Vv417 Lecture 24

Jing Liu

UM-SJTU Joint Institute

November 28, 2019

Theorem

The eigenvalues of the following real matrix are $\lambda_1 = a + b\text{i}$ and $\lambda_2 = a - b\text{i}$.

$$\mathbf{C} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \qquad \text{where} \quad b \geq 0$$

If a and b are *not both zero*, then this matrix can be factored as

$$\mathbf{C} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where θ is the argument of λ_1 and r is the modulus of λ_1,

$$\theta = \arg(\lambda_1), \qquad r = \operatorname{mod}(\lambda_1)$$

- For every vector $\mathbf{x} \in \mathbb{R}^2$, consider the matrix multiplication between $\mathbf{C}$ and $\mathbf{x}$

$$\mathbf{C}\mathbf{x} = r \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{x}$$

- Geometrically, this states that multiplication by a matrix of the given form is a rotation through the angle θ followed by a scaling with factor r.

Proof

- The first part is trivial,

$$(a-\lambda)^2 + b^2 = 0 \implies \lambda = a \pm b\mathrm{i}$$

- To prove the second part,

$$\mathbf{C} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \frac{a}{r} & -\frac{b}{r} \\ \frac{b}{r} & \frac{a}{r} \end{bmatrix}$$

- Consider the polar form,

$$a = \mathrm{Re}(\lambda_1) = r\cos\theta, \qquad b = \mathrm{Im}(\lambda_1) = r\sin\theta$$

- Thus

$$\mathbf{C} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \square$$

Theorem

Suppose $\mathbf{A}$ is a real 2×2 matrix with complex eigenvalues

$$\lambda = a \pm b\mathrm{i}, \qquad \text{where} \quad b > 0$$

If $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ corresponding to $\lambda = a - b\mathrm{i}$, then the matrix

$$\mathbf{P} = \begin{bmatrix} \mathrm{Re}(\mathbf{x}) & \mathrm{Im}(\mathbf{x}) \end{bmatrix} \quad \text{is invertiable and} \quad \mathbf{A} = \mathbf{P} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mathbf{P}^{-1}$$

Q: What do we know from this theorem?

- Every real 2×2 matrix $\mathbf{A}$ with complex eigenvalues is a matrix representation of a rotation and scaling operator. Let

$$\mathbf{S} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \qquad \text{and} \qquad \mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

by the last two theorems, we can factor a matrix $\mathbf{A}$ with complex eigenvalues

$$\mathbf{A} = \mathbf{P}\mathbf{S}\mathbf{R}_\theta\mathbf{P}^{-1}$$

- If we now view $\mathbf{P}$ as the transition matrix from the basis

$$\mathcal{B} = \{\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})\}$$

to the standard basis, then the last formula tells us that computing a product

$$\mathbf{A}\mathbf{x}_0$$

can be broken down into a three-step process:

Step 1. Map $\mathbf{x}_0$ from standard coordinates into $\mathcal{B}$-coordinates, that is, $\mathbf{P}^{-1}\mathbf{x}_0$.

Step 2. Rotate and scale the vector $\mathbf{P}^{-1}\mathbf{x}_0$, that is,

$$\mathbf{S}\mathbf{R}_\theta\mathbf{P}^{-1}\mathbf{x}_0$$

Step 3. Map the rotated and scaled vector back to standard coordinates,

$$\mathbf{A}\mathbf{x}_0 = \mathbf{P}\mathbf{S}\mathbf{R}_\theta\mathbf{P}^{-1}\mathbf{x}_0$$

Exercise

Find the matrix that diagonalizes $\mathbf{A}$. Explain the effect of applying $\mathbf{A}$ repeatedly.

