Eigenvalues & Eigenvectors — I

Finding eigenvalues ≡ solving polynomial Eqns.
Consider a monic (WLOG) polynomial

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

$$A = \begin{bmatrix} 0 & 1 & \cdots & & 0 \\ & & \ddots & 1 & \\ 0 & & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \rightarrow \text{companion matrix of } p(\lambda)$$

Take $u = (1, \lambda, \cdots, \lambda^{n-1})$. If λ is a root of $p(\lambda)$ then $Au = \lambda u \Rightarrow \lambda$ is an eig of A.
Conversely, if λ is an eigenvalue of A then $\det(\lambda I - A) = 0$.

FACT: There is no general formula (involving $+, -, \times, \div$ and $\sqrt{\cdot}$) for the roots of polynomials with degree > 4.
$\Rightarrow$ There is no direct method for solving the general eigenvalue problem.

Power Iteration: Let $A \in \mathbb{C}^{n\times n}$ with n linearly independent eigenvectors $v_1, \cdots, v_n$. Note
Assume: $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \lambda_4 \geq \cdots \geq \lambda_n$
dominant eigenvalue
Note strict inequality

Choose an arbitrary vector q.
$$q = c_1 v_1 + \cdots + c_n v_n \quad (\text{assume } c_1 \neq 0)$$

(If q is chosen randomly, $c_1 \neq 0$ almost surely)

$$Aq = c_1 A v_1 + \cdots + c_n A v_n$$
$$= c_1 \lambda_1 v_1 + \cdots + c_n \lambda_n v_n$$
$$A^2 q = c_1 \lambda_1^2 v_1 + \cdots + c_n \lambda_n^2 v_n$$
$$A^j q = c_1 \lambda_1^j v_1 + \cdots + c_n \lambda_n^j v_n$$

Define $q_j := \dfrac{A^j q}{\lambda_1^j} = \left(c_1 v_1 + c_2 \left(\dfrac{\lambda_2}{\lambda_1}\right)^j v_2 + \cdots + c_n \left(\dfrac{\lambda_n}{\lambda_1}\right)^j v_n\right)$

Then $\|q_j - c_1 v_1\| \leq \Big[\overbrace{|c_2| \|v_2\| + \cdots + |c_n| \|v_n\|}^{C}\Big] \left|\dfrac{\lambda_2}{\lambda_1}\right|^j$

$\leq C \left|\dfrac{\lambda_2}{\lambda_1}\right|^j$ [using $|\lambda_2| \geq \lambda_i, i \geq 3$]

Since $\left|\dfrac{\lambda_2}{\lambda_1}\right| < 1$, $\|q_j - c_1 v_1\| \to 0$ as $j \to \infty$

Note: We do not know λ_1. Hence q_j cannot be used.

$\to \|A^j q\| \to \infty$ if $|\lambda_1| > 1$

$\|A^j q\| \to 0$ if $|\lambda_1| < 1$

Hence it is better to use a scaled version of $A^j q$:

$$(*) \begin{cases} q_0 = q \\ q_{j+1} = \dfrac{A q_j}{\sigma_{j+1}} \end{cases}$$

$\sigma_{j+1} \to$ convenient scaling factor

e.g. $\sigma_{j+1} = \max_i |[A q_j]_i|$

or

$\sigma_{j+1} = \|A q_j\|_2$

When $(*)$ converges, $q_j = \dfrac{A q_j}{\sigma_j}$

$\Rightarrow$ 1) $q_j = v_1$ (or multiple)

2) $\sigma_j = \lambda_1$

Flops: For j-iterations, $2n^2 j$ flops. However this is an iterative method. Convergence can be slow (i.e. $j \to$ can be large before $\|q_{j+1} - q_j\|$ becomes ≈ 0.

Q. Does this work if A is not semisimple?

Rate of Convergence: x_j is said to converge linearly to x if $\exists\ 0 < r < 1$ s.t.

$$\lim_{j \to \infty} \frac{\|x_{j+1} - x\|}{\|x_j - x\|} = r$$

i.e $\|x_{j+1} - x\| = r \|x_j - x\|$ for j sufficiently large.

r ↑ convergence ratio/ contraction no.

