Lecture 24

Eigenvalue Problems

NLA Reading Group Spring'13

by Alp Kındıroğlu & İsmail Arı

Eigenvalues and eigenvectors

- Let $A \in \mathbb{C}^{m\,x\,m}$ be a square matrix, a nonzero $x \in \mathbb{C}^m$ is an eigenvector of A, and $\lambda \in \mathbb{C}$ is its corresponding eigenvalue if:

$$Ax = \lambda x.$$

- Idea: The action of a matrix A on a subspace S of $\mathbb{C}^m$ may sometimes mimic scalar multiplication.

- In such a case, the special subspace S is called an eigenspace, and any nonzero $x \in S$ is an eigenvector

- The set of all eigenvalues of a matrix A is the spectrum of A, a subset of $\mathbb{C}$ denoted by $\Lambda(A)$

Eigenvalues and eigenvectors

$$Ax = \lambda x.$$

- What are eigenvalues useful for?
  - Algorithmically: simplify solutions of certain problems by reducing a coupled system to a collection of scalar problems
  - Physically: given insight into the behavior of evolving systems governed by linear equations, e.g., resonance (of musical instruments when struck or plucked or bowed), stability (of fluid flows with small perturbations)

Eigenvalue decomposition

- An eigenvalue decomposition of a square matrix A is a factorization where X is a nonsingular and Λ is diagonal:

$$A = X\Lambda X^{-1}$$

- This can be rewritten as $AX = X\Lambda$ which is:

$$\left[\begin{array}{c} \\ \\ A \\ \\ \\ \end{array}\right]\left[\begin{array}{c|c|c|c} \\ x_1 & x_2 & \cdots & x_m \\ \\ \end{array}\right] = \left[\begin{array}{c|c|c|c} \\ x_1 & x_2 & \cdots & x_m \\ \\ \end{array}\right]\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

$$Ax_j = \lambda_j x_j$$

- λ_j is an eigenvalue and j-th column of X is the corresponding eigenvector

Eigenvalue decomposition

- Express a change of basis to eigenvector coordinates
- Let $Ax = b$ and $A = X\Lambda X^{-1}$, we have
$$(X^{-1}b) = \Lambda(X^{-1}x)$$
- Thus, to compute Ax,
  - we can expand x in the basis of columns of X, apply Λ, and interpret the result as a vector of coefficients of a linear combination of the columns of X

Geometric multiplicity

- The set of eigenvectors corresponding to a single eigenvalue, together with the zero vector, forms a subspace of $\mathbb{C}^m$ known as an eigenspace, E_λ
- An eigenspace E_λ is an invariant subspace of A, i.e., $AE_\lambda \subseteq E_\lambda$
- The geometric multiplicity of λ: The dimension of E_λ can be interpreted as the maximum number of linearly independent eigenvectors that can be found, all with the same eigenvalue λ
- Geometric multiplicity can also be described as the dimension of the null space of $A - \lambda I$ since the null space is again E_λ

Characteristic polynomial

- The characteristic polynomial of $A \epsilon \mathbb{C}^{m \times m}$, denoted by p_A, is the degree m polynomial
  - $p_A(z) = \det(zI - A)$
- Note p is monic (i.e., the coefficient of its degree m term is 1)

$$\det(\mathsf{A} - \lambda \mathsf{I}) = 0,$$

$$\mathsf{A} \equiv \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix},$$

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} - \lambda \end{vmatrix} = 0$$

Characteristic polynomial (cont'd)

Theorem 24.1. *λ is an eigenvalue of A if and only if $p_A(\lambda) = 0$.*

Proof. This follows from the definition of an eigenvalue:

$$\begin{aligned} \lambda \text{ is an eigenvalue} &\iff \text{there is a nonzero vector } x \text{ such that } \lambda x - Ax = 0 \\ &\iff \lambda I - A \text{ is singular} \\ &\iff \det(\lambda I - A) = 0. \end{aligned}$$ $\square$