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Solution

- We want $\mathbf{A}^n\mathbf{x}_0$, if $\mathbf{A}$ diagonalizable, then

$$\mathbf{A}^n\mathbf{x}_0 = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}\mathbf{x}_0$$

- So we need to solve the eigenvalue and eigenvector problem,

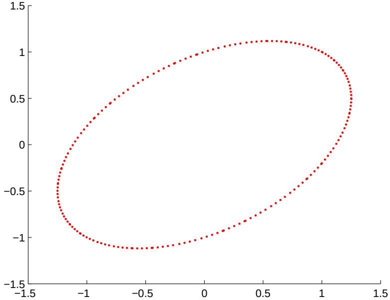
$$\mathbf{D} = \begin{bmatrix} 0.8000 + 0.6000\mathrm{i} & 0.0000 + 0.0000\mathrm{i} \\ 0.0000 + 0.0000\mathrm{i} & 0.8000 - 0.6000\mathrm{i} \end{bmatrix}$$

and

$$\mathbf{P} = \begin{bmatrix} 0.7454 + 0.0000\mathrm{i} & 0.7454 + 0.0000\mathrm{i} \\ 0.2981 + 0.5963\mathrm{i} & 0.2981 - 0.5963\mathrm{i} \end{bmatrix}$$

- Since $|\lambda| = 1$, perhaps you would expect a circular orbit.
- To understand why the points move along an elliptical path, we will need to examine the eigenvectors as well as the eigenvalues of $\mathbf{A}$.
- However, the basis $\mathcal{B}$ is skewed (not orthogonal), so when the points on the circular orbit are transformed back to standard coordinates, the effect is to distort the circular orbit into the elliptical orbit traced by $\mathbf{A}^n \mathbf{x}_0$.

```
>> A = [1/2 3/4; -3/5 11/10];
%Eigenvalues and Eigenvectors
>> [P_e D] = eig(A)
>> theta = angle(D) %in Radians
>> theta / (pi) * 180 %in Degrees
>> r = abs(D) %Modulus
%Graphically
>> x = [ 1; 1];
>> figure; xlim([-1.5,1.5]); ylim([-1.5,1.5]);
>> hold on; plot(x(1,1),x(2,1),'r.'); x = A*x;
>> hold off;
%Take the second eigenvalue
>> D(2,2)
>> P_e(:,2)
>> tmp = P_e(:,2) / P_e(2,2);
>> P = [ real(tmp) imag(tmp) ];
>> S = eye(2);
>> R = [cos(theta(1,1)) (- sin(theta(1,1)));...
    sin(theta(1,1)) cos(theta(1,1))];
>> A*x
>> P*S*R*P^(-1)*x
```



Q: Do you remember the definition of symmetric and orthogonal matrices?

Definition

If $\mathbf{u} = \left[u_1, \ldots, u_n\right]^{\mathrm{T}}$ and $\mathbf{v} = \left[v_1, \ldots, v_n\right]^{\mathrm{T}}$ are vectors in $\mathbb{C}^n$, then the complex Euclidean inner product of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= \mathbf{u}^{\mathrm{T}} \overline{\mathbf{v}} \\ &= u_1 \overline{v}_1 + u_2 \overline{v}_2 + \cdots + u_n \overline{v}_n = \overline{\mathbf{v}}^{\mathrm{T}} \mathbf{u}\end{aligned}$$

We also define the Euclidean norm on $\mathbb{C}^n$ to be

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \left(\overline{\mathbf{v}}^{\mathrm{T}} \mathbf{v}\right)^{1/2} = \sqrt{|v_1|^2 + \cdots + |v_n|^2} \qquad \text{where} \quad |v_i| = \mathrm{mod}(v_i)$$

- As a notational convenience, we write $\mathbf{v}^{\mathrm{H}}$ for the transpose of $\overline{\mathbf{v}}$, so

$$\overline{\mathbf{v}}^{\mathrm{T}} = \mathbf{v}^{\mathrm{H}}, \qquad \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^{\mathrm{H}} \mathbf{u}, \qquad \text{and} \qquad \|\mathbf{v}\| = \left(\mathbf{v}^{\mathrm{H}} \mathbf{v}\right)^{1/2}$$

and $\mathbf{v}^{\mathrm{H}}$ is known as the conjugate transpose of $\mathbf{v}$.

