# EIGENVALUE PROBLEMS

- **Background on eigenvalues/ eigenvectors / decompositions 7.1**
- **Perturbation analysis, condition numbers.. 7.2**
- **Power method 7.3**
- **The QR algorithm 7.5**
- **Practical QR algorithms: use of Hessenberg form and shifts 7.4**
- **The symmetric eigenvalue problem. 8-**

## Eigenvalue Problems. Introduction

Let $A$ an $n \times n$ real nonsymmetric matrix. The eigenvalue problem:

$$\boxed{Ax = \lambda x}$$

$\lambda \in \mathbb{C}$ : eigenvalue
$x \in \mathbb{C}^n$ : eigenvector

**Types of Problems:**

- Compute a few $\lambda_i$ 's with smallest or largest real parts;
- Compute all $\lambda_i$'s in a certain region of $\mathbb{C}$;
- Compute a few of the dominant eigenvalues;
- Compute all $\lambda_i$'s.

## Eigenvalue Problems. Their origins

- **Structural Engineering [$Ku = \lambda Mu$]**
- **Stability analysis [e.g., electrical networks, mechanical system,..]**
- **Bifurcation analysis [e.g., in fluid flow]**
- **Electronic structure calculations [Shrödinger equation..]**
- **Application of new era: page ranking on the world-wide web.**

## Basic definitions and properties

**A complex scalar $\lambda$ is called an eigenvalue of a square matrix $A$ if there exists a nonzero vector $u$ in $\mathbb{C}^n$ such that $Au = \lambda u$. The vector $u$ is called an eigenvector of $A$ associated with $\lambda$. The set of all eigenvalues of $A$ is the 'spectrum' of $A$. Notation: $\Lambda(A)$.**

➤ $\lambda$ is an eigenvalue iff the columns of $A - \lambda I$ are linearly dependent.

➤ ... equivalent to saying that its rows are linearly dependent. So: there is a nonzero vector $w$ such that

$$w^H(A - \lambda I) = 0$$

➤ $w^H$ is a left eigenvector of $A$ ($u$= right eigenvector)

➤ $\lambda$ is an eigenvalue iff $\boxed{\det(A - \lambda I) = 0}$

## Basic definitions and properties (cont.)

➤ An eigenvalue is a root of the **Characteristic polynomial:**

$$p_A(\lambda) = \det(A - \lambda I)$$

➤ So there are $n$ eigenvalues (counted with their multiplicities).

➤ The multiplicity of these eigenvalues as roots of $p_A$ are called **algebraic multiplicities**.

➤ The **geometric multiplicity** of an eigenvalue $\lambda_i$ is the number of linearly independent eigenvectors associated with $\lambda_i$.

➤ Geometric multiplicity is $\leq$ algebraic multiplicity.

➤ An eigenvalue is simple if its (algebraic) multiplicity is one.

➤ It is semi-simple if its geometric and algebraic multiplicities are equal.

✍ Consider

$$A = \begin{pmatrix} 1 & 2 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

Eigenvalues of $A$? their algebraic multiplicities? their geometric multiplicities? Is one a semi-simple eigenvalue?

✍ Same questions if $a_{33}$ is replaced by one.

✍ Same questions if $a_{12}$ is replaced by zero.

➤ Two matrices $A$ and $B$ are similar if there exists a nonsingular matrix $X$ such that

$$B = XAX^{-1}$$

➤ Definition: A is diagonalizable if it is similar to a diagonal matrix

➤ THEOREM: A matrix is diagonalizable iff it has $n$ linearly independent eigenvectors

➤ THEOREM (Schur form): Any matrix is unitarily similar to a triangular matrix, i.e., for any $A$ there exists a unitary matrix $Q$ and an upper triangular matrix $R$ such that

$$A = QRQ^H$$

➤ Any Hermitian matrix is unitarily similar to a real diagonal matrix, (i.e. its Schur form is real diagonal).

