \xi_2 t_{12}, \end{cases}$$

or

$$\begin{cases} \xi_2 = -\tau_{23}/(\tau_{22} - \lambda), \\ (T_{11} - \lambda I)x_1 = -t_{13} - \xi_2 t_{12}. \end{cases}$$ $\sim$ solve by back-substitution

Suppose

$$T = \begin{pmatrix} T_{11} & t_{12} & T_{13} \\ 0 & \tau_{22} & t_{23}^T \\ 0 & 0 & T_{33} \end{pmatrix}$$

where $T_{33} \in \mathbb{R}^{2\times 2}$. Write

$$\begin{pmatrix} T_{11} & t_{12} & T_{13} \\ 0 & \tau_{22} & t_{23}^T \\ 0 & 0 & T_{33} \end{pmatrix} \begin{pmatrix} X_1 \\ x_2^T \\ X_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ x_2^T \\ X_3 \end{pmatrix} L, \quad L \in \mathbb{R}^{2\times 2}$$

(I) Suppose X_3 is nonsingular. Then

$$T_{33}X_3 = X_3 L \Longrightarrow L = X_3^{-1} T_{33} X_3.$$

It follows that L is similar to T_{33}.

Let $x_3 = y_3 + iz_3$ be the right eigenvector of T_{33} and the corresponding eigenvalue be $\mu + i\nu$, i.e.,

$$\begin{aligned} T_{33}(y_3 + iz_3) &= (\mu + i\nu)(y_3 + iz_3) \\ &= (\mu y_3 - \nu z_3) + i(\nu y_3 + \mu z_3) \end{aligned}$$

which implies that

$$T_{33}y_3 = \mu y_3 - \nu z_3 \quad \text{and} \quad T_{33}z_3 = \nu y_3 + \mu z_3$$

or

$$\begin{aligned} T_{33}\begin{pmatrix} y_3 & z_3 \end{pmatrix} &= \begin{pmatrix} \mu y_3 - \nu z_3 & \nu y_3 + \mu z_3 \end{pmatrix} \\ &= \begin{pmatrix} y_3 & z_3 \end{pmatrix}\begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix}. \end{aligned}$$

Take $X_3 = \begin{pmatrix} y_3 & z_3 \end{pmatrix}$. Then

$$L = \begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix}.$$

(II) Since

$$\tau_{22}x_2^T - x_2^T L = -t_{23}^T X_3,$$

it implies that

$$x_2^T(\tau_{22}I - L) = -t_{23}^T X_3.$$

Since τ_{22} is not an eigenvalue of L, we get that $\tau_{22}I - L$ is nonsingular and

$$x_2^T = -t_{23}^T X_3(\tau_{22}I - L)^{-1}.$$

(III) On the other hand,

$$T_{11}X_1 - X_1 L = -T_{13}X_3 - t_{12}x_2^T.$$

This is a Sylvester equation, which we can solve for X_1 because T_{11} and L have no eigenvalues in common.

The generalized eigenvalue problem

$$Ax = \lambda Bx \quad \sim \text{generalized eigenvalue problem}$$

Definition

Let A and B be of order n. The pair (λ, x) is an eigenpair or right eigenpair of the pencil (A, B) if

$$Ax = \lambda Bx, \qquad x \neq 0$$

The pair (λ, y) is a left eigenpair of the pencil (A, B) if

$$y^* A = \lambda y^* B, \qquad y \neq 0$$

Remark

If B is singular, it is possible for any number λ to be an eigenvalue of the pencil (A, B).

1. *If A and B have a common null vector x, then (λ, x) is an eigenpair of (A, B) for any λ.*
2. *Example:*

$$0 = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 - \lambda x_2 \\ -\lambda x_3 \\ x_3 \end{pmatrix}$$

$$\Rightarrow x_3 = 0, x_1 = \lambda x_2 \quad \forall \lambda$$

The determinant of $A - \lambda B$ defined in (I) and (II) is identically zero.

Definition

A matrix pencil (A, B) is regular if $\det(A - \lambda B)$ is not identically zero.