- Even if matrix A is real, some of its eigenvalues may be complex
- Physically, related to the phenomenon that real dynamical systems can have motions that oscillate as well as grow or decay
- Algorithmically, even if the input to a matrix eigenvalue problem is real, the output may have to be complex

Algebraic multiplicity

- By the fundamental theorem of algebra, we can write p_A in terms of their roots

$$pA(x) = (x - \lambda_1)(x - \lambda_2) \dots (x - \lambda_m)$$

for some numbers $\lambda_j \epsilon\ \mathbb{C}$

- Algebraic multiplicity of an eigenvalue of A: its multiplicity as a root of p_A
- An eigenvalue is simple if is algebraic multiplicity is 1
- Algebraic multiplicity is always as great as its geometric multiplicity

Theorem 24.2. *If $A \in \mathbb{C}^{m\times m}$, then A has m eigenvalues, counted with algebraic multiplicity. In particular, if the roots of p_A are simple, then A has m distinct eigenvalues.*

Similarity transformation

- If $X \in \mathbb{C}^{m \times m}$ is nonsingular, then the map $A \Rightarrow X^{-1}AX$ is called a similarity transformation of A
- Two matrices A and B are similar if there is a similarity transformation relating one to the other, i.e., if there exists a nonsingular $X \in \mathbb{C}^{m \times m}$ s.t. $B = X^{-1}AX$
- It is a change of basis operation

Theorem 24.3. *If X is nonsingular, then A and $X^{-1}AX$ have the same characteristic polynomial, eigenvalues, and algebraic and geometric multiplicities.*

Proof. The proof that the characteristic polynomials match is a straightforward computation:

$$\begin{aligned} p_{X^{-1}AX}(z) &= \det(zI - X^{-1}AX) = \det(X^{-1}(zI - A)X) \\ &= \det(X^{-1})\det(zI - A)\det(X) = \det(zI - A) = p_A(z). \end{aligned}$$

24.4

Theorem 24.4. *The algebraic multiplicity of an eigenvalue λ is at least as great as its geometric multiplicity.*

Proof. Let n be the geometric multiplicity of λ for the matrix A. Form an $m \times n$ matrix $\hat{V}$ whose n columns constitute an orthonormal basis of the eigenspace $\{x : Ax = \lambda x\}$. Then, extending $\hat{V}$ to a square unitary matrix V, we obtain V^*AV in the form

$$B = V^*AV = \begin{bmatrix} \lambda I & C \\ 0 & D \end{bmatrix}, \tag{24.7}$$

where I is the $n \times n$ identity, C is $n \times (m-n)$, and D is $(m-n) \times (m-n)$. By the definition of the determinant, $\det(zI - B) = \det(zI - \lambda I)\det(zI - D) = (z-\lambda)^n \det(zI - D)$. Therefore the algebraic multiplicity of λ as an eigenvalue of B is at least n. Since similarity transformations preserve multiplicities, the same is true for A. □

Defective eigenvalue

Example 24.1. Consider the matrices

$$A = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix}$$

- Both A and B have characteristic polynomial $(z - 2)^3$, so there is a single eigenvalue $\lambda = 2$ of algebraic multiplicity 3
- In the case of A, we can choose three independent eigenvectors, e.g., e1, e2, e3 and so the geometric multiplicity of $\lambda = 2$ is 3
- For B, on the other hand, we can only have one single independent eigenvector, i.e., a scalar multiple of e1, so the geometric multiplicity of the eigenvalue is only 1
- It means that there are not sufficient number of independent eigenvectors to span B
- It also means that A can be diagonalized but not B

Defective eigenvalue

- An eigenvalue whose algebraic multiplicity exceeds its geometric multiplicity is a defective eigenvalue

- A matrix that has one or more defective eigenvalues is a defective matrix

- Any diagonal matrix is nondefective, and both the algebraic and the geometric multiplicities of an eigenvalue λ are equal to the number of its occurrences along the diagonal

Diagonalizability

Theorem 24.5. *An $m \times m$ matrix A is nondefective if and only if it has an eigenvalue decomposition $A = X\Lambda X^{-1}$.*

Proof. ($\Longleftarrow$) Given an eigenvalue decomposition $A = X\Lambda X^{-1}$, we know by Theorem 24.3 that Λ is similar to A, with the same eigenvalues and the same multiplicities. Since Λ is a diagonal matrix, it is nondefective, and thus the same holds for A.

