9. EIGENVALUES: important in engineering applications

such as structural engineering (cite example of bridge).
Associated with resonances of system, dynamical systems, stability ...

§9.1 Linear Algebra preliminaries

Def (eigenvalue, eigenvector) Let $A \in \mathbb{C}^{n\times n}$. A non-zero vector $x \in \mathbb{C}^n$ is an eigenvector of A and λ its corresponding eigenvalue if

$$Ax = \lambda x$$

Def (Linear independence) Let $\{v_1, v_2, \ldots, v_n\}$ be a set of vectors. The set is linearly independent if:

$$\sum_{i=1}^{k} \alpha_k v_k = 0 \iff \alpha_i = 0, \ i = 1..k.$$

Theorem (Basis): Let $\{v_1, \ldots, v_n\}$ be a set of n lin indep vectors of $\mathbb{R}^n$. Then:

$$\forall x \in \mathbb{R}^n \ \exists \beta_i \text{ s.t. } \underline{x} = \sum_{i=1}^{n} \beta_i v_i$$

proof: $A = [v_1 | v_2 | \cdots | v_n] \in \mathbb{R}^{n\times n}$

$$\{v_1, v_2, \ldots, v_n\} \text{ lin indep} \iff \left(A\underline{\alpha} = \sum_{i=1}^{n} \alpha_i v_i = 0 \iff \underline{\alpha} = 0 \right)$$

$$\iff A \text{ is invertible (or nonsingular)}$$

Thus $\underline{x} = A\underline{\beta}$, where $\underline{\beta} = A^{-1}x$

Def (Dimension) The dimension of a vector subspace is the max. number of linearly independent vectors spanning the set.

Eigenvalue decomposition

For a square matrix $A \in \mathbb{C}^{n\times n}$ the eigenvalue decomposition is a factorization:

$$A = X \Lambda X^{-1} \qquad \text{(note: is not always possible!)}$$

$$\Leftrightarrow \quad AX = X\Lambda$$

$$\Leftrightarrow \quad A\,[x_1|x_2|\dots|x_n] = [x_1|x_2|\dots|x_n]\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$$

Def (characteristic polynomial) The characteristic poly of some $A \in \mathbb{C}^{m\times m}$ is

$$p_A(z) = \det(zI - A) = m \text{ degree poly.}$$

Theorem: λ is an eigenvalue of $A \Leftrightarrow p_A(\lambda) = 0$

proof: λ eigenvalue $\Leftrightarrow \exists\, x \neq 0$ s.t. $\lambda x - Ax = 0$
$\Leftrightarrow \lambda I - A$ is singular
$\Leftrightarrow \det(\lambda I - A) = 0$.

This theorem means that even if $A \in \mathbb{R}^n$, spectrum may be complex.
(physically complex eigenvalues give oscillatory behaviour).

Def (Algebraic Multiplicity)

Fundamental theorem of algebra $\Rightarrow p_A(z) = (z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_m)$

The algebraic multiplicity is the multiplicity of λ as a root of p_A ie how many times $z - \lambda$ appears in $p_A(z)$.

Def (Geometric Multiplicity)

Let λ be an eigenvalue of A.

$E_\lambda = \{x \mid Ax = \lambda x\}$ = vector space

= eigenspace or invariant subspace of A. $[AE_\lambda \subset E_\lambda]$

geometric multiplicity of $\lambda = \dim E_\lambda = \dim \operatorname{null}(\lambda I - A)$.

Theorem: Let $A \in \mathbb{C}^{m \times m}$. A has m eigenvalues counted with algebraic multiplicity. (easy corollary of fundamental theorem of algebra)

Similarity transformations: If $X \in \mathbb{C}^m$ is non-singular, then the map $A \to X^{-1}AX$ is a similarity transformation.

A and B are similar if there is a similarity transf s.t.

$A = X^{-1}BX$. ($\sim$ change of basis)

Theorem (similar matrices) If X is nonsingular then A and $X^{-1}AX$ have the same $p_A(z)$, eigenvalues and algebraic / geom. multiplicities.

Proof:
$$p_{X^{-1}AX}(z) = \det(zI - X^{-1}AX) = \det(X^{-1}(zI-A)X)$$
$$= \det(zI-A) = p(z)$$

proves same eigenvalues and algebraic multiplicity

Also if E_λ is an eigenspace for A then $X^{-1}E_\lambda$ is an eigenspace for $X^{-1}AX$ as well.

Theorem. algebraic multiplicity $\geq$ geom. multiplicity.

Matrices for which geom. multiplicity $<$ algebraic mult are called *defective*.

Example.

$$A = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix} \qquad \begin{aligned} p_A(z) &= (z-2)^3 \\ p_B(z) &= (z-2)^3 \end{aligned}$$

$\Rightarrow$ same eigenvalues and algebraic mult.
$(\lambda = 2)$ (alg. mult of λ is 3)

Three lin indep eigenvectors of A are: $\Rightarrow$ geom mult is 3
$\underline{e}_1, \underline{e}_2, \underline{e}_3$.