Remark

A regular matrix pencil can have only a finite number of eigenvalues.

- To see this

$$Ax = \lambda Bx, \quad x \neq 0 \Longleftrightarrow \det(A - \lambda B) = 0$$

- Now, $P(\lambda) = \det(A - \lambda B)$ is a polynomial of degree $m \leq n$.
- If (A, B) is regular, then $P(\lambda)$ is not identically zero.
- Hence $P(\lambda)$ has m zeros.
- That is (A, B) has m eigenvalues.

If $P(\lambda) \equiv$ constant, then (A, B) has no eigenvalues. This can only occur if B is singular.

Example

Consider

$$A = I_3, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $\det(A - \lambda B) \equiv 1$ for all λ. From

$$(A - \lambda B)x = 0,$$

we have

$$x_1 - \lambda x_2 = 0, \; x_2 - \lambda x_3 = 0, \; x_3 = 0$$

which implies that

$$x_1 = x_2 = x_3 = 0.$$

Therefore, it does not exist (λ, x) with $x \neq 0$ such that $(A - \lambda B)x = 0$. It follows that (A, B) has no eigenvalues.

- λ is an eigenvalue of $(A\ ,\ B) \Longleftrightarrow \mu = \lambda^{-1}$ is an eigenvalue of $(B\ ,\ A)$
- If B is singular, then $Bx = 0$ for some $x \neq 0$.
  $\Rightarrow 0$ is an eigenvalue of $(B\ ,\ A)$
  $\Rightarrow \infty = 1/0$ is an eigenvalue of $(A\ ,\ B)$
  $\Rightarrow$ If $P(\lambda) \equiv$ constant, then the pencil has infinite eigenvalues.

Definition

Let (A, B) be a matrix pencil, U and V be nonsingular. Then the pencil $(U^*AV\ ,\ U^*BV)$ is said to be equivalent to (A, B).

Theorem

Let (λ, x) and (λ, y) be left and right eigenpairs of the regular pencil (A, B). If U and V are nonsingular, then $(\lambda, V^{-1}x)$ and $(\lambda, U^{-1}y)$ are eigenpairs of $(U^*AV\ ,\ U^*BV)$.

Since

$$\det(U^*AV - \lambda U^*BV) = \det(U^*)\det(V)\det(A - \lambda B),$$

it holds that the eigenvalues and their multiplicity are preserved by equivalence transformations.

Theorem (Generalized Schur form)

Let (A, B) be a regular pencil. Then $\exists$ unitary matrices U and V such that $S = U^*AV$ and $T = U^*BV$ are upper triangular.

Proof:

- Let v be an eigenvector of (A, B) normalized so that $\|v\|_2 = 1$, and let $\begin{pmatrix} v & V_\perp \end{pmatrix}$ be unitary.
- Since (A, B) is regular, we have $Av \neq 0$ or $Bv \neq 0$, said $Av \neq 0$.
- Moreover, if $Bv \neq 0$, then, from $Av = \lambda Bv$, it follows that $Av//Bv$.
- Let $u = Av/\|Av\|_2$ and $\begin{pmatrix} u & U_\perp \end{pmatrix}$ be unitary.

Then

$$\begin{pmatrix} u & U_\perp \end{pmatrix}^* A \begin{pmatrix} v & V_\perp \end{pmatrix} = \begin{pmatrix} u^*Av & u^*AV_\perp \\ U_\perp^*Av & U_\perp^*AV_\perp \end{pmatrix} \equiv \begin{pmatrix} \sigma_{11} & s_{12}^* \\ 0 & \hat{A} \end{pmatrix}.$$

$(\because U_\perp^*Av = U_\perp^*u = 0.)$ Similarly,

$$\begin{pmatrix} u & U_\perp \end{pmatrix}^* B \begin{pmatrix} v & V_\perp \end{pmatrix} = \begin{pmatrix} u^*Bv & u^*BV_\perp \\ U_\perp^*Bv & U_\perp^*BV_\perp \end{pmatrix} \equiv \begin{pmatrix} \tau_{11} & t_{12}^* \\ 0 & \hat{B} \end{pmatrix}.$$

$(\because U_\perp^*Bv = \lambda^{-1}U_\perp^*Av = \lambda^{-1}\|Av\|_2U_\perp^*u = 0.)$ The proof is completed by an inductive reduction of $(\hat{A}, \hat{B})$ to triangular form. ■

Definition

Let (A, B) be a regular pencil of order n.