For power method, $r = \left|\frac{\lambda_2}{\lambda_1}\right|$ (Exercise)

Power Iteration often works well. But limited to finding the largest eigenvalue.

```
qk = randn(n)
qk = qk / norm(qk)
for k = 1: n_iterations
    zk = A * qk
    e_val = dot(zk, qk)
    qk = zk / norm(zk)
end
```

Eigenvalue est. $= \frac{qk^* A\, qk}{qk^* qk}$ (see below)

Can be replaced with $|\max(zk)|$

Inverse Iteration : [Same assumptions on A]

Claim : If A is non-singular. If $Av = \lambda v$ then $\boxed{A^{-1}} v = \boxed{\frac{1}{\lambda}} v$ $\left[\begin{array}{l} v = \frac{1}{\lambda} A v \\ \Rightarrow A^{-1} v = \frac{1}{\lambda} v \end{array} \right]$

clearly, A^{-1} has lin. ind. eigenvectors $v_n, \dots, v_1$ corr to $\lambda_n^{-1}, \lambda_{n-1}^{-1}, \dots, \lambda_1^{-1}$.

If $|\lambda_n^{-1}| > |\lambda_{n-1}^{-1}|$ i.e $|\lambda_{n-1}| > |\lambda_n|$ same method can be used to compute v_n.

Convergence rate : $\|\lambda_n / \lambda_{n-1}\| \Rightarrow$ Faster convergence if $|\lambda_n| \ll \|\lambda_{n-1}\|$

Takeaway: We want λ_n very close to zero.

Shifting : $A \in \mathbb{C}^{n \times n}$, $\rho \in \mathbb{C}$. If $Av = \lambda v$, $(A - \rho I) v = (\lambda - \rho) v$

Choose $\rho \approx \lambda_i$ for any $i = 1, \dots, n$.

Then (most likely) $|(\lambda_i - \rho)| \ll |(\lambda_j - \rho)| \quad \forall j \neq i$

Let $(\lambda_K - \rho)$ be the second smallest (in absolute value) eigenvalue

Apply inverse iteration on $(A - \rho I)$.

→ will converge to eigenvector v_i with rate $\left| \frac{(\lambda_i - \rho)}{(\lambda_K - \rho)} \right|$.

Intuition:	Shift by 0.09
$0.1, 0.2$ $\lambda_n \quad \lambda_{n-1}$	$0.01, 0.11,$ $\lambda_n' \quad \lambda_{n-1}'$
$\left\|\frac{\lambda_n}{\lambda_{n-1}}\right\| = \frac{1}{2}$	$\left\|\frac{\lambda_n'}{\lambda_{n-1}'}\right\| = \frac{1}{11}$

$q_{j+1} = \dfrac{(A - \rho I)^{-1} q_j}{\sigma_{j+1}}$ | No need to invert.

\# Just solve using e.g LU

$(A - \rho I)\hat{q}_{j+1} = q_j$

\# $q_j = \dfrac{\hat{q}_{j+1}}{\sigma_{j+1}}$

\# Flops: $\underbrace{\frac{2}{3}n^3}_{\text{LU once (of } A-\rho I)} + \underbrace{2n^2 j}_{\text{for } j \text{ iterations}}$

Julia notation for inverse

```
for k = 1:n_iter
    zk = (A - mu*I)\qk      → computes (A - mu*I)^-1 qk
    qk = zk / norm(zk)

    ev = dot(qk, A*qk)
end
```

Q1> How do we know λ_i, in turn ρ?

Q2> If we take $\rho \approx \lambda_i$ then $(A - \rho I)^{-1}$ is ill-conditioned. Can we calculate $\hat{q}_{j+1}$ accurately?

\# We do know that even if $(A - \rho I)$ is ill-conditioned,

(B.E) $\rightarrow$ $(A + \delta A - \rho I)\hat{q}_{j+1} = q_j$ where $|\delta A|$ is small.

$\Rightarrow$ But what about sensitivity? (Later)

<u>Rayleigh Quotient</u> $\leftarrow$ Answer to Q1.

\# Estimate eigenvalue at each iteration using q_j

<u>FACT</u>: Let $A \in \mathbb{C}^{n\times n}$ & $q \in \mathbb{C}^n$. The unique complex no. that minimizes $\|Aq - \rho q\|_2$ is the

Rayliegh Quotient $\rho = \dfrac{q^* A q}{q^* q}$

Proof: Std. least sq. problem: $[q][\rho] = [\;]$ 1×1 Aq

Normal Eqn: $\rho[q^* q] = q^*(Aq)$ → Note Complex conjugates

$\Rightarrow \rho = \dfrac{q^* A q}{q^* q}$

Clearly if q is an eigenvector of A then ρ = eigenvalue corr. to q.