- Note that inner product in a complex inner product space is slightly different,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}, \text{ rather than } \langle \mathbf{v}, \mathbf{u} \rangle$$

- Another major difference between $\mathbb{R}^n$ and $\mathbb{C}^n$ is that

$$\alpha \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \overline{\alpha} \mathbf{v} \rangle, \text{ rather than } \langle \mathbf{u}, \alpha \mathbf{v} \rangle$$

Theorem

If $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{C}^n$, and if α is a scalar, then the complex Euclidean inner product has the following properties:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ [Antisymmetry property]
2. $\alpha \langle \mathbf{u}, \mathbf{v} \rangle = \langle \alpha \mathbf{u}, \mathbf{v} \rangle$ [Homogeneity property]
3. $\langle \mathbf{u}, \alpha \mathbf{v} \rangle = \overline{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle$ [Antihomogeneity property]
4. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$ [Distributive property]
5. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$ [Positivity property]

- If we make the proper modifications to allow for the difference, theorems on real inner product spaces will still be valid for complex inner product spaces.

- As in the real case, we call $\mathbf{v}$ a unit vector in $\mathbb{C}^n$ if

$$\|\mathbf{v}\| = 1$$

and we say two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

- Recall if $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for a real inner product space

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \text{ then } \alpha_i = \langle \mathbf{u}_i, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{u}_i \rangle \text{ and } \|\mathbf{x}\|^2 = \sum_{i=1}^{n} \alpha_i^2$$

- For a complex inner product space, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis

$$\mathbf{z} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \text{ then } \overline{\alpha_i} = \langle \mathbf{u}_i, \mathbf{z} \rangle, \ \alpha_i = \langle \mathbf{z}, \mathbf{u}_i \rangle, \text{ and } \|\mathbf{z}\|^2 = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i}$$

- The inner product space $\mathbb{C}^n$ is similar to the inner product space $\mathbb{R}^n$.

$\mathbb{R}^n$	$\mathbb{C}^n$
$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^{\mathrm{T}}\mathbf{u}$	$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^{\mathrm{H}}\mathbf{u}$
$\mathbf{u}^{\mathrm{T}}\mathbf{v} = \mathbf{v}^{\mathrm{T}}\mathbf{u}$	$\mathbf{u}^{\mathrm{H}}\mathbf{v} = \overline{\mathbf{v}^{\mathrm{H}}\mathbf{u}}$
$\|\mathbf{v}\|^2 = \mathbf{v}^{\mathrm{T}}\mathbf{v}$	$\|\mathbf{v}\|^2 = \mathbf{v}^{\mathrm{H}}\mathbf{v}$

- We can extend the notation for conjugate transpose to complex matrices.
- The transpose of a conjugate of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{\mathrm{H}}$, that is,

$$\overline{\mathbf{A}}^{\mathrm{T}} = \mathbf{A}^{\mathrm{H}}$$

Theorem

If $\mathbf{A}$ and $\mathbf{B}$ are complex matrices of $m \times n$ and $\mathbf{C}$ is a complex matrix of $n \times r$,

1. $\left(\mathbf{A}^{\mathrm{H}}\right)^{\mathrm{H}} = \mathbf{A}$
2. $(\alpha\mathbf{A} + \beta\mathbf{B})^{\mathrm{H}} = \overline{\alpha}\mathbf{A}^{\mathrm{H}} + \overline{\beta}\mathbf{B}^{\mathrm{H}}$
3. $(\mathbf{AC})^{\mathrm{H}} = \mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}$

Definition

A matrix $\mathbf{A}$ is said to be Hermitian if $\mathbf{A} = \mathbf{A}^{\mathrm{H}}$.

- Foe example, the matrix $\mathbf{A} = \begin{bmatrix} 3 & 2-\mathrm{i} \\ 2+\mathrm{i} & 4 \end{bmatrix}$ is Hermitian, since

$$\mathbf{A}^{\mathrm{H}} = \begin{bmatrix} 3 & 2+\mathrm{i} \\ 2-\mathrm{i} & 4 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 3 & 2-\mathrm{i} \\ 2+\mathrm{i} & 4 \end{bmatrix} = \mathbf{A}$$

- If $\mathbf{A}$ is a real matrix, then $\mathbf{A}^{\mathrm{H}} = \mathbf{A}^{\mathrm{T}}$.
- In particular, if $\mathbf{A}$ is real symmetric matrix, then $\mathbf{A}$ is Hermitian.
- Hermitian matrices can be viewed as the complex analogue of symmetric real matrices. Hermitian matrices have many nice properties.