1. $P_{(A,B)}(\lambda) \equiv \det(A - \lambda B)$: characteristic poly. of (A, B).
2. algebraic multiplicity of a finite eigenvalue of $(A, B) =$ multiplicity of a zero of $P_{(A,B)}(\lambda) = 0$.
3. $deg(P_{(A,B)}(\lambda)) = m < n$ then (A, B) has an infinite eigenvalue of algebraic multiplicity $n - m$.

Let (A, B) be a regular pencil and

$$U^* AV = [\alpha_{ij}], \quad U^* BV = [\beta_{ij}]$$

be a generalized Schur form of (A, B). Then

$$P_{(A,B)}(\lambda) = \prod_{\beta_{ii} \neq 0} (\alpha_{ii} - \lambda\beta_{ii}) \prod_{\beta_{ii}=0} \alpha_{ii} \cdot \det(U) \det(V^*).$$

If $\beta_{ii} \neq 0$, then $\lambda = \alpha_{ii}/\beta_{ii}$ is a finite eigenvalue of (A, B). Otherwise, the eigenvalue is infinite.

$$\begin{aligned} Ax = \lambda Bx \quad &\Leftrightarrow \quad \beta_{ii} Ax = \alpha_{ii} Bx \\ &\Leftrightarrow \quad (\tau\beta_{ii})Ax = (\tau\alpha_{ii})Bx, \; \tau \in \mathbb{C}. \end{aligned}$$

Definition

$< \alpha_{ii}, \beta_{ii} >= \{\tau(\alpha_{ii}, \beta_{ii}) : \tau \in \mathbb{C}\}$ is called the projective representation of the eigenvalue.

- $< 0, 1 >$: zero eigenvalue,
- $< 1, 0 >$: infinite eigenvalue,
- $< \lambda, 1 >$: ordinary eigenvalue.

If (λ, x) and (λ, y) are simple right and left eigenpair of A, respectively, then $x^* y \neq 0$. This allows us to compute the eigenvalue in the form of a Rayleigh quotient

$$y^* Ax / y^* x.$$

But, the left and right eigenvectors of a simple eigenvalue of $(A\ ,\ B)$ can be orthogonal.

Example

Consider

$$A - \lambda B = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$\det(A - \lambda B) = (1 - \lambda)(2 - \lambda).$$

It follows that $(1\ ,\ e_1)$ and $(1\ ,\ e_2)$ are right and left eigenpair of $(A\ ,\ B)$, respectively. Thus, $e_1^T e_2 = 0$.

Perturbation theory:

Let $(<\alpha\ ,\ \beta>\ ,\ x)$ be a simple eigenpair of regular pencil (A, B) and

$$(\tilde{A}, \tilde{B}) = (A + E, B + F)$$

with

$$\sqrt{\|E\|_F^2 + \|F\|_F^2} = \varepsilon.$$

If $(<\tilde{\alpha}, \tilde{\beta}>, \tilde{x})$ is an eigenpair of $(\tilde{A}, \tilde{B})$, then $(<\tilde{\alpha}, \tilde{\beta}>, \tilde{x}) \longrightarrow (<\alpha, \beta>, x)$ as $\varepsilon \to 0$.