FACT: $A \in \mathbb{C}^{n\times n}$, $A\lambda = \lambda v$. Assume $\|v\|_2 = 1$. Let $q \in \mathbb{C}^n$, $\|q\|_2 = 1$ & let $\rho = q^* A q$ be the Rayliegh Quotient of q. Then

$$|\lambda - \rho| \leq 2\|A\|_2 \|v - q\|_2$$

Proof: Clearly, $\lambda = v^* A v$.

$$\Rightarrow \lambda - \rho = v^* A v - q^* A q$$
$$= v^* A v - v^* A q + v^* A q - q^* A q$$
$$= v^* A (v - q) + (v - q)^* A q$$
$$\Rightarrow |\lambda - \rho| \leq |v^* A (v-q)| + |(v-q)^* A q|$$
$$\leq \|v\|_2 \|A\|_2 \|v - q\|_2 + \|v - q\|_2 \|A\|_2 \|q\|_2$$
$$= 2\|A\|_2 \|v - q\|_2$$

Rayliegh Quotient Iteration

Inverse Iteration with each shift = the current Rayliegh quotient.

At kth step, $\rho_j = \dfrac{q_j^* A q_j}{q_j^* q_j}$

$(A - \rho_j I)\hat{q}_{j+1} = q_j$ and $q_{j+1} = \dfrac{\hat{q}_{j+1}}{\sigma_{j+1}}$

```
for k = 1: n_iter
    zk = (A - mu*I) \ qk
    qk = zk / norm(zk)
    mu = dot(qk, A*qk)
end
```

→ final mu is our approximation of eigenvalue

Convergence is difficult to analyse since μ is changing at each step
→ It is roughly quadratic in practice.
→ Show example.

Flops : $O(n^3)$ since a new LU is req. at each step.

Review of Linear Algebra ($\mathbb{C}^n$)

Complex Analog of Orthogonal Matrix → Unitary Matrix

$U \in \mathbb{C}^{n\times n}$ is unitary if $U^* U = I$ i.e $U^* = U^{-1}$

Eg. $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} i & -\frac{1}{\sqrt{2}} i \end{bmatrix}$ with columns a_1, a_2 $\Rightarrow A^* = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} i \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} i \end{bmatrix}$ Check $A^* A = I_2$

$\langle a_1, a_1 \rangle = a_1^* a_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} i \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} i \end{bmatrix} = \frac{1}{2} - \frac{1}{2} i^2 = 1$

$\langle a_1, a_2 \rangle = a_1^* a_2 = a_2^* a_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} i \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} i \end{bmatrix} = \frac{1}{2} + \frac{1}{2} i^2 \overset{?}{=} 0$

Clearly:

1) If U, V are unitary, $[UV][UV]^* = I \Rightarrow UV$ is unitary

2) U^{-1} is unitary

3) $\langle Ux, Uy \rangle = \langle x, y \rangle \rightarrow x^* y$, $\|Ux\|_2 = \|x\|_2$ $\forall, x, y \in \mathbb{C}^n$

4) Rotators & Reflectors have complex analogs

5) Every $A \in \mathbb{C}^{n \times n}$, $A = QR$ where $Q \rightarrow$ unitary
$R \rightarrow$ upper triangular

6) U is unitary $\Leftrightarrow$ columns are orthonormal

7) A & B $\in \mathbb{C}^{n \times n}$ are "unitarily similar" if $\exists$ a unitary matrix $U \in \mathbb{C}^{n \times n}$ s.t.
$B = U^* A U$ $(= U^{-1} A U)$

8) U is unitary $\Rightarrow \|U\|_2 = 1$, $\kappa_2(U) = 1$

9) If $A = A^*$ & A is unitarily similar to B, then $B = B^*$
[Proof $B = U^* A U$, $B^* = U^* A^* U = U^* A U = B$]

Thm: (Schur's Thm) Let $A \in \mathbb{C}^{n \times n}$. Then there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ & a upper triangular matrix T s.t. $T = U^* A U$

Proof: Induction on n. Trivial for $n = 1$.
Assume true for $n = k-1$.
Let $A \in \mathbb{C}^{k \times k}$, $Av = \lambda v$ with $\|v\|_2 = 1$
Let $U_1 = [v \;\; W] \longrightarrow$ be unitary
$W \hookrightarrow \mathbb{C}^{k \times (k-1)}$ chosen in a way s.t. U_1 is unitary.
$\Rightarrow$ Clearly $W^* v = 0$.