# Schur Form – Proof

✍ Show that there is at least one eigenvalue and eigenvector of $A$: $Ax = \lambda x$, with $\|x\|_2 = 1$

✍ There is a unitary transformation $P$ such that $Px = e_1$. How do you define $P$?

✍ Show that $PAP^H = \left(\begin{array}{c|c} \lambda & ** \\ \hline 0 & A_2 \end{array}\right)$.

✍ Apply process recursively to $A_2$.

✍ What happens if $A$ is Hermitian?

✍ Another proof altogether: use Jordan form of $A$ and QR factorization

## Perturbation analysis

- General questions: If $A$ is perturbed how does an eigenvalue change? How about an eigenvector?
- Also: sensitivity of an eigenvalue to perturbations

**THEOREM [Gerschgorin]**

$$\forall \lambda \in \Lambda(A), \quad \exists\, i \quad \text{such that} \quad |\lambda - a_{ii}| \le \sum_{\substack{j=1 \\ j \ne i}}^{j=n} |a_{ij}| \, .$$

- In words: An eigenvalue $\lambda$ of $A$ is located in one of the closed discs of the complex plane centered at $a_{ii}$ and with radius $\rho_i = \sum_{j \neq i} |a_{ij}|$ .
- The proof is by contradiction.

## Gerschgorin's theorem - example

✍ **Find a region of the complex plane where the eigenvalues of the following matrix are located:**

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ -1 & -2 & -3 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 & -4 \end{pmatrix}$$

➤ **Refinement: if disks are all disjoint then each of them contains one eigenvalue**

➤ **Refinement: can combine row and column version of the theorem (column version obtained by applying theorem to $A^H$).**

## Bauer-Fike theorem

**THEOREM [Bauer-Fike]** Let $\tilde{\lambda}, \tilde{u}$ be an approximate eigenpair with $\|\tilde{u}\|_2 = 1$, and let $r = A\tilde{u} - \tilde{\lambda}\tilde{u}$ ('residual vector'). Assume $A$ is diagonalizable: $A = XDX^{-1}$, with $D$ diagonal. Then

$$\exists\, \lambda \in \Lambda(A) \quad \text{such that} \quad |\lambda - \tilde{\lambda}| \leq \text{cond}_2(X)\|r\|_2 \, .$$

➤ Very restrictive result - also not too sharp in general.

➤ Alternative formulation. If $E$ is a perturbation to $A$ then for any eigenvalue $\tilde{\lambda}$ of $A + E$ there is an eigenvalue $\lambda$ of $A$ such that:

$$|\lambda - \tilde{\lambda}| \leq \text{cond}_2(X)\|E\|_2 \, .$$

✍ Prove this result from the previous one.

## Conditioning of Eigenvalues

➤ Assume that $\lambda$ is a simple eigenvalue with right and left eigenvectors $u$ and $w^H$ respectively. Consider the matrices:

$$A(t) = A + tE$$

➤ Eigenvalue $\lambda(t)$, eigenvector $u(t)$.

➤ Conditioning of $\lambda$ of $A$ relative to $E$ is $\left|\frac{d\lambda(t)}{dt}\right|_{t=0}$.

➤ Write $\quad A(t)u(t) = \lambda(t)u(t)$

➤ Then multiply both sides to the left by $w^H$

$$w^H(A + tE)u(t) = \lambda(t)w^H u(t) \quad \rightarrow$$

$$\begin{aligned}\lambda(t)w^H u(t) &= w^H A u(t) + t w^H E u(t) \\ &= \lambda w^H u(t) + t w^H E u(t).\end{aligned}$$

$$\rightarrow \qquad \frac{\lambda(t)-\lambda}{t} w^H u(t) = w^H E u(t)$$

➤ Take the limit at $t = 0$,

$$\boxed{\lambda'(0) = \frac{w^H E u}{w^H u}}$$

➤ Note: the left and right eigenvectors associated with a simple eigenvalue cannot be orthogonal to each other.