Proof: Assume B is nonsingular $\Rightarrow$ $B + F$ is also nonsingular. Hence,

$$(A + E)\tilde{x} = \tilde{\lambda}(B + F)\tilde{x} \Rightarrow (B + F)^{-1}(A + E)\tilde{x} = \tilde{\lambda}\tilde{x}.$$

Similarly, for the left eigenvector $\tilde{y}$,

$$\tilde{y}^*(A + E)(B + F)^{-1} = \tilde{\lambda}\tilde{y}^*$$

By Theorem 3.13 in Chapter 1,

$$\sin\angle(x, \tilde{x}) = O(\varepsilon), \quad \sin\angle(y, \tilde{y}) = O(\varepsilon).$$

Suppose $\|x\|_2 = \|\tilde{x}\|_2 = \|y\|_2 = \|\tilde{y}\|_2 = 1$. Then

$$\cos\angle(x,\tilde{x}) = |x^*\tilde{x}|, \quad \cos\angle(y,\tilde{y}) = |y^*\tilde{y}|$$

or

$$|x^*\tilde{x}|^2 = \cos^2\angle(x,\tilde{x}) = 1 - \sin^2\angle(x,\tilde{x}) = 1 - O(\varepsilon),$$

which implies that

$$\tilde{x} = x + O(\varepsilon) \text{ and } \tilde{y} = y + O(\varepsilon).$$

Therefore,

$$\begin{aligned} < \tilde{\alpha}, \tilde{\beta} > &= < \tilde{y}^*\tilde{A}\tilde{x}, \tilde{y}^*B\tilde{x} > \\ &= < y^*Ax, y^*Bx > + O(\varepsilon) \\ &= < \alpha, \beta > + O(\varepsilon). \end{aligned}$$

Theorem

Let x and y be a simple eigenvectors of the regular pencil (A, B), and $< \alpha, \beta > = < y^*Ax, y^*Bx >$ be the corresponding eigenvalue. Then

$$< \tilde{\alpha}, \tilde{\beta} > = < \alpha + y^*Ex, \beta + y^*Fx > + O(\varepsilon^2). \tag{12}$$

Proof: Since (A, B) is regular, it holds that not both y^*A and y^*B can be zero. Assume $u^* \equiv y^*A \neq 0$. By (4.6), ($y^*Ax = 0 \Rightarrow Ax = 0$ and $y^*A = 0$)

$$u^*x = y^*Ax \neq 0.$$

Let U be an orthonormal basis for the orthogonal complement of u. Then $\begin{pmatrix} x & U \end{pmatrix}$ is nonsingular. Write $\tilde{x} = rx + Uc$ for some r and c. Since $\tilde{x} \to x$, it implies that $r \to 1$. Setting $e = Uc/r$, we may write $\tilde{x} = x + e$ with $\|e\|_2 = O(\varepsilon)$. Then

$$y^*Ae = u^*Uc/r = 0.$$

On the other hand, since

$$0 \neq y^* A = \lambda y^* B,$$

it holds that $\lambda \neq 0$. By the fact that

$$0 = y^* A e = \lambda y^* B e,$$

we get $y^* B e = 0$. Similarly, write

$$\tilde{y} = y + f, \text{ where } f^* A x = f^* B x = 0 \text{ and } \|f\|_2 = O(\varepsilon).$$

Now,

$$\begin{aligned} \tilde{\alpha} &= \tilde{y}^* \tilde{A} \tilde{x} = (y + f)^* (A + E)(x + e) \\ &= y^* A x + y^* E x + f^* A x + y^* A e + f^* A e + f^* E e + f^* E x + y^* E e \\ &= \alpha + y^* E x + f^* A e + f^* E e + f^* E x + y^* E e \\ &= \alpha + y^* E x + O(\varepsilon^2). \end{aligned}$$

Similarly,

$$\tilde{\beta} = \beta + y^* F x + O(\varepsilon^2)$$

∎

The expression (12) can be written in the form

$$< \tilde{\alpha}, \tilde{\beta} > = < y^* \tilde{A} x, y^* \tilde{B} x > + O(\varepsilon^2).$$

If λ is finite, then

$$\tilde{\lambda} = \frac{y^* \tilde{A} x}{y^* \tilde{B} x} + O(\varepsilon^2)$$

The chordal matric

$$< \alpha, \beta > = \{\tau(\alpha, \beta) : \tau \in \mathbb{C}\} = span\{(\alpha, \beta)\}$$

Question

How to measure the distance between two eigenvalues $< \alpha, \beta >$ and $< \gamma, \delta >$?