Theorem

The eigenvalues of a Hermitian matrix are all real, and eigenvectors belonging to distinct eigenvalues are orthogonal.

Proof

- Suppose $\mathbf{A}$ is a Hermitian matrix, and λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ is an eigenvector corresponding to λ.
- Let $\alpha = \mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x}$, and consider

$$\overline{\alpha} = \alpha^{\mathrm{H}} = \left(\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x}\right)^{\mathrm{H}} = \mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{x} = \mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} = \alpha$$

- Thus, α is a real number.
- If follows that

$$\alpha = \mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{H}}\lambda\mathbf{x} = \lambda\mathbf{x}^{\mathrm{H}}\mathbf{x} = \lambda\|\mathbf{x}\|^2$$

and hence

$$\lambda = \frac{\alpha}{\|\mathbf{x}\|^2}$$

is real.

Proof

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2,

$$(\mathbf{A}\mathbf{x}_1)^{\mathrm{H}}\mathbf{x}_2 = (\mathbf{A}\mathbf{x}_1)^{\mathrm{H}}\mathbf{x}_2 = \mathbf{x}_1^{\mathrm{H}}\mathbf{A}\mathbf{x}_2 = \mathbf{x}_1^{\mathrm{H}}\lambda_2\mathbf{x}_2 = \lambda_2\mathbf{x}_1^{\mathrm{H}}\mathbf{x}_2$$

However,

$$(\mathbf{A}\mathbf{x}_1)^{\mathrm{H}}\mathbf{x}_2 = (\lambda_1\mathbf{x}_1)^{\mathrm{H}}\mathbf{x}_2 = \overline{\lambda_1}\mathbf{x_1}^{\mathrm{H}}\mathbf{x}_2 = \lambda_1\mathbf{x_1}^{\mathrm{H}}\mathbf{x}_2$$

Thus

$$\lambda_2\mathbf{x}_1^{\mathrm{H}}\mathbf{x}_2 = \lambda_1\mathbf{x}_1^{\mathrm{H}}\mathbf{x}_2$$

Eigenvalues λ_1 and λ_2 are distinct,

$$\mathbf{x}_1^{\mathrm{H}}\mathbf{x}_2 = 0$$

therefore the two eigenvectors are orthogonal.

Definition

An $n \times n$ matrix $\mathbf{U}$ is said to be unitary if its columns are orthonormal in $\mathbb{C}^n$.

- Thus, $\mathbf{U}$ is unitary if and only if

$$\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$$

- Since the columns are orthonormal, $\mathbf{U}$ must have rank n, it follows that

$$\mathbf{U}^{-1} = \mathbf{I}\mathbf{U}^{-1} = \left(\mathbf{U}^{\mathrm{H}}\mathbf{U}\right)\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{H}}$$

- A real unitary matrix is an orthogonal matrix.

Theorem

If the eigenvalues of a Hermitian matrix $\mathbf{A}$ are distinct, then there exists a unitary matrix $\mathbf{U}$ that diagonalizes $\mathbf{A}$.

$$\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}$$

Exercise

Find a matrix $\mathbf{U}$ that unitarily diagonalizes $\mathbf{A} = \begin{bmatrix} 2 & 1-\mathrm{i} \\ 1+\mathrm{i} & 1 \end{bmatrix}$.

Solution

- Since $\mathbf{A}$ is Hermitian, it can be unitarily diagonalized,

$$\mathbf{A}^{\mathrm{H}} = \begin{bmatrix} \overline{2} & \overline{1-\mathrm{i}} \\ \overline{1+\mathrm{i}} & \overline{1} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 2 & 1-\mathrm{i} \\ 1+\mathrm{i} & 1 \end{bmatrix} = \mathbf{A}$$

- By solving the characteristic equation, we have

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \implies \lambda_1 = 3 \quad \text{and} \quad \lambda_2 = 0$$

- Find a basis for the null space of $\mathbf{A} - \lambda \mathbf{I}$, we have

$$\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2\} = \operatorname{span}\left\{ \begin{bmatrix} 1-\mathrm{i} \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1+\mathrm{i} \end{bmatrix} \right\}$$