For B the only eigenvector possible is $\underline{e}_1$. $\Rightarrow$ geom mult is 1.

Theorem A non defective $\Leftrightarrow$ $A = X \Lambda X^{-1}$ (diagonalizable)

Def A matrix Q is unitary if $Q^* Q = I$.

The columns of a unitary matrix are orthogonal.
$Q^{-1} = Q^*$

Theorem. Let $A \in \mathbb{C}^{n \times n}$.

$A = A^*$ $\Rightarrow$ $A = Q \Lambda Q^*$, Q unitary, Λ diagonal
(Hermitian)

and eigenvalues are real.

Theorem: A is unitarily diagonalizable iff it is normal, i.e. $A^* A = A A^*$.

Schur Factorization (behind famous QR algo)

$$A = QTQ^* \qquad T = (\nabla) \qquad Q = \text{unitary}$$

This is an eigenvalue revealing factorization because A and T are similar
$\Rightarrow$ eigenvalues of A appear on diagonal of T.

Theorem: $\forall A \in \mathbb{C}^{n \times n}$ admits a Schur factorization

proof: By induction on size of A.

$n = 1$ trivial

Assume any matrix of size n has a Schur factorization.

Consider $A \in \mathbb{C}^{n+1 \times n+1}$, Let x be an eigenvector of A w/ eigenvalue λ.

$$U = \Big[\underbrace{\tfrac{x}{\|x\|}, \;\;}_{\text{complete basis}}\Big] = \text{unitary}.$$

Then: $U^* A U = \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix}$. By induction hyp:

$$C = VTV^*, \quad T = (\nabla)$$

$V = \text{unitary}$

Let $Q = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$

Then
$$\begin{aligned} Q^* A Q &= \begin{bmatrix} 1 & 0 \\ 0 & V^* \end{bmatrix} U^* A U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & V^* \end{bmatrix} \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \\ &= \begin{bmatrix} \lambda & B \\ 0 & V^* C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \\ &= \begin{bmatrix} \lambda & BV \\ 0 & T \end{bmatrix} = (\nabla). \end{aligned}$$

QED

This is an existence proof! it does not tell us how to compute Schur facto!

<u>Note</u>: all eigenvalue algorithms need to be iterative.

The reason is that finding eigenvalues $\Leftrightarrow$ finding roots of a poly.

$(\Rightarrow)$ characteristic poly

$(\Leftarrow)$ <u>companion matrix</u> of poly:

$$p(z) = z^m + a_{m-1} z^{m-1} + \dots + a_1 z + a_0$$

$$A = \begin{bmatrix} 0 & & & & -a_0 \\ 1 & \ddots & & & -a_1 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & 0 & -a_{m-2} \\ & & & 1 & -a_{m-1} \end{bmatrix}$$

Can verify that, if z is a root of p: $v = (1, z, z^2, \dots, z^{m-1})$ is a left eigenvector of A with eigenvalue z:

$$vA = zv$$

And it is a well known fact that no general formula exists for computing the roots of a polynomial of degree ≥ 5, so we can only hope to approximate them through an iterative process.

§ 9.2 Power Method

<u>Power method (or iteration) algorithm</u>

$v^{(0)}$ = some vector with $\|v^{(0)}\|_\infty = 1 = v^{(0)}_{p_0}$

for $k = 1, 2, \dots$

$\quad w^{(k)} = A v^{(k-1)}$

$\quad v^{(k)} = w^{(k)} / \|w^{(k)}\|_\infty$

$\quad \lambda^{(k)} = w_p$, where p = smallest index s.t. $|v^{(k-1)}_p| = \|v^{(k-1)}\|_\infty = 1$

The iteration stops when two successive iterates become close to within some tolerance.

The power method finds the largest eigenvalue (in magnitude) and its associated eigenvector.

Here is a sketch of the convergence proof.
For simplicity we assume A is diagonalizable. We also assume that the largest eigenvalue (in magnitude) of A is <u>simple</u>.
Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A ordered s.t.

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$

Let $q_1, q_2, q_3, \ldots, q_n$ the associated eigenvectors (lin indep)

Then given $v^{(0)}$ (the starting vector) $\exists\, \beta_i$ s.t.

$$v^{(0)} = \sum_{i=1}^{n} \beta_i q_i$$

$$A v^{(0)} = \sum_{i=1}^{n} \beta_i \lambda_i q_i$$

$$A^k v^{(0)} = \sum_{i=1}^{n} \beta_i \lambda_i^k q_i = \lambda_1^k \beta_1 q_1 + \lambda_1^k \sum_{i=2}^{n} \beta_i \left(\frac{\lambda_i}{\lambda_1}\right)^k q_i$$

It's not hard to see that when we apply the power method, we have:

$$v^{(k)} = \frac{A^k v^{(0)}}{\|A^k v^{(0)}\|} = \frac{\lambda_1^k \left(\beta_1 q_1 + \sum_{i=1}^{2} \beta_i \left(\frac{\lambda_i}{\lambda_1}\right)^k q_i\right)}{|\lambda_1|^k \left\|\beta_1 q_1 + \sum_{i=2}^{n} \beta_i \left(\frac{\lambda_i}{\lambda_1}\right)^k q_i\right\|_\infty}$$

Now since $\lim_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k = 0, \; i = 2, \ldots, n,$

we have $v^{(k)} \to$ a multiple of q_1.
and the rate of convergence is <u>linear</u> with ratio $\left|\frac{\lambda_2}{\lambda_1}\right|$.
(for both vector & eigenvalue)

<u>Drawback of power method</u>:

- nobody guarantees initial sol has component in direction of largest eigenvector
- computes only largest eigenvalue & eigenvector (we will see how to fix this)
- convergence rate can be very slow if $\lambda_1 \sim \lambda_2$.