Answer: By the sine of the angle θ between them.

By the Cauchy inequality

$$\cos^2\theta = \frac{|\alpha\gamma + \beta\delta|^2}{(|\alpha|^2 + |\beta|^2)(|\gamma|^2 + |\delta|^2)}.$$

Hence,

$$\sin^2\theta = 1 - \cos^2\theta = \frac{|\alpha\delta - \beta\gamma|^2}{(|\alpha|^2 + |\beta|^2)(|\gamma|^2 + |\delta|^2)}.$$

Definition

The chordal distance between $< \alpha, \beta >$ and $< \gamma, \delta >$is the number

$$\chi(< \alpha, \beta >, < \gamma, \delta >) = \frac{|\alpha\delta - \beta\gamma|}{\sqrt{|\alpha|^2 + |\beta|^2}\sqrt{|\gamma|^2 + |\delta|^2}}.$$

Remark

1. *If* β *and* δ *are nonzero, set* $\lambda = \alpha/\beta$ *and* $\mu = \gamma/\delta$*, then*

$$\chi(<\alpha,\beta>,<\gamma,\delta>) = \frac{|\lambda-\mu|}{\sqrt{1+|\lambda|^2}\sqrt{1+|\mu|^2}} := \chi(\lambda,\mu).$$

   $\chi(\lambda,\mu)$ *defines a distance between numbers in the complex plane.*

2. *If* $|\lambda|,\ |\mu| \leq 1$*, then*

$$\frac{1}{2}|\lambda-\mu| \leq \chi(\lambda\ ,\ \mu) \leq |\lambda-\mu|.$$

   Hence, for eigenvalues that are not large, the chordal matric behaves like the ordinary distance between two points in the complex plane.

The condition of an eigenvalue

Since

$$< \alpha, \beta > \cong < \alpha + y^*Ex, \beta + y^*Fx >,$$

we have

$$\chi(< \alpha, \beta >, < \tilde{\alpha}, \tilde{\beta} >) \cong \frac{|\alpha y^*Fx - \beta y^*Ex|}{|\alpha|_2 + |\beta|_2}.$$

By the fact

$$\begin{aligned} \left| \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} y^*Fx \\ -y^*Ex \end{pmatrix} \right| &\leq \sqrt{|\alpha|_2 + |\beta|_2} \|x\|_2 \|y\|_2 \sqrt{\|E\|_F^2 + \|F\|_F^2} \\ &= \varepsilon \|x\|_2 \|y\|_2 \sqrt{|\alpha|_2 + |\beta|_2}, \end{aligned}$$

we get

$$\chi(< \alpha , \beta > , < \tilde{\alpha} , \tilde{\beta} >) \lesssim \frac{\|x\|_2 \|y\|_2}{\sqrt{|\alpha|_2 + |\beta|_2}} \cdot \varepsilon.$$

Theorem

Let λ be a simple eigenvalue (possibly infinite) of (A, B) and let x and y be its right and left eigenvectors. Let the projective representation of λ be $<\alpha, \beta>$, where

$$\alpha = y^* A x \quad \text{and} \quad \beta = y^* B x.$$

Let $\tilde{A} = A + E$ and $\tilde{B} = B + F$, and set

$$\varepsilon = \sqrt{\|E\|_F^2 + \|F\|_F^2}.$$

Then for ε sufficiently small, $\exists$ eigenvalue $\tilde{\lambda}$ of $(\tilde{A}\ ,\ \tilde{B})$ satisfying

$$\chi(\lambda\ ,\ \tilde{\lambda}) \leq \nu\varepsilon + O(\varepsilon^2)$$

where

$$\nu = \frac{\|x\|_2\|y\|_2}{\sqrt{|\alpha|_2 + |\beta|_2}}.$$

Remark

1. ν *is a condition number of eigenvalue.*
2. *If* $\|x\|_2 = \|y\|_2 = 1, \alpha$ *and* β *are both small, then the eigenvalue is ill conditioned, i.e., it is sensitive to the perturbation* E *and* F. *Otherwise, i.e., one of* α *or* β *is large, the eigenvalue is well conditioned.*

