10. Schur decomposition

- eigenvalues of nonsymmetric matrix
- Schur decomposition
- Sylvester equation

Eigenvalues and eigenvectors

a nonzero vector x is an *eigenvector* of the $n \times n$ matrix A, with *eigenvalue* λ, if

$$Ax = \lambda x$$

- the eigenvalues are the roots of the characteristic polynomial

$$\det(\lambda I - A) = 0$$

- eigenvectors are nonzero vectors in the nullspace of $\lambda I - A$

for most of the lecture, we assume that A is a complex $n \times n$ matrix

Linear independence of eigenvectors

suppose $x_1, \ldots, x_k$ are eigenvectors for k different eigenvalues:

$$Ax_1 = \lambda_1 x_1, \qquad \ldots, \qquad Ax_k = \lambda_k x_k$$

then $x_1, \ldots, x_k$ are linearly independent

- the result holds for $k = 1$ because eigenvectors are nonzero
- suppose it holds for $k - 1$, and assume $\alpha_1 x_1 + \cdots + \alpha_k x_k = 0$; then

$$0 = A(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k) = \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \cdots + \alpha_k \lambda_k x_k$$

- subtracting $\lambda_1(\alpha_1 x_1 + \cdots + \alpha_k x_k) = 0$ gives

$$\alpha_2(\lambda_2 - \lambda_1)x_2 + \cdots + \alpha_k(\lambda_k - \lambda_1)x_k = 0$$

- since $x_2, \ldots, x_k$ are linearly independent, $\alpha_2 = \cdots = \alpha_k = 0$
- $\alpha_1 x_1 = -(\alpha_2 x_2 + \cdots + \alpha_k x_k) = 0$, so α_1 is also zero
- we conclude that $\alpha_1 = \cdots = \alpha_k = 0$, so $x_1, \ldots, x_k$ are linearly independent

Multiplicity of eigenvalues

Algebraic multiplicity

- the multiplicity of the eigenvalue as a root of the characteristic polynomial
- the sum of the algebraic multiplicities of the eigenvalues of an $n \times n$ matrix is n

Geometric multiplicity

- the geometric multiplicity is the dimension of $\mathrm{null}(\lambda I - A)$
- the maximum number of linearly independent eigenvectors with eigenvalue λ
- sum is the maximum number of linearly independent eigenvectors of the matrix

Defective eigenvalue

- geometric multipicity never exceeds algebraic multiplicity (proof on page 10.7)
- eigenvalue is *defective* if geometric muliplicity is less than algebraic multiplicity
- a matrix is *defective* if some of its eigenvalues are defective

Example

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \qquad \det(\lambda I - A) = (\lambda - 1)^2 (\lambda - 2)^2$$

- two eigenvalues, $\lambda = 1$ and $\lambda = 2$, each with algebraic multiplicity two
- eigenvalue $\lambda = 1$ is defective and has geometric multiplicity one:

$$\text{null}(I - A) = \text{span}\,\{(1, 0, 0, 0)\}$$

- eigenvalue $\lambda = 2$ is not defective and has geometric multiplicity two:

$$\text{null}(2I - A) = \text{span}\{(0, 0, 1, 0),\ (0, 0, 0, 1)\}$$

- maximum number of linearly independent eigenvectors is three; for example,

$$(1, 0, 0, 0), \quad (0, 0, 1, 0), \quad (0, 0, 0, 1)$$

Similarity transformation

two matrices A and B are *similar* if

$$B = X^{-1}AX$$

for some nonsingular matrix X

- the matrices $B = X^{-1}AX$ and A have the same characteristic polynomial:

$$\det(\lambda I - B) = \det(\lambda I - X^{-1}AX) = \det(X^{-1}(\lambda I - A)X) = \det(\lambda I - A)$$

- similarity transformation preserves eigenvalues and algebraic multiplicities
- if x is an eigenvector of A then $y = X^{-1}x$ is an eigenvector of B:

$$By = (X^{-1}AX)(X^{-1}x) = X^{-1}Ax = X^{-1}(\lambda x) = \lambda y$$

- similarity transformation preserves geometric multiplicities:

$$\dim \operatorname{null}(\lambda I - B) = \dim \operatorname{null}(\lambda I - A)$$

Geometric and algebraic multiplicities

suppose α is an eigenvalue with geometric multiplicity r:

$$\dim \operatorname{null}(\alpha I - A) = r$$

- define an $n \times r$ matrix U with orthonormal columns that span $\operatorname{null}(\alpha I - A)$
- complete U to define a unitary matrix $W = \begin{bmatrix} U & V \end{bmatrix}$ and define $B = W^H A W$:

$$B = \begin{bmatrix} U^H A U & U^H A V \\ V^H A U & V^H A V \end{bmatrix} = \begin{bmatrix} \alpha U^H U & U^H A V \\ \alpha V^H U & V^H A V \end{bmatrix} = \begin{bmatrix} \alpha I & U^H A V \\ 0 & V^H A V \end{bmatrix}$$