($\Longrightarrow$) A nondefective matrix must have m linearly independent eigenvectors, because eigenvectors with different eigenvalues must be linearly independent, and each eigenvalue can contribute as many linearly independent eigenvectors as its multiplicity. If these m independent eigenvectors are formed into the columns of a matrix X, then X is nonsingular and we have $A = X\Lambda X^{-1}$. $\square$

Determinant and trace

Theorem 24.6. *The determinant* $\det(A)$ *and trace* $\operatorname{tr}(A)$ *are equal to the product and the sum of the eigenvalues of* A*, respectively, counted with algebraic multiplicity:*

$$\det(A) = \prod_{j=1}^{m} \lambda_j, \qquad \operatorname{tr}(A) = \sum_{j=1}^{m} \lambda_j. \tag{24.8}$$

Proof. From (24.5) and (24.6), we compute

$$\det(A) = (-1)^m \det(-A) = (-1)^m p_A(0) = \prod_{j=1}^{m} \lambda_j.$$

Unitary diagonalization

- A matrix $A \epsilon \mathbb{C}^{m\ x\ m}$ is unitarily diagonaizable when it has m linearly independent and orthogonal eigenvectors. i.e., there exists a unitary matrix Q such that

$$A = Q\Lambda Q^H$$

Theorem 24.7. *A hermitian matrix is unitarily diagonalizable, and its eigenvalues are real.*

- A matrix is normal if $A^H A = AA^H$.

Theorem 24.8. *A matrix is unitarily diagonalizable if and only if it is normal.*

Schur decomposition

- Every square matrix can be factorized in Schur decomposition

$$A = QTQ^H, A \epsilon \mathbb{C}^{m\,x\,m}$$
$$T = Q^H A Q$$

where Q is unitary and T is upper triangular, and the eigenvalues of A appear on the diagonal of T

- Any square matrix, defective or not, can be triangularized by unitary transformations
- The diagonal elements of a triangular matrix are its eigenvalues
- The unitary transformations preserve eigenvalues

Schur decomposition (cont'd)

Theorem 24.9. *Every square matrix A has a Schur factorization.*

Proof. We proceed by induction on the dimension m of A. The case $m = 1$ is trivial, so suppose $m \geq 2$. Let x be any eigenvector of A, with corresponding eigenvalue λ. Take x to be normalized and let it be the first column of a unitary matrix U. Then, just as in (24.7), it is easily checked that the product U^*AU has the form

$$U^*AU = \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix}.$$

By the inductive hypothesis, there exists a Schur factorization VTV^* of C. Now write

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

This is a unitary matrix, and we have

$$Q^*AQ = \begin{bmatrix} \lambda & BV \\ 0 & T \end{bmatrix}.$$

This is the Schur factorization we seek. $\square$

Eigenvalue revealing factorization

- diagonalization $A = X\Lambda X^{-1}$ exists if and only if A is nondefective
- unitary diagonalization $A = Q\Lambda Q^H$ exits if and only if A is normal
- unitary triangularization (Schur factorization) $A = QTQ^H$ always exists

- All three of these factorizations can be used to compute eigenvalues
- In general, Schur factorization is used as this applies without restriction
- If A is normal, then Schur form comes out diagonal and its eigenvalues are real
- If A is Hermitian, then we can take advantage of symmetry with half as much work or less than is required for general A

Lecture 25

Overview of Eigenvalue Algorithms

NLA Reading Group Spring'13
by Alp Kındıroğlu & İsmail Arı

Shortcomings of obvious algorithms

- Characteristic polynomial: Compute the coefficients of the characteristic polynomial and find the roots (an ill-conditioned problems in general)

- Power iteration: The sequence

$$\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{A\mathbf{x}}{\|A\mathbf{x}\|}, \frac{A^2\mathbf{x}}{\|A^2\mathbf{x}\|}, \frac{A^3\mathbf{x}}{\|A^3\mathbf{x}\|}, \cdots$$

- slowly converges, under certain assumptions, to an eigenvector corresponding to the largest eigenvalue of A

Shortcomings of obvious algorithms

- Best general purpose eigenvalue algorithms are based on a different principle: the computation of an eigenvalue revealing factorization of the matrix A, where the eigenvalues appear as entries of one of the factors.