Theorem

Let (A, B) be a real regular pencil. Then there are orthogonal matrices U and V such that

$$S = U^T A V = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ 0 & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_{kk} \end{pmatrix}$$

and

$$T = U^T B V = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1k} \\ 0 & T_{22} & \cdots & T_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_{kk} \end{pmatrix},$$

where $T_{ii}, S_{ii} \in \mathbb{R}$ or $\mathbb{R}^{2\times 2}$.

Remark

The pencils (T_{ii}, S_{ii}) *with* $T_{ii}, S_{ii} \in \mathbb{R}$ *contain the real eigenvalues of* (A, B). *The pencils* (T_{ii}, S_{ii}) *with* $T_{ii}, S_{ii} \in \mathbb{R}^{2\times 2}$ *contain a pair of complex conjugate eigenvalues of* (A, B). *The blocks can made to appear in any order.*

Sketch the procedure of the proof: Let $x = y + iz$ be the right eigenvector of $(A \;,\; B)$ corresponding to the eigenvalue $\lambda = \mu + i\nu$, i.e.,

$$
\begin{aligned}
A(y+iz) &= (\mu + i\nu)B(y+iz) \\
&= (\mu By - \nu Bz) + i(\nu By + \mu Bz)
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
A(y \quad z) &= (\mu By - \nu Bz \quad \nu By + \mu Bz) \\
&= B(\mu y - \nu z \quad \nu y + \mu z) \\
&= B(y \quad z)\begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix} \equiv BXL \qquad (13)
\end{aligned}
$$

Since $\{y, z\}$ is linearly independent, it holds that $\exists\, V$ with $V^T V = I_2$ and a nonsingular 2×2 matrix R such that

$$\begin{pmatrix} y & z \end{pmatrix} = VR. \tag{14}$$

Substituting (14) into (13), we get

$$AVR = BVRL \;\Rightarrow AV = BV(RLR^{-1}).$$

Let $U \in \mathbb{R}^{2\times 2}$ with $U^T U = I_2$ and $S \in \mathbb{R}^{2\times 2}$ such that

$$AV = US.$$

Then

$$BV = AV(RLR^{-1})^{-1} = USRL^{-1}R^{-1} \equiv UT, \quad T \in \mathbb{R}^{2\times 2}.$$

Let $(V \quad V_\perp)$ and $(U \quad U_\perp)$ be orthogonal. Then

$$
\begin{aligned}
& \begin{pmatrix} U^T \\ U_\perp^T \end{pmatrix} (A,\ B) \begin{pmatrix} V & V_\perp \end{pmatrix} \\
= & \left(\begin{pmatrix} U^T A V & U^T A V_\perp \\ U_\perp^T A V & U_\perp^T A V_\perp \end{pmatrix}, \begin{pmatrix} U^T B V & U^T B V_\perp \\ U_\perp^T B V & U_\perp^T B V_\perp \end{pmatrix} \right) \\
= & \left(\begin{pmatrix} S & G \\ 0 & \hat{A} \end{pmatrix}, \begin{pmatrix} T & H \\ 0 & \hat{B} \end{pmatrix} \right).
\end{aligned}
$$

∎

Hessenberg-triangular form

1. Determine an orthogonal matrix Q such that $Q^T B$ is upper triangular.
2. Apply Q^T to A : $Q^T A$.
3. Use plane rotations to reduce A to Hessenberg form while preserving the upper triangularity of B.