Solution

Normalizes $\mathbf{x}_1$ and $\mathbf{x}_2$,

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1-\mathrm{i} \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1+\mathrm{i} \end{bmatrix}$$

So, the unitary matrix diagonalizes $\mathbf{A}$ is $\mathbf{U} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1-\mathrm{i} & -1 \\ 1 & 1+\mathrm{i} \end{bmatrix}$, and

$$\mathbf{D} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U} = \frac{1}{3}\begin{bmatrix} 1+\mathrm{i} & 1 \\ -1 & 1-\mathrm{i} \end{bmatrix}\begin{bmatrix} 2 & 1-\mathrm{i} \\ 1+\mathrm{i} & 1 \end{bmatrix}\begin{bmatrix} 1-\mathrm{i} & -1 \\ 1 & 1+\mathrm{i} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$$

Q: Can we always find a diagonal matrix that is unitarily similar to arbitrary $\mathbf{A}$?

Schur's Theorem

For each $n \times n$ matrix $\mathbf{A}$, there exist a unitary matrix $\mathbf{U}$ such that

the matrix $\mathbf{R} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}$ is upper triangular.

The factorization $\mathbf{A} = \mathbf{U}\mathbf{R}\mathbf{U}^{\mathrm{H}}$ is known as the Schur decomposition of $\mathbf{A}$.

Proof

The result is obviously true if $n = 1$ since

$$1^{\mathrm{H}} \cdot a \cdot 1 = a$$

Assume the theorem is true for $n = k$, that is, there is a unitary matrix $\mathbf{W}_{k\times k}$

such that $\mathbf{T}_{k\times k} = \mathbf{W}_{k\times k}^{\mathrm{H}} \mathbf{M}_{k\times k} \mathbf{W}_{k\times k}$ is upper triangular for any $\mathbf{M}_{k\times k}$.

Now consider the case when $n = k + 1$. Let λ_1 be an eigenvalue of the matrix

$$\mathbf{A}_{(k+1)\times(k+1)}$$

and $\mathbf{b}_1$ be a unit eigenvector corresponding to λ_1.

We can construct an orthonormal basis for $\mathbb{C}^{k+1}$ by using Gram-Schmidt

$$\mathcal{B} = \{\mathbf{b}_1, \underbrace{\mathbf{b}_2, \mathbf{b}_3, \ldots, \mathbf{b}_{k+1}}_{\text{orthogonalization}}\}$$

Proof

Let $\mathbf{B}$ be the matrix such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_{k+1} \end{bmatrix}$$

The matrix $\mathbf{B}$ is unitary by construction, and the first column of $\mathbf{B}^{\mathrm{H}}\mathbf{A}\mathbf{B}$ will be

$$\mathbf{B}^{\mathrm{H}}\mathbf{A}\mathbf{b}_1 = \mathbf{B}^{\mathrm{H}}\lambda_1\mathbf{b}_1 = \lambda_1\mathbf{B}^{\mathrm{H}}\mathbf{b}_1 = \lambda_1\mathbf{e}_1$$

Thus the matrix $\mathbf{B}^{\mathrm{H}}\mathbf{A}\mathbf{B}$ is of the form,

$$\left[\begin{array}{c|ccc} \lambda_1 & \bullet & \cdots & \bullet \\ \hline 0 & & & \\ \vdots & & \mathbf{M} & \\ 0 & & & \end{array}\right]$$

where $\mathbf{M}$ is a $k \times k$ matrix.

We assumed the theorem is true for $n = k$, thus for this matrix $\mathbf{M}$, we have

$$\mathbf{W}_{k\times k}^{\mathrm{H}}\mathbf{M}\mathbf{W}_{k\times k} = \mathbf{T}_{k\times k}$$

Proof

Suppose $\mathbf{V} = \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W} & \\ 0 & & & \end{array}\right] \implies \mathbf{V}^{\mathrm{H}} = \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W}^{\mathrm{H}} & \\ 0 & & & \end{array}\right]$