➤ Actual conditioning of an eigenvalue, given a perturbation "in the direction of $E$" is $|\lambda'(0)|$.

➤ In practice only estimate of $\|E\|$ is available, so

$$|\lambda'(0)| \le \frac{\|Eu\|_2 \|w\|_2}{|(u,w)|} \le \|E\|_2 \frac{\|u\|_2 \|w\|_2}{|(u,w)|}$$

**Definition.** The condition number of a simple eigenvalue $\lambda$ of an arbitrary matrix $A$ is defined by

$$\text{cond}(\lambda) = \frac{1}{\cos \theta(u, w)}$$

in which $u$ and $w^H$ are the right and left eigenvectors, respectively, associated with $\lambda$.

*Example:* Consider the matrix

$$A = \begin{pmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{pmatrix}$$

➤ $\Lambda(A) = \{1, 2, 3\}$. Right and left eigenvectors associated with $\lambda_1 = 1$:

$$u = \begin{pmatrix} 0.3162 \\ -0.9487 \\ 0.0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} 0.6810 \\ 0.2253 \\ 0.6967 \end{pmatrix}$$

So:

$$\boxed{\text{cond}(\lambda_1) \approx 603.64}$$

➤ Perturbing $a_{11}$ to $-149.01$ yields the spectrum:

$$\{0.2287, 3.2878, 2.4735\}.$$

➤ as expected..

➤ For Hermitian (also normal matrices) every simple eigenvalue is well-conditioned, since $\text{cond}(\lambda) = 1$.

## Perturbations with Multiple Eigenvalues - Example

- $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = I_3 + \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} = I + 2J$
- Worst case perturbation is in 3,1 position: set $J_{31} = \epsilon$.
- Eigenvalues of perturbed $J$ are the roots of
$$p(\mu) = \mu^3 - 2 \cdot 2 \cdot \epsilon.$$
- Hence eigenvalues of perturbed $A$ are $1 + O(\sqrt[3]{\epsilon})$.
- In general, if index of eigenvalue (dimension of largest Jordan block) is $k$, then an $O(\epsilon)$ perturbation to $A$ can lead to $O(\sqrt[k]{\epsilon})$ change in eigenvalue. Simple eigenvalue case corresponds to $k = 1$.

## The power method

➤ Basic idea is to generate the sequence of vectors $A^k v_0$ where $v_0 \neq 0$ – then normalize.

➤ Most commonly used normalization: ensure that the largest component of the approximation is equal to one.

**The Power Method**

1. Choose a nonzero initial vector $v^{(0)}$.
2. For $k = 1, 2, \ldots,$ until convergence, Do:
3. $\quad v^{(k)} = \frac{1}{\alpha_k} A v^{(k-1)}$ where
4. $\quad \alpha_k = \text{argmax}_{i=1,\ldots,n} |(Av^{(k-1)})_i|$
5. EndDo

➤ $\text{argmax}_{\text{i}=1,..,\text{n}} |\text{x}_\text{i}| \equiv$ the component $x_i$ with largest modulus

## Convergence of the power method

**THEOREM** Assume there is one eigenvalue $\lambda_1$ of $A$, s.t. $\lambda_1 = \max_j |\lambda_j|$, and that $\lambda_1$ is semi-simple. Then either the initial vector $v^{(0)}$ has no component in Null$(A - \lambda_1 I)$ or $v^{(k)}$ converges to an eigenvector associated with $\lambda_1$ and $\alpha_k \to \lambda_1$.

**Proof in the diagonalizable case.**

➤ $v^{(k)}$ is = vector $A^k v^{(0)}$ normalized by a certain scalar $\hat{\alpha}_k$ in such a way that its largest component is 1.

➤ Decompose initial vector $v^{(0)}$ in the eigenbasis as:

$$v^{(0)} = \sum_{i=1}^{n} \gamma_i u_i$$

➤ Each $u_i$ is an eigenvector associated with $\lambda_i$.