  - diagonalization $A = X\Lambda X^{-1}$ exists if and only if A is nondefective
  - unitary diagonalization $A = Q\Lambda Q^H$ exits if and only if A is normal
  - unitary triangularization (Schur factorization) $A = QTQ^H$ always exists

- Goal of these methods: Apply a seqeunce of transformations to A to introduce zeros in necessary places.

A fundamental difficulty

- Any polynomial rootfinding problem can be represented as an eigenvalue problem.

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

$$\begin{bmatrix} -z & & & & & -a_0 \\ 1 & -z & & & & -a_1 \\ & 1 & -z & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & -z & -a_{m-2} \\ & & & & 1 & (-z - a_{m-1}) \end{bmatrix}$$

A fundamental difficulty

- A is called a companion matrix of p.

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

$$A = \begin{bmatrix} 0 & & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & 0 & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & 0 & -a_{m-2} \\ & & & & 1 & -a_{m-1} \end{bmatrix}.$$

- The roots of p are equal to the eigenvalues of A.

A fundamental difficulty

Theorem 25.1. *For any $m \geq 5$, there is a polynomial $p(z)$ of degree m with rational coefficients that has a real root $p(r) = 0$ with the property that r cannot be written using any expression involving rational numbers, addition, subtraction, multiplication, division, and kth roots.*

- Abel in 1824 proved that no analog of the quadratic formula can exist for polynomials of degree 5 or more

- Any eigenvalue solver must be iterative
  - The goal is to produce sequences of numbers that converge rapidly towards eigenvalues

Schur factorization and diagonalization

- Most of the general purpose eigenvalue algorithms in use today proceed by computing the Schur factorization

$$A = QTQ^H \quad T = Q^H AQ$$

- Schur factorization $A = QTQ^H$ can be computed by transforming A using a sequence of elementary unitary similarity transformation

$$X \Rightarrow Q_j^H X Q_j$$

so the product

$$Q_j^H \dots Q_2^H Q_1^H A Q_1 Q_2 \dots Q_j$$

converges to an upper triangular matrix T as $j \Rightarrow \infty$

Schur factorization and diagonalization

- If A is real but not symmetric, then in general it may have complex eigenvalues in conjugate pairs

- An algorithm that computes the Schur factorization will have to be capable of generating complex outputs from real inputs

- If A is Hermitian, then $Q_j^H \dots Q_2^H Q_1^H A Q_1 Q_2 \dots Q_j$ is also Hermitian, and thus the limit of the converging sequence is both triangular and Hermitian, hence diagonal

- This implies that the same algorithms that compute a unitary triangularization of a general matrix also compute a unitary diagonalization of a Hermitian matrix

Two phases of eigenvalue computation

- In the 1st phase, a direct method is applied to produce an upper Hessenberg matrix, i.e., a matrix with zeros below the 1st subdiagonal

- In the second phase, an iteration is applied to generate a formally infinite sequences of Hessenberg matrices that converge to a triangular form

$$\underset{A \neq A^*}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 1}} \underset{H}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 2}} \underset{T}{\begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ & & & & \times \end{bmatrix}}.$$

Two phases of eigenvalue computation

- If A is Hermitian, the two phase approach becomes faster
- The intermediate matrix is a Hermitian Hessenberg matrix, i.e.,tridiagonal
- The result is a Hermitian triangular matrix, i.e., diagonal

$$\underset{A = A^*}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 1}} \underset{T}{\begin{bmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 2}} \underset{D}{\begin{bmatrix} \times & & & & \\ & \times & & & \\ & & \times & & \\ & & & \times & \\ & & & & \times \end{bmatrix}}.$$