Advantages

- Algorithm works even when A is nondiagonalizable.
- simplicity.
- Uses only matrix vector prod.

It's possible to get faster linear convergence of the eigenvalues in the case where A is real & symmetric.

Power method (symmetric case)

$v^{(0)}$ = some vector with $\|v^{(0)}\|=1$

for $k = 1, 2, \dots$

$$w^{(k)} = A\,v^{(k-1)}$$
$$v^{(k)} = w^{(k)} / \|w^{(k)}\|$$
$$\lambda^{(k)} = (v^{(k)})^T A\, v^{(k)}$$

In this case we can expect: $|\lambda^{(k)} - \lambda_1| = O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)$ faster than in general case.

and $|v^{(k)} - q_1| = O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{k}\right)$

depends on sign of multiple of q_1 we converge to.

The reason behind the update in $\lambda^{(k)}$ is the so called Rayleigh quotient:

$$r(x) = \frac{x^T A x}{x^T x}.$$

For real symmetric matrices it can be shown that:

$$\lambda_{min} \leq r(x) \leq \lambda_{max}$$

So if $v^{(k)}$ is a good approx to q_1, then:

$$v^{(k)T} A\, v^{(k)} = \frac{v^{(k)T} A v^{(k)}}{v^{(k)T} v^{(k)}} \approx \frac{q_1^T A q_1}{q_1^T q_1} = \lambda_1 \cancel{\frac{q_1^T q_1}{q_1^T q_1}}$$

1

The power method can be modified to zoom in and find the closest eigenvalue to some μ. The idea is that:

if v is an eigenvector of A with eigenvalue λ

then v ——— $(A-\mu I)^{-1}$ ——— $(\lambda-\mu)^{-1}$

So if we apply power method to matrix $(A-\mu I)^{-1}$ the method will converge to the largest eigenvalue $|\lambda_j-\mu|^{-1}$, a in the closest to μ. This method is:

Inverse iteration

$v^{(0)}$ = some vector with $\|v^{(0)}\|=1$

for $k=1,2,\ldots$

solve $(A-\mu I)w^{(k)} = v^{(k-1)}$ for $w^{(k)}$ $\left(\equiv \text{apply } (A-\mu I)^{-1}\right)$

$v^{(k)} = w^{(k)}/\|w^{(k)}\|$

$\lambda^{(k)} = v^{(k)T} A v^{(k)}$

(here we wrote the version for symmetric matrices, but the same modification can be carried out in the original method)

Convergence is still linear, but since we control μ, we can control the convergence rate.

Suppose λ_{J_1} is the closest eigenvalue of A to μ

λ_{J_2} ——— next to closest ———

$$|\lambda_{J_1}-\mu| < |\lambda_{J_2}-\mu| \le |\lambda_j-\mu| \quad \text{for } j \neq J_1$$

then:

$$|\lambda^{(k)}-\lambda_{J_1}| = O\left(\left|\frac{\mu-\lambda_{J_1}}{\mu-\lambda_{J_2}}\right|^{2k}\right)$$

and
$$\|v^{(k)} - (\pm q_1)\| = O\left(\left| \frac{\mu - \lambda_{J_1}}{\mu - \lambda_{J_2}} \right|^k \right)$$

So why not use variable shifts μ that get closer to the eigenvalue of interest?
This gives the following algo:

Rayleigh quotient iteration

$v^{(0)}$ = some vector with $\|v^{(0)}\| = 1$

$\lambda^{(0)} = v^{(0)T} A v^{(0)}$ = Rayleigh quot.

for $k = 1, 2, \dots$

| solve $(A - \lambda^{(k-1)} I) w^{(k)} = v^{(k-1)}$ (apply $(A - \lambda^{(k-1)} I)^{-1}$)

| $v^{(k)} = w^{(k)} / \|w^{(k)}\|$

| $\lambda^{(k)} = v^{(k)T} A v^{(k)}$ (update guess with Rayleigh quot)

Convergence of RQI is one of the rare cases where one gets cubic convergence!
If the start vector $v^{(0)}$ is sufficiently close to the eigenvector q_J:

$$\|v^{(k+1)} - (\pm q_J)\| = O(\|v^{(k)} - (\pm) q_J\|^3)$$

$$|\lambda^{(k+1)} - \lambda_J| = O(|\lambda^{(k)} - \lambda_J|^3)$$