Let $A_1 = U_1^* A U_1 = \begin{bmatrix} v^* \\ w^* \end{bmatrix} A \begin{bmatrix} v & w \end{bmatrix} = \begin{bmatrix} v^*Av & \overbrace{v^*AW}^{\neq 0 \text{ necessarily}} \\ W^*Av & \underbrace{W^*AW}_{\hat{A}} \end{bmatrix}$

Since $Av = \lambda v$, $v^*Av = v^*\lambda v = \lambda$

$W^*Av = \lambda W^* v = 0$

$$\Rightarrow A_1 = \left[\begin{array}{c|ccc} \lambda & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \hat{A} & \\ 0 & & & \end{array}\right] \quad \text{where } \hat{A} = \mathbb{C}^{(k-1)\times(k-1)}$$

But $Av = \lambda v \Rightarrow v^*A^* = \lambda v^*$

But $v^*A \neq \lambda v^* \Rightarrow v^*AW \neq 0$

By induction hypothesis, $\exists$ unitary $\hat{U}_2$ & upper triangular $\hat{T}$ s.t. $\hat{T} = \hat{U}_2^* \hat{A} \hat{U}_2$

Define $U_2 = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \hat{U}_2 & \\ 0 & & & \end{array}\right] \Rightarrow U_2$ is unitary

& $U_2^* A_1 U_2 = \left[\begin{array}{c|ccc} \lambda & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \hat{U}_2^* \hat{A} \hat{U}_2 & \\ 0 & & & \end{array}\right] = \underbrace{\left[\begin{array}{c|ccc} \lambda & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \hat{T} & \\ 0 & & & \end{array}\right]}_{T:\text{ upper triar.}}$

Let $U = U_1 U_2$.

$\Rightarrow T = U_2^* A_1 U_2 = U_2^* U_1^* A U_1 U_2 = U^* A U$

Note: Construction of U requires the eigenvectors v. $\Rightarrow$ Hence cannot be used for numerical construction.

Schur Decomposition: $A = UTU^*$

Valid for all matrices.

$A[u_1 \cdots u_n] = [u_1 \cdots u_n]\begin{bmatrix} t_{11} & & \\ 0 & \ddots & \\ & & \ddots \end{bmatrix}$

$\Rightarrow Au_1 = u_1 t_{11}$

$\hookrightarrow$ Eigenvector with eig. value t_{11}

(Others are not so easy to find)

Example: $A = \begin{bmatrix} -1.06 & -0.61 \\ 2.35 & 0.74 \end{bmatrix}$ $v_1 = \begin{bmatrix} 0.34 - 0.31i \\ -0.89 \end{bmatrix}, \lambda_1 = -0.156 + 0.79i$

Schur Decomp $v_2 = \begin{bmatrix} 0.34 + 0.31i \\ -0.89 \end{bmatrix}, \lambda_2 = -0.15 - 0.79i$

$$A = \overbrace{\begin{bmatrix} -0.37 + 0.25i & -0.88 + 0.109i \\ 0.88 + 0.109i & -0.37 - 0.255i \end{bmatrix}}^{U} \times$$

Check: $UU^* = I$

$$\begin{bmatrix} -0.15 + 0.79i & -2.5 \\ 0 & -0.15 - 0.79i \end{bmatrix} \times U^*$$

Related Results

(Spectral Thm for Hermitian Matrices): Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. Then $\exists$ a unitary $U \in \mathbb{C}^{n\times n}$ & a diagonal $D \in \mathbb{R}^{n\times n}$ s.t. $D = U^* A U$. Columns of U are eigenvectors & diagonal entries of D are the eigenvalues.

$\Rightarrow$ $A = UDU^*$. $\leftarrow$ Spectral Decomp of A

$\Rightarrow$ Eigenvalues of A are real.

$\Rightarrow$ A has n-orthonormal eigenvectors

Eg: $A = \begin{bmatrix} 1 & 1+2i \\ 1-2i & 2 \end{bmatrix}$ Clearly $A = A^*$ (Hermitian)

Eigvectors: $v_1 = \begin{bmatrix} 0.34 + 0.69i \\ -0.6 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0.27 + 0.5i \\ 0.78 \end{bmatrix}$

with eig values $\{\overset{\lambda_1}{-0.79}, \overset{\lambda_2}{3.79}\}$

Schur Decomposition:

$$A = \underbrace{\begin{bmatrix} \underbrace{0.3+0.6i}_{} & \underbrace{0.2+0.5i}_{} \\ -0.6 & 0.7 \end{bmatrix}}_{v_1 \quad v_2} \begin{bmatrix} -0.79 & 0 \\ 0 & 3.79 \end{bmatrix} U^*$$

