EIGENVALUE PROBLEMS

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

Eigenvalue Problems. Introduction

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

$$Ax = \lambda x$$

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

Types of Problems:

- Compute a few λ_i 's with smallest or largest real parts;
- Compute all λ_i's in a certain region of $\mathbb{C}$;
- Compute a few of the dominant eigenvalues;
- Compute all λ_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 [Schrödinger equation..]
- Applications of new era: page rank (of the world-wide web) and many types of dimension reduction (SVD instead of eigenvalues)

Basic definitions and properties

A complex scalar λ 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 λ. The set of all eigenvalues of A is the 'spectrum' of A. Notation: $\Lambda(A)$.

➤ λ 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 is a left eigenvector of A (u= right eigenvector)

➤ λ 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 λ_i is the number of linearly independent eigenvectors associated with λ_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.

✍1 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?

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

✍3 Same questions if, in addition, a_{12} is replaced by zero.

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

$$A = XBX^{-1}$$

- $Av = \lambda v \iff B(X^{-1}v) = \lambda(X^{-1}v)$
  eigenvalues remain the same, eigenvectors transformed.

- Issue: find X so that B has a simple structure

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

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

- ... iff all its eigenvalues are semi-simple

- ... iff its eigenvectors form a basis of $\mathbb{R}^n$

Transformations that preserve eigenvectors

Shift $B = A - \sigma I$: $Av = \lambda v \iff Bv = (\lambda - \sigma)v$
eigenvalues move, eigenvectors remain the same.

Polynomial $B = p(A) = \alpha_0 I + \cdots + \alpha_n A^n$: $Av = \lambda v \iff Bv = p(\lambda)v$
eigenvalues transformed, eigenvectors remain the same.

Invert $B = A^{-1}$: $Av = \lambda v \iff Bv = \lambda^{-1}v$
eigenvalues inverted, eigenvectors remain the same.

Shift & Invert $B = (A - \sigma I)^{-1}$: $Av = \lambda v \iff Bv = (\lambda - \sigma)^{-1}v$
eigenvalues transformed, eigenvectors remain the same. spacing between eigenvalues can be radically changed.

➤ 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).

➤ It is easy to read off the eigenvalues (including all the multiplicities) from the triangular matrix R

➤ Eigenvectors can be obtained by back-solving

Schur Form – Proof

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

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

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

✍7 Apply process recursively to A_2.

✍8 What happens if A is Hermitian?

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

Localization theorems and perturbation analysis

➤ Localization: where are the eigenvalues located in $\mathbb{C}$?

➤ Perturbation analysis: If A is perturbed how does an eigenvalue change? How about an eigenvector?

➤ Also: sensitivity of an eigenvalue to perturbations

➤ Next result is a "localization" theorem

➤ We have seen one such result before. Let $\|.\|$ be a matrix norm.

Then:

$$\forall \lambda \in \Lambda(A) : |\lambda| \leq \|A\|$$

➤ All eigenvalues are located in a disk of radius $\|A\|$ centered at 0.

➤ More refined result: Gerschgorin

THEOREM [Gerschgorin]

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

➤ In words: eigenvalue λ is located in one of the closed discs of the complex plane centered at a_{ii} and with radius $\rho_i = \sum_{j \neq i} |a_{ij}|$.

Proof: By contradiction. If contrary is true then there is one eigenvalue λ that does not belong to any of the disks, i.e., such that $|\lambda - a_{ii}| > \rho_i$ for all i. Write matrix $A - \lambda I$ as:

$$A - \lambda I = D - \lambda I - [D - A] \equiv (D - \lambda I) - F$$

where D is the diagonal of A and $-F = -(D - A)$ is the matrix of off-diagonal entries. Now write

$$A - \lambda I = (D - \lambda I)(I - (D - \lambda I)^{-1} F).$$

From assumptions we have $\|(D - \lambda I)^{-1} F\|_\infty < 1$. (Show this). The Lemma in P. 5-3 of notes would then show that $A - \lambda I$ is nonsingular – a contradiction $\square$

Gerschgorin's theorem - example

✍10 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: apply 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

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

- 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 λ of A such that:

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

Conditioning of Eigenvalues

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

➤ Write $A(t)u(t) = \lambda(t)u(t)$ Then multiply both sides to the left by w^H:

$$\begin{aligned} w^H(A + tE)u(t) &= \lambda(t)w^H u(t) \quad \rightarrow \\ \lambda(t)w^H u(t) &= w^H Au(t) + tw^H Eu(t) \\ &= \lambda w^H u(t) + tw^H Eu(t). \end{aligned}$$

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

- Take the limit at $t = 0$,

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

- 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

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

Definition. The condition number of a simple eigenvalue λ 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 λ.

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: $\mathit{cond}(\lambda_1) \approx 603.64$

➤ 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 cond$(\lambda) = 1$.

Perturbations with Multiple Eigenvalues - Example

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

- Worst case perturbation is in 3,1 position: set $A_{31} = \epsilon$.

- Eigenvalues of perturbed A are the roots of

$$p(\mu) = (\mu - 1)^3 - 4 \cdot \epsilon.$$

- Roots: $\mu_k = 1 + (4\epsilon)^{1/3} e^{\frac{2ki\pi}{3}}, \quad k = 1, 2, 3$

- Hence eigenvalues of perturbed A are $1 + O(\sqrt[3]{\epsilon})$.

- If index of eigenvalue (dimension of largest Jordan block) is k, then an $O(\epsilon)$ perturbation to A leads to $O(\sqrt[k]{\epsilon})$ change in eigenvalue. Simple eigenvalue case corresponds to $k = 1$.

Basic algorithm: 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. $\quad \alpha_k = \text{argmax}_{i=1,\ldots,n} |(Av^{(k-1)})_i|$
4. $\quad v^{(k)} = \frac{1}{\alpha_k} A v^{(k-1)}$
5. EndDo

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

Convergence of the power method

THEOREM Assume there is one eigenvalue λ_1 of A, s.t. $|\lambda_1| > |\lambda_j|$, for $j \neq i$, and that λ_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 λ_1 and $\alpha_k \rightarrow \lambda_1$.

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 λ_i.

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

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

- 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 λ_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 σ

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 σ 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: If we shift A to $A - \sigma I$:

The shift σ that yields the best convergence factor is:

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

✍11 Plot a typical convergence factor $\phi(\sigma)$ as a function of σ. 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 σ.
- Rayleigh-Quotient Iteration: use $\sigma = \frac{v^T A v}{v^T v}$ (best approximation to λ given v).
- Advantages: fast convergence in general.
- Drawbacks: need to factor A (or $A - \sigma I$) into LU.