$$\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix}, \begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{pmatrix}$$

$$\xrightarrow{P_{34}(A\ ,\ B)} \begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & * & * \\ 0 & 0 & + & * \end{pmatrix}$$

$$\xrightarrow{(A\ ,\ B)\hat{P}_{43}} \begin{pmatrix} \times & \times & * & * \\ \times & \times & * & * \\ \times & \times & * & * \\ 0 & \times & * & * \end{pmatrix}, \begin{pmatrix} \times & \times & * & * \\ 0 & \times & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}$$

$$\xrightarrow{P_{23}(A\ ,\ B)} \begin{pmatrix} \times & \times & \times & \times \\ * & * & * & * \\ 0 & * & * & * \\ 0 & \times & \times & \times \end{pmatrix}, \begin{pmatrix} \times & \times & \times & \times \\ 0 & * & * & * \\ 0 & + & * & * \\ 0 & 0 & 0 & \times \end{pmatrix}$$

$$\xrightarrow{(A\ ,\ B)\hat{P}_{32}} \begin{pmatrix} \times & * & * & \times \\ \times & * & * & \times \\ 0 & * & * & \times \\ 0 & * & * & \times \end{pmatrix}, \begin{pmatrix} \times & * & * & \times \\ 0 & * & * & \times \\ 0 & 0 & * & \times \\ 0 & 0 & 0 & \times \end{pmatrix}$$

Deflation

A : upper Hessenberg matrix, B : upper triangular matrix

(I) If $a_{k+1,k} = 0$, then

$$A - \lambda B = \begin{pmatrix} A_{11} - \lambda B_{11} & A_{12} - \lambda B_{12} \\ 0 & A_{22} - \lambda B_{22} \end{pmatrix}$$

$\Rightarrow$ Solve two small problems $A_{11} - \lambda B_{11}$ and $A_{22} - \lambda B_{22}$

(II) If $b_{kk} = 0$ for some k, then it is possible to introduce a zero in A's $(n\ ,\ n-1)$ position and thereby deflate.

$$A = \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{pmatrix}, \quad B = \begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & \otimes & \times \\ 0 & 0 & 0 & 0 & \times \end{pmatrix}$$

$$A = P_{34}^{*}A = \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & + & + & + & + \\ 0 & \oplus & + & + & + \\ 0 & 0 & 0 & \times & \times \end{pmatrix}, B = P_{34}^{*}B = \begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & \times \end{pmatrix}$$

$$A = AQ_{23} = \begin{pmatrix} \times & + & + & \times & \times \\ \times & + & + & \times & \times \\ 0 & + & + & \times & \times \\ 0 & 0 & + & \times & \times \\ 0 & 0 & 0 & \times & \times \end{pmatrix}, B = BQ_{23} = \begin{pmatrix} \times & + & + & \times & \times \\ 0 & + & + & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 & \otimes \end{pmatrix}$$

Real Schur and Hessenberg-triangular forms

$$A = P_{45}^{*}A = \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & + & + & + \\ 0 & 0 & \oplus & + & + \end{pmatrix}, B = P_{45}^{*}B = \begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A = AQ_{34} = \begin{pmatrix} \times & \times & + & + & \times \\ \times & \times & + & + & \times \\ 0 & \times & + & + & \times \\ 0 & 0 & + & + & \times \\ 0 & 0 & 0 & \oplus & \times \end{pmatrix}, B = BQ_{34} = \begin{pmatrix} \times & \times & + & + & \times \\ 0 & \times & + & + & \times \\ 0 & 0 & + & + & \times \\ 0 & 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A = AQ_{45} = \begin{pmatrix} \times & \times & \times & + & + \\ \times & \times & \times & + & + \\ 0 & \times & \times & + & + \\ 0 & 0 & \times & + & + \\ 0 & 0 & 0 & 0 & + \end{pmatrix}, B = BQ_{45} = \begin{pmatrix} \times & \times & \times & + & + \\ 0 & \times & \times & + & + \\ 0 & 0 & \times & + & + \\ 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

doubly shifted QR algorithm

iterative reduction of a real Hessenberg matrix to real Schur form.

doubly shifted QZ algorithm

iterative reduction of a real Hessenberg-triangular pencil to real generalized Schur form.