$U \leftarrow$ orthogonal

Normal Matrix: A matrix is normal if $AA^* \overset{?}{=} A^*A$ (not necessarily $= I$)

Thm: (Spectral Thm for Normal Matrices)

Let $A \in \mathbb{C}^{n\times n}$. Then A is normal iff $\exists$ a unitary matrix $U \in \mathbb{C}^{n\times n}$ & a diagonal $D \in \mathbb{C}^{n\times n}$ s.t. $D = U^*AU$

$\Rightarrow$ A normal $\Longleftrightarrow$ n-orthonormal eigenvectors

$\Rightarrow$ Every skew-Hermitian matrix is normal

$(A^* = -A)$ $\qquad (AA^* = -A^2 = A^*A)$

Example: $A = \begin{bmatrix} -i & 2i \\ 2i & i \end{bmatrix}$ Check: $A^* = \begin{bmatrix} i & -2i \\ -2i & -i \end{bmatrix} = -A$

Also check $AA^* = A^*A = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$

Eig: $v_1 = \begin{bmatrix} -0.5 \\ -0.8 \end{bmatrix}$, $v_2 = \begin{bmatrix} -0.8 \\ 0.5 \end{bmatrix}$, $\lambda_1 = 2.23i$, $\lambda_2 = -2.23i$

Schur: $A = \underbrace{\begin{bmatrix} -0.85i & 0.42-0.31i \\ 0.52i & 0.69-0.49i \end{bmatrix}}_{U} \begin{bmatrix} -2.23i & 0 \\ 0 & 2.23i \end{bmatrix} U^*$

Subspaces, Distance between Subspaces, Invariant Subspaces

Orthogonal Projection: Let $S \subset \mathbb{R}^n$ be a subspace. Then $P \in \mathbb{R}^{n \times n}$ is the orthogonal projection onto S if $\mathcal{R}(P) = S$, $P^2 = P$ and $P^T = P$

$\Rightarrow$ If $x \in \mathbb{R}^n$, $Px \in S$, $(I-P)x \in S^{\perp}$

$\Rightarrow$ If P_1, P_2 are orthogonal projections onto S, then for any $z \in \mathbb{R}^n$

$$\|(P_1 - P_2)z\|_2^2 = (P_1 z)^T (I - P_2) z + (P_2 z)^T (I - P_1) z = 0$$

$\Rightarrow$ P is unique.

$\Rightarrow$ If $V = [v_1 | \cdots | v_k]$ is an orthonormal basis for S, then $P = VV^T$ is the unique orthogonal projection onto S.

(Recall SVD based projections)

Distance between Subspaces: If S_1, S_2 are subspaces of $\mathbb{R}^n / \mathbb{C}^n$, and $\dim(S_1) = \dim(S_2)$ then $\mathrm{Dist}(S_1, S_2) = \|P_1 - P_2\|_2$

where P_1, P_2 are orthogonal projection onto S_1, S_2 resp.

FACT: If $W = [\underset{k}{W_1} | \underset{n-k}{W_2}]$, $Z = [\underset{k}{Z_1} | \underset{n-k}{Z_2}]$ are $n \times n$ orthogonal matrices. If $S_1 = \mathcal{R}(W_1)$ & $S_2 = \mathcal{R}(Z_1)$ then $\mathrm{dist}(S_1, S_2) = \|W_1^T Z_2\|_2$

$$= \|Z_1^T W_2\|_2$$

Invariant Subspace : A subspace $S \subset \mathbb{C}^n$ is said to be invariant for A if

$$\forall x \in S \Rightarrow Ax \in S$$

Define $S_\lambda = \{ v \in \mathbb{C}^n \mid Av = \lambda v \}$ $\therefore$ S_λ is a subspace of $\mathbb{C}^n$

→ $S_\lambda = 0$ if $\lambda \neq$ eig. value of A,

→ For λ = eig. value of A, S_λ = eig. space of A associated with λ.

Let S be an invariant subspace of A. Define $\hat{A} = A|_S$. Then $\hat{A} : S \to S$. But eig. vec / eig. value of $\hat{A}$ & A are same.

→ Study $\hat{A}$ instead of A.

Next results show: If we know any invariant Subspace S of A, we can convert A to block triangular form using unitary sim. transforms

FACT : Let $S = sp\{ x_1, \dots x_k \}$ $\in \mathbb{C}^n$ & $\hat{X} = [x_1 \cdots x_k] \in \mathbb{C}^{n \times k}$ (be rank k)

Then S is invariant under $A \in \mathbb{C}^{n \times n}$ iff $\exists$ $\hat{B} \in \mathbb{C}^{k \times k}$ s.t. $A\hat{X} = \hat{X}\hat{B}$

Proof: Exercise.