If we consider the following matrix product,

$$\mathbf{V}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{A}\mathbf{B}\mathbf{V} = \mathbf{V}^{\mathrm{H}} \left[\begin{array}{c:ccc} \lambda_1 & \bullet & \cdots & \bullet \\ \hdashline 0 & & & \\ \vdots & & \mathbf{M} & \\ 0 & & & \end{array}\right] \mathbf{V} = \left[\begin{array}{c:ccc} \lambda_1 & \bullet & \cdots & \bullet \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W}^{\mathrm{H}}\mathbf{M}\mathbf{W} & \\ 0 & & & \end{array}\right]$$

$$= \left[\begin{array}{c:ccc} \lambda_1 & \bullet & \cdots & \bullet \\ \hdashline 0 & & & \\ \vdots & & \mathbf{T} & \\ 0 & & & \end{array}\right] = \mathbf{R}$$

Proof

Clearly, the matrix $\mathbf{R}$ is a triangular matrix, and if we let

$$\mathbf{U} = \mathbf{B}\mathbf{V}$$

then the only step left is to show the matrix $\mathbf{U}$ is unitary,

$$\begin{aligned}
\mathbf{U}^{\mathrm{H}}\mathbf{U} &= (\mathbf{B}\mathbf{V})^{\mathrm{H}}\mathbf{B}\mathbf{V} \\
&= \mathbf{V}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{B}\mathbf{V} \\
&= \mathbf{V}^{\mathrm{H}}\mathbf{V} \\
&= \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W}^{\mathrm{H}} & \\ 0 & & & \end{array}\right] \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W} & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{W}^{\mathrm{H}}\mathbf{W} & \\ 0 & & & \end{array}\right] \\
&= \mathbf{I}
\end{aligned}$$

Spectral Theorem

For Hermitian matrices $\mathbf{A}$, there exists a unitary matrix $\mathbf{U}$ that diagonalizes $\mathbf{A}$.

Proof

- The last theorem says there is a unitary matrix $\mathbf{U}$ for each square matrix $\mathbf{A}$

  such that $\quad \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U} = \mathbf{R}, \quad$ where $\mathbf{R}$ is upper triangular.

- So if use the fact that $\mathbf{A}$ is Hermitian, then

$$
\begin{aligned}
\mathbf{R}^{\mathrm{H}} &= \left(\mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}\right)^{\mathrm{H}} \\
&= \mathbf{U}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{U} \\
&= \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U} \\
&= \mathbf{R}
\end{aligned}
$$

- Therefore the triangular matrix $\mathbf{R}$ is also Hermitian, thus diagonal.

Exercise

Find a matrix $\mathbf{P}$ that orthogonally diagonalizes $\mathbf{A} = \begin{bmatrix} 0 & 2 & -1 \\ 2 & 3 & -2 \\ -1 & -2 & 0 \end{bmatrix}$.

Solution

- The eigenvalues of $\mathbf{A}$ are

$$\lambda_1 = \lambda_2 = -1 \text{ and } \lambda_3 = 5.$$

- Computing eigenvectors in the usual way, for $\lambda = 1$, we have

$$\mathcal{B}_1 = \{\mathbf{x}_1, \mathbf{x}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

- We need an orthonormal basis for

$$\operatorname{span}(\mathcal{B}_1)$$

Solution

- Apply Gram-Schmidt, we have

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\left\|\mathbf{x}_2 - \left(\mathbf{x}_2^{\mathrm{T}}\mathbf{u}_1\right)\mathbf{u}_1\right\|}\left(\mathbf{x}_2 - \left(\mathbf{x}_2^{\mathrm{T}}\mathbf{u}_1\right)\mathbf{u}_1\right) = \frac{1}{\sqrt{3}}\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

- The eigenspace corresponding to $\lambda_3 = 5$ is spanned by $\mathbf{x}_3 = \begin{bmatrix}-1\\-2\\1\end{bmatrix}$.

- $\mathbf{A}$ is symmetric (Hermitian), $\mathbf{x}_3 \perp \operatorname{span}(\mathcal{B}_1)$, so we only need to normalize

$$\mathbf{u}_3 = \frac{1}{\|\mathbf{x}_3\|}\mathbf{x}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}-1\\-2\\1\end{bmatrix}$$

Solution

- Thus $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set and

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

orthogonally diagonalizes $\mathbf{A}$,

$$\mathbf{D} = \mathbf{P}^{\mathrm{T}}\mathbf{A}\mathbf{P}$$

Q: Is Hermitian a necessary for a matrix to be unitarily diagonalizable?