- Note that $A^k u_i = \lambda_i^k u_i$

$$v^{(k)} = \frac{1}{scaling} \times \sum_{i=1}^{n} \lambda_i^k \gamma_i u_i$$
$$= \frac{1}{scaling} \times \left[ \lambda_1^k \gamma_1 u_1 + \sum_{i=2}^{n} \lambda_i^k \gamma_i^k u_i \right]$$
$$= \frac{1}{scaling'} \times \left[ u_1 + \sum_{i=2}^{n} \left( \frac{\lambda_i}{\lambda_1} \right)^k \frac{\gamma_i}{\gamma_1} u_i \right]$$

- Second term inside bracket converges to zero. QED

- Proof suggests that the convergence factor is given by

$$\rho_D = \frac{|\lambda_2|}{|\lambda_1|}$$

where $\lambda_2$ is the second largest eigenvalue in modulus.

*Example:* Consider a 'Markov Chain' matrix of size $n = 55$. Dominant eigenvalues are $\lambda = 1$ and $\lambda = -1$ ➤ the power method applied directly to $A$ fails. (Why?)

➤ We can consider instead the matrix $I + A$ The eigenvalue $\lambda = 1$ is then transformed into the (only) dominant eigenvalue $\lambda = 2$

| Iteration | Norm of diff. | Res. norm | Eigenvalue |
|---|---|---|---|
| 20 | 0.639D-01 | 0.276D-01 | 1.02591636 |
| 40 | 0.129D-01 | 0.513D-02 | 1.00680780 |
| 60 | 0.192D-02 | 0.808D-03 | 1.00102145 |
| 80 | 0.280D-03 | 0.121D-03 | 1.00014720 |
| 100 | 0.400D-04 | 0.174D-04 | 1.00002078 |
| 120 | 0.562D-05 | 0.247D-05 | 1.00000289 |
| 140 | 0.781D-06 | 0.344D-06 | 1.00000040 |
| 161 | 0.973D-07 | 0.430D-07 | 1.00000005 |

## The Shifted Power Method

➤ **In previous example shifted $A$ into $B = A + I$ before applying power method. We could also iterate with $B(\sigma) = A + \sigma I$ for any positive $\sigma$**

*Example:* **With $\sigma = 0.1$ we get the following improvement.**

| Iteration | Norm of diff. | Res. Norm | Eigenvalue |
|---|---|---|---|
| 20 | 0.273D-01 | 0.794D-02 | 1.00524001 |
| 40 | 0.729D-03 | 0.210D-03 | 1.00016755 |
| 60 | 0.183D-04 | 0.509D-05 | 1.00000446 |
| 80 | 0.437D-06 | 0.118D-06 | 1.00000011 |
| 88 | 0.971D-07 | 0.261D-07 | 1.00000002 |

➤ Question: What is the best shift-of-origin $\sigma$ to use?

➤ Easy to answer the question when all eigenvalues are real.

Assume all eigenvalues are real and labeled decreasingly:

$$\lambda_1 > \lambda_2 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

Then:

The value of $\sigma$ which yields the best convergence factor is:

$$\sigma_{opt} = \frac{\lambda_2 + \lambda_n}{2}$$

✍ Plot a typical function $\phi(\sigma) = \rho(A - \sigma I)$ as a function of $\sigma$. Determine the minimum value and prove the above result.

## Inverse Iteration

**Observation:** The eigenvectors of $A$ and $A^{-1}$ are identical.

➤ Idea: use the power method on $A^{-1}$.

➤ Will compute the eigenvalues closest to zero.

➤ **Shift-and-invert** Use power method on $\boxed{(A - \sigma I)^{-1}}$. ➤ will compute eigenvalues closest to $\sigma$.

➤ Advantages: fast convergence in general.

➤ Drawbacks: need to factor $A$ (or $A - \sigma I$) into LU.