Clearly, if $\hat{B}\hat{v} = \lambda \hat{v}$ $\in \mathbb{C}^k$, then $A\hat{X}\hat{v} = \hat{X}\hat{B}\hat{v} = \lambda \hat{X}\hat{v}$

$\Rightarrow$ $v = \hat{X}\hat{v}$ is an eigvector of A with eigval λ.

$\Rightarrow$ $v \in S$ $\Rightarrow$ v is an eig. vector of $A|_S$.

FACT: Under above assumptions, $\exists$ unitary $Q \in \mathbb{C}^{n\times n}$ s.t.

$$Q^* A Q = T = \begin{array}{cc} \left[\begin{array}{c|c} T_{11} & T_{12} \\ \hline 0 & T_{22} \end{array}\right] & \begin{array}{l} p \\ n-p \end{array} \\ \begin{array}{cc} p & n-p \end{array} & \end{array}$$

& $\lambda(T_{11}) = \lambda(A) \cap \lambda(\hat{B})$

Proof: Let $\underbrace{\hat{X}}_{n\times p} = \underbrace{Q}_{n\times n} \begin{bmatrix} \underbrace{R_1}_{p\times p} \\ 0 \end{bmatrix} \leftarrow$ QR fact.

$\# \; A\hat{X} = \hat{X}\hat{B} \rightarrow AQ\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix}\hat{B}$

$\Rightarrow \underbrace{Q^* A Q}_{\text{define } T} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}\hat{B}$

Let $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \Rightarrow \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}\hat{B}$

$\Rightarrow T_{21} R_1 = 0 \Rightarrow T_{21} = 0$ ($\because R_1$ non-sing)

Also, $T_{11} R_1 = R_1 \hat{B} \Rightarrow \lambda(T_{11}) = \lambda(\hat{B})$.

Reinterpreting Schur Decomposition using Invariant Subspaces

Thm: If $A \in \mathbb{C}^{n\times n}$, then $\exists$ a unitary $Q \in \mathbb{C}^{n\times n}$ s.t.

(Schur D. Repeat)

$$Q^* A Q = T = \underbrace{D}_{\mathrm{diag}(\lambda_1, \dots, \lambda_n)} + \underbrace{N}_{\text{strictly upper triangular}}$$

Proof: (using above ideas): True for $n=1$. Assume for $n-1$.

If $Ax = \lambda x$ $(x \neq 0)$, $S = sp\{x\}$ is inv. subsp. of A.
Then by above FACT;

$$A[x] = [x][\lambda] \quad (\text{for } B)$$

& $\exists$ unitary $U \in \mathbb{C}^{n \times n}$, s.t.

$$U^* A U = \begin{array}{c} \\ \left[\begin{array}{c|c} \lambda & w^* \\ \hline 0 & C \end{array}\right] \\ \begin{array}{cc} 1 & n-1 \end{array} \end{array} \begin{array}{l} 1 \\ n-1 \\ \\ \end{array}$$

Rest: Exercise

$Q = [q_1 | \cdots | q_n]$ in above Thm are called Schur vectors

$AQ = QT \Rightarrow Aq_k = \lambda_k q_k + \sum_{i=1}^{k-1} n_{ik} q_i \quad k = 1:n$

$\Rightarrow S_k = sp\{q_1, \cdots, q_k\} \quad k = 1:n$ are invariant.

If $Q_k = [q_1 | \cdots | q_k]$, then $\lambda(Q_k^* A Q_k) = \{\lambda_1, \cdots, \lambda_k\}$

$$Q^* A Q = \begin{bmatrix} Q_k^* \\ \bar{Q}_k^* \end{bmatrix} A \begin{bmatrix} Q_k & \bar{Q}_k \end{bmatrix}$$

$$= \left(\begin{array}{c|c} Q_k^* A Q_k & Q_k^* A \bar{Q}_k \\ \hline \bar{Q}_k^* A Q_k & \bar{Q}_k^* A \bar{Q}_k \end{array}\right) =: \left[\begin{array}{c|c} T_{11} & T_{12} \\ \hline 0 & T_{22} \end{array}\right]$$

$$\Rightarrow \lambda(T_{11}) = \lambda(Q_k^* A Q_k)$$

For each $\{\lambda_1, \cdots, \lambda_k\}$ there is a k-dim inv. subsp associated i.e. $R\{Q_k\} = sp\{q_1, \cdots, q_k\}$

Q. Is this unique?

QR Iteration

How to $\to$ without why for now

Given $A \in \mathbb{C}^{n\times n}$ and a unitary $U_0 \in \mathbb{C}^{n\times n}$.