Basic idea

Update A and B as follows:

$$(\hat{A} - \lambda \hat{B}) = \hat{Q}^T (A - \lambda B) \hat{Z},$$

where $\hat{A}$ is upper Hessenberg, $\hat{B}$ is upper triangular, $\hat{Q}$ and $\hat{Z}$ are orthogonal.

1. Let (A, B) be in Hessenberg-triangular form and B be nonsingular.
2. $C = AB^{-1}$ is Hessenberg.
3. a Francis QR step were explicitly applied to C.

Let a and b be the eigenvalues of

$$\begin{pmatrix} c_{n-1,n-1} & c_{n-1,n} \\ c_{n,n-1} & c_{n,n} \end{pmatrix},$$

v be the first column of $(C - aI)(C - bI)$. Then, there is only three nonzero components in v which requires $O(1)$ flops. Take Householder transformation H such that

$$H^T v = \alpha e_1.$$

Determine orthogonal matrices Q and Z with $Qe_1 = e_1$ such that

$$(\hat{A}\ ,\ \hat{B}) = Q^T(H^T A\ ,\ H^T B)Z$$

is in Hessenberg-triangular form. Then

$$\begin{aligned}\hat{C} &= \hat{A}\hat{B}^{-1} = (Q^T H^T A Z)(Q^T H^T B Z)^{-1} \\ &= (Q^T H^T A Z)(Z(TB^{-1}HQ) = (HQ)^T C(HQ).\end{aligned}$$

- Moreover, since $Qe_1 = e_1$, we have $(HQ)e_1 = He_1$.
- It follows that $\hat{C}$ is the result of performing an implicit double QR step on C.
- Consequently, at least one of the subdiagonal elements $c_{n,n-1}$ and $c_{n-1,n-2}$ converges to zero.
  Since (A, B) is in Hessenberg-triangular form and $A = CB$, we have

$$\begin{cases} a_{n,n-1} = c_{n,n-1} b_{n-1,n-1}, \\ a_{n-1,n-2} = c_{n-1,n-2} b_{n-2,n-2}. \end{cases}$$

Hence,

- if $b_{n-1,n-1}$ and $b_{n-2,n-2}$ do not approach zero, then at least one of the subdiagonal elements $a_{n,n-1}$ and $a_{n-1,n-2}$ must approach zero.
    - $a_{n,n-1} \rightarrow 0 \Rightarrow$ deflate with a real eigenvalue.
    - $a_{n-1,n-2} \rightarrow 0 \Rightarrow$ a 2×2 block, which may contain real or complex eigenvalues, is isolated.

  $\Rightarrow$ The iteration can be continued with a smaller matrix.
- On the other hand, if either $b_{n-1,n-1}$ or $b_{n-2,n-2}$ approach zero, the process converges to an infinite eigenvalue, which can be deflated.

The QZ step

- only the first three components of v are nonzero and H is Householder transformation such that

$$H^T v = \alpha e_1$$

$$A = H^T A = \begin{pmatrix} + & + & + & + & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix},$$

$$B = H^T B = \begin{pmatrix} + & + & + & + & + & + \\ \oplus & + & + & + & + & + \\ \oplus & \oplus & + & + & + & + \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \end{pmatrix},$$

The doubly shifted QZ algorithm

$$A = AZ_1Z_2 = \begin{pmatrix} + & + & + & \times & \times & \times \\ + & + & + & \times & \times & \times \\ \oplus & + & + & \times & \times & \times \\ \oplus & \oplus & + & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix},$$

$$B = BZ_1Z_2 = \begin{pmatrix} + & + & + & \times & \times & \times \\ 0 & + & + & \times & \times & \times \\ 0 & 0 & + & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \end{pmatrix}$$

The doubly shifted QZ algorithm

$$A = Q_2 Q_1 A = \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ + & + & + & + & + & + \\ 0 & + & + & + & + & + \\ 0 & 0 & + & + & + & + \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{pmatrix},$$

$$B = Q_2 Q_1 B = \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ 0 & + & + & + & + & + \\ 0 & \oplus & + & + & + & + \\ 0 & \oplus & \oplus & + & + & + \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \end{pmatrix}$$