- the characteristic polynomial of B is

$$\det(\lambda I - B) = (\lambda - \alpha)^r \det(\lambda I - V^H A V)$$

this shows that the algebraic multiplicity of eigenvalue α of B is at least r

Diagonalizable matrices

the following three properties are equivalent

1. A is diagonalizable by a similarity: there exists nonsingular X, diagonal Λ s.t.

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

2. A has a set of n linearly independent eigenvectors (for example, columns of X):

$$AX = X\Lambda$$

3. all eigenvalues of A are nondefective: for every eigenvalue λ,

$$\text{algebraic multiplicity} = \text{geometric multiplicity}$$

- not all square matrices are diagonalizable
- real symmetric matrices are an important class of diagonalizable matrices

Outline

- eigenvalues of nonsymmetric matrix
- **Schur decomposition**
- Sylvester equation

Schur decomposition

every $A \in \mathbf{C}^{n \times n}$ can be factored as

$$A = UTU^H \tag{1}$$

- subscript ^{H} denotes complex conjugate transpose
- U is unitary: $U^H U = UU^H = I$
- T is upper triangular, with the eigenvalues of A on its diagonal
- the eigenvalues can be chosen to appear in any order on the diagonal of T
- A is reduced to triangular form by unitary similarity transformation:

$$U^H AU = T$$

- in general, the matrices U, T are complex, even when A is real
- complexity of computing the factorization is order n^3

Proof by induction

- the decomposition (1) obviously exists if $n = 1$
- suppose it exists if $n = m$ and A is an $(m+1) \times (m+1)$ matrix
- let λ be any eigenvalue of A and u a corresponding eigenvector, with $\|u\| = 1$
- let V be an $(m+1) \times m$ matrix that makes the matrix $\begin{bmatrix} u & V \end{bmatrix}$ unitary; then

$$\begin{bmatrix} u^H \\ V^H \end{bmatrix} A \begin{bmatrix} u & V \end{bmatrix} = \begin{bmatrix} u^H A u & u^H A V \\ V^H A u & V^H A V \end{bmatrix} = \begin{bmatrix} \lambda u^H u & u^H A V \\ \lambda V^H u & V^H A V \end{bmatrix} = \begin{bmatrix} \lambda & u^H A V \\ 0 & V^H A V \end{bmatrix}$$

- $V^H A V$ is an $m \times m$ matrix, so by the induction hypothesis,

$$V^H A V = \tilde{U} \tilde{T} \tilde{U}^H \quad \text{for some unitary } \tilde{U} \text{ and upper triangular } \tilde{T}$$

- the matrix $U = \begin{bmatrix} u & V\tilde{U} \end{bmatrix}$ is unitary and satisfies

$$U^H A U = \begin{bmatrix} u^H \\ \tilde{U}^H V^H \end{bmatrix} A \begin{bmatrix} u & V\tilde{U} \end{bmatrix} = \begin{bmatrix} \lambda & u^H A V \tilde{U} \\ 0 & \tilde{Q}^H V^H A V \tilde{Q} \end{bmatrix} = \begin{bmatrix} \lambda & u^H A V \tilde{U} \\ 0 & \tilde{T} \end{bmatrix}$$

Real Schur decomposition

if A is real, a similar factorization with real matrices exists:

$$A = UTU^T$$

- U is orthogonal: $U^T U = UU^T = I$
- T is *quasi-triangular:*

$$T = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1m} \\ 0 & T_{22} & \cdots & T_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_{mm} \end{bmatrix}$$

  the diagonal blocks T_{ii} are 1×1 or 2×2
- the scalar diagonal blocks are real eigenvalues of A
- the eigenvalues of the 2×2 diagonal blocks are complex eigenvalues of A

Normal matrix

a square matrix A is *normal* if

$$A^H A = AA^H$$

Examples

- Hermitian (and symmetric real) matrices: $A^H = A$
- skew-Hermitian (and skew-symmetric real) matrices: $A^H = -A$
- unitary (and orthogonal real) matrices: $A^H A = AA^H = I$
- $A = I + B$ where B is a normal matrix