<u>set</u>: $T_0 = U_0^* A U_0$

for $k = 1, 2, \cdots$

$U_k R_k = T_{k-1}$ (QR factorization of T_{k-1})

$T_k = R_k U_k$ (Multiply R_k & U_k in opposite order to get T_k)

end

Clearly
$$T_k = R_k U_k = U_k^*(U_k R_k) U_k$$
$$= U_k^* T_{k-1} U_k$$
$$= [U_0 U_1 \cdots U_k]^* A [U_0 U_1 \cdots U_k]$$

Claim: T_k (which is unitarily similar to A) almost always converge to the Schur decomposition (upper triangular) of A.

Proof: Over the next several results.

<u>Power Iteration</u> (and Invariant Subspaces)

Recall $A^k q^{(0)} = a_1 \lambda_1^k \left(v_1 + \sum \frac{a_j^0}{a_1} \cdot \left(\frac{\lambda_j}{\lambda_1}\right)^k v_j\right)$

$$q^{(k)} = \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|} \in sp\{A^k q^{(0)}\}$$

$$dist\left(sp\{q^{(k)}\}, sp\{x_1\}\right) = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$$

Also, $\left|\lambda_1 - \lambda^{(k)}\right| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$

Power Iteration generalizes to higher dimensional invariant subspaces:

<u>Orthogonal Iteration</u> : Let $1 \leq r \leq n$

Let $A \in \mathbb{C}^{n\times n}$, $Q_0 \in \mathbb{C}^{n\times r}$ with orthonormal columns:

for $k = 1, 2, \ldots$

$Z_k = A Q_{k-1}$

$Q_k R_k = Z_k$ (QR factorization of Z_k)

end

Optional: $\lambda(Q_k^* A Q_k) = \{\lambda_1^{(k)}, \ldots, \lambda_r^{(k)}\}$

For $r=1$, O.I. = P.I. exactly

Even for $r > 1$, the sequence $\{Q_k e_1\}$ is exactly the seq. produced by P.I with $q^{(0)} = Q_0 e_1$

[For simplicity assume A semisimple]

Let the Schur decomposition of $A \in \mathbb{C}^{n\times n}$

$(*)$ $Q^* A Q = T = \mathrm{diag}(\lambda_i) + N$ $\quad |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$

Let $Q = [\underbrace{Q_\alpha}_{r} \mid \underbrace{Q_\beta}_{n-r}]$ & $T = \begin{array}{cc} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} & \begin{array}{l} \}r \\ \}n-r \end{array} \\ \begin{array}{cc} \;\; r & \;\; n-r \end{array} & \end{array}$

If $|\lambda_r| > |\lambda_{r+1}|$ then

$D_r(A) = R(Q_\alpha)$ is called the <u>dominant inv. subsp.</u>

$\mathcal{D}_r(A)$ ← unique invariant subspace associated with eig. values $\{\lambda_1, \dots, \lambda_r\}$

Thm: Let the Schur decomposition of $A \in \mathbb{C}^{n\times n}$ be given by ⊛ above, $n \geq 2$. Assume $|\lambda_r| > |\lambda_{r+1}|$ and $\mu > 0$ satisfies

$$(1+\mu)|\lambda_r| > \|N\|_F$$

Suppose $Q_0 \in \mathbb{C}^{n\times r}$ has orthonormal columns and define

$$d_k = \text{dist}\left(\mathcal{D}_r(A), \mathcal{R}(Q_k)\right), \quad k \geq 0$$

If $d_0 < 1$, then Q_k (generated by O.I.) satisfies:

$$d_k \leq (1+\mu)^{n-2}\left(1 + \frac{\|T_{12}\|_F}{\text{sep}(T_{11}, T_{22})}\right)\left[\frac{|\lambda_{r+1}| + \frac{\|N\|_F}{1+\mu}}{|\lambda_r| - \frac{\|N\|_F}{1+\mu}}\right]^k \frac{d_0}{\sqrt{1-d_0^2}}$$

or.

$$d_k \leq c \left|\frac{\lambda_{r+1}}{\lambda_r}\right|^k$$

$$\text{sep}(T_{11}, T_{22}) = \min_{X \neq 0} \frac{\|T_{11}X - XT_{22}\|_F}{\|X\|_F}$$