- A matrix $\mathbf{A}$ is skew-Hermitian if

$$\mathbf{A}^{\mathrm{H}} = -\mathbf{A}$$

- For example,

$$\mathbf{A} = \begin{bmatrix} \mathrm{i} & 1-\mathrm{i} & 5 \\ -1-\mathrm{i} & 2\mathrm{i} & \mathrm{i} \\ -5 & \mathrm{i} & 0 \end{bmatrix}$$

- It can be shown that a skew-Hermitian matrix is also unitarily diagonalizable.
- Thus non-Hermitian matrices may be unitarily diagonalizable,

$$\mathbf{D}^{\mathrm{H}} \neq \mathbf{D} \implies \mathbf{A}^{\mathrm{H}} = \left(\mathbf{U}^{\mathrm{H}}\right)^{\mathrm{H}} \mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}} \neq \mathbf{A}$$

- In general, if $\mathbf{A}$ is a matrix with a complete orthonormal set of eigenvectors,

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$$

  where $\mathbf{U}$ is unitary and $\mathbf{D}$ is a diagonal matrix, then

$$\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}} = \mathbf{U}\mathbf{D}\mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}$$

  and

$$\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{U}\mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}} = \mathbf{U}\mathbf{D}^{\mathrm{H}}\mathbf{D}\mathbf{U}^{\mathrm{H}}$$

- Since

$$\mathbf{D}^{\mathrm{H}}\mathbf{D} = \begin{bmatrix} \lambda_1\overline{\lambda_1} & & & \\ & \lambda_2\overline{\lambda_2} & & \\ & & \ddots & \\ & & & \lambda_n\overline{\lambda_n} \end{bmatrix} = \mathbf{D}\mathbf{D}^{\mathrm{H}}$$

it follows that

$$\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$$

Definition

A matrix $\mathbf{A}$ is said to be normal if $\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$.

- The above argument shows that if a matrix has a complete orthonormal set of eigenvectors, then the matrix is normal. The converse is also true.

Theorem

A matrix has a complete orthonormal set of eigenvectors if and only if it is normal.

Proof

By Schur's theorem, there is a unitary matrix $\mathbf{U}$ and a triangular matrix $\mathbf{R}$

$$\text{such that} \qquad \mathbf{R} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U} \implies \begin{matrix} \mathbf{R}^{\mathrm{H}}\mathbf{R} = \mathbf{U}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U} = \mathbf{U}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{U} \\ \mathbf{R}\mathbf{R}^{\mathrm{H}} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{U} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{A}^{\mathrm{H}}\mathbf{U} \end{matrix}$$

Since $\mathbf{A}$ is normal, $\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$,

$$\mathbf{R}^{\mathrm{H}}\mathbf{R} = \mathbf{R}\mathbf{R}^{\mathrm{H}}$$

$$\begin{bmatrix} \overline{r}_{11} & 0 & \cdots & 0 \\ \overline{r}_{12} & \overline{r}_{22} & \cdots & 0 \\ \vdots & & \ddots & \\ \overline{r}_{1n} & \overline{r}_{2n} & \cdots & \overline{r}_{nn} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} \begin{bmatrix} \overline{r}_{11} & 0 & \cdots & 0 \\ \overline{r}_{12} & \overline{r}_{22} & \cdots & 0 \\ \vdots & & \ddots & \\ \overline{r}_{1n} & \overline{r}_{2n} & \cdots & \overline{r}_{nn} \end{bmatrix}$$

Compare the diagonal elements, we see $r_{ij} = 0$ whenever $i \neq j$, e.g.

$$\overline{r}_{11}r_{11} = \overline{r}_{11}r_{11} + \overline{r}_{12}r_{12} + \cdots + \overline{r}_{1n}r_{1n}$$
$$\|r_{11}\|^2 = \|r_{11}\|^2 + \|r_{12}\|^2 + \cdots + \|r_{1n}\|^2 \quad \square$$