Schur decomposition of normal matrix

a matrix is normal if and only if it is diagonalizable by a unitary similarity

$$A = UDU^H, \quad \text{with } U \text{ unitary, } D \text{ diagonal} \tag{2}$$

Proof

- if A satisfies (2), then it is normal:

$$A^H A = UD^H DU^H = UDD^H U^H = AA^H$$

  $D^H D = DD^H$ is the diagonal matrix with diagonal entries $|D_{ii}|^2$

- if A is normal with Schur decomposition $A = UTU^H$, then

$$A^H A = UT^H TU^H = UTT^H U^H = AA^H \quad \Longrightarrow \quad T^H T = TT^H$$

  a triangular matrix that satisfies $T^H T = TT^H$ is diagonal

Example: circulant matrix

recall from 133A the definition of a *circulant matrix*

$$A = \begin{bmatrix} a_1 & a_n & a_{n-1} & \cdots & a_3 & a_2 \\ a_2 & a_1 & a_n & \cdots & a_4 & a_3 \\ a_3 & a_2 & a_1 & \cdots & a_5 & a_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_1 & a_n \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_2 & a_1 \end{bmatrix} = \frac{1}{n} W^H \,\mathbf{diag}(Wa) W$$

W is the DFT matrix

- A is diagonalizable by a unitary similarity transformation

$$A = UDU^H, \qquad \text{where } U = \frac{1}{\sqrt{n}} W^H \text{ and } D = \mathbf{diag}(Wa)$$

- hence, A is normal
- eigenvalues of A are given by the DFT of a, columns of W are eigenvectors

Exercises

Exercise 1

1. show that the eigenvalues of a unitary matrix are on the unit circle
2. show that the eigenvalues of a Hermitian matrix are on the real axis
3. show that the eigenvalues of a skew-Hermitian matrix are on the imaginary axis

Exercise 2: consider the transfer function

$$H(s) = c^T (sI - A)^{-1} b$$

where $c, b \in \mathbf{R}^n$, $A \in \mathbf{R}^{n \times n}$, and s is a complex number

explain how to evaluate $H(s)$ at $m \gg n$ points s with order mn^2 complexity

Outline

- eigenvalues of nonsymmetric matrix
- **Schur decomposition**
- Sylvester equation

Sylvester equation

Sylvester equation: a matrix equation

$$AX + XB = C$$

where A is $m \times m$, B is $n \times n$, C is $m \times n$

- the variable is an $m \times n$ matrix X
- a set of mn linear linear equations in mn variables X_{ij}
- standard algorithms for linear equations of this size have order $m^3 n^3$ complexity
- we'll see that the Schur decomposition provides a much more efficient algorithm

Lyapunov equation

- special case with $B = A^T$ and symmetric C
- important in theory of linear dynamical systems

Solving Sylvester equations

$$AX + XB = C$$

Step 1: reduce to a Sylvester equation with upper triangular matrices

- compute Schur decomposition of A, B:

$$A = USU^H, \qquad B = VTV^H, \qquad \text{with } U, V \text{ unitary and } S, T \text{ upper triangular}$$

- complexity is order m^3 for $m \times m$ matrix A and order n^3 for $n \times n$ matrix B
- substitute Schur decompositions in Sylvester equation:

$$USU^H X + XVTV^H = C$$

- change of variables $Y = U^H XV$ gives Sylvester equation

$$SY + YT = D, \quad \text{where } D = U^H CV$$

Solving Sylvester equations

Step 2: solve triangular Sylvester equation $SY + YT = D$ column by column

$$S \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} + \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} T = \begin{bmatrix} D_1 & D_2 & \cdots & D_n \end{bmatrix}$$

- a set of n upper triangular equations in the m-vectors $Y_1, \ldots, Y_n$:

$$\begin{aligned}
(S + T_{11}I)Y_1 &= D_1 \\
(S + T_{22}I)Y_2 &= D_2 - T_{12}Y_1 \\
(S + T_{33}I)Y_3 &= D_3 - T_{13}Y_1 - T_{23}Y_2 \\
&\vdots \\
(S + T_{nn}I)Y_n &= D_n - T_{1n}Y_1 - \cdots - T_{n-1,n}Y_{n-1}
\end{aligned}$$

- solvable (by back substitution) if $S_{ii} + T_{jj} \neq 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$
- complexity of calculating right-hand sides is order mn^2
- complexity of the n back substitutions is order nm^2

Solving Sylvester equations

Step 3: compute solution X from Y:

$$X = UYV^H$$

two matrix–matrix products, with complexity $2m^2n + 2mn^2$

Overall complexity

- highest-order terms are cubic: m^3, m^2n, mn^2, n^3
- much more efficient than standard linear equation solver (m^3n^3 complexity)

this algorithm is known the *Bartels–Stewart method*