→ a measure of the distance between $\lambda(T_{11})$ & $\lambda(T_{22})$

→ smallest singular value of the linear transformation $X \longrightarrow T_{11}X - XT_{22}$

(Proof is skipped in this course. Refer to Golub pg 368 for details)

At the k^{th} step one can calculate

$$Q_k^* A Q_k = \begin{bmatrix} T_{11}^{(k)} & T_{12}^{(k)} \\ T_{21}^{(k)} & T_{22}^{(k)} \end{bmatrix}$$

Under above assumptions $T_{21}^{(k)} \to 0$
& $\lambda(T_{11}^{(k)}) = \{\lambda_1^{(k)}, \dots, \lambda_r^{(k)}\}$ are the estimates of $\lambda_1, \dots, \lambda_r$ at the k^{th} iteration.

If $|\lambda_r| > |\lambda_{r+1}|$ holds for all $1 \leq r < n$ then, the above convergence happens simultaneously for all r.

$\Rightarrow$

$$Q_k^* A Q_k = \begin{bmatrix} & & \\ \approx 0 & & \end{bmatrix}$$

$\to$ eigenvalues can be read off.

Q. What happens for complex conjugate eigenvalue pairs?

<u>Orthogonal Iteration to QR Iteration</u>

for $k = 1, 2, \dots$ $\boxed{Q_0}$ $Z_k = A Q_{k-1}$ $Q_k R_k = Z_k$ end	for $k = 1, 2, \dots$ $\boxed{T_0 = Q_0^* A Q_0}$ $Q_k R_k = T_{k-1}$ $T_k = R_k Q_k$

Assume $Q_0 = I$ in O.I. Then

$$AI = Z_1 = Q_1 R_1 \Rightarrow A = Q_1 R_1 \quad \text{—} (1)$$

Immediately estimate the eigenvalues:

$$\begin{aligned} T_1 &= Q_1^* A Q_1 \\ &= Q_1^*(Q_1 R_1) Q_1 \quad (\text{using } (1)) \\ &= R_1 Q_1 \end{aligned}$$

(Optimistically this should be upper triangular immediately)

In O.I. we would have continued as $\hat{Z}_2 = A Q_1$

Instead we do the same computation in the coordinate basis of T_1 (similar to A)

$$Z_2 := \underbrace{Q_1^* \hat{Z}_2}_{\hat{Z}_2 \text{ expressed in basis of } T_1 = Q_1^* A Q_1} = Q_1^*(A Q_1) = T_1$$

$\Rightarrow$ Next step: $Z_2 = Q_2 R_2$ (QR fact in O.I.) is equivalent to $T_1 = Q_2 R_2$ ...

Current guess of eigenvalues (or the upper tr. matrix T)

$$\begin{aligned} T_2 &= Q_2^* T_1 Q_2 \\ &= Q_2^*(Q_2 R_2) Q_2 = R_2 Q_2 \end{aligned}$$

QR Iteration is same as O.I. with change of basis at each step.

Alternate method to derive QR from O.I.

From O.I. $T_{k-1} = Q_{k-1}^* \underbrace{A Q_{k-1}}_{} = [Q_{k-1}^* Q_k] R_k$

and $T_k = Q_k^* A Q_k = Q_k^* \underbrace{A Q_{k-1}}_{} Q_{k-1}^* Q_k$

$= Q_k^* (Q_k R_k) Q_{k-1}^* Q_k = R_k [Q_{k-1}^* Q_k]$

Properties of QR Iteration

T_k $\overline{\text{At } k\text{-th step,}}$ generated by QR $= Q_k^* A Q_k$ generated by O.I. if both started from $Q_0 = I$.

R_k at kth step are same for both QR & O.I.

$\underbrace{Q_k}_{\text{O.I.}} = \underbrace{Q_1 Q_2 \cdots Q_k}_{\text{Q.R.}}$

In QR : $A = Q_1 R_1$

$A^2 = Q_1 \underbrace{R_1 Q_1} R_1 = Q_1 Q_2 R_2 R_1 = \hat{Q}_2 \hat{R}_2$

$\vdots$

$A^m = \underbrace{Q_1 \cdots Q_m}_{\hat{Q}_m} \underbrace{R_m \cdots R_1}_{\hat{R}_m} = \hat{Q}_m \hat{R}_m$

Hence QR is computing $A^m [e_1 \cdots e_n]$ and finding an orthonormal basis using $\hat{Q}_m$. $\rightarrow$ Power iteration.